The Annals of Statistics
2000, Vol. 28, No. 4, 1206–1218

A LIKELIHOOD RATIO TEST FOR MTP_2 WITHIN BINARY VARIABLES[1]

BY FRANCESCO BARTOLUCCI AND ANTONIO FORCINA

Università di Perugia

Multivariate Totally Positive (MTP_2) binary distributions have been studied in many fields, such as statistical mechanics, computer storage and latent variable models. We show that MTP_2 is equivalent to the requirement that the parameters of a saturated log-linear model belong to a convex cone, and we provide a Fisher-scoring algorithm for maximum likelihood estimation. We also show that the asymptotic distribution of the log-likelihood ratio is a mixture of chi-squares (a distribution known as chi-bar-squared in the literature on order restricted inference); for this we derive tight bounds which turn out to have very simple forms. A potential application of this method is for *Item Response Theory* (IRT) models, which are used in educational assessment to analyse the responses of a group of subjects to a collection of questions (*items*): an important issue within IRT is whether the joint distribution of the manifest variables is compatible with a single latent variable representation satisfying local independence and monotonicity which, in turn, imply that the joint distribution of item responses is MTP_2.

1. Introduction. Let $\mathbf{X}$ be a $J \times 1$ random vector of binary variables with elements X_j taking vales in $\{0, 1\}$ and joint probability distribution $p(\mathbf{X})$ defined on the set $\mathscr{X}$ of the $s = 2^J$ possible response configurations. For any two vectors $\mathbf{x}_1$ and $\mathbf{x}_2 \in \mathscr{X}$, let the functions $\min(\mathbf{x}_1, \mathbf{x}_2)$ and $\max(\mathbf{x}_1, \mathbf{x}_2)$ act elementwise on the two vectors.

DEFINITION 1. *If, for any pair of vectors $\mathbf{x}_1$ and $\mathbf{x}_2 \in \mathscr{X}$,*

$$p[\min(\mathbf{x}_1, \mathbf{x}_2)]\,p[\max(\mathbf{x}_1, \mathbf{x}_2)] \geq p(\mathbf{x}_1)\,p(\mathbf{x}_2)$$

then the random vector $\mathbf{X}$ is Multivariate Totally Positive (MTP_2).

This defines a stochastic ordering amounting to a strong form of positive dependence, since it implies *Association* (A) and *Strongly Positive Orthant Dependence* (SPOD) [e.g., Holland and Rosenbaum (1986)]. This condition is relevant in many fields [see examples in Karlin and Rinott (1980)]. In the case of binary variables, it is applied within statistical mechanics [where it is called FKG; see Fortuin, Kasteleyn and Ginibgre (1971) and van den Berg and Gandolfi (1995)]; to the replacements algorithm in computer storage [van den Berg and Gandolfi (1992)]; and to *Item Response Theory* (IRT) models.

Received January 1999; revised June 2000.
[1]Supported by grants from the Italian MURST.
AMS 1991 *subject classifications*. Primary 62H20; secondary 62G10, 62H15, 62H17.
Key words and phrases. Chi-bar-squared distribution, conditional association, item response models, order-restricted inference, stochastic ordering.

IRT models are latent variable models developed for the analysis of the results of an aptitude test, made-up of dichotomously-scored *items*, assigned to a group of subjects [see Hambleton and Swaminathan (1985) for a general review and Junker and Ellis (1997) for a deeper discussion of the properties of these models]. In IRT, assumptions are typically made on the conditional distribution of the vector of responses $\mathbf{X}$ to a set of J items by a randomly chosen subject, given the levels of a set of latent abilities. Underlying these assumptions, are usually the non-parametric assumptions of *Local Independence* (LI) and *Unidimensional Monotonicity* (UM). An approach to maximum likelihood estimation in IRT models, utilizing *Latent Class Models* subject to the explicit constraints of LI and UM, has been attempted by Hoijtink and Molenaar (1997), who also propose several statistics for model checking. The connection between Latent Class Models and IRT models has also been investigated by Lindsay, Clogg and Grego (1991). Holland and Rosenbaum (1986) have shown that MTP_2 is a weaker version of *Conditional Association* (CA): violation of MTP_2, and hence of CA, implies that no monotone latent variable model can exist.

In this paper we provide a method for detecting *any possible* violation of MTP_2; in doing so we also outline a general approach for handling inequality constraints in log-linear models. In particular, after recalling that MTP_2 is equivalent to the constraint that all possible 2×2 conditional subtables within the multivariate binary distribution of $\mathbf{X}$ have non negative log-odds ratio, in Section 2 we show the equivalence of MTP_2 with the condition that the parameters of a saturated log-linear model belong to a convex cone. We also describe the additional constraints needed to obtain the CA ordering. Then, in Section 3, an extension of the approach of Dardanoni and Forcina (1998), based on iteratively reweighted least squares with linear inequality constraints, is used to construct an algorithm for maximum likelihood estimation. In Section 4, we show that the usual G^2 test statistics for independence can be partitioned into two components, one measuring departure from independence in the direction of the MTP_2 ordering and the other measuring violations of the same. We also derive the asymptotic distributions of these statistics which turn out to be distributions of the chi-bar-squared type, a mixture of chi-squared variables well known in the literature on multivariate one-sided testing [e.g., Perlman (1969) and Shapiro (1988)]. Though these distributions depend on nuisance parameters, we derive tight bounds on the resulting *p-values* which are very simple to use. Finally, in section 5, we use a small dataset to illustrate how our approach can be used within IRT models.

Any analysis of the joint probability distribution of $\mathbf{X}$, such as that described in this paper, is limited by the fact that the set of all possible configurations and the number of constraints induced by MTP_2 are of order 2^J; thus our approach is feasible only for small J. As such, it is still useful within *latent class models* where the number of items is usually small. This is also true of *capture-recapture data*, in the modeling of which LI and UM are also relevant issues [e.g., Darroch et al. (1993)]; our approach could handle these data with minor adjustments (see Section 2). With a large number of items one could select certain important violations and rely on multiple testing, as in the original work of Rosenbaum (1984). For instance, one could test for TP_2 in the bivariate distribution of the scores associated with two subsets of items [Dardanoni and Forcina (1998)].

