PSYCHOMETRIKA—VOL. 75, NO. 3, 393–419
SEPTEMBER 2010
DOI: 10.1007/S11336-010-9165-5

A GENERAL FAMILY OF LIMITED INFORMATION GOODNESS-OF-FIT STATISTICS FOR MULTINOMIAL DATA

HARRY JOE

DEPARTMENT OF STATISTICS, UNIVERSITY OF BRITISH COLUMBIA

ALBERTO MAYDEU-OLIVARES

FACULTY OF PSYCHOLOGY, UNIVERSITY OF BARCELONA

Maydeu-Olivares and Joe (J. Am. Stat. Assoc. 100:1009–1020, 2005; Psychometrika 71:713–732, 2006) introduced classes of chi-square tests for (sparse) multidimensional multinomial data based on low-order marginal proportions. Our extension provides general conditions under which quadratic forms in linear functions of cell residuals are asymptotically chi-square. The new statistics need not be based on margins, and can be used for one-dimensional multinomials. We also provide theory that explains why limited information statistics have good power, regardless of sparseness. We show how quadratic-form statistics can be constructed that are more powerful than X^2 and yet, have approximate chi-square null distribution in finite samples with large models. Examples with models for truncated count data and binary item response data are used to illustrate the theory.

Key words: categorical data analysis, cell-focusing, discrete data, item response theory, overdispersion, overlapping cells, Poisson models, quadratic form statistics, Rasch models, score test, sparse contingency tables, zero-inflation.

1. Introduction

Goodness-of-fit testing under multinomial sampling becomes increasingly difficult as the number of multinomial categories C increases. The empirical variance of Pearson's X^2 and its variance under its reference asymptotic distribution differ by a term that depends on the inverse of the category probabilities (Cochran, 1952). As the number of categories increases with category probabilities becoming smaller, the discrepancy between the empirical and asymptotic variances of X^2 can be large; and the type I error for X^2 will be larger than the α level based on its asymptotic critical value. However, for C fixed, the accuracy of the asymptotic p-values for X^2 depends also on sample size N. As N becomes smaller, some of the observed proportions increasingly become more poorly estimated (their estimates can be zero) and the empirical Type I errors of X^2 will become inaccurate. The degree of sparseness N/C summarizes the relationship between sample size and model size.

For categorical data from multinomial samples that may be summarized in n-dimensional tables, Maydeu-Olivares and Joe (2005, 2006) introduced classes of quadratic-form goodness-of-fit statistics in low-order marginal cell residuals: quadratic forms in univariate cell residuals only, quadratic forms in univariate and bivariate residuals only, in univariate, bivariate and trivariate residuals only, and so forth. They also showed that Pearson's X^2 is a special case of these classes, in which residuals up to order n, the number of categorical variables, are employed. As a way to

This research has been by grant SEJ2006-08204 from the Spanish Ministry of Education, and an NSERC Canada Discovery Grant. We are grateful to the referees and associate editor for comments leading to improvements. Also, we thank Virginia Yue Chen for some early numerical investigations.

Requests for reprints should be sent to Alberto Maydeu-Olivares, Faculty of Psychology, University of Barcelona, P. Valle de Hebrón, 171, 08035 Barcelona, Spain. E-mail: amaydeu@ub.edu

handle sparsity, they suggested using quadratic forms in low-order marginal cell residuals. They used these quadratic forms to test item response theory (IRT) models, showing empirically that when only low-order margins are used in the quadratic form, accurate p-values can be obtained even in tables of more than 10^7 cells with as few as 300 observations. Also, when only low-order margins were used, the statistics were shown empirically in some cases to have better power in nonsparse tables than statistics based on higher order residuals (including X^2).

Maydeu-Olivares and Joe's approach is based on linear maps of the cell residuals taking advantage of the multidimensional structure of the data. Although very large models can be tested with their approach, it is limited in three important aspects: (1) it cannot be used with unidimensional tables, (2) it can not be used for models for multidimensional tables that can not be identified from margins, and (3) there is a computational limit in the size of the models for multidimensional tables that can be tested due to the need to store very large matrices.

To overcome these limitations, in this article we extend Maydeu-Olivares and Joe's approach by providing conditions for quadratic forms in arbitrary linear combinations of cell residuals to be asymptotically chi-square under simple null and composite null hypotheses. These results can be applied not only to multidimensional tables, but also to unidimensional ones. For multidimensional tables, univariate, bivariate, and higher-order marginal residuals are special cases of these linear combinations. These results enable researchers to construct test statistics that may be well approximated by asymptotic methods even in sparse tables based on summary statistics that are not necessarily marginal residuals.

Given that for any given application there will be a number of candidate statistics within this family that could be used, we also provide a number of properties for these test statistics that may be used to choose the most suitable statistic within this class. As a special case, these properties may also be used to choose among the statistics originally introduced in Maydeu-Olivares and Joe (2005, 2006). Consider two statistics within this class, T_2 and T_1, with T_2 based on a further reduction of the data than T_1 (for example, univariate and bivariate residuals versus univariate, bivariate, and trivariate residuals). We show that the statistic based on a further reduction of the data (i.e., T_2) is algebraically smaller than or equal to the other statistic (i.e., T_1). Thus, if the non-centrality parameter of T_2 is the same or slightly smaller than that for T_1 while having fewer degrees of freedom, T_2 will be asymptotically more powerful than T_1. Because Pearson's X^2 statistic is a member of this class of statistics, these results explain why one can obtain statistics based on limited information (linear maps of the cell residuals) with much better power than X^2.

The present extension of Maydeu-Olivares and Joe's framework to quadratic forms in arbitrary linear combinations of cell residuals enables us to relate their original family of test statistics to the test statistics proposed by Glas (1988) and Glas and Verhelst (1989). Also, we can relate limited information statistics based on margins to the literature (a) on statistics for directional tests (i.e., tests with a specified alternative hypothesis; see, e.g., Rayner & Best, 1989, 1990), and (b) on the concept of dilution (expending degrees of freedom that are not helpful in detecting the alternative; see, e.g., Eubank, 1997). In particular, for testing simple nulls, our present extensions and Rayner and Best's (1989) score tests for smooth alternatives lead to the same family of statistics, except we do not require an orthonormality condition. For composite nulls, our approach and that of Rayner and Best are different. But for some choices of statistics in the quadratic form, the difference between our family and theirs reduces to a choice of weight matrix in the quadratic form.

The remainder of this paper is as follows. In Section 2, the classes of quadratic-form statistics are introduced and a subsection has theorems that show how the statistics compare under different sets of summaries of the data. The proofs of the theorems are deferred to the Appendix. Section 3 applies the theorems to power comparisons under sequences of local alternatives. Section 4 has two examples. The first example involves truncated count data (a form of multinomial data). We compare the performance of different test statistics within our family in testing a null hypothesis of Poisson when the true model involves zero-inflation or overdispersion. In the second example, we consider a log-linear model for binary item response data and compare different statistics for testing exchangeability among the items. Section 5 discusses in detail the relationships between our statistics and related ones, such as the quadratic-form statistics of Glas (1988) and Glas and Verhelst (1989), as well as to the smooth tests of Rayner and Best (1989). In particular, we provide advantages of the present approach over existing approaches, such as the theoretical results for power comparisons. Section 6 provides some concluding remarks and directions for future research.

2. Quadratic-form Statistics for Multinomial Data

In this section, we introduce classes of quadratic-form statistics for testing the goodness-of-fit of a model for discrete multinomial data with a finite number of categories. There are two classes: one for a simple null hypothesis and one for a composite null. For the latter, we assume there is a parametric model and that maximum likelihood estimation has been used (the latter is used for ease of exposition—see Section 5). Hence, we assume that the usual regularity assumptions for maximum likelihood estimation apply. In practice, the composite null hypothesis is the more useful one, but the results for the simple null help in understanding the results for the composite null.

Let $\mathcal{C}$ denote that set of categories and let C denote its cardinality; this set might be multidimensional. The probability distribution is denoted as $\{\pi_c\}$ for $c \in \mathcal{C}$. Assuming that the categories have been indexed linearly, the vector of all probabilities is denoted as the column vector $\boldsymbol{\pi}$. If there is a parametric family for the probability distribution, we let $\boldsymbol{\theta}$ be the parameter vector and assume that it is q-dimensional and belongs to a set Θ. For a parametric model, the probability vector is denoted $\boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\theta}) = \{\pi_c(\boldsymbol{\theta}) : c \in \mathcal{C}\}$. Other notation used below are the following: $\mathbf{0}$ denotes a zero column vector or matrix, with dimension clear from context, and $\mathbf{1}$ is a column vector of 1s with the dimension shown as a subscript if necessary for clarity.

The goodness-of-fit test statistics are based on a random multinomial sample of size N. Let $\mathbf{p}$ be the vector of sample proportions. For a simple null hypothesis $H_0 : \boldsymbol{\pi} = \boldsymbol{\pi}_0$, Pearson's statistic is

$$X^2 = N(\mathbf{p} - \boldsymbol{\pi}_0)'\mathbf{D}^{-1}(\mathbf{p} - \boldsymbol{\pi}_0),$$

where $\mathbf{D} = \text{diag}(\boldsymbol{\pi}_0)$. The asymptotic null distribution of X^2 is χ^2_{C-1}. For a composite null hypothesis $H_0 : \boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\theta})$ for some $\boldsymbol{\theta}$, we use $\widehat{X}^2$ for Pearson's statistic to indicate that it involves the maximum likelihood estimate (MLE) $\widehat{\boldsymbol{\theta}}$:

$$\widehat{X}^2 = N\big(\mathbf{p} - \boldsymbol{\pi}(\widehat{\boldsymbol{\theta}})\big)'\big[\mathbf{D}(\widehat{\boldsymbol{\theta}})\big]^{-1}\big(\mathbf{p} - \boldsymbol{\pi}(\widehat{\boldsymbol{\theta}})\big),$$

where $\mathbf{D}(\boldsymbol{\theta}) = \text{diag}(\boldsymbol{\pi}(\boldsymbol{\theta}))$. The asymptotic null distribution of $\widehat{X}^2$ is χ^2_{C-1-q}.

We introduce and discuss properties of two general classes of quadratic-form statistics based on a vector of summary statistics $\boldsymbol{\kappa}$. Each element of $\boldsymbol{\kappa}$ is a linear combination of the probabilities in $\boldsymbol{\pi}$. That is, $\boldsymbol{\kappa} = \mathbf{T}_\kappa \boldsymbol{\pi}$, where $\mathbf{T}_\kappa$ is an $s \times C$ matrix with full row rank, and $1 \le s \le C - 1$. Correspondingly, $\hat{\boldsymbol{\kappa}} = \mathbf{T}_\kappa \mathbf{p}$ is the sample version for a random sample, and if estimation with a parametric model is used, $\boldsymbol{\kappa}(\widehat{\boldsymbol{\theta}}) = \mathbf{T}_\kappa \boldsymbol{\pi}(\widehat{\boldsymbol{\theta}})$ is the model-based summary based on estimator $\widehat{\boldsymbol{\theta}}$. The subscript on $\mathbf{T}$ and other variables is mainly needed to compare two different summary vectors such as $\boldsymbol{\kappa}_1$ and $\boldsymbol{\kappa}_2$. The subscript will be suppressed if no confusion results.

Examples of summaries that could be elements of the vector $\boldsymbol{\kappa}$ are (a) $\kappa_{\mathcal{S}} = \sum_{c \in \mathcal{S}} \pi_c$, a pooling of the cells in $\mathcal{S}$ where $\mathcal{S}$ is a proper subset of $\mathcal{C}$; (b) $\kappa_1 = \sum_{j \in \mathcal{C}} j\pi_j$, the mean ordinal score with index set $\mathcal{C} = \{0, \dots, C-1\}$ representing scores of an ordinal variable, and (c) with a

3-dimensional index set of ordinal categories or scores $\mathcal{C} = \{(i_1, i_2, i_3) : i_j = 0, \ldots, K-1,\ j = 1, 2, 3\}$, a bivariate probability for the first two variables has the form $\kappa_{12:i_1,i_2} = \sum_{j_3} \pi_{i_1,i_2,j_3}$ and a bivariate moment of the scores has the form $\kappa_{12} = \sum_{j_1,j_2,j_3} j_1 j_2 \pi_{j_1,j_2,j_3}$. Each summary would be matched to a row of the matrix $\mathbf{T}_\kappa$. In another example, studied in Maydeu-Olivares and Joe (2006), $\boldsymbol{\kappa}$ consists of the vector of all bivariate marginal probabilities in an n-dimensional discrete distribution with support on $\{0, 1, \ldots, K-1\}^n$. This corresponds to a case of pooling categories with overlap that make use of the multidimensional structure.

The first class of statistics, L_κ, is suitable for testing simple null hypotheses. The second class of statistics, M_κ, is suitable for testing composite null hypotheses. The quadratic form goodness-of-fit statistics proposed by Maydeu-Olivares and Joe (2005, 2006) for multidimensional tables are special cases of L_κ and M_κ, respectively, as are X^2 and $\widehat{X}^2$.

A statistic of the form L_κ is well defined if $\mathbf{T}_\kappa$ satisfies Condition T, and a statistic of the form M_κ is well defined if $\mathbf{T}_\kappa$ satisfies Conditions T and D. These conditions are given next.

Condition T. The matrix $\mathbf{T}_\kappa$ has full row rank and $\mathbf{1}'_C$ is not in its row span.

Condition D. With a parametric model $\boldsymbol{\pi}(\boldsymbol{\theta})$ with q-dimensional $\boldsymbol{\theta}$, the matrix $\mathbf{T}_\kappa$ has row dimension $s > q$ and leads to

$$\boldsymbol{\Delta}_\kappa = \boldsymbol{\Delta}_\kappa(\boldsymbol{\theta}) = \frac{\partial \boldsymbol{\kappa}}{\partial \boldsymbol{\theta}'} = \mathbf{T}_\kappa \frac{\partial \boldsymbol{\pi}}{\partial \boldsymbol{\theta}'} \tag{1}$$

being an $s \times q$ matrix with full column rank q.

Consequently, Condition T is verified if

$$\operatorname{rank}(\mathbf{T}_\kappa) = s \neq \operatorname{rank}\begin{pmatrix} \mathbf{1}'_C \\ \mathbf{T}_\kappa \end{pmatrix},$$

and Condition D is verified if (1) is of full column rank. Also, the covariance matrix of $\sqrt{N}\,[\mathbf{p} - \boldsymbol{\pi}]$ is $\boldsymbol{\Gamma} = \operatorname{diag}(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}'$, so that the covariance matrix of $\sqrt{N}\,(\hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa})$ is $\boldsymbol{\Xi}_\kappa = \mathbf{T}_\kappa \boldsymbol{\Gamma} \mathbf{T}'_\kappa$. Condition T ensures that $\boldsymbol{\Xi}_\kappa$ is positive definite (see Appendix A.1). Condition D ensures the identifiability of the parametric model $\boldsymbol{\pi}(\boldsymbol{\theta})$ from the summary vector $\boldsymbol{\kappa}(\boldsymbol{\theta})$, that is, if $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$, then $\mathbf{T}_\kappa \boldsymbol{\pi}(\boldsymbol{\theta}_1) \neq \mathbf{T}_\kappa \boldsymbol{\pi}(\boldsymbol{\theta}_2)$.

We next present the quadratic-form statistics, first for the simple null hypothesis and then for the composite null hypothesis.

2.1. Statistics for Testing Simple Null Hypotheses: The L_κ Family

The simple null hypothesis is $H_0 : \boldsymbol{\pi} = \boldsymbol{\pi}_0$. Let $\mathbf{T}_\kappa$ be an $s \times C$ matrix, $\boldsymbol{\kappa}_0 = \mathbf{T}_\kappa \boldsymbol{\pi}_0$ be the transform of the probability vector, and $\hat{\boldsymbol{\kappa}} = \mathbf{T}_\kappa \mathbf{p}$ be the sample counterpart.

The quadratic form statistic, denoted as L_κ, is:

$$L_\kappa = N(\hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}_0)' \boldsymbol{\Xi}_\kappa^{-1} (\hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}_0), \tag{2}$$

where $\boldsymbol{\Xi}_\kappa = \mathbf{T}_\kappa \boldsymbol{\Gamma} \mathbf{T}'_\kappa$, $\boldsymbol{\Gamma} = \operatorname{diag}(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}'$, and $\mathbf{T}_\kappa$ satisfies Condition T. The matrix of the quadratic form $\boldsymbol{\Xi}_\kappa^{-1}$ is an $s \times s$ matrix. From standard results for multivariate normal distributions, the asymptotic null distribution of L_κ is χ^2_s. Note that if $s = 1$ and $\mathbf{T}$ has $C - 1$ 0s and one 1, then L_κ is just the square of the z-statistic for testing a population proportion.

2.2. *Statistics for Testing Composite Null Hypotheses: The M_κ Family*

We assume a parametric family $\{\pi(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$, where the dimension of $\boldsymbol{\theta}$ is $q \geq 1$. The composite null hypothesis is $H_0 : \boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\theta})$ for some $\boldsymbol{\theta} \in \Theta$.

Let $\boldsymbol{\Delta} = \partial\boldsymbol{\pi}/\partial\boldsymbol{\theta}'$ be the $C \times q$ matrix of derivatives of the probabilities with respect to the parameters, and let $\boldsymbol{\kappa}$ be a s-dimensional vector of statistics that satisfies Conditions T and D. Then from (1), $\boldsymbol{\Delta}_\kappa = \boldsymbol{\Delta}_\kappa(\boldsymbol{\theta}) = \mathbf{T}_\kappa \boldsymbol{\Delta}$, is an $s \times q$ matrix.

