Testing for positive association in contingency tables with fixed margins

F. Bartolucci[a,*], L. Scaccia[b]

[a]*Dipartimento di Economia, Università di Urbino, Via A. Saffi 42, Urbino 61029, Italy*
[b]*Dipartimento di Scienze Statistiche, Università di Perugia, Via A. Pascoli, Perugia 06100, Italy*

Abstract

An exact conditional approach is developed to test for certain forms of positive association between two ordinal variables (e.g. positive quadrant dependence, total positivity of order 2). The approach is based on the use of a test statistic measuring the goodness-of-fit of the model formulated according to the type of positive association of interest. The nuisance parameters, corresponding to the marginal distributions of the two variables, are eliminated by conditioning the inference on the observed margins. This, in turn, allows to remove the uncertainty on the conclusion of the test, which typically arises in the unconditional context where the null distribution of the test statistic depends on such parameters. Since the multivariate generalized hypergeometric distribution, which results from conditioning, is normally intractable, Markov chain Monte Carlo methods are used to obtain maximum likelihood estimates of the parameters of the constrained model. The Pearson's chi-squared statistics is used as a test statistic; a p-value for this statistic is computed through simulation, when the data are sparse, or exploiting the asymptotic theory based on the chi-bar squared distribution. The extension of the present approach to deal with bivariate contingency tables, stratified according to one or more explanatory discrete variables, is also outlined. Finally, three applications based on real data are presented.
© 2003 Elsevier B.V. All rights reserved.

Keywords: Chi-bar squared distribution; Conditional inference; Fisher-scoring algorithm; Fisher's exact test; Markov chain Monte Carlo; Pearson's chi-squared statistic

---

* Corresponding author. Tel.: +39-0722-305502; fax: +39-0722-305550.
  E-mail address: Francesco.Bartolucci@uniurb.it (F. Bartolucci).
  URL: http://www.stat.unipg.it/~bart

1. Introduction

A relevant problem in categorical data analysis is that the inference on the association between two variables is affected by the presence of the so-called *nuisance parameters*, corresponding to the marginal distributions. This is due to the fact that the distribution of the test statistic used to measure the discrepancy between the model incorporating the hypothesis of interest and the saturated model typically depends on the marginal distributions of the two variables, which are obviously unknown. Several strategies have been proposed to cope with this problem; the most well-known strategy goes back to the Fisher's exact test (see Fisher, 1934, 1935; Agresti, 1992) and consists of conditioning the inference on the observed margins which are sufficient statistics for the nuisance parameters.

Obviously, the problem also exists when we want to test the hypothesis that a certain type of positive association exists between the variables of interest, an important issue that finds applications in reliability theory, operational research, economics, finance and many other fields; see, for example Douglas et al. (1991), Robertson et al. (1988), Agresti and Coull (1998), Shaked and Shanthikumar (1994) and Dykstra et al. (1995). This topic has been recently dealt with by Bartolucci et al. (2001) who proposed a general framework for fitting and testing models incorporating several forms of positive association, such as *positive quadrant dependence* (PQD), introduced by Lehmann (1966), and *total positivity of order 2* (TP$_2$), introduced by Karlin (1968). In particular, Bartolucci et al. (2001) showed that the deviance between the saturated model and the model incorporating a certain notion of positive dependence and the deviance between this model and that of independence have asymptotic *chi-bar squared* distribution (a mixture of chi-squared distributions). However, the null distributions of such statistics depend on the (unknown) marginal distributions of the two variables and this confers uncertainty to the results of the testing procedure. To get rid of these nuisance parameters, Bartolucci et al. (2001) also explored the conditional approach. However, the distribution resulting from conditioning on the observed margins, known as *multivariate generalized hypergeometric*, is almost intractable whenever the frequencies or the dimension of the table are moderately large; in fact, computing the probability of observing a certain table requires enumerating all the possible tables with the same margins. This strongly limits the applicability of the conditional approach and so Bartolucci et al. (2001) suggested an approximation based on maximizing the product multinomial likelihood, that derives from conditioning only on the row margin, under the constraint that the marginal distribution of the column variable is equal to the observed one.