2. Preliminary results. Assume that $p(\mathbf{X})$ is strictly positive for $\mathbf{X} \in \mathcal{X}$. The MTP_2 condition can be restated as follows. Denote, by $\mathcal{S}$, a nonempty subset of $\mathcal{J} = \{1, 2, \ldots, J\}$, by $\bar{\mathcal{S}}$, its complement and, by $\mathbf{X}(\mathcal{S})$ and $\mathbf{X}(\bar{\mathcal{S}})$, respectively, the vectors made of the elements of $\mathbf{X}$ in $\mathcal{S}$ and in $\bar{\mathcal{S}}$. It is well known [Karlin and Rinott (1980), page 469] that $\mathbf{X}$ is MTP_2 if and only if, for any $\mathcal{P} = \{j_1, j_2\} \subset \mathcal{J}$, the conditional distribution of $\mathbf{X}(\mathcal{P})$, given any value of $\mathbf{X}(\bar{\mathcal{P}})$, has non negative log-odds ratio, namely

$$\log \frac{p(X_{j_1} = 0, X_{j_2} = 0|\mathbf{X}(\bar{\mathcal{P}}))\, p(X_{j_1} = 1, X_{j_2} = 1|\mathbf{X}(\bar{\mathcal{P}}))}{p(X_{j_1} = 0, X_{j_2} = 1|\mathbf{X}(\bar{\mathcal{P}}))\, p(X_{j_1} = 1, X_{j_2} = 0|\mathbf{X}(\bar{\mathcal{P}}))} \geq 0.$$

It is useful to have a concise notation for the conditional log-odds ratios. For any given $\mathcal{P}$ and any subset $\mathcal{U}$ of $\bar{\mathcal{P}}$ (including the empty set and $\bar{\mathcal{P}}$ itself), we denote, by $\rho(\mathcal{P}, \mathcal{U})$, the log-odds ratio in the 2×2 subtable corresponding to the conditional distribution of $\mathbf{X}(\mathcal{P})$ given $X_j = 1, \forall j \in \mathcal{U}$ and $X_j = 0, \forall j \notin \mathcal{P} \cup \mathcal{U}$.

We can show that the MTP_2 condition can be expressed directly in terms of a suitable parameterization of a saturated log-linear model. Write $p(\mathbf{x})$ for $p(\mathbf{X} = \mathbf{x})$. Then, because $\sum_{\mathbf{x}} p(\mathbf{x}) = 1$, one of the $p(\mathbf{x})$, say $p(\mathbf{0})$, is redundant and so, by $\mathbf{p}$, we will denote the $(s-1) \times 1$ vector with elements arranged so that the components of $\mathbf{x}$ on the right run faster from 0 to 1, with the $\mathbf{x} = \mathbf{0}$ entry removed. A similar convention will be used for all related vectors and matrices. However, because of symmetry, certain calculations are simpler if based on the s-dimensional space so write $\dot{\mathbf{p}} = (\, p(\mathbf{0}) \quad \mathbf{p}'\,)'$ for the corresponding extended vector. More generally, we will use the convention that, if $\dot{\mathbf{v}}$ is a vector on the s-dimensional space, then $\mathbf{v}$ is the corresponding vector obtained by deleting the first element. This notation will apply also to matrices where the dot on top will be used to denote the addition of an initial row (or column or both, in a way made clear by the context) to conform with the full s-dimensional space.

Now let $\dot{\boldsymbol{\lambda}} = \log(\dot{\mathbf{p}})$ and define $\dot{\boldsymbol{\beta}}$ as

$$\dot{\boldsymbol{\beta}} = \dot{\mathbf{C}}\dot{\boldsymbol{\lambda}} \quad \text{where} \quad \dot{\mathbf{C}} = \underbrace{\mathbf{D} \otimes \cdots \otimes \mathbf{D}}_{J \text{ times}} \quad \text{and} \quad \mathbf{D} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}. \tag{2.1}$$

Each row of $\dot{\mathbf{C}}$, and the corresponding element of $\dot{\boldsymbol{\beta}}$, may be indexed with the subset $\mathcal{S}$ of $\mathcal{J}$ whose elements specify the positions of the *contrast vectors* $\mathbf{d}' = (\,-1 \quad 1\,)$ within the Kronecker product. Hence $\beta(\mathcal{S})$ may be interpreted as interactions among the variables with index in $\mathcal{S}$. The order of such interactions is given by the cardinality of $\mathcal{S}$.

LEMMA 1. *On the basis of the parametrization* (2.1),

$$\rho(\mathscr{P}, \mathscr{U}) = \left[\bigotimes_j \mathbf{v}_j(\mathscr{P}, \mathscr{U})\right] \dot{\boldsymbol{\beta}}, \tag{2.2}$$

where

$$\mathbf{v}_j(\mathscr{P}, \mathscr{U}) = \begin{cases} \mathbf{e}' = (0 \quad 1) = \mathbf{d}'\mathbf{E}, & \textit{if } j \in \mathscr{P}, \\ (1 \quad 1) = \mathbf{e}'\mathbf{E}, & \textit{if } j \in \mathscr{U}, \\ (1 \quad 0), & \textit{otherwise}, \end{cases} \quad \textit{and} \quad \mathbf{E} = \mathbf{D}^{-1} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

PROOF. From the definition of log-odds ratios and the ordering of the elements of $\dot{\boldsymbol{\lambda}}$,

$$\rho(\mathscr{P}, \mathscr{U}) = \left[\bigotimes_j \mathbf{u}_j(\mathscr{P}, \mathscr{U})\right] \dot{\boldsymbol{\lambda}}, \tag{2.3}$$

where

$$\mathbf{u}_j(\mathscr{P}, \mathscr{U}) = \begin{cases} \mathbf{d}' = (-1 \quad 1), & \text{if } j \in \mathscr{P}, \\ \mathbf{e}', & \text{if } j \in \mathscr{U}, \\ (1 \quad 0), & \text{otherwise.} \end{cases}$$

Thus, after inverting the linear transformation defining $\dot{\boldsymbol{\beta}}$, which gives

$$\dot{\boldsymbol{\lambda}} = \dot{\mathbf{Z}}\dot{\boldsymbol{\beta}} \quad \text{where} \quad \dot{\mathbf{Z}} = \dot{\mathbf{C}}^{-1} = \underbrace{\mathbf{E} \otimes \cdots \otimes \mathbf{E}}_{J \text{ times}},$$

the result follows by substitution. □

The vector premultiplying $\dot{\boldsymbol{\beta}}$ in (2.2) has all elements equal to 0 except for $2^{|\mathscr{U}|}$ elements equal to 1, where $|\mathscr{U}|$ is the cardinality of $\mathscr{U}$. So (2.2) may also be written as

$$\rho(\mathscr{P}, \mathscr{U}) = \sum_{\mathscr{L} \subseteq \mathscr{U}} \beta(\mathscr{P} \cup \mathscr{L})$$

[recall that $\beta(\mathscr{P} \cup \mathscr{L})$ represents the interaction among the variables with index in $\mathscr{P}$ or $\mathscr{L}$]. To see this let, without loss of generality, $\mathscr{P} = (J-1, J)$. Then $\rho(\mathscr{P}, \varnothing)$ is simply $\beta(\mathscr{P})$, which is the 4th element of $\dot{\boldsymbol{\beta}}$, $\rho(\mathscr{P}, \{J-2\}) = \beta(\mathscr{P}) + \beta[\mathscr{P}, (J-2)]$ and so on. Thus the MTP_2 condition places restrictions only on second and higher order interactions.