With $\widehat{\boldsymbol{\theta}}$ being the MLE, $\boldsymbol{\kappa}(\widehat{\boldsymbol{\theta}}) = \mathbf{T}_\kappa \boldsymbol{\pi}(\widehat{\boldsymbol{\theta}})$ is based on the model-based estimated proportions and $\hat{\boldsymbol{\kappa}} = \mathbf{T}_\kappa \mathbf{p}$ is based on the sample proportions. The quadratic-form statistic, denoted as M_κ, is

$$M_\kappa = N\big(\hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}(\widehat{\boldsymbol{\theta}})\big)' \mathbf{U}_\kappa(\widehat{\boldsymbol{\theta}}) \big(\hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}(\widehat{\boldsymbol{\theta}})\big), \tag{3}$$

where

$$\mathbf{U}_\kappa = \mathbf{U}_\kappa(\boldsymbol{\theta}) = \boldsymbol{\Delta}_\kappa^{(c)}(\boldsymbol{\theta})\big[\big(\boldsymbol{\Delta}_\kappa^{(c)}(\boldsymbol{\theta})\big)' \boldsymbol{\Xi}_\kappa(\boldsymbol{\theta}) \boldsymbol{\Delta}_\kappa^{(c)}(\boldsymbol{\theta})\big]^{-1} \big(\boldsymbol{\Delta}_\kappa^{(c)}(\boldsymbol{\theta})\big)', \tag{4}$$

$\boldsymbol{\Xi}_\kappa(\boldsymbol{\theta}) = \mathbf{T}_\kappa \boldsymbol{\Gamma}(\boldsymbol{\theta}) \mathbf{T}_\kappa'$, $\boldsymbol{\Gamma}(\boldsymbol{\theta}) = \mathrm{diag}(\boldsymbol{\pi}(\boldsymbol{\theta})) - \boldsymbol{\pi}(\boldsymbol{\theta})\boldsymbol{\pi}'(\boldsymbol{\theta})$, and $\boldsymbol{\Delta}_\kappa^{(c)}$ is an orthogonal complement of $\boldsymbol{\Delta}_\kappa$ (satisfying $\boldsymbol{\Delta}_\kappa^{(c)\prime}\boldsymbol{\Delta}_\kappa = \mathbf{0}$), with dimension $s \times (s-q)$. Note that $\mathbf{U}_\kappa$ is an $s \times s$ matrix. The orthogonal complement is nonunique but M_κ is invariant to the choice of the orthogonal complement. Also, according to an identity of Khatri (1966)—see problem 33 of Section 1f of Rao (1973), we can write (4) as

$$\mathbf{U}_\kappa = \boldsymbol{\Xi}_\kappa^{-1} - \boldsymbol{\Xi}_\kappa^{-1}\boldsymbol{\Delta}_\kappa\big(\boldsymbol{\Delta}_\kappa' \boldsymbol{\Xi}_\kappa^{-1}\boldsymbol{\Delta}_\kappa\big)^{-1}\boldsymbol{\Delta}_\kappa' \boldsymbol{\Xi}_\kappa^{-1}. \tag{5}$$

The asymptotic null distribution of M_κ is χ^2_{s-q}. This follows from standard results for quadratic forms in normal random variables, based on the following construction of M_κ:

Consider the $(s-q)$-dimensional vector

$$\mathbf{Z}_\kappa = \sqrt{N}\,\boldsymbol{\Delta}_\kappa^{(c)\prime}\big[\hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}(\widehat{\boldsymbol{\theta}})\big] = \sqrt{N}\,\boldsymbol{\Delta}_\kappa^{(c)\prime}\mathbf{T}_\kappa\big[\mathbf{p} - \boldsymbol{\pi}(\widehat{\boldsymbol{\theta}})\big],$$

that has asymptotic covariance matrix equal to

$$\boldsymbol{\Delta}_\kappa^{(c)\prime}\mathbf{T}_\kappa \boldsymbol{\Sigma} \mathbf{T}_\kappa' \boldsymbol{\Delta}_\kappa^{(c)}, \tag{6}$$

where $\boldsymbol{\Sigma} = \boldsymbol{\Gamma} - \boldsymbol{\Delta}\mathcal{I}^{-1}\boldsymbol{\Delta}'$ and $\mathcal{I} = \boldsymbol{\Delta}'\mathbf{D}^{-1}(\boldsymbol{\theta})\boldsymbol{\Delta}$ is the $q \times q$ Fisher information matrix (see Appendix A.2 for some standard background results for maximum likelihood that are relevant to Pearson's $\widehat{X}^2$). Let

$$\boldsymbol{\Sigma}_\kappa = \mathbf{T}_\kappa \boldsymbol{\Sigma} \mathbf{T}_\kappa'. \tag{7}$$

Then with the above definitions of $\boldsymbol{\Delta}_\kappa$ and $\boldsymbol{\Xi}_\kappa$, (6) is the same as

$$\boldsymbol{\Delta}_\kappa^{(c)\prime}\boldsymbol{\Sigma}_\kappa\boldsymbol{\Delta}_\kappa^{(c)} = \boldsymbol{\Delta}_\kappa^{(c)\prime}\boldsymbol{\Xi}_\kappa\boldsymbol{\Delta}_\kappa^{(c)} - \boldsymbol{\Delta}_\kappa^{(c)\prime}\boldsymbol{\Delta}_\kappa\mathcal{I}^{-1}\boldsymbol{\Delta}_k'\boldsymbol{\Delta}_\kappa^{(c)} = \boldsymbol{\Delta}_\kappa^{(c)\prime}\boldsymbol{\Xi}_\kappa\boldsymbol{\Delta}_\kappa^{(c)}.$$

Note that M_κ is just the quadratic form

$$M_\kappa = \mathbf{Z}_k'\big[\boldsymbol{\Delta}_\kappa^{(c)\prime}(\widehat{\boldsymbol{\theta}})\,\boldsymbol{\Xi}_\kappa(\widehat{\boldsymbol{\theta}})\boldsymbol{\Delta}_\kappa^{(c)}(\widehat{\boldsymbol{\theta}})\big]^{-1}\mathbf{Z}_k,$$

and this is asymptotically χ^2_{s-q} under the null hypothesis. Also, note that $\boldsymbol{\Sigma}_\kappa$ is a generalized inverse of $\mathbf{U}_\kappa$, but in general $\mathbf{U}_\kappa$ is not a generalized inverse of $\boldsymbol{\Sigma}_\kappa$. That is, it is straightforward to show that

$$\mathbf{U}_\kappa = \mathbf{U}_\kappa\boldsymbol{\Sigma}_\kappa\mathbf{U}_\kappa = \mathbf{U}_\kappa\boldsymbol{\Xi}_\kappa\mathbf{U}_\kappa.$$

We will be comparing two different M_κ's, based on $\boldsymbol{\kappa}_1, \boldsymbol{\kappa}_2$, in the next section. To do this, we next define the notion of $\boldsymbol{\kappa}_2$ being a reduction of $\boldsymbol{\kappa}_1$, and then introduce two examples to make the idea more concrete.

Definition (Reduction). For $j = 1, 2$, let $\boldsymbol{\kappa}_j = \mathbf{T}_{\kappa_j}\boldsymbol{\pi}$ with dimension s_j, where $\mathbf{T}_{\kappa_1}, \mathbf{T}_{\kappa_2}$ satisfy Condition T. Suppose $s_1 > s_2$. Then $\boldsymbol{\kappa}_2$ is a further summary *reduction* of $\boldsymbol{\kappa}_1$ if there is a full row-rank transform matrix $\mathbf{T}_{21}$ of dimension $s_2 \times s_1$ such that $\boldsymbol{\kappa}_2 = \mathbf{T}_{21}\boldsymbol{\kappa}_1$.

Examples. (a) Suppose there are 6 categories indexed as $0, 1, 2, 3, 4, 5$. Let $\boldsymbol{\kappa}_1$ with $s_1 = 4$ correspond to the individual categories 1, 2, 3 and the pooling of categories 4, 5. Let $\boldsymbol{\kappa}_2$ with $s_2 = 2$ correspond to a pooling of categories 1, 2, 3 and a second pooling of categories 4, 5. Then

$$\mathbf{T}_{\kappa_1} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}, \qquad \mathbf{T}_{\kappa_2} = \begin{pmatrix} 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix},$$
$$\mathbf{T}_{21} = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

(b) Consider an n-way table with dimension $n > 3$ and m categories per dimension, labeled $0, \dots, m-1$. Let $\boldsymbol{\kappa}_1$ be vector of dimension $s_1 = \binom{n}{3}(m-1)^3$ from all 3-dimensional margins with category indices (j_1, j_2, j_3), $1 \le j_1, j_2, j_3 \le m$, and let $\boldsymbol{\kappa}_2$ be vector of dimension $s_2 = \binom{n}{2}(m-1)^2$ from all 2-dimensional margins with category indices (j_1, j_2), $1 \le j_1, j_2 \le m$. Then $\boldsymbol{\kappa}_2$ is a summary reduction of $\boldsymbol{\kappa}_1$. The matrices $\mathbf{T}_{\kappa_1}$, $\mathbf{T}_{\kappa_2}$ and $\mathbf{T}_{21}$ depend on the linear arrangement of the s_1 bivariate and s_2 trivariate category vectors.

For $j = 1, 2$, let $\boldsymbol{\kappa}_j = \mathbf{T}_{\kappa_j}\boldsymbol{\pi}$ have dimension s_j, and let $\hat{\boldsymbol{\kappa}}_j = \mathbf{T}_{\kappa_j}\mathbf{p}$. Suppose $\boldsymbol{\kappa}_2$ is a reduction of $\boldsymbol{\kappa}_1$ with transform matrix $\mathbf{T}_{21}$ and notice that $\mathbf{T}_{21} = \mathbf{T}_2\mathbf{T}_1'(\mathbf{T}_1\mathbf{T}_1')^{-1}$. With the above definition, we can express M_{κ_2} in terms of $\mathbf{T}_{21}$ and $\boldsymbol{\kappa}_1$, and this is a key to comparing M_{κ_1} and M_{κ_2}. Let $\mathbf{U}_{\kappa_j}$, $j = 1, 2$, be the matrices of the quadratic forms, with the dependence on $\widehat{\boldsymbol{\theta}}$ suppressed. Then $\hat{\boldsymbol{\kappa}}_2 - \boldsymbol{\kappa}_2(\widehat{\boldsymbol{\theta}}) = \mathbf{T}_{21}[\hat{\boldsymbol{\kappa}}_1 - \boldsymbol{\kappa}_1(\widehat{\boldsymbol{\theta}})]$,

$$M_{\kappa_1} = N\big(\hat{\boldsymbol{\kappa}}_1 - \boldsymbol{\kappa}_1(\widehat{\boldsymbol{\theta}})\big)'\mathbf{U}_{\kappa_1}\big(\hat{\boldsymbol{\kappa}}_1 - \boldsymbol{\kappa}_1(\widehat{\boldsymbol{\theta}})\big),$$
$$M_{\kappa_2} = N\big(\hat{\boldsymbol{\kappa}}_2 - \boldsymbol{\kappa}_2(\widehat{\boldsymbol{\theta}})\big)'\mathbf{U}_{\kappa_2}\big(\hat{\boldsymbol{\kappa}}_2 - \boldsymbol{\kappa}_2(\widehat{\boldsymbol{\theta}})\big) = N\big(\hat{\boldsymbol{\kappa}}_1 - \boldsymbol{\kappa}_1(\widehat{\boldsymbol{\theta}})\big)'\mathbf{T}_{21}'\mathbf{U}_{\kappa_2}\mathbf{T}_{21}\big(\hat{\boldsymbol{\kappa}}_1 - \boldsymbol{\kappa}_1(\widehat{\boldsymbol{\theta}})\big).$$

Then

$$M_{\kappa_1} - M_{\kappa_2} = N\big(\hat{\boldsymbol{\kappa}}_1 - \boldsymbol{\kappa}_1(\widehat{\boldsymbol{\theta}})\big)'(\mathbf{U}_{\kappa_1} - \mathbf{T}_{21}'\mathbf{U}_{\kappa_2}\mathbf{T}_{21})\big(\hat{\boldsymbol{\kappa}}_1 - \boldsymbol{\kappa}_1(\widehat{\boldsymbol{\theta}})\big).$$

Several theorems will be proved in the next section with the main ones concerning the matrix of the quadratic form in the difference $M_{\kappa_1} - M_{\kappa_2}$, i.e., $\mathbf{U}_{\kappa_1} - \mathbf{T}_{21}'\mathbf{U}_{\kappa_2}\mathbf{T}_{21}$.

We end this subsection with a comment on the numerical computation of the orthogonal complement. Suppose $\boldsymbol{\Delta}_\kappa$ has dimension $s \times q$ with $q < s$. Using a singular value decomposition,

$$\boldsymbol{\Delta}_\kappa = \begin{pmatrix} \mathbf{W}_1 & \mathbf{W}_2 \end{pmatrix} \begin{pmatrix} \mathbf{D}_1 \\ \mathbf{0} \end{pmatrix} \mathbf{V}' = \mathbf{W}_1\mathbf{D}_1\mathbf{V}',$$

where $\mathbf{W}_1$ is $s \times q$, $\mathbf{W}_2$ is $s \times (s-q)$, $\mathbf{W}_2'\mathbf{W}_1 = \mathbf{0}$, $\mathbf{D}_1$ is a diagonal $q \times q$ matrix of singular values, the $\mathbf{0}$ matrix is $(s-q) \times q$ and $\mathbf{V}$ is $q \times q$. Then an orthogonal complement of $\boldsymbol{\Delta}_\kappa$ is $\boldsymbol{\Delta}_\kappa^{(c)} = \mathbf{W}_2$ since $\mathbf{W}_2'\boldsymbol{\Delta}_\kappa = \mathbf{W}_2'\mathbf{W}_1\mathbf{D}_1\mathbf{V}' = \mathbf{0}$.

2.3. Some Theorems on the L_κ and M_κ Families of Statistics

This subsection has some theoretical results on the relations between different members of the L_κ and M_κ family of statistics. In particular, we prove some theorems that (a) cover special cases of the $\mathbf{T}$ matrix leading to X^2 or $\widehat{X}^2$, and (b) show that $L_{\kappa_2} \le L_{\kappa_1}$ and $M_{\kappa_2} \le M_{\kappa_1}$ when κ_2 is a summary reduction of κ_1 and Conditions T and D (for composite nulls) are satisfied. Proofs are given in Appendix A.3. Theorems 5 and 7 are used for the power comparisons in Section 3. Throughout this subsection, $\mathbf{I}_m$ denotes an $m \times m$ identity matrix.

Theorem 1. *Let $\mathbf{T}_\kappa = (\mathbf{0}\ \mathbf{I}_{C-1})$ be a $(C-1) \times C$ matrix. Then $L_\kappa = X^2$ and $M_\kappa = \widehat{X}^2$ if ML estimation is used.*

Thus, our first result states that when the summary statistics $\hat{\boldsymbol{\kappa}}$ consists of the set of all C cell proportions except one, $L_\kappa = X^2$ and, for the MLE, $M_\kappa = \widehat{X}^2$. Our next result states that if two sets of statistics, $\boldsymbol{\kappa}_1$ and $\boldsymbol{\kappa}_2$, are linearly related and their relationship is one-to-one, then $L_{\kappa_2} = L_{\kappa_1}$ and $M_{\kappa_2} = M_{\kappa_1}$.

Theorem 2. *Let $\mathbf{T}_{\kappa_1}$ and $\mathbf{T}_{\kappa_2}$ be $s \times C$ matrices satisfying Condition T. If there exists an $s \times s$ invertible matrix $\mathbf{B}$ and an $s \times 1$ vector $\boldsymbol{\beta}$ such that $\boldsymbol{\kappa}_2 = \mathbf{B}\boldsymbol{\kappa}_1 + \boldsymbol{\beta}$ and $\mathbf{T}_{\kappa_2}\boldsymbol{\Gamma}\mathbf{T}'_{\kappa_2} = \mathbf{B}\mathbf{T}_{\kappa_1}\boldsymbol{\Gamma}\mathbf{T}'_{\kappa_1}\mathbf{B}'$, then $L_{\kappa_2} = L_{\kappa_1}$. If further, $\boldsymbol{\Delta}_{\kappa_2} = \mathbf{B}\boldsymbol{\Delta}_{\kappa_1}$ and Condition D is satisfied for both $\mathbf{T}_{\kappa_j}$, then $M_{\kappa_2} = M_{\kappa_1}$.*

Together, Theorems 1 and 2 imply Theorem 3, which states that any L_κ and M_κ based on a proper set of statistics $\boldsymbol{\kappa}$ of dimension $C-1$ equals X^2 and $\widehat{X}^2$, respectively.

Theorem 3. *Let $\mathbf{T}_\kappa$ be a $(C-1) \times C$ matrix satisfying Condition T. Then $L_\kappa = X^2$. If in addition Condition D is satisfied by $\mathbf{T}_\kappa$, then $M_\kappa = \widehat{X}^2$ for the MLE.*

For the remaining theorems that compare two different L_κ or M_κ, the following are assumed: the vector $\boldsymbol{\kappa}_2 = \mathbf{T}_{\kappa_2}\boldsymbol{\pi} = \mathbf{T}_{21}\boldsymbol{\kappa}_1$ is a reduction of $\boldsymbol{\kappa}_1 = \mathbf{T}_{\kappa_1}\boldsymbol{\pi}$, where $\mathbf{T}_{\kappa_1}$ and $\mathbf{T}_{\kappa_2}$ satisfy Condition T, $\mathbf{T}_{\kappa_j}$ has row dimension s_j, $j = 1, 2$, with $s_1 > s_2$, and $\mathbf{T}_{21}$ is a $s_2 \times s_1$ matrix with full row rank.

Theorem 4. *The matrix*

$$\mathbf{L}_{21} = \boldsymbol{\Xi}_{\kappa_1}^{-1} - \mathbf{T}'_{21}\boldsymbol{\Xi}_{\kappa_2}^{-1}\mathbf{T}_{21} \tag{8}$$

is non-negative definite with rank $s_1 - s_2$.

Using the above, the next theorem states that $L_{\kappa_2} \le L_{\kappa_1}$ with equality possible.

Theorem 5.

$$L_{\kappa_1} - L_{\kappa_2} = (\hat{\boldsymbol{\kappa}}_1 - \boldsymbol{\kappa}_1)'\mathbf{L}_{21}(\hat{\boldsymbol{\kappa}}_1 - \boldsymbol{\kappa}_1) \ge 0.$$

Equality occurs if $\mathbf{L}_{21}(\hat{\boldsymbol{\kappa}}_1 - \boldsymbol{\kappa}_1) = \mathbf{0}$.

The next two theorems establish that $M_{\kappa_2} \le M_{\kappa_1}$ with equality possible.

Theorem 6. *Assume that Condition D is satisfied for both $\mathbf{T}_{\kappa_j}$. The matrix*

$$\mathbf{M}_{21} = \mathbf{U}_{\kappa_1} - \mathbf{T}'_{21}\mathbf{U}_{\kappa_2}\mathbf{T}_{21} \tag{9}$$

is non-negative definite with rank $s_1 - s_2$.