In this paper we show how the conditional approach at issue may be exploited also when the hypergeometric distribution is intractable. Basically, for estimating the parameters of the model under the constraints due to a certain type of positive association, we rely on the *Monte Carlo maximum likelihood* approach; see Geyer (1991), and Geyer and Thompson (1992). Instead of maximizing the likelihood, which cannot be computed exactly, we maximize an estimate of the likelihood ratio, with respect to a fixed point of the parameter space, obtained by the joint use of the importance sampling (Hammersley and Handscomb, 1964, p. 57–59) and the Metropolis–Hastings

algorithm (Metropolis et al., 1953; Hastings, 1970); in practice, this strategy avoids the enumeration of all the contingency tables with the given margins by using an appropriate random subset of such tables. The estimated likelihood ratio is maximized by means of the constrained Fisher-scoring algorithm used in Bartolucci et al. (2001). As a discrepancy measure between nested models, we use the Pearson's chi-squared statistic that, in this context, proved to perform better than the likelihood ratio test and to be easily implementable. A p-value for this statistic may be simply computed on the basis of Monte Carlo simulations, when the observed table is sparse, or by exploiting the asymptotic theory based on the chi-bar squared distribution.

The paper is organized as follows. In Section 2 we briefly describe the class of models of interest, while in Section 3 we illustrate the conditional approach and the resulting hypergeometric distribution. The algorithm for fitting these models and the connected inference are described in Sections 4–6; in the latter, in particular, we discuss the case in which the data are stratified according to one or more explanatory categorical variables. Finally, in Section 7, three applications involving real data sets are discussed.

2. Models of positive association

Let A and B be two categorical variables having, respectively, I and J categories; let also p_{ij} be the joint probability of the ith category of A and the jth category of B and $\boldsymbol{p}$ be the IJ-dimensional column vector with elements p_{ij} arranged by letting the index j run faster than i. Douglas et al. (1991), in a study of concepts of *positive dependence*, show that 16 different types of log-odds ratios may be used to describe the association between A and B. These log-odds ratios are identified by a type for A and a type for B that may be: local (l), continuation (c), reverse continuation (r) or global (g). Some examples are given below, where the log-odds ratio referred to the ijth cell of the contingency table ($i = 1, \ldots, I-1$, $j = 1, \ldots, J-1$) is denoted by η_{ij} (see also Agresti and Coull, 1998):

- *Local*: type l for both A and B

$$\eta_{ij} = \log \frac{p_{ij} \, p_{i+1,j+1}}{p_{i,j+1} \, p_{i+1,j}};$$

- *Local–global*: type l for A and g for B

$$\eta_{ij} = \log \frac{\left(\sum_{k \leqslant j} p_{ik}\right) \left(\sum_{k > j} p_{i+1,k}\right)}{\left(\sum_{k > j} p_{ik}\right) \left(\sum_{k \leqslant j} p_{i+1,k}\right)},$$

- *Continuation*: type c for both A and B

$$\eta_{ij} = \log \frac{p_{ij} \left(\sum_{h > i, k > j} p_{hk}\right)}{\left(\sum_{k > j} p_{ik}\right) \left(\sum_{h > i} p_{hj}\right)},$$

- *Global*: type g for both A and B

$$\eta_{ij} = \log \frac{\left(\sum_{h \leqslant i, k \leqslant j} p_{hk}\right) \left(\sum_{h > i, k > j} p_{hk}\right)}{\left(\sum_{h \leqslant i, k > j} p_{hk}\right) \left(\sum_{h > i, k \leqslant j} p_{hk}\right)}.$$

The most well-known log-odds ratios are those of type local, also because they are the only ones that may be expressed through a log-linear parameterization. However, the literature on ordinal categorical variables is more closely related to log-odds ratios of type g or c. In particular, global log-odds ratios seem to be more appropriate for dealing with ordered categorical variables of a general type and have been used in the global cross-ratio models of Dale (1986). Continuation log-odds, instead, are more appropriate when categories correspond to levels of achievement that may be entered only if the previous level has also been achieved. For example, they have been used for survival variables, for which categories correspond to successive time intervals (see Dykstra et al., 1991).

Different types of log-odds ratios determine different notions of positive association. Each of these notions may be expressed through the constraint $\eta_{ij} \geqslant 0$ for all i and j. For instance, TP$_2$ and PQD may be expressed by using local and global log-odds, respectively. Note that there is a hierarchy among these notions of positive dependence in the sense that, for instance, TP$_2$ implies that all the continuation log-odds ratios are non-negative which, in turn, implies PQD (see Douglas et al., 1991, p. 197 for details). Note also that, regardless of the type of log-odds ratios, independence between A and B may be expressed through the constraint $\eta_{ij} = 0$, for all i and j.