To express the relation between each $\rho(\mathscr{P}, \mathscr{U})$ and $\boldsymbol{\beta}$ in matrix notation, it is convenient to construct first a vector, say $\boldsymbol{\rho}(\mathscr{P})$ with $\mathscr{P} = \{j_1, j_2\}$, containing all the conditional log-odds ratios between X_{j_1} and X_{j_2} arranged so that values of the remaining variables go from 0 to 1, with those on the right moving faster. This is equivalent to stacking the vectors $(1 \quad 0)$ and $(1 \quad 1)$ one below the other, so that

$$\boldsymbol{\rho}(\mathscr{P}) = \dot{\mathbf{R}}(\mathscr{P})\dot{\boldsymbol{\beta}} \quad \text{with} \quad \dot{\mathbf{R}}(\mathscr{P}) = \bigotimes_j \mathbf{V}_j(\mathscr{P}),$$

where $\mathbf{V}_j(\mathscr{P})$ is $\mathbf{e}'$ if $j \in \mathscr{P}$ and $\mathbf{E}$ otherwise.

Now place the $u = J(J-1)/2$ vectors $\boldsymbol{\rho}(\mathscr{P})$ one below the other, ordered so that the second element of $\mathscr{P}$ is greater than the first and runs faster from J to $j_1 + 1$, into the single vector $\boldsymbol{\rho}$. Consider then the matrix $\dot{\mathbf{R}}$ with blocks of rows $\dot{\mathbf{R}}(\mathscr{P})$ arranged in the same way used for $\boldsymbol{\rho}$. In the case of $J = 4$, for instance, we have that

$$\boldsymbol{\rho}(\mathscr{P}) = \begin{pmatrix} \boldsymbol{\rho}(\{3,4\}) \\ \boldsymbol{\rho}(\{2,4\}) \\ \boldsymbol{\rho}(\{2,3\}) \\ \boldsymbol{\rho}(\{1,4\}) \\ \boldsymbol{\rho}(\{1,3\}) \\ \boldsymbol{\rho}(\{1,2\}) \end{pmatrix}, \qquad \dot{\mathbf{R}} = \begin{pmatrix} \mathbf{E} & \otimes & \mathbf{E} & \otimes & \mathbf{e}' & \otimes & \mathbf{e}' \\ \mathbf{E} & \otimes & \mathbf{e}' & \otimes & \mathbf{E} & \otimes & \mathbf{e}' \\ \mathbf{E} & \otimes & \mathbf{e}' & \otimes & \mathbf{e}' & \otimes & \mathbf{E} \\ \mathbf{e}' & \otimes & \mathbf{E} & \otimes & \mathbf{E} & \otimes & \mathbf{e}' \\ \mathbf{e}' & \otimes & \mathbf{E} & \otimes & \mathbf{e}' & \otimes & \mathbf{E} \\ \mathbf{e}' & \otimes & \mathbf{e}' & \otimes & \mathbf{E} & \otimes & \mathbf{E} \end{pmatrix}.$$

Then the MTP_2 condition can be expressed as

$$\boldsymbol{\rho} = \dot{\mathbf{R}}\dot{\boldsymbol{\beta}} \geq \mathbf{0}. \tag{2.4}$$

Note that the first column of $\dot{\mathbf{R}}$ is $\mathbf{0}$ and so condition (2.4) can be restated as $\mathbf{R}\boldsymbol{\beta} \geq \mathbf{0}$ which defines the convex cone $\mathscr{C}$. Although the number of inequalities is much larger than the dimension of the $\boldsymbol{\beta}$ space, different inequalities are active in different regions of the space and so there are no redundancies. To see this, for any given $\mathscr{P}_a$ and $\mathscr{U}_a$, take $0 < k < l$ and set

$$\beta(\mathscr{P} \cup \mathscr{L}) = \begin{cases} -k, & \text{if } \mathscr{P} = \mathscr{P}_a \text{ and } \mathscr{L} = \mathscr{U}_a, \\ 0, & \text{if } \mathscr{P} = \mathscr{P}_a \text{ and } \mathscr{L} \subset \mathscr{U}_a, \\ l, & \text{otherwise.} \end{cases}$$

Then $\rho(\mathscr{P}_a, \mathscr{U}_a) = \beta(\mathscr{P}_a \cup \mathscr{U}_a) = -k < 0$, whereas any other $\rho(\mathscr{P}, \mathscr{U})$ is non negative.

For binary variables, the stronger notion of CA requires non negative log-odds ratios for any pair of items $\mathscr{P}$, conditional on the values of $\mathbf{h}[\mathbf{X}(\bar{\mathscr{P}})]$, where $\mathbf{h}$ is any arbitrary function [Holland and Rosenbaum (1986)]. In practice, for any $\mathscr{P}$, in addition to the 2^{J-2} distinct configurations of the conditioning set (which are used in MTP_2), one has to condition also on the distribution obtained by marginalizing with respect to all subsets of the conditioning set with size greater than 1. These additional constraints can be written as $[\mathbf{I}_v \otimes (1 \quad -1 \quad -1 \quad 1)] \log(\mathbf{B}\dot{\mathbf{p}}) \geq \mathbf{0}$, where $v = \frac{J(J-1)}{2}[2^{2^{J-2}} - (2^{J-2}+1)]$ and $\mathbf{B}\dot{\mathbf{p}}$ are v different two-way marginals stacked one below the other. The number of constraints is huge for even small values of J (almost 1 million with $J = 6$).

The above approach could be extended to the capture-recapture context by letting $X_j = 0$ if a subject is captured and 1 otherwise. In this way the missing cell corresponds to the $\mathbf{X} = \mathbf{1}$ configuration and, because $p(\mathbf{1})$ is used only in the computation of $\beta(\mathscr{J})$, which is not identified, to define the cone of identifiable restrictions induced by MTP_2, one need simply remove from $\mathbf{R}$ the last column and all the rows involving $\beta(\mathscr{J})$.

3. Maximum likelihood estimation. With a sample of n observations, let $y(\mathbf{x})$ be the frequency of the response pattern $\mathbf{x}$ and let $\mathbf{y}$ be the vector with elements $\{y(\mathbf{x})\}$ arranged in the same way as in $\mathbf{p}$. It is convenient to introduce the canonical parameters $\phi(\mathbf{x}) = \lambda(\mathbf{x}) - \lambda(\mathbf{0})$; thus $p(\mathbf{0}) = [1 + \mathbf{1}' \exp(\boldsymbol{\phi})]^{-1}$ and $p(\mathbf{x}) = p(\mathbf{0}) \exp[\phi(\mathbf{x})]$, where $\boldsymbol{\phi} = \{\phi(\mathbf{x})\}$ with elements arranged in the usual way.