Theorem 7.

$$M_{\kappa_1} - M_{\kappa_2} = \big(\hat{\boldsymbol{\kappa}}_1 - \boldsymbol{\kappa}_1(\widehat{\boldsymbol{\theta}})\big)'\mathbf{M}_{21}\big(\hat{\boldsymbol{\kappa}}_1 - \boldsymbol{\kappa}_1(\widehat{\boldsymbol{\theta}})\big) \geq 0.$$

Equality occurs if $\mathbf{M}_{21}(\hat{\boldsymbol{\kappa}}_1 - \boldsymbol{\kappa}_1(\widehat{\boldsymbol{\theta}})) = \mathbf{0}$.

Our results imply the following corollary.

Corollary 8. *If* $\mathbf{T}_\kappa$ *satisfies Condition T, then* $L_\kappa \leq X^2$. *If* $\mathbf{T}_\kappa$ *satisfies Conditions T and D, then* $M_\kappa \leq \widehat{X}^2$.

3. Asymptotic Power under a Sequence of Local Alternatives

In this section, we apply Theorems 4 and 6 to obtain results to compare the noncentrality parameters of the asymptotic distributions $L_{\kappa_1}, L_{\kappa_2}$ and $M_{\kappa_1}, M_{\kappa_2}$ when κ_2 is a reduction of κ_1, assuming there is a sequence of local alternatives indexed by the sample size N. Also, in Section 3.1 we state some results on the form of the limiting direction $\boldsymbol{\delta}$ of the sequence of local alternatives, when there is a nesting of the null model into a larger parametric model. Then in Section 3.2, a condition is established for the limiting direction $\boldsymbol{\delta}$. A reference for the noncentral chi-square distribution under sequence of local alternatives is Bishop, Fienberg, and Holland (1975).

For a simple null hypothesis with $H_0 : \boldsymbol{\pi} = \boldsymbol{\pi}_0$, a sequence of local alternatives is H_{1N} : $\boldsymbol{\pi}_N = \boldsymbol{\pi}_0 + \boldsymbol{\delta}/\sqrt{N}$, where $\boldsymbol{\delta}$ is a directional vector such that $\boldsymbol{\delta}'\mathbf{1}_C = 0$. With $\boldsymbol{\delta}_\kappa = \mathbf{T}_\kappa\boldsymbol{\delta}$, the limiting nonnull distribution of L_κ is noncentral $\chi^2_{s_\kappa}$ with noncentrality parameter $\text{ncp}(L_\kappa) = \boldsymbol{\delta}'_\kappa \boldsymbol{\Xi}_\kappa \boldsymbol{\delta}_\kappa = \boldsymbol{\delta}'\mathbf{T}'_\kappa \boldsymbol{\Xi}_\kappa \mathbf{T}_\kappa \boldsymbol{\delta}$, where s_κ is the dimension of $\boldsymbol{\kappa}$. Under the assumptions of Theorem 4, the difference of two noncentrality parameters for L_{κ_1} versus L_{κ_2}, is

$$\text{ncp}(L_{\kappa_1}) - \text{ncp}(L_{\kappa_2}) = \boldsymbol{\delta}'_{\kappa_1} \boldsymbol{\Xi}^{-1}_{\kappa_1} \boldsymbol{\delta}_{\kappa_1} - \boldsymbol{\delta}'_{\kappa_1} \mathbf{T}'_{21} \boldsymbol{\Xi}^{-1}_{\kappa_2} \mathbf{T}_{21} \boldsymbol{\delta}_{\kappa_1} = \boldsymbol{\delta}'_{\kappa_1} \mathbf{L}_{21} \boldsymbol{\delta}_{\kappa_1},$$

with $\mathbf{L}_{21}$ as in (8). Theorem 4 implies that $\text{ncp}(L_{\kappa_1}) - \text{ncp}(L_{\kappa_2}) \geq 0$ with equality possible if $\mathbf{L}_{21}\boldsymbol{\delta}_{\kappa_1} = \mathbf{0}$.

For a composite null hypothesis $H_0 : \boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\theta})$, we assume that the sequence of local alternatives $\{\boldsymbol{\pi}_N\}$ gets closer to $\{\boldsymbol{\pi}(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$ at a rate $N^{-1/2}$. More specifically, let $\widehat{\boldsymbol{\theta}}_N$ be the parameter such that $\boldsymbol{\pi}(\widehat{\boldsymbol{\theta}}_N)$ is closest to $\boldsymbol{\pi}_N = (\pi_{Nc})$ in Kullback–Leibler divergence (see White, 1982), i.e., $\widehat{\boldsymbol{\theta}}_N$ maximizes $L(\boldsymbol{\theta}) = \sum_{c \in C} \pi_{Nc} \log \pi_c(\boldsymbol{\theta})$ over $\theta \in \Theta$. Further suppose that $\widehat{\boldsymbol{\theta}}_N$ approaches a $\boldsymbol{\theta}_0 \in \Theta$ at a rate $N^{-1/2}$. We define $\boldsymbol{\delta} = \lim_{N\to\infty} \sqrt{N}\,[\boldsymbol{\pi}_N - \boldsymbol{\pi}(\widehat{\boldsymbol{\theta}}_N)]$, so that $\boldsymbol{\delta}$ is the local direction of the sequence of alternatives. The vector $\boldsymbol{\delta}$ satisfies $\boldsymbol{\delta}'\mathbf{1}_C = 0$ and another condition (introduced later as Equation (11) in Section 3.2). With $\boldsymbol{\delta}_\kappa = \mathbf{T}_\kappa \boldsymbol{\delta}$, the limiting nonnull distribution of M_κ is noncentral $\chi^2_{s_\kappa - q}$ with noncentrality parameter $\text{ncp}(M_\kappa) = \boldsymbol{\delta}'_\kappa \mathbf{U}_\kappa \boldsymbol{\delta}_\kappa = \boldsymbol{\delta}'\mathbf{T}'_\kappa \mathbf{U}_\kappa \mathbf{T}_\kappa \boldsymbol{\delta}$. Under the assumptions of Theorem 6, the difference of two noncentrality parameters for M_{κ_1} versus M_{κ_2} is

$$\text{ncp}(M_{\kappa_1}) - \text{ncp}(M_{\kappa_2}) = \boldsymbol{\delta}'_{\kappa_1} [\mathbf{U}_{\kappa_1} - \mathbf{T}'_{21} \mathbf{U}_{\kappa_2} \mathbf{T}_{21}] \boldsymbol{\delta}_{\kappa_1} = \boldsymbol{\delta}'_{\kappa_1} \mathbf{M}_{21} \boldsymbol{\delta}_{\kappa_1},$$

with $\mathbf{M}_{21}$ as in (9). Theorem 6 implies that $\text{ncp}(M_{\kappa_1}) - \text{ncp}(M_{\kappa_2}) \geq 0$ with equality possible if $\mathbf{M}_{21}\boldsymbol{\delta}_{\kappa_1} = \mathbf{0}$.

The comments below are for the M_κ family of test statistics, but similar comments could be made for the L_κ family. For M_κ, based on $\mathbf{T}_\kappa$ with row dimension $q < s < C - 1$, it is possible to have a direction $\boldsymbol{\delta}$ from the null, with $\boldsymbol{\pi}$ not belonging to $\{\boldsymbol{\pi}(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$, but with $\boldsymbol{\kappa}$ belonging to $\{\boldsymbol{\kappa}(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$. In this case M_κ has power equal to the size of the test for this direction $\boldsymbol{\delta}$. If $\boldsymbol{\delta}$ is such that $\text{ncp}(M_{\kappa_1}) - \text{ncp}(M_{\kappa_2}) = 0$, then M_{κ_2} with fewer degrees of freedom is more powerful in this direction. This follows because if $\chi^2_{1-\alpha,\nu}$ is the upper α quantile of the χ^2_ν distribution and $F_{\chi^2}(\cdot;\nu,\lambda)$ is the noncentral chi-square distribution function with ν degrees of freedom and noncentrality parameter λ, then $1 - F_{\chi^2}(\chi^2_{1-\alpha,\nu};\nu,\lambda)$ is decreasing as ν increases for a constant λ. More generally, M_{κ_2} could be more powerful than M_{κ_1} in local directions $\boldsymbol{\delta}$ such that $\text{ncp}(M_{\kappa_2})/\,\text{ncp}(M_{\kappa_1})$ is sufficiently large (e.g., exceeding 0.9). Hence, it would not surprising to find M_κ statistics that are more powerful than $\widehat{X}^2$ over a variety of local directions.

Below in Section 3.1, we say more about the calculation of $\boldsymbol{\delta}$ for M_κ for local directions arising from nesting $\{\boldsymbol{\pi}(\boldsymbol{\theta})\}$ in a larger parameter family. In Section 4, these results are used in power comparisons for some multinomial goodness-of-fit situations.

3.1. Sequence of Local Alternatives: Derivation of $\boldsymbol{\delta}$

We derive $\boldsymbol{\delta}$ under a sequence of local alternatives that follows from the special case of embedding $\{\boldsymbol{\pi}(\boldsymbol{\theta})\}$ into a larger parametric family $\{\boldsymbol{\pi}(\boldsymbol{\theta},\boldsymbol{\eta})\}$ with another parameter $\boldsymbol{\eta}$. Suppose $\boldsymbol{\eta}$ is r-dimensional, and the composite null corresponds to a specified second parameter $\boldsymbol{\eta} = \boldsymbol{\eta}_0$ and unspecified $\boldsymbol{\theta}$. The derivation in this case is similar to that given in Maydeu-Olivares and Joe (2005) (in which the derivation was for the case of a null hypothesis consisting of a subset of the parameters set equal), so we will just state the final result and interpret it. Note that although the null hypothesis allows an arbitrary $\boldsymbol{\theta}$, the power will generally depend on the value of $\boldsymbol{\theta}$, which we denote as $\boldsymbol{\theta}_0$.

For a nested model, consider the sequence of local alternatives with parameters $(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0 + N^{-1/2}\boldsymbol{\epsilon})$ where $\boldsymbol{\epsilon}$ is an r-dimensional column vector. Then the limit of $\sqrt{N}[\boldsymbol{\pi}(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0 + N^{-1/2}\boldsymbol{\epsilon}) - \boldsymbol{\pi}(\widehat{\boldsymbol{\theta}}_N, \boldsymbol{\eta}_0)]$ leads to

$$\boldsymbol{\delta} = \frac{\partial\boldsymbol{\pi}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0)}{\partial\boldsymbol{\eta}'}\boldsymbol{\epsilon} - \frac{\partial\boldsymbol{\pi}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0)}{\partial\boldsymbol{\theta}'}\cdot\left[\mathcal{I}(\boldsymbol{\theta}_0)\right]^{-1}\sum_{c\in\mathcal{C}}\frac{\log\pi_c(\boldsymbol{\theta}_0,\eta_0)}{\partial\boldsymbol{\theta}}\cdot\frac{\partial\pi_c(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0)}{\partial\boldsymbol{\eta}'}\boldsymbol{\epsilon}, \tag{10}$$

where

$$\mathcal{I}(\boldsymbol{\theta}) = \sum_{c\in\mathcal{C}}\frac{\partial\pi_c(\boldsymbol{\theta},\boldsymbol{\eta}_0)}{\partial\boldsymbol{\theta}}\frac{\partial\pi_c(\boldsymbol{\theta},\boldsymbol{\eta}_0)}{\partial\boldsymbol{\theta}'}\Big/\pi_c(\boldsymbol{\theta},\boldsymbol{\eta}_0)$$

is the Fisher information matrix for the null model, and $\widehat{\boldsymbol{\theta}}_N$ is defined above. The expression (10) is a form that is useful for computations, especially when $\mathcal{C}$ is a multidimensional set. In matrix form, (10) can be written as

$$\boldsymbol{\delta} = \frac{\partial\boldsymbol{\pi}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0)}{\partial\boldsymbol{\eta}'}\boldsymbol{\epsilon} - \frac{\partial\boldsymbol{\pi}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0)}{\partial\boldsymbol{\theta}'}\left[\mathcal{I}(\boldsymbol{\theta}_0)\right]^{-1}\left[\frac{\partial\boldsymbol{\pi}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0)}{\partial\boldsymbol{\theta}'}\right]'\left[\text{diag}\big(\boldsymbol{\pi}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0)\big)\right]^{-1}\frac{\partial\boldsymbol{\pi}(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0)}{\partial\boldsymbol{\eta}'}\boldsymbol{\epsilon}.$$

To compute $\boldsymbol{\delta}$, what are needed are the first order derivatives of the probabilities with respect to (a) parameters in the null model and (b) additional parameters in the nonnull model that nests the null model.

3.2. Condition Satisfied by a Limiting Direction Vector $\boldsymbol{\delta}$ *for Composite Null*

We show that the $\boldsymbol{\delta}$ vector for a composite null satisfies another condition besides $\boldsymbol{\delta}'\mathbf{1}_C = 0$. This result is needed for generating random local directions $\boldsymbol{\delta}$ from the composite null hypothesis for comparing noncentrality parameters of different M_κ. That is, M_κ can be analyzed as an omnibus goodness-of-fit statistic without embedding $\{\pi(\boldsymbol{\theta})\}$ into a larger parametric family.

With the sample of proportions $\mathbf{p}_N$ from distribution $\boldsymbol{\pi}_N$, $\boldsymbol{\delta} = \boldsymbol{\delta}(\boldsymbol{\theta}_0)$ essentially comes from a limit of the expectation of $\sqrt{N}[\mathbf{p}_N - \boldsymbol{\pi}(\widehat{\boldsymbol{\theta}}_N)] = \sqrt{N}\,\mathbf{e}_N$, where $\widehat{\boldsymbol{\theta}}_N$ is the maximum likelihood estimate based on assuming the null model. Therefore, from a result in Appendix A.2,

$$\boldsymbol{\Delta}'(\widehat{\boldsymbol{\theta}}_N)\big[\mathbf{D}(\widehat{\boldsymbol{\theta}}_N)\big]^{-1}\mathbf{e}_N = 0,$$

and, as $N \to \infty$, $\boldsymbol{\delta}$ must satisfy

$$\boldsymbol{\Delta}'(\boldsymbol{\theta}_0)\big[\mathbf{D}(\boldsymbol{\theta}_0)\big]^{-1}\boldsymbol{\delta}(\boldsymbol{\theta}_0) = \mathbf{0}. \tag{11}$$

An alternative proof of the condition is based on the noncentrality parameter of $\widehat{X}^2$ and $M_{\kappa 0}$ with $\mathbf{T}_{\kappa 0} = (\mathbf{0}\ \mathbf{I}_{C-1})$. Because $\widehat{X}^2 = M_{\kappa 0}$ for this $\mathbf{T}_{\kappa 0}$, with $\mathbf{D} = \mathrm{diag}(\boldsymbol{\pi}(\boldsymbol{\theta}_0))$,

$$\mathrm{ncp}\big(\widehat{X}^2\big) = \boldsymbol{\delta}'\mathbf{D}^{-1}\boldsymbol{\delta} = \boldsymbol{\delta}'\mathbf{T}'_{\kappa 0}\mathbf{U}_{\kappa 0}\mathbf{T}_{\kappa 0}\boldsymbol{\delta} = \check{\boldsymbol{\delta}}'\mathbf{U}_{\kappa 0}\check{\boldsymbol{\delta}},$$

with the partitioning $\boldsymbol{\delta} = \binom{\delta_0}{\check{\boldsymbol{\delta}}}$. Using (5), this is the same as

$$\check{\boldsymbol{\delta}}'\boldsymbol{\Xi}_{\kappa 0}^{-1}\check{\boldsymbol{\delta}} - \check{\boldsymbol{\delta}}'\boldsymbol{\Xi}_{\kappa 0}^{-1}\boldsymbol{\Delta}_{\kappa 0}\big(\boldsymbol{\Delta}'_{\kappa 0}\boldsymbol{\Xi}_{\kappa 0}^{-1}\boldsymbol{\Delta}_{\kappa 0}\big)^{-1}\boldsymbol{\Delta}'_{\kappa 0}\boldsymbol{\Xi}_{\kappa 0}^{-1}\check{\boldsymbol{\delta}}. \tag{12}$$

From the proof of Theorem 1, $\mathbf{1}'\check{\boldsymbol{\delta}} = -\delta_0$ and $\boldsymbol{\delta}'\mathbf{D}^{-1}\boldsymbol{\delta} = \check{\boldsymbol{\delta}}'\boldsymbol{\Xi}_{\kappa 0}^{-1}\check{\boldsymbol{\delta}}$. Therefore, the second term in (12) must be zero or

$$\boldsymbol{\Delta}'_{\kappa 0}\boldsymbol{\Xi}_{\kappa 0}^{-1}\check{\boldsymbol{\delta}} = \boldsymbol{\Delta}\mathbf{D}^{-1}\boldsymbol{\delta} = \mathbf{0},$$

evaluated at $\boldsymbol{\theta}_0 \in \Theta$.

4. Examples

The primary application of the theory presented above is for developing new goodness-of-fit statistics, particularly for sparse high-dimensional tables for which summaries are based on low-order margins. For high-dimensional tables, there are generally no good alternatives for goodness-of-fit besides quadratic-form statistics in summary statistics (see Cai, Maydeu-Olivares, Coffman, & Thissen, 2006; Mavridis, Moustaki, & Knott, 2007; Reiser, 1996), but for univariate data there are other classes of statistics that could be considered. Applications of this theory to models for high-dimensional tables will be presented in separate reports.

Here, instead, we present two simple examples that do not involve sparse data to illustrate the theory. Both examples involve a composite null. In each of the examples, we (1) describe the model of interest (i.e., the null model), (2) describe the alternative models of interest, (3) construct a series of test statistics within the M_κ family that could be used to test the null with the alternative models in mind (the set always includes X^2 for comparison), and (4) compute the asymptotic power of each test statistic under consideration for each alternative of interest to determine which test statistic yields higher power.

The first example involves a unidimensional multinomial. The model of interest is a truncated Poisson model. The alternatives of interest are (a) the zero inflated Poisson model, and (b) the generalized Poisson model allowing for overdispersion. The second example involves a multidimensional multinomial. The null model is the log linear counterpart of Rasch's (1960) model with additional constraints leading to the exchangeability of the items. The alternative of interest is the same model without the exchangeability constraints. For this example, in addition to (1) to (4) above, we also (5) perform simulation studies to investigate the small sample null distributions of the statistics considered under conditions of increasing sparseness, and (6) fit the models of interest to the well known LSAT6 data of Bock and Lieberman (1970) to illustrate the theory.

4.1. Count Data: Poisson Versus Zero-Inflation and/or Overdispersion

In modeling count data using a Poisson model, often zero-inflation (greater presence of zeros) or overdispersion (greater variability) than expected under a Poisson model is encountered. In recent years, there has been a number of papers on models for count data with either zero-inflation or overdispersion relative to Poisson. For example, see Böckenholt (1999), Van Duijn and Jansen (1995), Lee, Wang, and Yau (2001), Joe and Zhu (2005) and references therein.