Following Bartolucci et al. (2001), all the log-odds ratios of a certain type may be easily expressed making use of the matrix notation. Let $\boldsymbol{\eta}$ be the vector with elements η_{ij} arranged by letting the index j run faster than i; this vector may be simply expressed as

$$\boldsymbol{\eta} = \boldsymbol{K} \log (\boldsymbol{Mp}),$$

where $\boldsymbol{K}$ is a matrix of contrasts and $\boldsymbol{M}$ a marginalization matrix. More precisely, let $\boldsymbol{K}_{h+1} = (-\boldsymbol{I}_h \ \boldsymbol{I}_h)$ and

$$\boldsymbol{M}_{h+1}(l) = \left(\begin{array}{c|c} \boldsymbol{I}_h & \boldsymbol{0}_{h,1} \\ \hline \boldsymbol{0}_{h,1} & \boldsymbol{I}_h \end{array} \right), \quad \boldsymbol{M}_{h+1}(g) = \left(\begin{array}{c|c} \boldsymbol{T}_h & \boldsymbol{0}_{h,1} \\ \hline \boldsymbol{0}_{h,1} & \boldsymbol{T}'_h \end{array} \right),$$

$$\boldsymbol{M}_{h+1}(c) = \left(\begin{array}{c|c} \boldsymbol{I}_h & \boldsymbol{0}_{h,1} \\ \hline \boldsymbol{0}_{h,1} & \boldsymbol{T}'_h \end{array} \right), \quad \boldsymbol{M}_{h+1}(r) = \left(\begin{array}{c|c} \boldsymbol{T}_h & \boldsymbol{0}_{h,1} \\ \hline \boldsymbol{0}_{h,1} & \boldsymbol{I}_h \end{array} \right),$$

where $\boldsymbol{T}_h$ is a $h \times h$ lower triangular matrix of ones (notice that the sub-matrices which are one on the top of the other in each $\boldsymbol{M}_{h+1}$ have not the same number of columns); then, we have

$$\boldsymbol{K} = \boldsymbol{K}_I \otimes \boldsymbol{K}_J \quad \text{and} \quad \boldsymbol{M} = \boldsymbol{M}_I(a) \otimes \boldsymbol{M}_J(b),$$

with $\otimes$ denoting the Kronecker product, a the type of logit used for A and b that used for B.

As indicated above, the constraint of positive association may be expressed as $\boldsymbol{\eta} \geqslant \mathbf{0}$. However, in several circumstances, it may be interesting to use further linear equality and/or inequality constraints on the log-odds ratios. Therefore, we deal with a general class of models that may be expressed in the form $\boldsymbol{C}\boldsymbol{\eta} = \mathbf{0}$, $\boldsymbol{D}\boldsymbol{\eta} \geqslant \mathbf{0}$. Let, for instance, $\boldsymbol{C} = (\boldsymbol{I} \ \ \mathbf{0}) - (\mathbf{0} \ \ \boldsymbol{I})$; when local log-odds ratios are used, this model corresponds to the linear-by-linear association model with equispaced scores (Goodman, 1979), while, when global log-odds ratios are used, this is the Plackett (1965) model. Similarly we can express the hypothesis that the $(I-1) \times (J-1)$ matrix of all the log-odds ratios is symmetric, $\eta_{ij} = \eta_{ji}$, $\forall i, j$, which corresponds to the quasi-symmetric model introduced by Caussinus (1966) or the hypothesis that all the upper diagonal entries of this matrix are greater or equal to the corresponding lower diagonal entries (see also McCullagh, 2000).