By construction, the first row of $\dot{\mathbf{C}}$ is $(1 \quad \mathbf{0}')$, while the block of remaining rows being contrasts may be written as $(\mathbf{c} \quad \mathbf{C})$, with $\mathbf{c} = -\mathbf{C}\mathbf{1}$. So, by substitution,

$$\boldsymbol{\beta} = \lambda(\mathbf{0})\mathbf{c} + \mathbf{C}\boldsymbol{\lambda} = -\lambda(\mathbf{0})\mathbf{C}\mathbf{1} + \mathbf{C}\boldsymbol{\lambda} = \mathbf{C}[\boldsymbol{\lambda} - \mathbf{1}\lambda(\mathbf{0})] = \mathbf{C}\boldsymbol{\phi}$$

and the log-linear model can be rewritten in terms of the canonical parameters, as $\boldsymbol{\phi} = \mathbf{Z}\boldsymbol{\beta}$ with $\mathbf{Z} = \mathbf{C}^{-1}$. Maximum likelihood estimates of such a model, subject to the MTP_2 condition, can be obtained by a constrained Fisher-scoring algorithm similar to the one described in Dardanoni and Forcina (1998). The basic idea is to approximate the proper log-likelihood locally at each step using a quadratic function with the same first derivative and the same information matrix.

More precisely, we want to maximize the multinomial log-likelihood

$$L(\boldsymbol{\beta}; \mathbf{y}) = \dot{\mathbf{y}}' \log(\dot{\mathbf{p}}) + \text{constant} = \mathbf{y}'\boldsymbol{\phi} - n \log[1 + \mathbf{1}' \exp(\boldsymbol{\phi})] + \text{constant}$$

subject to the constraint $\boldsymbol{\beta} \in \mathscr{C}$.

The score vector and the information matrix have, respectively, the form

$$\mathbf{s} = \frac{\partial L}{\partial \boldsymbol{\beta}} = \frac{\partial \boldsymbol{\phi}'}{\partial \boldsymbol{\beta}} \frac{\partial L}{\partial \boldsymbol{\phi}} = \mathbf{Z}'(\mathbf{y} - n\mathbf{p}),$$

$$\mathbf{H} = -\frac{1}{n} \frac{\partial}{\partial \boldsymbol{\beta}} \left(\frac{\partial L}{\partial \boldsymbol{\beta}} \right)' = \mathbf{Z}'\mathbf{F}\mathbf{Z},$$

where $\mathbf{F} = \operatorname{diag}(\mathbf{p}) - \mathbf{p}\mathbf{p}'$. To construct a quadratic approximation to $L(\boldsymbol{\beta}; \mathbf{y})$ at $\boldsymbol{\beta} = \boldsymbol{\beta}_m$, first define the *pseudo dependent variable* $\mathbf{b}_m = \boldsymbol{\beta}_m + \mathbf{Z}^{-1}\mathbf{F}_m^{-1}(\mathbf{y}/n - \mathbf{p}_m)$ and then let

$$Q(\mathbf{b}_m, \boldsymbol{\beta}) = -\frac{n}{2}(\mathbf{b}_m - \boldsymbol{\beta})'\mathbf{H}_m(\mathbf{b}_m - \boldsymbol{\beta}).$$

The *Constrained Fisher-Scoring algorithm* (*CFS*) works as follows: (i) set the starting value $\boldsymbol{\beta}_0 = \mathbf{Z}^{-1} \log[\mathbf{y}/y(\mathbf{0})]$ (the unrestricted maximum likelihood estimate), (ii) at step $m + 1$, maximize $Q(\mathbf{b}_m, \boldsymbol{\beta})$ subject to $\boldsymbol{\beta} \in \mathscr{C}$, (iii) iterate until convergence.

By an argument similar to the one used in Dardanoni and Forcina (1998), it can be shown that, if the elements of $\dot{\mathbf{p}}$ are strictly positive, CFS converges to the maximum of L, which is unique because $\mathbf{R}\boldsymbol{\beta} \geq \mathbf{0}$ defines a convex cone and $-n\mathbf{H}$, the matrix of second derivatives, is negative definite. If $\boldsymbol{\beta}_0 \in \mathscr{C}$, then $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_0$ and the algorithm stops. Otherwise updated estimates will lie on the boundary and satisfy the orthogonality condition $\boldsymbol{\beta}'_{m+1}\mathbf{H}_m(\mathbf{b}_m - \boldsymbol{\beta}_{m+1}) = 0$ so that, at convergence, we have

$$\boldsymbol{\beta}'_m\mathbf{H}_m(\mathbf{b}_m - \boldsymbol{\beta}_m) = \boldsymbol{\beta}'_m\mathbf{s}_m = 0$$

because the first derivative of $Q(\mathbf{b}_m, \boldsymbol{\beta})$ at $\boldsymbol{\beta}_m$ equals $\mathbf{s}_m$. This implies that $\boldsymbol{\beta}_m$ is the maximum of L within the linear subspace orthogonal to the direction of steepest ascent. To implement the algorithm one has to solve, at each step, a constrained least squares problem; for this purpose, several reliable procedures are available in the literature [e.g., Lawson and Hanson (1995) or Dykstra (1983)].

A computer program that implements the algorithm in MATLAB is available from the authors on request. Our experience is that the algorithm is extremely stable and capable of reaching convergence in a few steps. Special care is needed in dealing with zeros in the vector $\mathbf{y}$. The main issue concerns the computation of the starting values. In this respect recall that, due to the concavity of the likelihood function and the convexity of the parameter space, starting values can only affect the efficiency of the algorithm, not the results at convergence. However, especially with sparse data, replacing zeros with a very small value would be equivalent to starting the algorithm near the boundary of the parameter space and this, apart from being rather inefficient, is likely to cause numerical difficulties. So, our recommendation in this respect is to replace zeros with something close to 1 (0.5 or larger, depending on the sparseness of the data).

4. Hypothesis testing.

Let H_0 denote the hypothesis of independence in the distribution of $\mathbf{X}$ which is equivalent to $\mathbf{R}\boldsymbol{\beta} = \mathbf{0}$, H_P the hypothesis that $\boldsymbol{\beta} \in \mathscr{C}$, H_U the hypothesis that $\boldsymbol{\beta}$ is unrestricted. The approach proposed here for testing H_P against H_U is based on the likelihood ratio test.

Let $L_h(\mathbf{y})$, with $h = 0, P, U$, denote the maximum log-likelihood under H_h. Then the G^2 statistics for testing independence may be partitioned as

$$\begin{aligned} G^2 = 2[L_U(\mathbf{y}) - L_0(\mathbf{y})] &= 2[L_U(\mathbf{y}) - L_P(\mathbf{y})] + 2[L_P(\mathbf{y}) - L_0(\mathbf{y})] \\ &= T_{PU} + T_{0P} \text{ (say)} \end{aligned}$$

where T_{0P} and T_{PU} are two measure of discrepancy, respectively, against H_0 in the direction of H_P and against H_P in the direction of H_U. Notice also that the log-likelihood for the unrestricted model is undefined when one or more observations are 0 and the usual approach is to replace zeros with a small value ϵ, say 10^{-8}. However, when the data are very sparse, the asymptotic results that we derive in this section have to be used carefully as they will no longer provide a reliable approximation to the distribution of the test statistics.