In this subsection, we examine the performance of various M_κ statistics in testing a composite null hypothesis of truncated Poisson. That is, we consider count data with heavy tails with some small counts so the categories in the tail need to be pooled or weighted if one is to do goodness-of-fit tests.

The null model is $\pi_y = e^{-\theta}y^\theta/y!$ for $y = 0,\dots,C-2$ and $\pi_{C-1} = \sum_{j=C-1}^{\infty} e^{-\theta} j^\theta/j!$. Power of several M_κ statistics is examined for three sets of local alternatives: (a) generalized Poisson, (b) zero-inflated Poisson, and (c) random $\boldsymbol{\delta}$'s in (10).

The generalized Poisson (GP) distribution (see Consul, 1989; Joe & Zhu, 2005) is a two-parameter family allowing for overdispersion. With parameters $\theta > 0$ and $\eta \ge 0$, the GP probability distribution is

$$f_{\mathrm{GP}}(y;\theta,\eta) = \theta(\theta+\eta y)^{y-1}e^{-\theta-\eta y}/y!, \quad y = 0, 1, \dots.$$

The parameter η for GP is the overdispersion parameter. For the truncated GP distribution, we use $\pi_y(\theta,\eta) = f_{\mathrm{GP}}(y;\theta,\eta)$ for $y = 0,\dots,C-2$ and $\pi_{C-1}(\theta,\eta) = \sum_{y=C-1}^{\infty} f_{\mathrm{GP}}(y;\theta,\eta)$.

The zero-inflated Poisson (ZIP) distribution (see Lee et al., 2001) is a mixture of a Poisson and a degenerate distribution at zero. With parameters $\theta > 0$ and $0 \le \eta \le 1$, the ZIP probability distribution is

$$f_{\mathrm{ZIP}}(y;\theta,\eta) = \begin{cases} \eta + (1-\eta)e^{-\theta}, & y = 0, \\ (1-\eta)\theta^y e^{-\theta}/y!, & y > 0. \end{cases}$$

The parameter η for ZIP is the extra probability mass at 0. For the truncated ZIP distribution, we use $\pi_y(\theta,\eta) = f_{\mathrm{ZIP}}(y;\theta,\eta)$ for $y = 0,\dots,C-2$ and $\pi_{C-1}(\theta,\eta) = \sum_{y=C-1}^{\infty} f_{\mathrm{ZIP}}(y;\theta,\eta)$. Results similar to those presented here were obtained for the negative binomial distribution, another overdispersed Poisson distribution.

In both cases, $\eta = \eta_0 = 0$ leads to the Poisson distribution, so that the null model is nested within alternatives (a) and (b) above as specified in (10). For (c), random $\boldsymbol{\delta}$'s based on a perturbation of a Geometric(p) distribution with $1/p$ random in the interval $(1, 2.2)$ were used. For random $\boldsymbol{\delta}$, from Section 3.2, conditions to satisfy are $\sum_{c=0}^{C-1}\delta_c = 0$ and $\sum_{c=0}^{C-1}[\partial \log \pi_c(\theta)/\partial\theta]\delta_c = 0$ (latter from $\boldsymbol{\Delta}'\mathbf{D}^{-1}\boldsymbol{\delta} = \mathbf{0}$ with $q = 1$). So for the random $\boldsymbol{\delta}$, we start with random δ_c for $c \ge 2$, and then solve for δ_0, δ_1 (these are not too much different for $\theta_0 \in [1, 5]$).

For the GP probability distribution and $0 \le y < C-1$, functions needed for the calculation of $\boldsymbol{\delta}$ are:

(i) $\ell(\theta,\eta;y) = \log \pi_y(\theta,\eta) = \log\theta + (y-1)\log(\theta+\eta y) - \theta - \eta y - \log(y!)$.
(ii) $\frac{\partial \ell}{\partial\theta}(\theta,\eta;y) = \theta^{-1} + (y-1)/(\theta+\eta y) - 1$, which becomes $y/\theta - 1$ for $\eta = 0$.
(iii) $\frac{\partial \ell}{\partial\eta}(\theta,\eta;y) = y(y-1)/(\theta+\eta y) - y$, which becomes $y(y-1)/\theta - y$ for $\eta = 0$.

For the ZIP probability distribution and $0 \le y < C-1$, functions needed for the calculation of $\boldsymbol{\delta}$ are

$$\ell(\theta,\eta;y) = \log \pi_y(\theta,\eta) = \begin{cases} \log[\eta + (1-\eta)e^{-\theta}] & \text{for } y = 0, \\ \log(1-\eta) + y\log\theta - \theta - \log(y!) & \text{for } y > 0; \end{cases}$$

$$\frac{\partial \ell}{\partial \theta}(\theta,\eta;y) = \begin{cases} -(1-\eta)e^{-\theta}/[\eta + (1-\eta)e^{-\theta}] & \text{for } y = 0, \\ y/\theta - 1 & \text{for } y > 0; \end{cases}$$

$$\frac{\partial \ell}{\partial \eta}(\theta,\eta;y) = \begin{cases} [1-e^{-\theta}]/[\eta + (1-\eta)e^{-\theta}] & \text{for } y = 0, \\ -(1-\eta)^{-1} & \text{for } y > 0. \end{cases}$$

Finally, for the Poisson null model, the derivatives of π_y for $0 \le y < C-1$ needed for the power computations are $\partial\pi_y/\partial\theta = \pi_y[\partial\ell(\theta,\eta)/\partial\theta]$ and $\partial\pi_y/\partial\eta = \pi_y[\partial\ell(\theta,\eta)/\partial\eta]$. The derivatives for π_{C-1} are obtained from one minus the sum of the other probabilities.

For the results reported in this example, we assume that the sample size is such that we can consider $C = 15$ categories $(0, 1, \dots, 13, 14+)$, that is, $\mathcal{C} = \{0, 1, \dots, 14\}$ with the last category indicating truncation to 14. With the above categorization, we are not expecting a large expected count, and we will do some power comparisons for θ in the range of 0 to 5. The null probabilities of exceeding 10 for $\theta \in [1, 5]$ are virtually zero, but these probabilities are not negligible with some overdispersion.

The $\mathbf{T}_{\kappa_i}$ $(i = 1, \dots, 7)$ that will be compared are based on the following operations.

(1) Pooling categories 10–14; categories for $1, \dots, 9$ separate; $s = 10$.
(2) Pooling 10–14, 7–9; $1, \dots, 6$ separate; $s = 8$.
(3) Pooling 10–14, 7–9, 4–6; $1, \dots, 3$ separate; $s = 6$.
(4) Pooling 10–14, 7–9, 4–6, 1–3; $s = 4$.
(5) Pooling 3–14; 1, 2 separate; $s = 3$.
(6) Pooling 2–14, 1 separate; $s = 2$.
(7) Category 0, mean, second moment; $s = 3$.

For example,

$$\mathbf{T}_{\kappa_4} = \begin{pmatrix} 0&1&1&1&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&1&1&1&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&1&1&1&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&1&1&1&1&1 \end{pmatrix},$$

$$\mathbf{T}_{\kappa_7} = \begin{pmatrix} 1&0&0&0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&1&2&3&4&5&6&7&8&9&10&11&12&13&14 \\ 0&1&4&9&16&25&36&49&64&81&100&121&144&156&196 \end{pmatrix}.$$

A priori, we think that $\mathbf{T}_{\kappa_7}$ and some of the others $\mathbf{T}_{\kappa_i}$ with poolings of neighboring categories might provide good power to detect a heavier tail than Poisson (overdispersion) and zero inflation (larger zero count than expected under a Poisson model).

Table 1 summarizes some results on ratios of noncentrality parameters and power from the three sets of local alternatives, (a), (b), and (c) above. In order that the values of power are not all close to 0.05 or 1, we scale the ϵ in (10) so that the power is 0.5 for $\widehat{X}^2$ with a 0.05 significance level. For (c), the power depends on the actual $\boldsymbol{\delta}$ and Table 1 shows the average over 10 different random $\boldsymbol{\delta}$'s. It follows from results in Section 3 that if $\text{ncp}(M_\kappa)/\text{ncp}(\widehat{X}^2) = 1$, then M_κ is more powerful than $\widehat{X}^2$ in the given local direction $\boldsymbol{\delta}$, and if this ratio is large, the same conclusion can be expected to hold.

TABLE 1.
Ratio of noncentrality parameters and power relative to $\widehat{X}^2$ for M_{κ_i}, $i = 1, \ldots, 7$.

Ratio of ncp(M_{κ_i}) with ncp($\widehat{X}^2$)								
Alternative	θ_0	ncpr1	ncpr2	ncpr3	ncpr4	ncpr5	ncpr6	ncpr7
GP	1	1.000	0.999	0.911	0.770	0.662	0.291	1.000
GP	2	1.000	0.982	0.817	0.735	0.321	0.094	1.000
GP	3	0.994	0.938	0.852	0.674	0.143	0.032	1.000
GP	4	0.967	0.891	0.879	0.710	0.061	0.011	1.000
GP	5	0.902	0.843	0.836	0.719	0.026	0.004	0.999
ZIP	1	1.000	1.000	0.994	0.458	0.962	0.810	1.000
ZIP	2	1.000	1.000	0.986	0.848	0.925	0.757	1.000
ZIP	3	1.000	0.999	0.988	0.958	0.914	0.755	1.000
ZIP	4	1.000	0.999	0.993	0.987	0.918	0.772	1.000
ZIP	5	1.000	0.999	0.996	0.995	0.930	0.797	1.000
random	1	0.64	0.63	0.63	0.05	0.59	0.47	0.45
random	2	0.99	0.99	0.99	0.30	0.97	0.87	0.71
random	3	1.00	1.00	1.00	0.43	0.99	0.89	0.66
random	4	1.00	1.00	1.00	0.51	0.99	0.90	0.63
random	5	1.00	1.00	1.00	0.57	0.99	0.91	0.63
Power of M_{κ_i} when ϵ chosen to achieve power of 0.5 for $\widehat{X}^2$ at 0.05 significance level								
Alternative	θ_0	power1	power2	power3	power4	power5	power6	power7
GP	1	0.576	0.626	0.641	0.647	0.640	0.407	0.825
GP	2	0.576	0.617	0.587	0.624	0.348	0.166	0.825
GP	3	0.573	0.593	0.608	0.583	0.174	0.089	0.825
GP	4	0.559	0.567	0.623	0.607	0.100	0.063	0.825
GP	5	0.523	0.539	0.598	0.613	0.070	0.055	0.824
ZIP	1	0.576	0.626	0.686	0.415	0.809	0.821	0.825
ZIP	2	0.576	0.626	0.682	0.694	0.793	0.795	0.825
ZIP	3	0.576	0.626	0.683	0.752	0.787	0.794	0.825
ZIP	4	0.576	0.626	0.685	0.766	0.790	0.803	0.825
ZIP	5	0.575	0.625	0.687	0.769	0.794	0.815	0.824
random	1	0.38	0.41	0.45	0.08	0.52	0.51	0.43
random	2	0.57	0.62	0.68	0.29	0.81	0.83	0.65
random	3	0.58	0.63	0.69	0.39	0.82	0.85	0.62
random	4	0.58	0.63	0.69	0.46	0.82	0.85	0.61
random	5	0.58	0.63	0.69	0.50	0.82	0.86	0.61

Note in Table 1 that M_{κ_5} and M_{κ_6} pool too many categories together to be useful for overdispersion, but they do fine for zero inflation. Over the two given local directions of overdispersion and zero inflation, M_{κ_7} has the best power. Similar results were obtained for the negative binomial distribution, another overdispersed Poisson distribution. Thus, our results suggest that when fitting a (possibly truncated) Poisson model, the M_{κ_7} is the statistic of choice if the alternatives of interest are overdispersion and zero-inflation.

These results should not be taken to imply that M_{κ_7} is the best statistic over all possible alternatives. Indeed, we included condition (c) precisely to show that this is not the case. Table 1 shows that M_{κ_7} does less well over random local directions, and indeed its power would be much lower for directions corresponding to bimodal count distributions with a variance to mean ratio of 1 and the same ratio of mean to zero proportion as for Poisson distributions. In closing this example, note that on average over random local directions, none of the M_{κ_i} dominates $\widehat{X}^2$ for all θ_0 in the interval $[1, 5]$.

4.2. *Binary Item Response Data: Testing for Exchangeability*

In this section, we apply the theory to a log linear item response theory (IRT) model for binary data. The null model under consideration is so highly constrained that it is not of interest in applications. However, it was chosen because (a) its simplicity allows us to illustrate the theory rather easily, and (b) it can not be tested with the statistics based on low-order margins proposed in Maydeu-Olivares and Joe (2005).

Consider an n-dimensional binary random vector $\mathbf{Y} = (Y_1, \ldots, Y_n)$ with joint distribution:

$$\pi_{\mathbf{y}} = \Pr(Y_k = y_k, k = 1, \ldots, n), \quad \mathbf{y} = (y_1, \ldots, y_n), \ y_k \in \{0, 1\}. \tag{13}$$

For a parametric model with parameter vector $\boldsymbol{\theta}$, we write $\pi_{\mathbf{y}}(\boldsymbol{\theta})$ for an individual probability and $\boldsymbol{\pi}(\boldsymbol{\theta})$ for the vector of $C = 2^n$ joint probabilities (with the $\mathbf{y}$'s ordered lexicographically).

Consider the following parametric (exponential-family) model

$$\pi_{\mathbf{y}} = \alpha^{-1} \exp\{\mu_{\mathbf{y}}(\boldsymbol{\theta})\}, \tag{14}$$

$$\mu_{\mathbf{y}}(\boldsymbol{\theta}) = \gamma_{y_1+\cdots+y_n} + \sum_j y_j \sigma_j, \quad \boldsymbol{\theta} = (\gamma_1, \ldots, \gamma_n, \sigma_1, \ldots, \sigma_n), \tag{15}$$

where $\alpha = \sum_{\mathbf{y}} \exp\{\mu_{\mathbf{y}}(\boldsymbol{\theta})\}$. Because of the normalizing via α, we assume $\gamma_0 = 0$ without loss of generality. Also, for identification, we take $\gamma_n = 0$ since $\sigma_k' = \sigma_k + \gamma_n/n$ $(k = 1, \ldots, n)$, $\gamma_k' = \gamma_k - \gamma_n/k$ $(k = 1, \ldots, n-1)$ and $\gamma_n' = 0$ lead to exactly the same probabilities. Therefore, for the null model (14) with (15) and $n = 3$, the binary response patterns (in lexicographic order) and $\mu_{\mathbf{y}}(\boldsymbol{\theta})$ are given by

$$\begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \qquad \begin{pmatrix} 0 \\ \gamma_1 + \sigma_1 \\ \gamma_1 + \sigma_2 \\ \gamma_1 + \sigma_3 \\ \gamma_2 + \sigma_1 + \sigma_2 \\ \gamma_2 + \sigma_1 + \sigma_3 \\ \gamma_2 + \sigma_2 + \sigma_3 \\ \sigma_1 + \sigma_2 + \sigma_3 \end{pmatrix}.$$

Tjur (1982) showed that this model is equivalent to Rasch's (1960) latent trait model (provided the log of the gamma parameters be the first moments of a positive random variable, see Cressie & Holland, 1983).

The null model we consider in this example assumes that items are exchangeable. This is obtained by introducing the constraint $\sigma_1 = \cdots = \sigma_n = \sigma$ in the above model, so that (15) becomes:

$$\mu_{\mathbf{y}}(\boldsymbol{\theta}) = \gamma_{\Sigma_j y_j} + \sigma \sum_j y_j, \quad \boldsymbol{\theta} = (\gamma_1, \ldots, \gamma_{n-1}, \sigma) \tag{16}$$

where $\boldsymbol{\theta}$ has dimension $q = n$.

We shall examine the performance of a series of M_κ statistics when fitting the null model given by Equations (14) and (16) when the alternative of interest is model (14) with (15). Different sets of statistics $\boldsymbol{\kappa}$ may be used for testing this null model. For instance, Maydeu-Olivares and Joe (2005) proposed testing multivariate binary models using the vector of joint moments of the multivariate Bernoulli distribution up to order $r \le n$. For summaries from low-order margins of (13), we let $\boldsymbol{\pi}_r = (\dot{\boldsymbol{\pi}}_1', \dots, \dot{\boldsymbol{\pi}}_r')'$, where $\dot{\boldsymbol{\pi}}_1' = (\dot{\pi}_1, \dots, \dot{\pi}_n)'$ consists of the $\mathrm{E}(Y_k) = \Pr(Y_k = 1)$, $\dot{\boldsymbol{\pi}}_2$ is the $\binom{n}{2}$-dimensional vector of bivariate non-central moments with elements $\mathrm{E}(Y_k Y_\ell) = \Pr(Y_k = 1, Y_\ell = 1) = \dot{\pi}_{k\ell}$, $k < \ell$, $\dot{\boldsymbol{\pi}}_3$ is the $\binom{n}{3}$-dimensional vector of trivariate noncentral moments with elements $\mathrm{E}(Y_{k_1} Y_{k_2} Y_{k_3}) = \Pr(Y_{k_1} = 1, Y_{k_2} = 1, Y_{k_3} = 1) = \dot{\pi}_{k_1 k_2 k_3}$, $k_1 < k_2 < k_3$ and so on. Yet, the M_κ statistics of Maydeu-Olivares and Joe (2005) based on $\boldsymbol{\pi}_r$ (for $r = 2, 3$) may not be used to test the null model, because Condition D is not satisfied. For these summary statistics, $\boldsymbol{\Delta}_\kappa$ is not of full rank. This is because under this null model $\mathrm{E}(Y_1) = \cdots = \mathrm{E}(Y_n)$, $\mathrm{E}(Y_1 Y_2) = \cdots = \mathrm{E}(Y_{n-1} Y_n)$, etc.