3. The conditional approach

For an observed contingency table of dimension $I \times J$, let x_{ij} be the frequency in the ijth cell and $\boldsymbol{x}$ be the vector with elements x_{ij}, for $i = 1, \ldots, I$ and $j = 1, \ldots, J$, arranged as in $\boldsymbol{p}$. If we assume, as usual, that the (unconditional) distribution of $\boldsymbol{x}$ is multinomial, the conditional distribution of $\boldsymbol{x}$, given the row and column margins, is *multivariate generalized hypergeometric*. Such a distribution, which belongs to the exponential family, may be expressed as

$$\pi(\boldsymbol{y}; \boldsymbol{\theta}) = \frac{m(\boldsymbol{y})\exp(\boldsymbol{y}'\boldsymbol{\theta})}{c(\boldsymbol{\theta})},$$

where $\boldsymbol{y}$ is the vector with the elements in $\boldsymbol{x}$ without those referred to the last row and column of the contingency table and $\boldsymbol{\theta}$ is the vector of canonical parameters with entries given by the log-odds ratios

$$\theta_{ij} = \log \frac{p_{ij}\, p_{IJ}}{p_{iJ}\, p_{Ij}}, \quad i = 1, \ldots, I-1, \ \ j = 1, \ldots, J-1.$$

Moreover, $m(\boldsymbol{y}) = n! / \sum_{ij} x_{ij}!$ is the multinomial factor, with n being the sample size, and $c(\boldsymbol{\theta})$ is the normalizing constant given by the sum of $m(\boldsymbol{y}) \exp(\boldsymbol{y}'\boldsymbol{\theta})$ for all possible tables $\boldsymbol{y}$ with the given margins $r_i = \sum_j x_{ij}$ and $c_j = \sum_i x_{ij}$; hereafter, the set of these tables will be denoted by $\mathcal{Y}$. Note that we may simply express $\boldsymbol{\theta}$ as $(\bar{\boldsymbol{K}}_I \otimes \bar{\boldsymbol{K}}_J)\log(\boldsymbol{p})$, where $\bar{\boldsymbol{K}}_h = (\boldsymbol{I}_{h-1} \ \ -\mathbf{1}_{h-1})$, but, apart from trivial cases, we cannot compute exactly $c(\boldsymbol{\theta})$ and, consequently, the maximum likelihood estimate (MLE) of $\boldsymbol{\theta}$ under a certain model. However, we can compute an approximate MLE by maximizing a suitable estimate of the likelihood ratio

$$\lambda(\boldsymbol{\theta}; \boldsymbol{y}) = \log \frac{\pi(\boldsymbol{y}; \boldsymbol{\theta})}{\pi(\boldsymbol{y}; \tilde{\boldsymbol{\theta}})},$$

for a fixed parameter vector $\tilde{\boldsymbol{\theta}}$ appropriately chosen (Gelman and Meng, 1998). This estimate is obtained through a Monte Carlo based technique known as *importance sampling* (Hammersley and Handscomb, 1964, p. 57–59). First of all, we have that

$$\lambda(\boldsymbol{\theta}; \boldsymbol{y}) = \boldsymbol{y}'(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}) - \log\{c(\boldsymbol{\theta})/c(\tilde{\boldsymbol{\theta}})\},$$

where, as it may be easily shown, the ratio $c(\boldsymbol{\theta})/c(\tilde{\boldsymbol{\theta}})$ is equal to the expected value, with respect to the measure $\pi(\boldsymbol{y}; \tilde{\boldsymbol{\theta}})$, of $\exp\{\boldsymbol{y}'(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})\}$. So, this ratio may be estimated through $(1/T) \sum_{t=1}^{T} u_t$, where $u_t = \exp\{\boldsymbol{y}_t'(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})\}$ and $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T$ are T tables drawn from the distribution $\pi(\boldsymbol{y}; \tilde{\boldsymbol{\theta}})$. To draw such tables, we can use the MCMC algorithm developed by Diaconis and Sturmfels (1998), which is outlined in Appendix A.1. Finally, $\lambda(\boldsymbol{\theta}; \boldsymbol{y})$ may be estimated as

$$\tilde{\lambda}(\boldsymbol{\theta}; \boldsymbol{y}) = \boldsymbol{y}'(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}) - \log \frac{1}{T} \sum_{t=1}^{T} u_t,$$

and maximized, under a certain model, using a constrained version of the Fisher-scoring algorithm that will be described in Section 4. Such an algorithm makes use of the expected value and the variance of $\boldsymbol{y}$ that may be again estimated on the basis of MCMC samples. More precisely, the expected value of $\boldsymbol{y}$ may be expressed as

$$\boldsymbol{\mu}(\boldsymbol{\theta}) = \sum_{\boldsymbol{y} \in \mathcal{Y}} \boldsymbol{y}\pi(\boldsymbol{y}; \tilde{\boldsymbol{\theta}}) \exp\{\lambda(\boldsymbol{y}; \boldsymbol{\theta})\}$$