To derive the asymptotic distribution of T_{PU} and T_{0P} under H_0 recall that, in such a case, the asymptotic distribution of $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$, where $\hat{\boldsymbol{\beta}}$ is the unconstrained ML estimate of $\boldsymbol{\beta}$, is $\mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = \mathbf{C}\mathbf{F}^{-1}\mathbf{C}'$. For any cone (or linear space) $\mathscr{V}$, let $\bar{\mathscr{V}}$ denote its dual (or orthogonal complement) in the $\boldsymbol{\Sigma}^{-1}$ metric. Moreover, for $\mathbf{v} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma})$, let $\hat{\mathbf{v}}(\mathscr{V}, \boldsymbol{\Sigma})$ be its projection onto $\mathscr{V}$ again in the $\boldsymbol{\Sigma}^{-1}$ metric, namely, the solution to the problem $\min_{\hat{\mathbf{v}} \in \mathscr{V}} (\mathbf{v} - \hat{\mathbf{v}})'\boldsymbol{\Sigma}^{-1}(\mathbf{v} - \hat{\mathbf{v}})$. Let $\mathscr{H}$ denote the linear space $\{\boldsymbol{\beta} : \mathbf{R}\boldsymbol{\beta} = \mathbf{0}\}$ and $\mathscr{K} = \mathscr{C} \cap \bar{\mathscr{H}}$.

THEOREM 1. *Under independence, when n increases while J remains constant, T_{PU} converges in distribution to*

$$Q_{\bar{\mathscr{C}}} = [\mathbf{v} - \hat{\mathbf{v}}(\mathscr{C}, \boldsymbol{\Sigma})]'\boldsymbol{\Sigma}^{-1}[\mathbf{v} - \hat{\mathbf{v}}(\mathscr{C}, \boldsymbol{\Sigma})] = \hat{\mathbf{v}}(\bar{\mathscr{C}}, \boldsymbol{\Sigma})'\boldsymbol{\Sigma}^{-1}\hat{\mathbf{v}}(\bar{\mathscr{C}}, \boldsymbol{\Sigma}) \sim \bar{\chi}^2(\bar{\mathscr{C}}, \boldsymbol{\Sigma}),$$

while T_{0P} converges in distribution to

$$\begin{aligned} Q_{\mathscr{K}} &= [\mathbf{v} - \hat{\mathbf{v}}(\mathscr{H}, \boldsymbol{\Sigma})]'\boldsymbol{\Sigma}^{-1}[\mathbf{v} - \hat{\mathbf{v}}(\mathscr{H}, \boldsymbol{\Sigma})] - Q_{\bar{\mathscr{C}}} \\ &= \hat{\mathbf{v}}(\mathscr{K}, \boldsymbol{\Sigma})'\boldsymbol{\Sigma}^{-1}\hat{\mathbf{v}}(\mathscr{K}, \boldsymbol{\Sigma}) \sim \bar{\chi}^2(\mathscr{K}, \boldsymbol{\Sigma}). \end{aligned}$$

PROOF. The convergence of T_{PU} to $Q_{\bar{\mathscr{C}}}$ may be derived from Theorem 2.1 in Shapiro (1985) because the log-likelihood ratio, as can be easily verified, belongs to the general class of *discrepancy functions* for which the result holds. A more direct derivation is contained in Wolak [(1989a), Section 3] and is based on a second order Taylor series expansion of the log-likelihood. The distribution of the quadratic form $Q_{\bar{\mathscr{C}}}$ follows from standard results on the $\bar{\chi}^2$ distribution and the theory concerning projection of normal vectors onto convex cones [see, e.g., Shapiro (1988) and Wolak (1989b) for details].

The convergence of T_{0P} to $Q_{\mathscr{K}}$ can be deduced from Shapiro [(1988), page 52] by recalling that G^2 converges in distribution to

$$Q_{\mathscr{H}} = [\mathbf{v} - \hat{\mathbf{v}}(\mathscr{H}, \boldsymbol{\Sigma})]'\boldsymbol{\Sigma}^{-1}[\mathbf{v} - \hat{\mathbf{v}}(\mathscr{H}, \boldsymbol{\Sigma})] \sim \chi^2_t,$$

where $t = s - J - 1$. □

This is how the *p-value* corresponding to an observed value of T_{PU}, say c, could be computed. The first part of Theorem 1 and the survival function of the $\bar{\chi}^2$ distribution imply that

$$\lim_{n\to\infty} P(T_{PU} > c) = \sum_0^t w_j(\bar{\mathscr{C}}, \boldsymbol{\Sigma}) P(\chi^2_j > c),$$

where $w_j(\bar{\mathscr{C}}, \boldsymbol{\Sigma})$ is the probability that the projection of $\mathbf{v} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma})$ onto $\bar{\mathscr{C}}$, the dual of $\mathscr{C}$, is contained on a face of dimension j. Assume, for the moment, that the value of $\boldsymbol{\beta}$, and hence $\mathbf{p}$ under H_0, was known; though there is no explicit formula for computing the probability weights w_j with $t > 3$, they can be estimated to any desired accuracy by projecting a sufficiently large number of pseudo-randomly generated normal vectors onto $\bar{\mathscr{C}}$ [Dardanoni and Forcina (1998), page 1117].

The dependence of the null distribution on the value of unknown parameters poses a more serious problem. One possible approach would be to replace the unknown parameters with their ML estimates. Although this procedure does not guarantee that the actual significance level will not exceed the size of the test, we would expect that, if n is reasonably large relative to J, the difference between the nominal and the actual significance level will not be appreciable in most cases.

A formally correct approach would require knowledge of the least favorable distribution of T_{PU}, which is the one giving the largest *p-value*. Such a distribution, which is determined in Theorem 2, by being conservative relative to the null H_P, leads to the most frequent acceptance of MTP_2. In contexts where a greater caution relative to H_P is required, one should search for a distribution of T_{PU} which is the smallest distribution under H_0; this is also derived in Theorem 2. One could go even further in this direction by searching for the smallest distribution of T_{PU} within the entire boundary of H_P. Such an extreme distribution would be achieved when all elements in $\mathbf{R}\boldsymbol{\beta}$ are strictly positive except for one which is exactly 0; although it can be shown that such a $\boldsymbol{\beta}$ exists, in this case T_{PU} would be distributed as a mixture with weights (1/2) to the constant 0 and a χ_1^2 variable; this would lead to a too severe procedure with the MTP_2 condition almost always being rejected.