In discrete exponential family or log linear models (14),

$$\boldsymbol{\Delta} = \frac{\partial \boldsymbol{\pi}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = \boldsymbol{\Gamma} \mathbf{X}'$$

where $\mathbf{X}'$ is a $C \times q$ matrix such that the (i, j) element of $\mathbf{X}'$ is the coefficient of θ_j in $\mu_{\mathbf{y}_i}(\boldsymbol{\theta})$ in the ith probability in $\boldsymbol{\pi}$. $\mathbf{X}'$ has full column rank when the model is identified. Therefore, $\boldsymbol{\kappa}_x = \mathbf{X}\boldsymbol{\pi}$ provides a minimal set of statistics that identify the model. For the null model (14) with (16), $\mathbf{X}$ is an $n \times C$ matrix. In particular, for $n = 3$,

$$\mathbf{X} = \begin{pmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 2 & 2 & 2 & 3 \end{pmatrix}.$$

Thus, the summary statistics in $\boldsymbol{\kappa}_x$ for this null model are $\mathrm{E}(Y_+)$ and $\Pr(Y_+ = z)$, $z = 1, \dots, n-1$, where $Y_+ = \sum_{k=1}^n Y_k$ is the sum score. Testing may not proceed solely on $\boldsymbol{\kappa}_x$ as in this case there are zero degrees of freedom. However, $s' = s - q$ linear combinations, $\boldsymbol{\kappa}_a = \mathbf{T}_a \boldsymbol{\pi}$, can be added to $\boldsymbol{\kappa}_x$ such that $\mathbf{T}_{xa} = \binom{\mathbf{X}}{\mathbf{T}_a}$ satisfies Conditions T and D. Thus, a test statistic with s' degrees of freedom is obtained.

One way to proceed in constructing $\boldsymbol{\kappa}_a$ is to compute a basis for the null space of $\mathbf{X}$, say $\mathbf{N}_x$, that satisfies $\mathbf{N}_x \mathbf{X}' = \mathbf{0}$. For the null model (14) with (16) and $n = 3$,

$$\mathbf{N}_x = \begin{pmatrix} 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Then by construction, any subsets of rows of $\mathbf{N}_x$ can be added to $\mathbf{X}$ to yield a $\mathbf{T}_{xa}$ such that Conditions T and D are satisfied. Another way to proceed in constructing $\boldsymbol{\kappa}_a$ is to select simple summaries that are not redundant to $\boldsymbol{\kappa}_x$.

For the null model under consideration, any of the sets $\dot{\boldsymbol{\pi}}_1$ to $\dot{\boldsymbol{\pi}}_3$ is spanned by the row space of $\mathbf{N}_x$ provided the first element within each set is removed. Therefore, they constitute a suitable set $\boldsymbol{\kappa}_a$. Four statistics, $M_{\kappa_1}, \dots, M_{\kappa_4}$, obtained in this fashion were compared for testing (14) satisfying (16), with (14) satisfying (15) as the alternative. The statistics are based on summaries obtained by adding to $\boldsymbol{\kappa}_x$,

(1) $\dot{\boldsymbol{\pi}}_1$ without its first element. In this case, $s = n + (n-1) = 2n - 1$.
(2) $\dot{\boldsymbol{\pi}}_2$ without first element. In this case, $s = n + \binom{n}{2} - 1 = \frac{1}{2}(n^2 + 3n - 4)$.
(3) $\dot{\boldsymbol{\pi}}_3$ without first element. In this case, $s = n + \binom{n}{3} - 1 = (n^3 - 3n^2 + 8n - 6)/6$.
(4) $\dot{\boldsymbol{\pi}}_2$ and $\dot{\boldsymbol{\pi}}_3$ without their first elements. In this case, $s = n + \binom{n}{2} + \binom{n}{3} - 2 = (n^3 + 5n - 12)/6$.

The $\mathbf{T}_{\kappa_i}$ $(i = 1, \dots, 4)$ matrices and corresponding M_{κ_i} are based on the above.

TABLE 2.

Ratio of noncentrality parameters and power relative to $\widehat{X}^2$ for M_{κ_i}, $i = 1, \ldots, 4$; some examples for $n = 5$ and 8. For $n = 5$, $\boldsymbol{\gamma} = (-1, -1.3, -1.2, -0.9)$ and $\boldsymbol{\epsilon} \propto (-0.5, -0.25, 0, 0.25, 0.5)$. For $n = 8$, $\boldsymbol{\gamma} = (-1, -1.3, -1.2, -0.9, -0.8, -0.7, -0.6)$ and $\boldsymbol{\epsilon} \propto (-0.5, -0.3, -0.2 - 0.1, 0.1, 0.2, 0.3, 0.5)$. The multiplying factor for $\boldsymbol{\epsilon}$ was chosen so that the asymptotic power of X^2 is 0.5. σ_0 is the common σ for the null model (15).

Ratio of ncp(M_{κ_i}) with ncp($\widehat{X}^2$)					
$n = 5$	$2^n - 1 = 31$	$s = 9$	$s = 14$	$s = 14$	$s = 23$
n	σ_0	ncpr1	ncpr2	ncpr3	ncpr4
5	0.00	1.000	0.732	0.396	0.835
5	0.45	1.000	0.824	0.540	0.917
5	0.90	1.000	0.888	0.670	0.962
5	1.35	1.000	0.930	0.773	0.984
5	1.80	1.000	0.957	0.849	0.993
$n = 8$	$2^n - 1 = 255$	$s = 15$	$s = 35$	$s = 63$	$s = 90$
n	σ_0	ncpr1	ncpr2	ncpr3	ncpr4
8	0.00	1.000	0.875	0.624	0.961
8	0.45	1.000	0.918	0.724	0.983
8	0.90	1.000	0.945	0.803	0.993
8	1.35	1.000	0.963	0.864	0.997
8	1.80	1.000	0.976	0.909	0.999
Power of M_{κ_i}					
n	σ_0	power1	power2	power3	power4
5	0.00	0.859	0.566	0.306	0.489
5	0.45	0.859	0.629	0.421	0.537
5	0.90	0.859	0.670	0.521	0.564
5	1.35	0.859	0.695	0.595	0.576
5	1.80	0.859	0.710	0.645	0.581
8	0.00	0.999	0.946	0.635	0.795
8	0.45	0.999	0.957	0.727	0.808
8	0.90	0.999	0.963	0.789	0.814
8	1.35	0.999	0.967	0.830	0.816
8	1.80	0.999	0.970	0.856	0.817

Table 2 provides some representative results for asymptotic powers and noncentrality parameter ratios for $n = 5$ and 8 items. The patterns are similar for other choices of $\boldsymbol{\gamma}$, $\boldsymbol{\epsilon}$ and σ_0. Note that all four statistics generally have better asymptotic power than $\widehat{X}^2$, because of getting only a slighter smaller noncentrality parameter with fewer degrees of freedom. For the cases in Table 2, only M_{κ_1} and M_{κ_2} are more powerful than $\widehat{X}^2$ for all of the listed parameter vectors. Combining bivariate and trivariate information to $\boldsymbol{\kappa}_x$ to get M_{κ_4} leads to reduced power compared to when only bivariate information is added to $\boldsymbol{\kappa}_x$ to get M_{κ_2}, because not a lot is gained for the extra degrees of freedom.

Because the null model assumes exchangeable items, it makes sense that M_{κ_1}, which has a small value for s, has the most power for some alternatives. Generally, lack of exchangeability is most easily discovered with a few "summary" statistics concerning univariate margins. For an exchangeable null model with directional alternatives where univariate but not bivariate margins are all the same, then something like M_{κ_2} and M_{κ_3} should be more powerful.

However, test statistics should not only be chosen based on power. One important factor should be accuracy of empirical p-values to the reference asymptotic distribution. A simulation study was performed to investigate the performance of the small sample null distributions of

TABLE 3.
Maximum likelihood estimates and SEs for model (15) applied to LSAT6 dataset.

Parameter	Estimate	SE
γ_1	−0.99	0.51
γ_2	−1.33	0.37
γ_3	−1.21	0.25
γ_4	−0.85	0.14
σ_1	2.18	0.16
σ_2	0.44	0.14
σ_3	−0.32	0.13
σ_4	0.75	0.14
σ_5	1.54	0.14

TABLE 4.
M_κ statistics for model (16).

Statistic	Value	df	Value/df
$\widehat{X}^2$	575.1	26	22
M_{κ_1}	490.6	4	123
M_{κ_2}	503.0	9	56
M_{κ_3}	384.6	9	43
M_{κ_4}	553.0	18	31

the statistics considered. Consistent with the results reported in Maydeu-Olivares and Joe (2005, 2006), all four M_κ considered maintained their nominal rates as model size increased and sample size decreased (to 100 for $n = 5$ and to 250 for $n = 8$, for example). X^2 maintained its nominal rates only in nonsparse conditions. In very sparse conditions, only the statistics where low order marginal statistics are added to $\boldsymbol{\kappa}_x$ are well approximated by the asymptotic null distribution. Among them, $\boldsymbol{\kappa}_1$ performed best in the most extreme sparse conditions considered. Taking together the simulation results for the small sample behavior of the statistics and the power results, we conclude that among the M_κ considered, M_{κ_1} is the best choice for testing the null (14) with (16) when the alternative hypothesis of interest is (14) with (15).

In closing this section, we point out that in applications, more powerful statistics show higher statistics to degrees of freedom ratios. To illustrate this point, we shall use the well-known LSAT6 dataset of Bock and Lieberman (1970). These data consist of 1000 observations on five binary variables. Model (14) with (15) (i.e., the log-linear version of Rasch's model) fits these data well: $X^2 = 17.77$ on 22 degrees of freedom, $p = 0.72$. The MLEs and standard errors (SE's) for this example are shown in Table 3, and do not suggest exchangeability of the items. Indeed, model (14) with (16) assumes that items are exchangeable, and it fits rather poorly, as shown in Table 4; the MLEs are $(\hat{\gamma}_1, \hat{\gamma}_2, \hat{\gamma}_3, \hat{\gamma}_4, \hat{\sigma}) = (-0.63, -0.80, -0.69, -0.51, 0.92)$. The statistics $\widehat{X}^2$ and M_{κ_i} reported in this table are all large with extremely small p-values. Table 4 lists also the test statistics to degrees of freedom ratios ($df = s - 5$), to show that order of the ratios is comparable to the ordering of power in Table 3. That is, appropriately chosen M_κ statistics can have quadratic form values that are almost as large as $\widehat{X}^2$ but with much fewer degrees of freedom (Theorems 1 and 7). Thus, if in any given application several M_{κ_i} statistics have been applied and previous evidence (obtained via simulations) suggests that finite sample behavior of the statistics is closely approximated by asymptotic methods, the theory presented here suggests that the statistic M_{κ_i}, which is the highest after dividing by the degrees of freedom, is likely to be the most powerful.

5. Some Alternatives to the L_κ and M_κ Families of Statistics

The classical procedures for assessing goodness-of-fit in multinomial models via asymptotic methods are Pearson's X^2 statistic and the asymptotically equivalent likelihood ratio test G^2. However, the asymptotic approximations to the sampling distributions of these procedures break down when the variances of the residual cell proportions involve small cell probabilities. This problem may be overcome by pooling cells, because pooled cells must have higher expected probabilities.

Pooling cells amounts to performing a linear transformation on the residual cell proportions before the analysis. Indeed, the classical procedure of using X^2 with pooled cells amounts to using a quadratic form in summary statistic of the type $\boldsymbol{\kappa} = \mathbf{T}\boldsymbol{\pi}$, where $\mathbf{T}$ linearly maps the original C categories into a set of disjoint C' pooled categories. Because the categories of $\boldsymbol{\kappa}$ are disjoint, the asymptotic covariance matrix of $\sqrt{N}(\hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa})$ is simply $\boldsymbol{\Gamma}_\kappa = \mathbf{D}_\kappa - \boldsymbol{\kappa}\boldsymbol{\kappa}'$, where $\mathbf{D}_\kappa = \mathrm{diag}(\boldsymbol{\kappa})$. As a result, $\mathbf{D}_\kappa^{-1}$ is a generalized inverse for $\boldsymbol{\Gamma}_\kappa$, and for simple nulls the quadratic form $X^2 = N(\hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa})'\mathbf{D}_\kappa^{-1}(\hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa})$ is asymptotically $\chi^2_{C'-1}$. For composite nulls, if the pooling is done before data are seen, and if the maximum likelihood estimation is based on the C' pooled categories and not the original categories, then the asymptotic null distribution of the resulting Pearson statistic is $\chi^2_{C'-1-q}$.

Yet, there is a limit in the amount of pooling into disjoint categories that can be performed without distorting the purpose of the analysis. To overcome this problem, and for n-dimensional contingency tables, Maydeu-Olivares and Joe (2005, 2006) proposed pooling cells into their marginals. This amounts to pooling cells with overlap. The idea of goodness-of-fit tests based on merging neighboring categories with overlap has been studied in Hall (1985) for univariate distributions. Because the variance of a marginal residual of order r depends on expected marginal probabilities of order $\min(2r, n)$, the distribution of quadratic forms in low order marginal residuals are well approximated by asymptotic methods even in very large models, assuming these marginal probabilities are not too small.

In this paper, we have extended these quadratic forms of linear transforms of multinomial cell residuals by considering linear transforms that are not necessarily marginal residuals. Thus, we have provided general conditions for quadratic forms in linear transforms of the type $\boldsymbol{\kappa} = \mathbf{T}\boldsymbol{\pi}$, with $\mathbf{T}$ a fixed matrix, to be asymptotically chi-square. These results may be applied not only to multidimensional tables, but also to unidimensional ones, and include cases of pooling cells with overlap and/or into disjoint categories.

Related ideas are given in Glas (1988) and Glas and Verhelst (1989). Glas (1988) proposed a statistic, R_1, aimed at testing the one-parameter logistic model (i.e., a Rasch model with a normally distributed trait) for binary item responses $\mathbf{Y} = (Y_1, \ldots, Y_n)$. This model has n item-specific parameters and one random effect parameter for a total of $q = n + 1$ parameters. The summaries used in R_1 are linear functions of the $\pi_{\mathbf{y}}$ corresponding to the $2 + (n-1)n$ aggregate probabilities $\Pr(Y_+ = 0), \Pr(Y_+ = n)$, $\Pr(Y_+ = z, Y_j = 1)$ for $j = 1, \ldots, n$, $z = 1, \ldots, n-1$, where $Y_+ = Y_1 + \cdots + Y_n$. Because these summary statistics are disjoint, their asymptotic covariance matrix has a simple form.

Yet, it can be shown that R_1 can be rewritten as an M_κ statistic based on a $\boldsymbol{\kappa}$ of dimension $s = 1 + (n-1)n$. To do so, the last category for $\{Y_+ = n\}$ is omitted in order for $\boldsymbol{\Delta}_\kappa$ to be of full column rank. Then the M_κ has $s - q = 1 + (n-1)n - (n+1) = n(n-2)$ degrees of freedom.

In Glas and Verhelst (1989; see also Glas & Verhelst, 1995), the theory of Glas (1988) is extended by considering more general linear transforms of the $\pi_{\mathbf{y}}$'s that are similar to our $\mathbf{T}_\kappa$. The distinctions with our theory are the following: (a) their quadratic form matrix only depends on $\mathbf{T}_\kappa$ and the probability vector $\boldsymbol{\pi}$, (b) in contrast, there is a strong condition that depends on $\boldsymbol{\Delta}$ and $\mathbf{T}_\kappa$ required for an asymptotic chi-square distribution. That is, our theory works for more general $\boldsymbol{\kappa}$, but in contrast our weight matrix does not have the simple form of Glas and Verhelst (1989).

Glas and Verhelst (1989) statistic can be written as

$$N\big[\hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}(\widehat{\boldsymbol{\theta}})\big]'\big(\mathbf{T}_\kappa \mathbf{D}(\widehat{\boldsymbol{\theta}})\mathbf{T}_\kappa'\big)^{-}\big[\hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}(\widehat{\boldsymbol{\theta}})\big], \tag{17}$$

where $\mathbf{A}^-$ denotes a generalized inverse of a matrix $\mathbf{A}$, that is $\mathbf{A}^-$ satisfies $\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}$. For the MLE and under the conditions set forth by Glas and Verhelst (1989), this statistic is asymptotically chi-square with degrees of freedom equal to $\text{rank}(\mathbf{T}_\kappa \mathbf{D}(\widehat{\boldsymbol{\theta}})\mathbf{T}_\kappa') - q - 1$.

A related statistic for testing composite null hypotheses is

$$N\big[\hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}(\widehat{\boldsymbol{\theta}})\big]'\boldsymbol{\Sigma}_\kappa^{-}(\widehat{\boldsymbol{\theta}})\big[\hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}(\widehat{\boldsymbol{\theta}})\big], \tag{18}$$

where recall that $\boldsymbol{\Sigma}_\kappa$, given in (7), is the asymptotic covariance matrix of $\sqrt{N}[\hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}(\widehat{\boldsymbol{\theta}})]$. This quadratic form is asymptotically chi-square distributed with degrees of freedom less than or equal than the minimum of s and $C - q - 1$ (see Appendix A.2). Therefore, use of (18) requires determining the rank of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ analytically for the model of interest. This is usually feasible for simple models. Alternatively, the degrees of freedom available for testing may be determined empirically by evaluating the rank of $\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})$ numerically. If the model requires numerical integration to obtain the pattern probabilities, as in many IRT models, it may be difficult to determine the rank of $\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})$ numerically (for an example, see Maydeu-Olivares & Joe, 2008). Nevertheless, simulation results by Mavridis et al. (2007) suggest that even in this situation statistics of the type (18) work well in practice. Reiser (1996) used statistics of this type for testing IRT models for binary data using univariate and bivariate marginal residuals, Reiser and Lin (1999) used bivariate marginal residuals to test latent class models, and in Reiser (2008) an extension of the results of Reiser (1996) for binary data is introduced that allows the inclusion of marginals of any order where the statistic is decomposed into a sum of orthogonal components.

The results presented here are also related to previous work by Rayner and Best (1989) who considered the problem of obtaining score tests for smooth alternatives to multinomial models. For simple nulls $H_0 : \boldsymbol{\pi} = \boldsymbol{\pi}_0$, they considered embedding $\boldsymbol{\pi}_0$ within the parametric family

$$\pi_c(\boldsymbol{\eta}) = \big[B(\boldsymbol{\eta})\big]^{-1}\pi_{0c}\exp\{\eta_1 t_{1c} + \cdots + \eta_s t_{sc}\}, \quad c \in \mathcal{C},\ \boldsymbol{\eta} = (\eta_1, \ldots, \eta_s), \tag{19}$$

where $B(\boldsymbol{\eta})$ is a normalizing constant. For testing null hypotheses against the directional alternative (19) they proposed the class of statistics

$$S_\kappa = N(\mathbf{p} - \boldsymbol{\pi}_0)'\mathbf{T}_\kappa' \boldsymbol{\Xi}_\kappa^{-1}\mathbf{T}_\kappa(\mathbf{p} - \boldsymbol{\pi}_0), \tag{20}$$

and this is the same as L_κ given in (2). Thus, Rayner and Best (1989) showed that $L_\kappa = S_\kappa$ is the score test for testing $\boldsymbol{\eta} = \mathbf{0}$ when the null is embedded within the parametric family (19). Also, note that Theorem 5.1.2 in Rayner and Best (1989) has a special case of the first part of Theorem 3 with L_κ since their conditions for $\mathbf{T}_\kappa$ are (i) orthonormal rows and (ii) $\mathbf{1}_C'$ not in its row span.