and so estimated as

$$\tilde{\boldsymbol{\mu}}(\boldsymbol{\theta}) = \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{y}_t \exp\{\tilde{\lambda}(\boldsymbol{y}_t; \boldsymbol{\theta})\} = \sum_{t=1}^{T} w_t \boldsymbol{y}_t,$$

where $w_t = u_t / \sum_h u_h$; instead, the estimate of the expected value of $\boldsymbol{x}$ may be simply expressed as $\tilde{\boldsymbol{v}}(\boldsymbol{\theta}) = \boldsymbol{a} + \boldsymbol{A}\tilde{\boldsymbol{\mu}}(\boldsymbol{\theta})$, with $\boldsymbol{a}$ and $\boldsymbol{A}$ appropriately chosen. Similarly, the variance of $\boldsymbol{y}$ may be estimated as

$$\tilde{\boldsymbol{V}}(\boldsymbol{\theta}) = \sum_{t=1}^{T} w_t \boldsymbol{y}_t \boldsymbol{y}_t' - \tilde{\boldsymbol{\mu}}(\boldsymbol{\theta})\tilde{\boldsymbol{\mu}}(\boldsymbol{\theta})'.$$

Finally, note that under independence, i.e. when $\boldsymbol{\theta} = \boldsymbol{0}$, we may compute exactly the probability of $\boldsymbol{y}$ and the corresponding moments. We have (see, for instance, Agresti, 1990)

$$\pi(\boldsymbol{y}; \boldsymbol{0}) = \frac{\left(\prod_i r_i!\right)\left(\prod_j c_j!\right)}{n! \prod_{ij} x_{ij}!},$$

whereas the expected value of y_{ij} is given by the expected frequency under independence, i.e. $r_i c_j / n$, and the covariance between y_{ij} and y_{hk} is

$$\frac{r_i r_h' c_j c_k'}{n^2(n-1)},$$

where r_h' is equal to $(n - r_h)$ if $h = i$ and to r_h otherwise and similarly for c_k'.

4. Parameter estimation

Dardanoni and Forcina (1998) described a general Fisher-scoring algorithm for the maximum likelihood estimation under equality and inequality constraints (see also

Bartolucci et al., 2001, Section 2.1). This algorithm may be also exploited in the present context with slight adjustments to allow for the fact that we cannot compute exactly, but only through simulation, the moments of the hypergeometric distribution. More precisely, let $\mathscr{H} = \{\boldsymbol{\eta} : \boldsymbol{C\eta} = \mathbf{0}, \boldsymbol{D\eta} \geqslant \mathbf{0}\}$ and suppose that we could compute exactly $\pi(\boldsymbol{y}; \boldsymbol{\theta})$. Then, to maximize the likelihood under the constraint $\boldsymbol{\eta} \in \mathscr{H}$, the algorithm at issue would solve, at step $s+1$, the following quadratic optimization problem

$$\min_{\boldsymbol{\eta} \in \mathscr{H}} (\boldsymbol{z}^s - \boldsymbol{\eta})' \boldsymbol{F}(\boldsymbol{\eta}^s)(\boldsymbol{z}^s - \boldsymbol{\eta}),$$

where $\boldsymbol{z}^s = \boldsymbol{\eta}^s + \boldsymbol{F}(\boldsymbol{\eta}^s)^{-1} \boldsymbol{s}(\boldsymbol{\eta}^s)$ and $\boldsymbol{s}(\boldsymbol{\eta})$ and $\boldsymbol{F}(\boldsymbol{\eta})$ are, respectively, the score vector and the information matrix with respect to $\boldsymbol{\eta}$ which are given by

$$\boldsymbol{s}(\boldsymbol{\eta}) = \boldsymbol{J}(\boldsymbol{\eta})'(\boldsymbol{y} - \boldsymbol{\mu}) \quad \text{and} \quad \boldsymbol{F}(\boldsymbol{\eta}) = \boldsymbol{J}(\boldsymbol{\eta})' \boldsymbol{V}(\boldsymbol{\theta}) \boldsymbol{J}(\boldsymbol{\eta}), \tag{1}$$