Before introducing Theorem 2, we need two intermediate results, in the first of which, under H_0, an explicit Choleski decomposition for $\boldsymbol{\Sigma}$ under H_0 is derived. This exploits the fact that $\dot{\mathbf{p}} = \otimes \mathbf{p}_j$, where $p_j = P(X_j = 1)$, $q_j = 1 - p_j$ and $\mathbf{p}_j = (\, q_j \quad p_j\,)'$. Moreover, to work with Kronecker products, we formally define $\dot{\boldsymbol{\Sigma}} = \dot{\mathbf{C}}\,\mathrm{diag}(\dot{\mathbf{p}})^{-1}\dot{\mathbf{C}}$ and show that $\boldsymbol{\Sigma}$ may be obtained from $\dot{\boldsymbol{\Sigma}}$ by removing the first row and the first column.

LEMMA 2. *Under H_0, $\boldsymbol{\Sigma}$ may be obtained from $\dot{\boldsymbol{\Sigma}}$ by removing the first row and the first column. Moreover $\dot{\boldsymbol{\Sigma}}$ factorizes as $\dot{\mathbf{L}}\dot{\mathbf{L}}'$ where the upper triangular matrix $\dot{\mathbf{L}}$ is equal to $\bigotimes_j \mathbf{L}_j$ with*

$$\mathbf{L}_j = \begin{pmatrix} 1 & -\sqrt{p_j/q_j} \\ 0 & 1/\sqrt{p_j q_j} \end{pmatrix};$$

by removing the first row and first column from $\dot{\mathbf{L}}$, we have $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}'$.

PROOF. Recall that the last $s-1$ rows of $\dot{\mathbf{C}}$ may be written as $\mathbf{C}\,(\,-\mathbf{1} \quad \mathbf{I}\,)$ and so the matrix obtained by removing the first row and the first column from $\dot{\boldsymbol{\Sigma}}$ is

$$\mathbf{C}\,(\,-\mathbf{1} \quad \mathbf{I}\,)\begin{pmatrix} 1/p(\mathbf{0}) & \mathbf{0}' \\ \mathbf{0} & \mathrm{diag}(\mathbf{p})^{-1} \end{pmatrix}\begin{pmatrix} -\mathbf{1}' \\ \mathbf{I} \end{pmatrix}\mathbf{C}' = \mathbf{C}\mathbf{F}^{-1}\mathbf{C}' = \boldsymbol{\Sigma}.$$

Under H_0, both $\dot{\mathbf{C}}$ and $\dot{\mathbf{p}}$ factorize into a Kronecker product so that $\dot{\boldsymbol{\Sigma}} = \bigotimes \boldsymbol{\Sigma}_j$, where $\boldsymbol{\Sigma}_j = \mathbf{D}\,\mathrm{diag}(\mathbf{p}_j)^{-1}\mathbf{D}' = \mathbf{L}_j\mathbf{L}_j'$. Thus $\dot{\boldsymbol{\Sigma}}$ is also equal to $\dot{\mathbf{L}}\dot{\mathbf{L}}'$. Since the first column of $\dot{\mathbf{L}}$ is $(\,1 \quad \mathbf{0}'\,)'$, $\boldsymbol{\Sigma}$ may be written as $\mathbf{L}\mathbf{L}'$ where $\mathbf{L}$ is also upper triangular. $\square$

The following lemma concerns transformation of the cone $\mathscr{C}$ into the Euclidean metric:

LEMMA 3. *The matrix $\dot{\mathbf{R}}\dot{\mathbf{L}}$ may be written as $\mathrm{diag}(\mathbf{a})\dot{\mathbf{R}}_p$ where $\mathbf{a}$ is a vector of positive constants and $\dot{\mathbf{R}}_p$ has the same structure as $\dot{\mathbf{R}}$ except that, in the*

Kronecker product within each block of rows, any $\mathbf{E}$ *in the* j*th position is replaced by the orthogonal matrix*

$$\mathbf{M}_j = \begin{pmatrix} \sqrt{q_j} & -\sqrt{p_j} \\ \sqrt{p_j} & \sqrt{q_j} \end{pmatrix}.$$

Then

$$Q_{\bar{\mathscr{C}}} = [\mathbf{z} - \hat{\mathbf{z}}(\mathscr{E})]'[\mathbf{z} - \hat{\mathbf{z}}(\mathscr{E})] = \hat{\mathbf{z}}(\bar{\mathscr{E}})'\hat{\mathbf{z}}(\bar{\mathscr{E}}) \;\sim\; \bar{\chi}^2(\bar{\mathscr{E}}),$$

where $\mathbf{z} \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$, $\mathscr{E}$ *is the cone defined by* $\{\mathbf{z} : \mathbf{R}_p\mathbf{z} \geq \mathbf{0}\}$ *and* $\hat{\mathbf{z}}(\mathscr{E})$, $\hat{\mathbf{z}}(\bar{\mathscr{E}})$ *are the projections of* $\mathbf{z}$ *onto* $\mathscr{E}$ *and* $\bar{\mathscr{E}}$, *respectively, in the Euclidean metric.*

PROOF. Each block of rows of $\dot{\mathbf{R}}\dot{\mathbf{L}}$ is the Kronecker product of two terms of the form $\mathbf{e}'\mathbf{L}_j = (q_j p_j)^{-1/2}\mathbf{e}'$ and $s-2$ terms of the form

$$\mathbf{E}\mathbf{L}_j = \begin{pmatrix} 1 & -\sqrt{p_j/q_j} \\ 1 & \sqrt{q_j/p_j} \end{pmatrix} = \operatorname{diag}(\mathbf{p}_j)^{-\frac{1}{2}}\mathbf{M}_j.$$

Thus, apart from scaling constants, $\mathbf{e}'$ turns into itself and $\mathbf{E}$ into $\mathbf{M}_j$.