For composite nulls, $H_0 : \boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\theta})$, Rayner and Best (1989) considered obtaining score tests for smooth alternatives to multinomial models by embedding $\boldsymbol{\pi}(\boldsymbol{\theta})$ within the parametric family

$$\pi_c(\boldsymbol{\theta}, \boldsymbol{\eta}) = \big[B(\boldsymbol{\theta}, \boldsymbol{\eta})\big]^{-1}\pi_c(\boldsymbol{\theta})\exp\big\{\eta_1 t_{1c}(\boldsymbol{\theta}) + \cdots + \eta_s t_{sc}(\boldsymbol{\theta})\big\}, \tag{21}$$

where $[B(\boldsymbol{\theta}, \boldsymbol{\eta})]^{-1}$ is again a normalizing constant and $\boldsymbol{\eta}$ is a vector of nuisance parameters. For testing composite nulls against the directional alternative (21) they proposed the class of statistics

$$\hat{S}_\kappa = N(\mathbf{p} - \widehat{\boldsymbol{\pi}})'\widehat{\mathbf{T}}_\kappa' \widehat{\boldsymbol{\Sigma}}_\kappa^{-1}\widehat{\mathbf{T}}_\kappa(\mathbf{p} - \widehat{\boldsymbol{\pi}}), \tag{22}$$

with $\widehat{\boldsymbol{\pi}} = \boldsymbol{\pi}(\widehat{\boldsymbol{\theta}})$, $\widehat{\mathbf{T}}_\kappa = \mathbf{T}_\kappa(\widehat{\boldsymbol{\theta}})$, and $\widehat{\boldsymbol{\Sigma}}_\kappa = \boldsymbol{\Sigma}_\kappa(\widehat{\boldsymbol{\theta}})$. $\hat{S}_\kappa$ is a score test for $H_0 : \boldsymbol{\eta} = \mathbf{0}$ for models embedded within (21).

Differences between their approach and ours are due to different motivations. Rayner and Best suggest choosing $\mathbf{T}_\kappa$, which may depend on model parameters, so that $\boldsymbol{\Sigma}_\kappa$ be of full rank. Conditions for $\boldsymbol{\Sigma}_\kappa$ to be nonsingular are given on pp. 111–115 of Rayner and Best (1989). When $\boldsymbol{\Sigma}_\kappa$ is not of full rank, they suggest employing a generalized inverse, in which case the rank of $\boldsymbol{\Sigma}_\kappa$ needs to be determined to obtain the degrees of freedom. In contrast, in our approach $\mathbf{T}_\kappa$ is a matrix of constants, we provide conditions for M_κ to be asymptotically chi-square with known degrees of freedom regardless of whether $\boldsymbol{\Sigma}_\kappa$ is of full rank, and our alternative is omnibus. Also, we emphasize selecting $\mathbf{T}_\kappa$ based on robustness to sparseness, and we provide theoretical results for choosing among members of M_κ based on power to detect alternatives of interest.

In closing this section, when omnibus alternatives are considered, and $\mathbf{T}_\kappa$ is a fixed matrix, not necessarily leading to a $\boldsymbol{\Sigma}_\kappa$ of full rank, Rayner and Best's approach leads to the family of statistics (18). Clearly, further research should investigate via simulation which weight matrix, $\boldsymbol{\Sigma}_\kappa^-$ or $\mathbf{U}_\kappa$, yields better results for power and null chi-square approximation in small samples. Our use of $\mathbf{U}_\kappa$ in the quadratic form is standard in covariance structure analysis (see Browne, 1984; Yuan & Bentler, 1997). Also, Theorems 6 and 7, which are used in the power comparisons in Section 3, depend on $\mathbf{U}_\kappa$ as weight matrix in the quadratic form. We have checked numerically for a simple case that these results do not hold with Moore–Penrose generalized inverses $\boldsymbol{\Sigma}_{\kappa_j}^-$ in place of $\mathbf{U}_{\kappa_j}$ for $j = 1, 2$. In addition, we have focused for ease of exposition on the maximum likelihood estimator. However, quadratic forms in summary statistics satisfying Conditions T and D are asymptotically chi-square more generally, for Fisher-consistent estimators. That is, it is straightforward to show that the asymptotic chi-square distribution of M_κ holds for any Fisher-consistent estimator (see Maydeu-Olivares & Joe, 2005). This is true as well for the statistic (18). Yet, (3) with (4) or (5) remains invariant to the choice of estimator. One implementation suits all Fisher-consistent estimators. In contrast, $\boldsymbol{\Sigma}_\kappa$ depends on the estimator used (for details see Maydeu-Olivares & Joe, 2005, 2008; Maydeu-Olivares, 2001, 2006) and as a result in implementing (18) each estimator needs its own programming. That is, the main advantages of using $\mathbf{U}_\kappa$ over the generalized inverse $\boldsymbol{\Sigma}_\kappa^-$ are: (a) known degrees of freedom, (b) theoretical results for power comparisons, and (c) one implementation suits any Fisher-consistent estimator.

6. Concluding Remarks

Maydeu-Olivares and Joe (2005, 2006) introduced classes of goodness-of-fit statistics for sparse multidimensional multinomials. They showed by simulation that very large models can be tested with their approach via asymptotic methods, even in extraordinarily sparse tables. Their approach is based on linear maps of the cell residuals taking advantage of the multidimensional structure of the data. Yet, their approach is limited in three important aspects: (1) it can not be used with unidimensional tables, (2) it can not be used for models for multidimensional tables that can not be identified from margins, (3) there is a computational limit in the size of the models for multidimensional tables that can be tested due to the need to store very large matrices. To overcome these limitations, in this paper we have described two general families of test statistics for multinomial data. L_κ is to be used for simple nulls (no parameters being estimated), and M_κ for composite nulls (estimated parameters). The theory provided here can be used to propose new test statistics for a null model that effectively overcome the limitations of their existing approach. Thus, in our first example, we applied the theory to a unidimensional multinomial. In our second example, we applied the theory to a multidimensional model that is not identified from low order (univariate, bivariate, ...) margins. In a separate report, we will apply the theory to propose test statistics that can overcome the computational limitations of current proposals for extremely large multidimensional models.

The test statistics described here "concentrate" the information available in the multinomial cells into some summaries so that the resulting statistic is better approximated in small samples by large sample theory methods. Also, when interest lies in testing the null with some competing models in mind, the summary statistics may be selected so that the test statistic is more powerful than classical tests such as X^2. In this respect, the present paper also provides completely general theory that explains why, for testing purposes, in most cases one wants to concentrate the information available as much as possible.

The theory and methods presented here are closely related to previous theory. Kendall and Stuart (1979, Chapter 30) provide a review of some of the relevant theory. Eubank (1997) explains that the X^2 test expends additional degrees of freedom on components that are not helpful in detecting the alternative. Also, in the context of smooth tests of goodness-of-fit for exponential family models, Rayner and Best (1989) describe a test statistic termed S_κ for simple nulls of the same form of L_κ (see their Chapter 5); for composite nulls, they describe the test statistics $\hat{S}_\kappa$ given in (22) that are related to M_κ (see their Chapter 7). Also, the merging of cells along with a test on certain cells as used in Example 4.1 is known as a cell-focusing test in the literature on directional tests. Notice, however, that the test statistics proposed here are omnibus tests.

How to choose which summary statistics to use? We have provided two conditions, Conditions T and D, that must be satisfied in the case of composite nulls and they are not difficult to check. For simple nulls, only Condition T must be satisfied. There are of course many summary statistics that satisfy these conditions. A general recommendation based on the theory is to choose the statistics that summarize as much as possible, but not too much, in such a way that potential models can be discriminated by the summaries. On the other hand, if the summaries are insufficient, such as with only first and second order univariate moments and second order mixed moments for multidimensional multinomial data, then power might only be roughly equal to the significance level in many local directions of alternatives. Also, for composite nulls, if the summaries are insufficient, the model may be (locally) identified but nearly nonidentified, i.e., $\boldsymbol{\Delta}_\kappa$ may have singular values near zero. In this case, $\boldsymbol{\Delta}_\kappa^{(c)}$ is not "stable," since the selection of the columns for $\boldsymbol{\Delta}_\kappa^{(c)}$ can come from the singular vectors associated with q zero singular values or from the singular vectors of singular values that are near zero. For this case, our experience is that with small changes to the maximum likelihood estimate $\widehat{\boldsymbol{\theta}}$, M_κ can change a lot. In looking at the second form of $\mathbf{U}_\kappa$ in (5), when $\boldsymbol{\Delta}_\kappa$ has singular values near zero, $\boldsymbol{\Delta}_k' \boldsymbol{\Xi}_k \boldsymbol{\Delta}_\kappa$ is nearly singular so that its inverse is "large" and $\mathbf{U}_\kappa$ can have many elements near zero. Hence, the noncentrality parameter is typically smaller and such $\boldsymbol{\kappa}$ will not have good power. To avoid this problem, we recommend that the singular values of $\widehat{\boldsymbol{\Delta}}_\kappa$ be checked, and if there are small singular values, then another statistic within the M_κ family be used that does not concentrate the information so much.

The theory and examples show that it is possible to construct M_κ statistics with a small number s of summary statistics to have good power against potential (realistic) alternatives. The primary applications are for sparse high-dimensional tables for which summaries are based on low-order margins. In this regard, note that the calculation of M_κ does not require $\mathbf{T}_\kappa$ explicitly or the calculation of $\boldsymbol{\pi}$. Rather, $\hat{\boldsymbol{\kappa}}$, $\boldsymbol{\kappa}(\widehat{\boldsymbol{\theta}})$, $\boldsymbol{\Xi}_\kappa$, $\boldsymbol{\Delta}_\kappa$ should be computed directly without involving the large matrix $\mathbf{T}_\kappa$. In closing, in most applications, no M_κ can be expected to be uniformly most powerful over all possible directions of alternatives. But with some thought, as shown in our examples, one can come up with M_κ statistics that are much more powerful than Pearson's $\widehat{X}^2$ and also avoid the low cell count or sparsity that affects the adequacy of the asymptotic χ^2 distribution as an approximation for small samples and/or large models.

Appendix

A.1. Some Remarks on Condition T

For the theory presented in this paper, $\boldsymbol{\Xi}_\kappa$ must be invertible and this means that $\mathbf{1}'_C$ *cannot* be in the row span of $\mathbf{T}_\kappa$, or equivalently $\mathbf{1}_C$ cannot be in the column span of $\mathbf{T}'_\kappa$. Note that $\boldsymbol{\Gamma}$ has rank $C-1$ and $\boldsymbol{\Gamma}\mathbf{1}_C = \mathbf{0}$. This is one reason for $s \le C-1$. If $\mathbf{1}_C$ is in the column span of $\mathbf{T}'_\kappa$, then there is an $s \times 1$ vector $\mathbf{x}$ such that $\mathbf{T}'_\kappa \mathbf{x} = \mathbf{1}_C$ and $\mathbf{x}'\mathbf{T}_\kappa \boldsymbol{\Gamma}\mathbf{T}'_\kappa \mathbf{x} = \mathbf{1}'_C \boldsymbol{\Gamma}\mathbf{1}_C = 0$; that is, $\boldsymbol{\Xi}_\kappa = \mathbf{T}_\kappa \boldsymbol{\Gamma}\mathbf{T}'_\kappa$ has a zero eigenvalue and is singular.

For the converse, if $\boldsymbol{\Xi}_\kappa$ is not of full rank then $\mathbf{1}'_C$ is in the row span of $\mathbf{T}_\kappa$. The proof is as follows. $\boldsymbol{\Gamma}$ has rank $C-1$ and the eigenspace for the zero eigenvalue is $\{b\mathbf{1}_C : b \text{ is real}\}$. If $\boldsymbol{\Xi}_\kappa$ is not of full rank, then there is a vector $\mathbf{z}$ such that $\mathbf{z}'\boldsymbol{\Xi}\mathbf{z} = 0$ or $\mathbf{z}'\mathbf{T}_\kappa \boldsymbol{\Gamma}\mathbf{T}'_\kappa \mathbf{z} = 0$. Then $\mathbf{T}'_\kappa \mathbf{z}$ is in the eigenspace of the zero eigenvalue of $\boldsymbol{\Gamma}$ so that $\mathbf{T}'_\kappa \mathbf{z}$ is a multiple of $\mathbf{1}_C$, or $\mathbf{1}_C$ is in the column span of $\mathbf{T}'_\kappa$.

As an example, suppose one is considering the merging of neighboring categories with overlap, using a $\mathbf{T}_\kappa = \mathbf{T} = (T_{ic})$ of dimension $(C-1) \times C$ that satisfies $T_{i,i} = T_{i,i+1} = 1$, $i = 1, \ldots, C-1$, and $T_{i,j} = 0$ otherwise. This $\mathbf{T}$ satisfies Condition T only if C is odd; if C is even, $\mathbf{T}'\mathbf{x} = \mathbf{1}_C$ for the $(C-1)$-dimensional column vector $\mathbf{x}$ that has $x_i = I$ (i odd) and the last row of $\mathbf{T}$ is redundant.

If $\mathbf{1}'_C$ is in the row span of $\mathbf{T}$, and $\mathbf{T}$ has full row rank, then one of the components of $\boldsymbol{\kappa}$ is redundant (that is, it can be derived from the others). The redundancy might be easier to see after eliminating one of the probabilities, for example, $\pi_d = 1 - \sum_{c \in \mathcal{C}, c \ne d} \pi_c$.

For some results, it might be convenient to suppose that $\boldsymbol{\kappa}$ has been reexpressed so that we can assume that $\mathbf{T}_\kappa$ can have a zero column (say in the first column denoted with index 0). The reasoning is as follows. If $\mathbf{T} = (T_{ic})$ is an $s \times C$ matrix, then the ith linear function of the probabilities is $\sum_{c \in \mathcal{C}} T_{ic}\pi_c = T_{i0}[1 - \sum_{c \ne 0} \pi_c] + \sum_{c \ne 0} T_{ic}\pi_c = T_{i0} + \sum_{c \ne 0}(T_{ic} - T_{i0})\pi_c$. The constant T_{i0} does not affect the difference $\hat{\kappa}_i - \kappa_i$ or the variance of $\hat{\kappa}_i$. Hence, we can use $\boldsymbol{\kappa} = \mathbf{T}_\kappa \boldsymbol{\pi}$ where for $c \ne 0$, the (i, c) element is $T_{ic} - T_{i0}$. If $\mathbf{T}_\kappa$ is such that the first column is a zero vector, then Condition T is the same as $\mathbf{T}_\kappa$ having full row rank (since the condition of $\mathbf{1}'_C$ not in row span of $\mathbf{T}_\kappa$ is satisfied). The requirement of $\mathbf{T}_\kappa$ having a zero column is mainly useful for proving some results. We show in the example of Section 4.1 one instance where we are better off using a statistic, named $\mathbf{T}_{\kappa_7}$ in that example, where the first column does not consist of all zeros.

A.2. Results for the Maximum Likelihood Estimate, $\widehat{X}^2$, and $\boldsymbol{\Sigma}_\kappa$

We use the notation $\widehat{\boldsymbol{\Delta}} = \boldsymbol{\Delta}(\widehat{\boldsymbol{\theta}})$, $\widehat{\boldsymbol{\pi}} = \boldsymbol{\pi}(\widehat{\boldsymbol{\theta}})$, $\widehat{\mathbf{D}} = \mathbf{D}(\widehat{\boldsymbol{\theta}})$, $\widehat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})$, $\hat{\mathcal{I}} = \mathcal{I}(\widehat{\boldsymbol{\theta}})$, when the MLE $\widehat{\boldsymbol{\theta}}$ is substituted as an argument.

Any equation below with the vector of sample proportions $\mathbf{p}$ depends on the MLE of $\boldsymbol{\theta}$. Otherwise, equations and identities are valid over all $\boldsymbol{\theta}$ in the parameter space.

- Fisher information matrix: $\mathcal{I} = \boldsymbol{\Delta}'\mathbf{D}^{-1}\boldsymbol{\Delta}$, a $q \times q$ matrix.
- Equation for MLE: $\widehat{\boldsymbol{\Delta}}'\widehat{\mathbf{D}}^{-1}\mathbf{p} = \mathbf{0}$. Since $\boldsymbol{\Delta}'\mathbf{D}^{-1}\boldsymbol{\pi} = \mathbf{0}$ is an identity for any $\boldsymbol{\theta}$, the MLE equation can be written as $\widehat{\boldsymbol{\Delta}}'\widehat{\mathbf{D}}^{-1}(\mathbf{p} - \widehat{\boldsymbol{\pi}}) = \mathbf{0}$.
- Covariance matrix of $\sqrt{N}\,(\mathbf{p} - \widehat{\boldsymbol{\pi}})$: $\boldsymbol{\Sigma} = \boldsymbol{\Gamma} - \boldsymbol{\Delta}\mathcal{I}^{-1}\boldsymbol{\Delta}'$.
- Rank of $\boldsymbol{\Sigma}$: $\boldsymbol{\Sigma}\mathbf{D}^{-1}$ is idempotent with trace or rank $C - 1 - q$; therefore, $\boldsymbol{\Sigma} = (\boldsymbol{\Sigma}\mathbf{D}^{-1})\mathbf{D}$ has rank $C - 1 - q$.
- $\mathbf{D}^{-1}$ is a generalized inverse of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Gamma}$.
- Rank of $\boldsymbol{\Sigma}_\kappa$: $\boldsymbol{\Sigma}$ can be written as $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}'$, with $\mathbf{L}$ a $C \times (C - q - 1)$ matrix of rank $C - q - 1$. Therefore, the rank of $\boldsymbol{\Sigma}_\kappa = \mathbf{T}_\kappa \mathbf{L}\mathbf{L}'\mathbf{T}'_\kappa$ equals $\operatorname{rank}(\mathbf{T}_\kappa \mathbf{L}) \le \min\{s, C - q - 1\}$.