with $\boldsymbol{J}(\boldsymbol{\eta})$ denoting the Jacobian of the transformation from $\boldsymbol{\eta}$ to $\boldsymbol{\theta}$, illustrated in Appendix A.2. In our context, even if we cannot compute exactly $\pi(\boldsymbol{y}; \boldsymbol{\theta})$ and its moments, we can still use the above algorithm to maximize the likelihood ratio $\tilde{\lambda}(\boldsymbol{\theta}; \boldsymbol{y})$, once a suitable value of the parameter vector $\tilde{\boldsymbol{\theta}}$ has been chosen: it is enough to replace $\boldsymbol{\mu}(\boldsymbol{\theta})$ and $\boldsymbol{V}(\boldsymbol{\theta})$ in (1) with the respective estimates defined at the end of Section 3 and obtained on the basis of a random sample of T contingency tables (30,000 in our applications) with the given margins. We experimented that a sensible strategy consists in setting $\tilde{\boldsymbol{\theta}}$ equal to the canonical parameter vector corresponding to the estimate of $\boldsymbol{\eta}$ for the multinomial scheme under $\mathscr{H}$ and the further constraint that the marginal probabilities are equal to the observed frequencies; this estimate, that will be denoted by $\tilde{\boldsymbol{\eta}}$, may be again computed through the constrained Fisher-scoring algorithm indicated above (see Bartolucci et al., 2001, Section 3.1). We also use $\tilde{\boldsymbol{\eta}}$ as the starting value for $\boldsymbol{\eta}$ in the maximization algorithm.

A final point concerns the presence of frequencies equal to zero in the observed table. We suggest to replace all such frequencies in the table with 10^{-6}, whose effect is negligible on the final estimate, but using a larger value, say 0.25, when $\tilde{\boldsymbol{\eta}}$ is computed. In this way we limit the risk of having always null frequencies in the corresponding cells of the tables generated under $\pi(\boldsymbol{y}; \tilde{\boldsymbol{\theta}})$.

5. Hypothesis testing

In the following we illustrate how we can test for a set of equality and inequality constraints on the parameters $\boldsymbol{\eta}$ on the basis of a goodness-of-fit statistic, as the Pearson's chi-squared or the likelihood ratio.

Let S denote the saturated model, H the model formulated as $\boldsymbol{\eta} \in \mathscr{H}$ and H_0 that formulated as $\boldsymbol{\eta} \in \mathscr{H}_0 = \{\boldsymbol{\eta} : \boldsymbol{C\eta} = \mathbf{0}, \boldsymbol{D\eta} = \mathbf{0}\}$, namely by turning all the inequality constraints into equality constraints. Obviously, H_0 is nested in H, which, in turn, is nested in S. Then, let $\dot{\boldsymbol{\eta}}$, $\hat{\boldsymbol{\eta}}$ and $\bar{\boldsymbol{\eta}}$ be the MLE computed through the algorithm in Section 4 under the models S, H and H_0, respectively, and $\dot{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\theta}}$ and $\bar{\boldsymbol{\theta}}$ be the corresponding

values of the canonical parameter vector. Following Dardanoni and Forcina (1998) and Bartolucci et al. (2001), the likelihood ratio statistic between S and H_0,

$$Q = 2 \log \frac{\pi(y; \dot{\theta})}{\pi(y; \bar{\theta})},$$

may be expressed as the sum of

$$Q_1 = 2 \log \frac{\pi(y; \dot{\theta})}{\pi(y; \hat{\theta})} \quad \text{and} \quad Q_2 = 2 \log \frac{\pi(y; \hat{\theta})}{\pi(y; \bar{\theta})},$$