Then apply the linear transformation $\mathbf{z} = \mathbf{L}^{-1}\mathbf{v} \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$ and recall that $\boldsymbol{\Sigma}^{-1} = (\mathbf{L}^{-1})'\mathbf{L}^{-1}$,

$$Q_{\bar{\mathscr{E}}} = \min_{\mathbf{R}\hat{\mathbf{v}}\geq\mathbf{0}}(\mathbf{v} - \hat{\mathbf{v}})'\boldsymbol{\Sigma}^{-1}(\mathbf{v} - \hat{\mathbf{v}}) = \min_{\mathbf{R}_p\hat{\mathbf{z}}\geq\mathbf{0}}(\mathbf{z} - \hat{\mathbf{z}})'(\mathbf{z} - \hat{\mathbf{z}}) = [\mathbf{z} - \hat{\mathbf{z}}(\mathscr{E})]'[\mathbf{z} - \hat{\mathbf{z}}(\mathscr{E})].$$

The result follows because, in a system of linear inequalities defining a convex cone, each row may be multiplied by arbitrary positive constants. □

THEOREM 2. *Under* H_0 *and for any* $c > 0$,

$$P(\bar{\chi}^2(\mathscr{O}_t) \geq c) \leq \lim_{n\to\infty} P(T_{PU} \geq c) \leq P(\chi^2_{t-u} + \bar{\chi}^2(\mathscr{O}_u) \geq c),$$

where $u = J(J-1)/2$, $t = s - J - 1$, $\mathscr{O}_t$ *and* $\mathscr{O}_u$ *are the positive orthants in* $\mathbb{R}^t$ *and* $\mathbb{R}^u$, *respectively, and the covariance matrix in the* $\bar{\chi}^2$ *distributions is the identity matrix.*

REMARK 1. Recall that, when the cone is the positive orthant and the metric is Euclidean, the probability weights of the $\bar{\chi}^2$ distribution are the probabilities of the symmetric binomial distribution with index equal to the dimension of the space.

PROOF OF THEOREM 2. Asymptotically, T_{PU} tends to $Q_{\bar{\mathscr{C}}} = \hat{\mathbf{z}}(\bar{\mathscr{E}})'\hat{\mathbf{z}}(\bar{\mathscr{E}})$, where $\mathbf{z} \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$ and $\mathscr{E} = \{\mathbf{z} : \mathbf{R}_p\mathbf{z} \geq \mathbf{0}\}$. Because $\bar{\mathscr{E}}$, the dual cone of $\mathscr{E}$, is spanned by the columns of $-\mathbf{R}_p'$ and this is equivalent to the cone spanned by the columns of $-\dot{\mathbf{R}}_p'$ (obtained simply by adding a first row of zeros), we can switch temporarily to the s-dimensional space.

To show that $\bar{\mathscr{E}}$ always contains a (suitably rotated) orthant of dimension t, we give a rule to construct, for every $\boldsymbol{\beta}$, a matrix of non negative constants $\mathbf{P}$ and an orthogonal matrix $\mathbf{Q}$, such that $\mathbf{P}\dot{\mathbf{R}}_p\mathbf{Q} = (\mathbf{O} \quad \mathbf{I}_t)$. The matrix $\mathbf{P}$ will be block-diagonal with u blocks. Let $\mathscr{P} = \{j_1, j_2\}$ be a pair of indices from $\mathscr{J}$, with $j_1 < j_2$. The block $\mathbf{P}(\mathscr{P})$ is defined as $\bigotimes_j \mathbf{P}_j(\mathscr{P})$, where

$$\mathbf{P}_j(\mathscr{P}) = \begin{cases} 1, & \text{if } j \in \mathscr{P}, \\ \mathbf{I}_2, & \text{if } j > j_2, \\ \mathbf{a}'_j = (\sqrt{q_j} \quad \sqrt{p_j}), & \text{otherwise.} \end{cases}$$

The block $\mathbf{P}(\mathscr{P})$ has 2^{J-j_2} rows and, because for a given element j of $\mathscr{J}$ there are $j-1$ pairs of indices which have j as second element, the number of rows of $\mathbf{P}$ is

$$\sum_{j=2}^{J} 2^{J-j}(j-1) = \sum_{k=1}^{J-1}(2^k - 1) = 2^J - J - 1 = t.$$

$\mathbf{P}\dot{\mathbf{R}}_p$ also has u blocks of rows; let $\mathbf{S}(\mathscr{P})$ be any of these which has the form $\bigotimes_j \mathbf{S}_j(\mathscr{P})$, where

$$\mathbf{S}_j(\mathscr{P}) = \begin{cases} \mathbf{e}', & \text{if } j \in \mathscr{P}, \\ \mathbf{M}_j, & \text{if } j > j_2, \\ \mathbf{u}' = (1 \quad 0), & \text{otherwise.} \end{cases}$$

Then it is sufficient to set $\mathbf{Q} = (\mathbf{Q}_1 \quad \mathbf{Q}_2)$, where $\mathbf{Q}_2 = \dot{\mathbf{R}}'_p\mathbf{P}'$ and $\mathbf{Q}_1$ is any matrix such that $\mathbf{Q}'_1\mathbf{Q}_2 = \mathbf{O}$ and $\mathbf{Q}'_1\mathbf{Q}_1 = \mathbf{I}$. To see why the columns of $\dot{\mathbf{R}}'_p\mathbf{P}'$ are orthogonal, let $\mathbf{S}(\mathscr{P})\mathbf{S}(\mathscr{P}')'$ be equal to $\bigotimes_j \mathbf{T}_j(\mathscr{P}, \mathscr{P}')$. It is easily verified that, for $\mathscr{P} = \mathscr{P}'$,

$$\mathbf{T}_j(\mathscr{P}, \mathscr{P}') = \begin{cases} \mathbf{I}_2, & \text{if } j > j_2, \\ 1, & \text{otherwise,} \end{cases}$$

while, for $\mathscr{P} \neq \mathscr{P}'$, $\mathbf{T}_j(\mathscr{P}, \mathscr{P}') = 0$ whenever $j \in (\mathscr{P} \cup \mathscr{P}')/(\mathscr{P} \cap \mathscr{P}')$, a set which is never empty.

To show that $\bar{\mathscr{E}}$ is always contained is a cone spanned by u halflines and $t-u$ axes let, $\mathscr{G}_k$ be the set of indices of the positions in the vector $\boldsymbol{\beta}$ of the interactions of order k. It follows easily from the construction of $\dot{\mathbf{R}}_p$ that its lth column is equal to $\mathbf{0}$ when $l \in \mathscr{G}_1$, has only positive and zero elements for $j \in \mathscr{G}_2$ (involving the first column of $\mathbf{M}_j$) and has both positive and negative elements otherwise. This completes the proof. □

COROLLARY 1. *The lower bound can be achieved in the limit if either* $p_j \to 1, \forall j$ *or* $p_j \to 0, \forall j$. *The upper bound cannot be achieved, except when* $J \leq 4$; *a tight upper bound is given by* $\chi^2_{t-g} + \bar{\chi}^2(\mathscr{O}_g)$, *where*

$$g = u + [2^G - 1 - G - G(G-1)/2] \\ + [2^{J-G} - 1 - J + G - (J-G)(J-G-1)/2],$$

with $G = J/2$ *if* J *is even and* $(J-1)/2$ *otherwise.*

PROOF. Both extreme distributions require that the columns of $-\dot{\mathbf{R}}_p'$ (which span $\bar{\mathscr{E}}$) converge to vectors with only one non zero element; let us call these *unitary vectors*. For this to happen, each p_j must tend to either 0 or 1 so that the corresponding $\mathbf{M}_j'$ matrix converges to $(\mathbf{u} \quad \mathbf{e})$ or $(-\mathbf{e} \quad \mathbf{u})$, respectively. If we recall the construction of $\dot{\mathbf{R}}$ (and hence $\dot{\mathbf{R}}_p$), it is clear that each unitary vector, being essentially a Kronecker product of $\mathbf{u}$ and $\mathbf{e}$ vectors, is characterized by the pair $\mathscr{P}$ indexing the block of columns to which it belongs and the set of indices $\mathscr{S}$ for which an $\mathbf{e}$ vector is used; so write $\mathbf{u}(\mathscr{P}, \mathscr{S})$ with $\mathscr{P} \subseteq \mathscr{S}$ for a single unitary vector. Then two cases are possible for a given $\mathscr{S}$: (i) all $\mathbf{u}(\mathscr{P}, \mathscr{S})$ have the same sign irrespective of $\mathscr{P}$ and so they define a half-line or (ii) unitary vectors with plus and minus signs appear in different blocks defining a full line. Also, by construction, $\mathbf{u}(\mathscr{P}, \mathscr{P})$ appears in a single block and is negative.