A.3. Proofs of Theorems

If not given, then $\boldsymbol{\theta}$ is implicitly an argument of all relevant vectors and matrices for $\widehat{X}^2$, M_κ, M_{κ_1} and M_{κ_2},

Proof of Theorem 1: We use 0 as the index for the first category and consider the M_κ statistic with this category omitted. Let $\mathbf{1}$ denote a $(C-1)$-dimensional column vector of 1's. Write $\mathbf{p} = \binom{p_0}{\check{\mathbf{p}}}$, $\boldsymbol{\pi} = \binom{\pi_0}{\check{\boldsymbol{\pi}}}$, and $\mathbf{e} = \mathbf{p} - \boldsymbol{\pi}(\widehat{\boldsymbol{\theta}}) = \binom{e_0}{\check{\mathbf{e}}}$. $\boldsymbol{\Delta}' = (\boldsymbol{\Delta}_0' \ \widehat{\boldsymbol{\Delta}}')$, $\boldsymbol{\Gamma} = \begin{pmatrix} \gamma_{00} & \boldsymbol{\gamma}' \\ \boldsymbol{\gamma} & \check{\mathbf{D}} - \check{\boldsymbol{\pi}}\check{\boldsymbol{\pi}}' \end{pmatrix}$, and $\check{\boldsymbol{\Gamma}} = \check{\mathbf{D}} - \check{\boldsymbol{\pi}}\check{\boldsymbol{\pi}}'$. Let $\check{\mathbf{D}} = \operatorname{diag}(\check{\boldsymbol{\pi}})$ and $D_0 = \pi_0$. Then $\boldsymbol{\Delta}_\kappa = \mathbf{T}_\kappa \boldsymbol{\Delta} = \widehat{\boldsymbol{\Delta}}$, $\boldsymbol{\Xi}_\kappa = \mathbf{T}_\kappa \boldsymbol{\Gamma} \mathbf{T}_\kappa' = \check{\boldsymbol{\Gamma}}$ and $\boldsymbol{\Xi}_\kappa^{-1} = \check{\boldsymbol{\Gamma}}^{-1} = \check{\mathbf{D}}^{-1} + \mathbf{1} D_0^{-1} \mathbf{1}'$, and $\mathbf{T}_\kappa \mathbf{e} = \check{\mathbf{e}}$.

Since $\mathbf{1}'\check{\mathbf{e}} = -e_0$, then

$$
\begin{aligned}
N\check{\mathbf{e}}'\big[\boldsymbol{\Xi}_\kappa(\widehat{\boldsymbol{\theta}})\big]^{-1}\check{\mathbf{e}} &= N\check{\mathbf{e}}'\big\{\big[\check{\mathbf{D}}(\widehat{\boldsymbol{\theta}})\big]^{-1} + \mathbf{1}\big(D_0(\widehat{\boldsymbol{\theta}})\big)^{-1}\mathbf{1}'\big\}\check{\mathbf{e}} \\
&= N\check{\mathbf{e}}'\big[\check{\mathbf{D}}(\widehat{\boldsymbol{\theta}})\big]^{-1}\check{\mathbf{e}} + N e_0\big(D_0(\widehat{\boldsymbol{\theta}})\big)^{-1} e_0 = N\mathbf{e}'\big[\mathbf{D}(\widehat{\boldsymbol{\theta}})\big]^{-1}\mathbf{e} = \widehat{X}^2. \qquad (23)
\end{aligned}
$$

Using the second form of $\mathbf{U}_\kappa$ in (5), $M_\kappa = \widehat{X}^2$ if $\boldsymbol{\Delta}_\kappa'(\widehat{\boldsymbol{\theta}})[\boldsymbol{\Xi}_\kappa(\widehat{\boldsymbol{\theta}})]^{-1}\check{\mathbf{e}} = \mathbf{0}$. From the above,

$$
\big[\boldsymbol{\Xi}_\kappa(\widehat{\boldsymbol{\theta}})\big]^{-1}\check{\mathbf{e}} = \big[\check{\mathbf{D}}(\widehat{\boldsymbol{\theta}})\big]^{-1}\check{\mathbf{e}} - \mathbf{1}\big(D_0(\widehat{\boldsymbol{\theta}})\big)^{-1} e_0
$$

and

$$
\begin{aligned}
\boldsymbol{\Delta}_\kappa'(\widehat{\boldsymbol{\theta}})\big[\boldsymbol{\Xi}_\kappa(\widehat{\boldsymbol{\theta}})\big]^{-1}\check{\mathbf{e}} &= \widehat{\boldsymbol{\Delta}}'(\widehat{\boldsymbol{\theta}})\big[\check{\mathbf{D}}(\widehat{\boldsymbol{\theta}})\big]^{-1}\check{\mathbf{e}} - \widehat{\boldsymbol{\Delta}}'(\widehat{\boldsymbol{\theta}})\mathbf{1}\big(D_0(\widehat{\boldsymbol{\theta}})\big)^{-1} e_0 \\
&= \widehat{\boldsymbol{\Delta}}'(\widehat{\boldsymbol{\theta}})\big[\check{\mathbf{D}}(\widehat{\boldsymbol{\theta}})\big]^{-1}\check{\mathbf{e}} + \boldsymbol{\Delta}_0'(\widehat{\boldsymbol{\theta}})\big(D_0(\widehat{\boldsymbol{\theta}})\big)^{-1} e_0 = \boldsymbol{\Delta}'(\widehat{\boldsymbol{\theta}})\big[\mathbf{D}(\widehat{\boldsymbol{\theta}})\big]^{-1}\mathbf{e} = \widehat{\boldsymbol{\Delta}}'\widehat{\mathbf{D}}^{-1}\mathbf{e} = \mathbf{0},
\end{aligned}
$$

from the likelihood score equation (see Appendix A.1).

The result for $X^2 = L_\kappa$ follows from (23) with $\mathbf{e} = \mathbf{p} - \boldsymbol{\pi} = \binom{e_0}{\check{\mathbf{e}}}$ and $\widehat{\boldsymbol{\theta}}$ omitted. □

Proof of Theorem 2: We will prove the second result $M_{\kappa_2} = M_{\kappa_1}$ since the first part of its proof covers the first result. From the given assumptions,

$$
\begin{aligned}
\hat{\boldsymbol{\kappa}}_2 - \boldsymbol{\kappa}_2(\widehat{\boldsymbol{\theta}}) &= \mathbf{B}\hat{\boldsymbol{\kappa}}_1 + \boldsymbol{\beta} - \mathbf{B}\boldsymbol{\kappa}_1(\widehat{\boldsymbol{\theta}}) - \boldsymbol{\beta} = \mathbf{B}\big[\hat{\boldsymbol{\kappa}}_2 - \boldsymbol{\kappa}_2(\widehat{\boldsymbol{\theta}})\big], \\
\boldsymbol{\Xi}_{\kappa_2} &= \mathbf{T}_{\kappa_2}\boldsymbol{\Gamma}\mathbf{T}_{\kappa_2}' = \mathbf{B}\boldsymbol{\Xi}_{\kappa_1}\mathbf{B}'.
\end{aligned}
$$

From (5), it follows that $\mathbf{U}_{\kappa_2} = (\mathbf{B}')^{-1}\mathbf{U}_{\kappa_1}\mathbf{B}^{-1}$. Hence, the conclusion follows. □

Proof of Theorem 3: Let $\mathbf{T}_{\kappa 0} = (\mathbf{0}\ \mathbf{I}_{C-1})$ as used in Theorem 1. Because of Condition T, $\binom{\mathbf{1}_C'}{\mathbf{T}_\kappa}$ and $\binom{\mathbf{1}_C'}{\mathbf{T}_{\kappa 0}}$ are invertible $C \times C$ matrices. Hence, there is a nonsingular $C \times C$ matrix $\mathbf{B}^*$ such that

$$
\begin{pmatrix} \mathbf{1}_C' \\ \mathbf{T}_\kappa \end{pmatrix} = \mathbf{B}^* \begin{pmatrix} \mathbf{1}_C' \\ \mathbf{T}_{\kappa 0} \end{pmatrix}.
$$

Partition $\mathbf{B}^*$ as $\begin{pmatrix} b_0 & \mathbf{b}_2' \\ \mathbf{b}_1 & \mathbf{B} \end{pmatrix}$, where $\mathbf{B}$ is a $(C-1)\times(C-1)$ matrix, and $\mathbf{b}_1, \mathbf{b}_2$ are $(C-1)\times 1$ vectors. Hence,

$$
\mathbf{T}_\kappa = \mathbf{b}_1\mathbf{1}' + \mathbf{B}\mathbf{T}_{\kappa 0}.
$$

We next show that the conditions of Theorem 2 hold, and then the conclusion follows from Theorem 1.

(i) $\boldsymbol{\kappa} = \mathbf{T}_{\kappa}\boldsymbol{\pi} = [\mathbf{b}_1\mathbf{1}' + \mathbf{B}\mathbf{T}_{\kappa 0}]\boldsymbol{\pi} = \mathbf{b}_1 + \mathbf{B}\check{\boldsymbol{\pi}}$, with $\check{\boldsymbol{\pi}} = \boldsymbol{\kappa}_0$.

(ii) $\mathbf{T}_{\kappa}\boldsymbol{\Gamma}\mathbf{T}'_{\kappa} = \mathbf{b}_1\mathbf{1}'\boldsymbol{\Gamma}\mathbf{1}\mathbf{b}'_1 + \mathbf{b}_1\mathbf{1}'\boldsymbol{\Gamma}\mathbf{T}'_{\kappa 0}\mathbf{B}' + \mathbf{B}\mathbf{T}_{\kappa 0}\boldsymbol{\Gamma}\mathbf{1}\mathbf{b}'_1 + \mathbf{B}\mathbf{T}_{\kappa 0}\boldsymbol{\Gamma}\mathbf{T}'_{\kappa 0}\mathbf{B}' = \mathbf{0} + \mathbf{B}\check{\boldsymbol{\Gamma}}\mathbf{B}'$, with $\check{\boldsymbol{\Gamma}} = \boldsymbol{\Xi}_{\kappa 0}$, since $\boldsymbol{\Gamma}\mathbf{1} = \mathbf{0}$.

(iii) $\boldsymbol{\Delta}_{\kappa} = \mathbf{T}_k\boldsymbol{\Delta} = \mathbf{b}_1\mathbf{1}'\boldsymbol{\Delta} + \mathbf{B}\mathbf{T}_{\kappa 0}\boldsymbol{\Delta} = \mathbf{0} + \mathbf{B}\boldsymbol{\Delta}_{\kappa 0}$, since $\boldsymbol{\Delta}'\mathbf{1} = \mathbf{0}$. □

Proof of Theorem 4: The covariance matrix of $\sqrt{N}\binom{\hat{\boldsymbol{\kappa}}_1 - \boldsymbol{\kappa}_1}{\hat{\boldsymbol{\kappa}}_2 - \boldsymbol{\kappa}_2}$ is

$$\begin{pmatrix} \mathbf{T}_{\kappa_1}\boldsymbol{\Gamma}\mathbf{T}'_{\kappa_1} & \mathbf{T}_{\kappa_1}\boldsymbol{\Gamma}\mathbf{T}'_{\kappa_2} \\ \mathbf{T}_{\kappa_2}\boldsymbol{\Gamma}\mathbf{T}'_{\kappa_1} & \mathbf{T}_{\kappa_2}\boldsymbol{\Gamma}\mathbf{T}'_{\kappa_2} \end{pmatrix}.$$

This is a square matrix with dimension $s_1 + s_2$ and rank s_1. The assumptions imply that $\mathbf{T}_{\kappa_j}\boldsymbol{\Gamma}\mathbf{T}'_{\kappa_j}$, $j = 1, 2$, is non-singular, so that

$$\begin{aligned} &\mathbf{T}_{\kappa_1}\boldsymbol{\Gamma}\mathbf{T}'_{\kappa_1} - \mathbf{T}_{\kappa_1}\boldsymbol{\Gamma}\mathbf{T}'_{\kappa_2}[\mathbf{T}_{\kappa_2}\boldsymbol{\Gamma}\mathbf{T}'_{\kappa_2}]^{-1}\mathbf{T}_{\kappa_2}\boldsymbol{\Gamma}\mathbf{T}'_{\kappa_1} \\ &\quad = \mathbf{T}_{\kappa_1}\boldsymbol{\Gamma}\mathbf{T}'_{\kappa_1} - \mathbf{T}_{\kappa_1}\boldsymbol{\Gamma}\mathbf{T}'_{\kappa_1}\mathbf{T}'_{21}(\mathbf{T}_{\kappa_2}\boldsymbol{\Gamma}\mathbf{T}'_{\kappa_2})^{-1}\mathbf{T}_{21}\mathbf{T}_{\kappa_1}\boldsymbol{\Gamma}\mathbf{T}'_{\kappa_1} \\ &\quad = \mathbf{T}_{\kappa_1}\boldsymbol{\Gamma}\mathbf{T}'_{\kappa_1}\big[(\mathbf{T}_{\kappa_1}\boldsymbol{\Gamma}\mathbf{T}'_{\kappa_1})^{-1} - \mathbf{T}'_{21}(\mathbf{T}_{\kappa_2}\boldsymbol{\Gamma}\mathbf{T}'_{\kappa_2})^{-1}\mathbf{T}_{21}\big]\mathbf{T}_{\kappa_1}\boldsymbol{\Gamma}\mathbf{T}'_{\kappa_1} \end{aligned}$$

is a nonnegative definite (conditional covariance) matrix with rank $s_1 - s_2$. Hence, the $s_1 \times s_1$ matrix

$$\mathbf{L}_{21} = (\mathbf{T}_{\kappa_1}\boldsymbol{\Gamma}\mathbf{T}'_{\kappa_1})^{-1} - \mathbf{T}'_{21}(\mathbf{T}_{\kappa_2}\boldsymbol{\Gamma}\mathbf{T}'_{\kappa_2})^{-1}\mathbf{T}_{21} = \boldsymbol{\Xi}_{\kappa_1}^{-1} - \mathbf{T}'_{21}\boldsymbol{\Xi}_{\kappa_2}^{-1}\mathbf{T}_{21}$$

is non-negative definite, and its rank is $s_1 - s_2$. □

Proof of Theorem 5: From Theorem 4,

$$L_{\kappa_1} - L_{\kappa_2} = (\hat{\boldsymbol{\kappa}}_1 - \boldsymbol{\kappa}_1)'\big[\boldsymbol{\Xi}_{\kappa_1}^{-1} - \mathbf{T}'_{21}\boldsymbol{\Xi}_{\kappa_2}^{-1}\mathbf{T}_{21}\big](\hat{\boldsymbol{\kappa}}_1 - \boldsymbol{\kappa}_1) = (\hat{\boldsymbol{\kappa}}_1 - \boldsymbol{\kappa}_1)'\mathbf{L}_{21}(\hat{\boldsymbol{\kappa}}_1 - \boldsymbol{\kappa}_1) \geq 0.$$

Equality occurs if $\hat{\boldsymbol{\kappa}}_1 - \boldsymbol{\kappa}_1$ is in the null space of (8) or $\mathbf{L}_{21}(\hat{\boldsymbol{\kappa}}_1 - \boldsymbol{\kappa}_1) = \mathbf{0}$. □

Proof of Theorem 6: For $j = 1, 2$, let

$$\boldsymbol{\Sigma}_{\kappa_j} = \mathbf{T}_{\kappa_j}\boldsymbol{\Sigma}\mathbf{T}'_{\kappa_j} = \boldsymbol{\Xi}_{\kappa_j} - \boldsymbol{\Delta}_{\kappa_j}\mathcal{I}^{-1}\boldsymbol{\Delta}'_{\kappa_j}.$$

By the definition of $\boldsymbol{\Delta}^{(c)}_{\kappa_j}$, for $j = 1, 2$,

$$\boldsymbol{\Delta}^{(c)\prime}_{\kappa_j}\boldsymbol{\Sigma}_{\kappa_j}\boldsymbol{\Delta}^{(c)}_{\kappa_j} = \boldsymbol{\Delta}^{(c)\prime}_{\kappa_j}\boldsymbol{\Xi}_{\kappa_j}\boldsymbol{\Delta}^{(c)}_{\kappa_j}. \tag{24}$$