Clearly, if $p_j \to 0$, $\forall j$, all unitary vectors are negative and the lower bound follows because $\bar{\chi}^2(\mathscr{O}_t) = \bar{\chi}^2(-\mathscr{O}_t)$. For the upper bound, we need to establish when the largest number of full lines may be obtained. Let $\mathscr{J}_0$ and $\mathscr{J}_1$ be, respectively, the set of indices for which $p_j \to 0$ and $p_j \to 1$. It is easy to see that the sign of $\mathbf{u}(\mathscr{P}, \mathscr{S})$ is independent of $\mathscr{P}$ whenever $\mathscr{S} \subseteq \mathscr{J}_0$ (negative sign) or $\mathscr{S} \subseteq \mathscr{J}_1$ [the sign depends on the number of times that the first column of $(-\mathbf{e} \quad \mathbf{u})$ is used and this equals the cardinality of $\mathscr{S}/\mathscr{P}$]. Otherwise the sign will be negative or positive depending on whether the cardinality of $(\mathscr{S} \cap \mathscr{J}_1)/\mathscr{P}$ is even or odd and this depends on the number of elements of $\mathscr{P}$ that belong to $\mathscr{J}_1$. Thus the number of full lines is maximized by minimizing the number of sets $\mathscr{S}$ which are entirely contained in either $\mathscr{J}_0$ or $\mathscr{J}_1$ and this happens when they have $J/2$ elements each with J even or $(J+1)/2$ and $(J-1)/2$ with J odd. $\square$

5. An application. As an illustration, we apply the methods discussed in this paper to a dataset concerning the responses of $n = 150$ students to a test made-up of $J = 4$ items. These data are collected within a computerized system of assessment for a basic course in Statistics at Perugia University. The frequencies of the corresponding contingency table are

$$\dot{\mathbf{y}}' = (0 \quad 0 \quad 1 \quad 1 \quad 4 \quad 24 \quad 0 \quad 3 \quad 0 \quad 0 \quad 4 \quad 10 \quad 0 \quad 3 \quad 10 \quad 90).$$

The value of T_{PU} equals 12.0603 and, on this basis, the MTP_2 condition cannot be rejected at the 5% significance level since the *p-value* is bounded between 0.0599 and 0.1564. The *local* estimate, based on replacing the unknown parameters with their maximum likelihood estimates, equals 0.1114.

Acknowledgments. We would like to thank Elena Stanghellini, Alessandra Giovagnoli, an Associate Editor and two referees for stimulating comments.

REFERENCES

DARDANONI, V. and FORCINA, A. (1998). A unified approach to likelihood inference on stochastic orderings in a nonparametric context. *J. Amer. Statist. Assoc.* **93** 1112–1123.

DARROCH, J. N., FIENBERG, S. E., GLONEK, G. F. V. and JUNKER, B. W. (1993). A three-sample multiple-recapture approach to census population estimation with heterogeneous catchability. *J. Amer. Statist. Assoc.* **88** 1137–1148.

DYKSTRA, R. L. (1983). An Algorithm for Restricted least Squares. *J. Amer. Statist. Assoc.* **78** 837–842.

FORTUIN, C. M., KASTELEYN, P. W. and GINIBGRE, J. (1971). Correlation inequalities on some partially ordered sets. *Comm. Math. Phys.* **22** 89–103.

HAMBLETON, R. K. and SWAMINATHAN, H. (1985). *Item Response Theory: Principles and Applications*. Kluwer, Boston.

HOIJTINK, H. and MOLENAAR, I. W. (1997). A multidimensional item response model: constrained latent class analysis using the Gibbs sampler and posterior predictive checkes. *Psychometrika* **62** 171–189.

HOLLAND, P. W. and ROSENBAUM, P. R. (1986). Conditional association and unidimensionality in monotone latent variable models. *Ann. Statist.* **14** 1523–1543.

JUNKER, B. W. and ELLIS, J. L. (1997). A characterization of monotone unidimensional latent variable models. *Ann. Statist.* **25** 1327–1343.

KARLIN, S. and RINOTT, Y. (1980). Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions. *J. Multivariate Anal.* **10** 467–498.

LAWSON, L. L. and HANSON, R. J. (1995). *Solving Least Squares Problems*. SIAM, Philadelphia.

LINDSAY, B., CLOGG, C. and GREGO, J. (1991). Semiparametric estimation in the Rasch model and related exponential response models, including a simple latent class model for item analysis, *J. Amer. Statist. Assoc.* **86** 86–107.

PERLMAN, M. D. (1969). One-sied testing problems in multivariate analysis. *Ann. Math. Statist.* **40** 549–567.

ROSENBAUM, P. R. (1984). Testing the conditional independence and monotonicity assumption of item response theory. *Psychometrika* **49** 425–435.

SHAPIRO, A. (1985). Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints. *Biometrika* **72** 13–144.

SHAPIRO, A. (1988). Towards a unified theory of inequality constrained testing in multivariate analysis. *Internat. Statist. Rev.* **56** 49–62.

VAN DEN BERG, J. and GANDOLFI, A. (1992). LRU is better than FIFO under the independent reference model. *J. Appl. Probab.* **29** 239–243.

VAN DEN BERG, J. and GANDOLFI, A. (1995). A triangle inequality for covariances of binary FKG random variables. *Ann. Appl. Probab.* **5** 322–326.

WOLAK, F. A. (1989a). Local and global testing of linear and nonlinear inequality constraints in nonlinear econometric models. *Econom. Theory* **5** 1–35.

WOLAK, F. A. (1989b). Testing inequality constraints in linear econometric models. *J. Economtrics* **41** 205–235.

DIPARTIMENTO DI SCIENZE STATISTICHE
UNIVERSITÀ DI PERUGIA
PERUGIA
ITALY 06100
E-MAIL: bart@stat.unipg.it
forcina@stat.unipg.it