Also,

$$\boldsymbol{\Delta}^{(c)\prime}_{\kappa_1}\boldsymbol{\Sigma}_{\kappa_1}\mathbf{T}'_{21}\boldsymbol{\Delta}^{(c)}_{\kappa_2} = \boldsymbol{\Delta}^{(c)\prime}_{\kappa_1}\boldsymbol{\Xi}_{\kappa_1}\mathbf{T}'_{21}\boldsymbol{\Delta}^{(c)}_{\kappa_2} \tag{25}$$

because $\boldsymbol{\Delta}^{(c)\prime}_{\kappa_1}\boldsymbol{\Delta}_{\kappa_1}\mathcal{I}^{-1}\boldsymbol{\Delta}'_{\kappa_1}\mathbf{T}'_{21}\boldsymbol{\Delta}^{(c)}_{\kappa_2} = \mathbf{0}$. The covariance matrix of

$$\sqrt{N}\begin{pmatrix} \boldsymbol{\Delta}^{(c)\prime}_{\kappa_1}[\hat{\boldsymbol{\kappa}}_1 - \boldsymbol{\kappa}_1(\widehat{\boldsymbol{\theta}})] \\ \boldsymbol{\Delta}^{(c)\prime}_{\kappa_2}[\hat{\boldsymbol{\kappa}}_2 - \boldsymbol{\kappa}_2(\widehat{\boldsymbol{\theta}})] \end{pmatrix} = \sqrt{N}\begin{pmatrix} \boldsymbol{\Delta}^{(c)\prime}_{\kappa_1}[\hat{\boldsymbol{\kappa}}_1 - \boldsymbol{\kappa}_1(\widehat{\boldsymbol{\theta}})] \\ \boldsymbol{\Delta}^{(c)\prime}_{\kappa_2}\mathbf{T}_{21}[\hat{\boldsymbol{\kappa}}_1 - \boldsymbol{\kappa}_1(\widehat{\boldsymbol{\theta}})] \end{pmatrix}$$

is

$$\begin{pmatrix} \boldsymbol{\Delta}^{(c)\prime}_{\kappa_1}\boldsymbol{\Sigma}_{\kappa_1}\boldsymbol{\Delta}^{(c)}_{\kappa_1} & \boldsymbol{\Delta}^{(c)\prime}_{\kappa_1}\boldsymbol{\Sigma}_{\kappa_1}\mathbf{T}'_{21}\boldsymbol{\Delta}^{(c)}_{\kappa_2} \\ \boldsymbol{\Delta}^{(c)\prime}_{\kappa_2}\mathbf{T}_{21}\boldsymbol{\Sigma}_{\kappa_1}\boldsymbol{\Delta}^{(c)}_{\kappa_1} & \boldsymbol{\Delta}^{(c)\prime}_{\kappa_2}\boldsymbol{\Sigma}_{\kappa_2}\boldsymbol{\Delta}^{(c)}_{\kappa_2} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Delta}^{(c)\prime}_{\kappa_1}\boldsymbol{\Xi}_{\kappa_1}\boldsymbol{\Delta}^{(c)}_{\kappa_1} & \boldsymbol{\Delta}^{(c)\prime}_{\kappa_1}\boldsymbol{\Xi}_{\kappa_1}\mathbf{T}'_{21}\boldsymbol{\Delta}^{(c)}_{\kappa_2} \\ \boldsymbol{\Delta}^{(c)\prime}_{\kappa_2}\mathbf{T}_{21}\boldsymbol{\Xi}_{\kappa_1}\boldsymbol{\Delta}^{(c)}_{\kappa_1} & \boldsymbol{\Delta}^{(c)\prime}_{\kappa_2}\boldsymbol{\Xi}_{\kappa_2}\boldsymbol{\Delta}^{(c)}_{\kappa_2} \end{pmatrix},$$

with the equalities in the matrices coming from (24) and (25). This is a square matrix with row dimension $(s_1 - q) + (s_2 - q)$. The assumptions imply that $\boldsymbol{\Delta}_{\kappa_j}^{(c)\prime} \boldsymbol{\Xi}_{\kappa_j} \boldsymbol{\Delta}_{\kappa_j}^{(c)}$ is non-singular $(j = 1, 2)$. Hence,

$$\begin{aligned}\boldsymbol{\Delta}_{\kappa_1}^{(c)\prime} \boldsymbol{\Xi}_{\kappa_1} \boldsymbol{\Delta}_{\kappa_1}^{(c)} - \boldsymbol{\Delta}_{\kappa_1}^{(c)\prime} \boldsymbol{\Xi}_{\kappa_1} \mathbf{T}_{21}' \boldsymbol{\Delta}_{\kappa_2}^{(c)} \big(\boldsymbol{\Delta}_{\kappa_2}^{(c)\prime} \boldsymbol{\Xi}_{\kappa_2} \boldsymbol{\Delta}_{\kappa_2}^{(c)}\big)^{-1} \boldsymbol{\Delta}_{\kappa_2}^{(c)\prime} \mathbf{T}_{21} \boldsymbol{\Xi}_{\kappa_1} \boldsymbol{\Delta}_{\kappa_1}^{(c)} \\ = \boldsymbol{\Delta}_{\kappa_1}^{(c)\prime} \boldsymbol{\Xi}_{\kappa_1} [\mathbf{U}_{\kappa_1} - \mathbf{T}_{21}' \mathbf{U}_{\kappa_2} \mathbf{T}_{21}] \boldsymbol{\Xi}_{\kappa_1} \boldsymbol{\Delta}_{\kappa_1}^{(c)} \end{aligned} \tag{26}$$

is a nonnegative definite (conditional) covariance matrix. Let its rank be m, where $1 \le m \le s_1 - q$; the actual value of m is $s_1 - s_2$ and this is shown in a remark below. The dimension of $\boldsymbol{\Delta}_{\kappa_1}^{(c)\prime} \boldsymbol{\Xi}_{\kappa_1}$ is $(s_1 - q) \times s_1$, and its rank is $s_1 - q$ (because of Condition D).

Let

$$\mathbf{M}_{21} = \mathbf{U}_{\kappa_1} - \mathbf{T}_{21}' \mathbf{U}_{\kappa_2} \mathbf{T}_{21};$$

this is an $s_1 \times s_1$ matrix. Note that $\mathbf{M}_{21}\mathbf{y} = \mathbf{0}$ for any $\mathbf{y}$ in the column span of $\boldsymbol{\Delta}_{\kappa_1}$, due to the following argument. If $\mathbf{y}$ is in the column span of $\boldsymbol{\Delta}_{\kappa_1}$, then there exists a $\mathbf{x}$ such that $\boldsymbol{\Delta}_{\kappa_1}\mathbf{x} = \mathbf{y}$. Hence, $\boldsymbol{\Delta}_{\kappa_1}^{(c)\prime}\mathbf{y} = \mathbf{0}$ and, therefore, $\mathbf{U}_{\kappa_1}\mathbf{y} = \mathbf{0}$. Also $\mathbf{U}_{\kappa_2}\mathbf{T}_{21}\mathbf{y} = \mathbf{U}_{\kappa_2}\mathbf{T}_{21}\boldsymbol{\Delta}_{\kappa_1}\mathbf{x} = \mathbf{U}_{\kappa_2}\boldsymbol{\Delta}_{\kappa_2}\mathbf{x} = \mathbf{0}$ since $\mathbf{U}_{\kappa_2}\boldsymbol{\Delta}_{\kappa_2} = \mathbf{0}$ from the form of $\mathbf{U}_{\kappa_2}$ in (4).

Next, write (26) as $\mathbf{B}'\mathbf{M}_{21}\mathbf{B}$, where $\mathbf{B} = \boldsymbol{\Xi}_{\kappa_1}\boldsymbol{\Delta}_{\kappa_1}^{(c)}$, an $s_1 \times (s_1 - q)$ matrix. Consider $(\boldsymbol{\Delta}_{\kappa_1}\ \mathbf{B}) = (\boldsymbol{\Delta}_{\kappa_1}\ \boldsymbol{\Xi}_{\kappa_1}\boldsymbol{\Delta}_{\kappa_1}^{(c)})$. This is an $s_1 \times s_1$ matrix. We claim that it is nonsingular (its columns are linearly independent). Suppose $\boldsymbol{\Delta}_{\kappa_1}\mathbf{a}_1 + \boldsymbol{\Xi}_{\kappa_1}\boldsymbol{\Delta}_{\kappa_1}^{(c)}\mathbf{a}_2 = \mathbf{0}$ where $\mathbf{a}_1, \mathbf{a}_2$ are respectively column vectors of dimensions q and $(s_1 - q)$. Then

$$\boldsymbol{\Delta}_{\kappa_1}^{(c)\prime}\boldsymbol{\Delta}_{\kappa_1}\mathbf{a}_1 + \boldsymbol{\Delta}_{\kappa_1}^{(c)\prime}\boldsymbol{\Xi}_{\kappa_1}\boldsymbol{\Delta}_{\kappa_1}^{(c)}\mathbf{a}_2 = \mathbf{0} + \boldsymbol{\Delta}_{\kappa_1}^{(c)\prime}\boldsymbol{\Xi}_{\kappa_1}\boldsymbol{\Delta}_{\kappa_1}^{(c)}\mathbf{a}_2 = \mathbf{0}.$$

Since $\boldsymbol{\Delta}_{\kappa_1}^{(c)\prime}\boldsymbol{\Xi}_{\kappa_1}\boldsymbol{\Delta}_{\kappa_1}^{(c)}$ is positive definite, $\mathbf{a}_2 = \mathbf{0}$. Then since $\boldsymbol{\Delta}_{\kappa_1}$ has full column rank, $\mathbf{a}_1 = \mathbf{0}$. This establishes the claim.

Let $\mathbf{Z} = (\mathbf{z}_1\ \cdots\ \mathbf{z}_{s_1-q})$ be a matrix of orthogonal $(s_1 - q)$-dimensional eigenvectors of $\mathbf{B}'\mathbf{M}_{21}\mathbf{B}$ with the last $s_1 - q - m$ corresponding to zero eigenvalues and the first m corresponding to positive eigenvalues, denoted as $\omega_1, \ldots, \omega_m$. With a similar argument to above, $(\boldsymbol{\Delta}_{\kappa_1}\ \boldsymbol{\Xi}_{\kappa_1}\boldsymbol{\Delta}_{\kappa_1}^{(c)}\mathbf{Z})$ is nonsingular.

Since $\mathbf{M}_{21}\boldsymbol{\Delta}_{\kappa_1} = \mathbf{0}$ and $\mathbf{M}_{21}\mathbf{B}\mathbf{z}_j = \mathbf{0}$ for $j = m + 1, \ldots, s_1 - q$, we have exhibited a null space of dimension $q + (s_1 - q - m) = s_1 - m$ for $\mathbf{M}_{21}$. Hence, $\text{rank}(\mathbf{M}_{21}) \le m$. From $\text{rank}(\mathbf{B}'\mathbf{M}_{21}\mathbf{B}) = m \le s_1 - q$, $\text{rank}(\mathbf{M}_{21}) \ge m$. The two inequalities imply $\text{rank}(\mathbf{M}_{21}) = m$.

Finally, we show that $\mathbf{M}_{21}$ is nonnegative definite. From the above, $\mathbf{B}\mathbf{z}_j = \boldsymbol{\Xi}_{\kappa_1}\boldsymbol{\Delta}_{\kappa_1}^{(c)}\mathbf{z}_j$ $(j = 1, \ldots, m)$ is a basis for the nonzero eigenvectors of $\mathbf{M}_{21}$. An arbitrary s_1-dimensional vector $\mathbf{y}$ can be written as $\sum_{j=1}^m a_j\mathbf{B}\mathbf{z}_j + \mathbf{z}_0$ where $\mathbf{z}_0$ is in the null space of $\mathbf{M}_{21}$. Then

$$\begin{aligned}\mathbf{y}'\mathbf{M}_{21}\mathbf{y} &= \left[\mathbf{z}_0' + \sum_{i=1}^m a_i\mathbf{z}_i'\mathbf{B}'\right]\mathbf{M}_{21}\left[\mathbf{z}_0 + \sum_{j=1}^m a_i\mathbf{B}\mathbf{z}_j\right] = \sum_{i=1}^m\sum_{j=1}^m a_ia_j\mathbf{z}_i'\mathbf{B}'\mathbf{M}_{21}\mathbf{B}\mathbf{z}_j \\ &= \sum_{i=1}^m\sum_{j=1}^m a_ia_j\omega_j\mathbf{z}_i'\mathbf{z}_j = \sum_{i=1}^m a_i^2\omega_i\mathbf{z}_i'\mathbf{z}_i \ge 0,\end{aligned}$$

since the $\mathbf{z}_i$'s are orthogonal.

Remark. We outline a proof that the value of m, defined above, is $s_1 - s_2$. From the orthogonal complements, $\binom{\boldsymbol{\Delta}_{\kappa_1}^{(c)\prime}}{\boldsymbol{\Delta}_{\kappa_2}^{(c)\prime}\mathbf{T}_{21}}\boldsymbol{\Delta}_{\kappa_1} = \mathbf{0}$, so that

$$\begin{pmatrix} \boldsymbol{\Delta}_{\kappa_1}^{(c)\prime} \\ \boldsymbol{\Delta}_{\kappa_2}^{(c)\prime}\mathbf{T}_{21} \\ \boldsymbol{\Delta}_{\kappa_1}^{\prime} \end{pmatrix}$$

is a $(s_1+s_2-q)\times s_1$ matrix with rank s_1, the same as the rank of $\binom{\boldsymbol{\Delta}_{\kappa_1}^{(c)\prime}}{\boldsymbol{\Delta}_{\kappa_1}^{\prime}}$. Since $\boldsymbol{\Delta}_{\kappa_1}^{\prime}$ and $\boldsymbol{\Delta}_{\kappa_1}^{(c)\prime}$ have respective ranks q and s_1-q, the above orthogonality implies that $\binom{\boldsymbol{\Delta}_{\kappa_1}^{(c)\prime}}{\boldsymbol{\Delta}_{\kappa_2}^{(c)\prime}\mathbf{T}_{21}}$ has rank s_1-q, and there is an $(s_2-q)\times(s_1-q)$ matrix $\mathbf{H}$ of rank (s_2-q) such that $\boldsymbol{\Delta}_{\kappa_2}^{(c)\prime}\mathbf{T}_{21} = \mathbf{H}\boldsymbol{\Delta}_{\kappa_1}^{(c)\prime}$. Therefore, the conditional covariance matrix of $\sqrt{N}\,\boldsymbol{\Delta}_{\kappa_1}^{(c)\prime}[\hat{\boldsymbol{\kappa}}_1 - \boldsymbol{\kappa}_1(\widehat{\boldsymbol{\theta}})]$ given $\sqrt{N}\,\boldsymbol{\Delta}_{\kappa_2}^{(c)\prime}\mathbf{T}_{21}[\hat{\boldsymbol{\kappa}}_1 - \boldsymbol{\kappa}_1(\widehat{\boldsymbol{\theta}})]$ has rank $(s_1-q)-(s_2-q) = s_1 - s_2$.

□

Proof of Theorem 7: Since $\mathbf{M}_{21}$ is nonnegative definite,

$$M_{\kappa_1} - M_{\kappa_2} = \bigl(\hat{\boldsymbol{\kappa}}_1 - \boldsymbol{\kappa}_1(\widehat{\boldsymbol{\theta}})\bigr)^{\prime}\mathbf{M}_{21}\bigl(\hat{\boldsymbol{\kappa}}_1 - \boldsymbol{\kappa}_1(\widehat{\boldsymbol{\theta}})\bigr) \geq 0.$$

Equality occurs if $\hat{\boldsymbol{\kappa}}_1 - \boldsymbol{\kappa}_1(\widehat{\boldsymbol{\theta}})$ is in the null space of $\mathbf{M}_{21}$ or $\mathbf{M}_{21}(\hat{\boldsymbol{\kappa}}_1 - \boldsymbol{\kappa}_1(\widehat{\boldsymbol{\theta}})) = \mathbf{0}$. □

Proof of Corollary 8: Referring to the discussion at the beginning of Section 2, we assume that $\mathbf{T}_{\kappa}$ has been converted to any equivalent $\mathbf{T}_{\kappa_2}$ that has zeros in its first column. Let $\mathbf{T}_{\kappa_1}$ be the matrix in Theorem 1 that leads to either X^2 or $\widehat{X}^2$. To apply Theorem 5, take $\mathbf{T}_{12}$ be the $s\times(C-1)$ matrix that derives from $\mathbf{T}_{\kappa_2}$ by omitting the first column. □

References

Bishop, Y.M.M., Fienberg, S.E., & Holland, P.W. (1975). *Discrete multivariate analysis*. Cambridge: MIT Press.

Bock, R.D., & Lieberman, M. (1970). Fitting a response model for n dichotomously scored items. *Psychometrika*, *35*, 179–197.

Böckenholt, U. (1999). An INAR(1) negative multinomial regression model for longitudinal count data. *Psychometrika*, *64*, 53–67.

Browne, M.W. (1984). Asymptotically distribution free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, *37*, 62–83.

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.

Cochran, W.G. (1952). The X^2 test of goodness of fit. *Annals of Mathematical Statistics*, *23*, 315–345.

Consul, P.C. (1989). *Generalized Poisson distribution: properties and applications*. New York: Dekker.

Cressie, N., & Holland, P.W. (1983). Characterizing the manifest probabilities of latent trait models. *Psychometrika*, *48*, 129–141.

Eubank, R. (1997). Testing goodness of fit with multinomial data. *Journal of the American Statistical Association*, *92*, 1084–1093.

Glas, C.A.W. (1988). The derivation of some tests for the Rasch model from the multinomial distribution. *Psychometrika*, *53*, 525–546.

Glas, C.A.W., & Verhelst, N.D. (1989). Extensions of the partial credit model. *Psychometrika*, *54*, 635–659.

Glas, C.A.W., & Verhelst, N.D. (1995). Testing the Rasch model. In G.H. Fischer & I.W. Molenaar (Eds.), *Rasch models. Their foundations, recent developments and applications* (pp. 69–96). New York: Springer.

Hall, P. (1985). Tailor-made tests of goodness of fit. *Journal of the Royal Statistical Society B*, *47*, 125–131.

Joe, H., & Zhu, R. (2005). Generalized Poisson distribution: the property of mixture of Poisson and comparison with negative binomial distribution. *Biometrical Journal*, *47*, 219–229.

Kendall, M., & Stuart, A. (1979). *The advanced theory of statistics*. London: Griffin.

Khatri, C.G. (1966). A note on a MANOVA model applied to problems in growth curve. *Annals of the Institute of Statistical Mathematics*, *18*, 75–86.
Lee, A.H., Wang, K., & Yau, K.K.W. (2001). Analysis of zero-inflated Poisson data incorporating extent of exposure. *Biometrical Journal*, *43*, 963–975.
Mavridis, D., Moustaki, I., & Knott, M. (2007). Goodness-of-fit measures for latent variable models for binary data. In S.-Y. Lee (Ed.), *Handbook of latent variable and related models* (pp. 135–161). Amsterdam: Elsevier.
Maydeu-Olivares, A. (2001). Multidimensional item response theory modeling of binary data: Large sample properties of NOHARM estimates. *Journal of Educational and Behavioral Statistics*, *26*, 49–69.
Maydeu-Olivares, A. (2006). Limited information estimation and testing of discretized multivariate normal structural models. *Psychometrika*, *71*, 57–77.
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.
Maydeu-Olivares, A., & Joe, H. (2008). An overview of limited information goodness-of-fit testing in multidimensional contingency tables. In K. Shigemasu, A. Okada, T. Imaizumi, & T. Hoshino (Eds.), *New trends in psychometrics* (pp. 253–262). Tokyo: Universal Academy Press.
Rao, C.R. (1973). *Linear statistical inference and its applications*. New York: Wiley.
Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danmarks Pedagogiske Institut.
Rayner, J.C.W., & Best, D.J. (1989). *Smooth tests of goodness of fit*. New York: Oxford University Press.
Rayner, J.C.W., & Best, D.J. (1990). Smooth tests of goodness of fit: An overview. *International Statistical Review*, *58*, 9–17.
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*, 331–360.
Reiser, M., & Lin, Y. (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* (pp. 81–111). Boston: Blackwell.
Tjur, T. (1982). A connection between Rasch's item analysis model and a multiplicative Poisson model. *Scandinavian Journal of Statistics*, *9*, 23–30.
Van Duijn, M.A.J., & Jansen, M.G.H. (1995). Repeated count data: some extensions of the rasch Poisson counts model. *Journal of Educational and Behavioral Statistics*, *20*, 241–258.
White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, *50*, 1–26.
Yuan, K.-H., & Bentler, P.M. (1997). Mean and covariance structure analysis: Theoretical and practical improvements. *Journal of the American Statistical Association*, *92*, 767–774.

Manuscript Received: 13 AUG 2007
Final Version Received: 4 NOV 2009
Published Online Date: 16 APR 2010