where Q_1 provides a measure of the discrepancy between S and H, while Q_2 provides a measure of the discrepancy between H and H_0. In our context we cannot compute exactly the hypergeometric distribution and so it is natural to use, instead of Q_1 and Q_2, the statistics $\tilde{Q}_1 = 2\tilde{\lambda}(\dot{\theta}; y)$ and $\tilde{Q}_2 = 2\tilde{\lambda}(\hat{\theta}; y)$, where the first is computed on the basis of a sample of tables drawn under $\theta = \hat{\theta}$, and the second is computed on the basis of a sample of tables drawn under $\theta = \bar{\theta}$, respectively. However, we believe that, for at least two reasons, it is preferable to use the Pearson's chi-squared statistic, X^2, for measuring the discrepancy between nested models. Firstly, $\tilde{Q}_1$ is a function of the estimate of θ under the saturated model, $\dot{\theta}$, which may not be computed explicitly on the basis of the observed frequencies, but with a Fisher-scoring algorithm, with consequent waste of time. Secondly, if $\dot{\theta}$ and $\hat{\theta}$ are far from each other, $\tilde{Q}_1$ is seldom a reliable estimate of Q_1; the same may happen for $\tilde{Q}_2$. This is a typical problem of the importance sampling method that, intuitively, may be explained as follows. Suppose that the elements of $\dot{\theta}$ are much different from the corresponding elements of $\hat{\theta}$. Then in an MCMC sample generated from $\pi(y; \hat{\theta})$ to compute $\tilde{\lambda}(\dot{\theta}; y)$, we will typically have a few tables y_t for which the statistic $\exp\{y_t'(\dot{\theta} - \hat{\theta})\}$ is very large and many tables for which this statistic is very small and so the resulting estimate will be inaccurate (for a deeper discussion see Geyer, 1996, Section 14.7). In principle, this problem could be overcome by sampling in correspondence of one or more values of θ between $\dot{\theta}$ and $\hat{\theta}$ (or between $\hat{\theta}$ and $\bar{\theta}$); see, for instance, Gelman and Meng (1998). This solution, however, is too computationally expensive to be taken into consideration in our context.

The Pearson's chi-squared statistic between S and H_0 may be expressed as

$$X^2 = (x - \bar{v})' \text{diag}(\bar{v})^{-1} (x - \bar{v}),$$

where $\bar{v}$ is the expected value of x when $\eta = \bar{\eta}$ (see Section 3). As the likelihood ratio, X^2 may be expressed as the sum of two components. The first one, given by

$$X_1^2 = (x - \hat{v})' \text{diag}(\hat{v})^{-1} (x - \hat{v}),$$

where $\hat{v}$ denotes the expected value of x when $\eta = \hat{\eta}$, is similar to Q_1, being a measure of the discrepancy between S and H; note, however, that X_1^2 does not depend on $\dot{\theta}$. The second one, given by $X_2^2 = X^2 - X_1^2$ corresponds to Q_2 and measures the discrepancy between H and H_0. Clearly, the difference between X^2 and X_1^2 might be negative. However, in our applications, we observed that the chance of this is very small especially, if a reasonable number of MCMC samples is used for the parameter estimation.

We can compute p-values for X_1^2 and X_2^2 on the basis of which we can decide to reject H in favor of S, implying that the positive association of interest does not hold, or H_0 in favor of H, having in this way a directed test for independence which has more power against a narrower set of alternatives. As usual these p-values are computed under H_0 that, in our context, normally corresponds to the independence model; see Dardanoni and Forcina (1998) and Bartolucci et al. (2001). The only parameter vector satisfying this model is $\boldsymbol{\eta} = \mathbf{0}$ and, given the margins, this uniquely determines the distributions of X_1^2 and X_2^2. This is the major advantage of the conditional approach with respect to the unconditional approach where such distributions vary according to the nuisance parameters and this determines uncertainty on the interpretation of the test (for a deeper discussion on the opportunity of conditioning see Agresti, 1992). In this regard, we have to distinguish between small and large samples. In the first case we have to rely on a standard Monte Carlo simulation which consists in generating a certain number of tables, say 1000, under H_0 and then compute the p-values for X_1^2 and X_2^2 as the proportions of tables with values of the two statistics larger than the observed table. When H_0 corresponds to the independence model, we can use an exact Monte Carlo algorithm to generate these tables, such as the one due to Darroch and described by Diaconis and Sturmfels (1998, p. 372). In presence of large samples, instead, we can rely on the asymptotic theory. Under H_0, the asymptotic distribution of X_1^2 is $\bar{\chi}^2(\boldsymbol{F}_0^{-1}, \mathscr{H}^o)$ and that of X_2^2 is $\bar{\chi}^2(\boldsymbol{F}_0^{-1}, \mathscr{H})$, where $\bar{\chi}^2(\boldsymbol{\Sigma}, \mathscr{C})$ denotes the *chi-bar squared* distribution, that is the distribution of the squared norm of the projection of a random vector with distribution $N(\mathbf{0}, \boldsymbol{\Sigma})$ onto $\mathscr{C}$, $\boldsymbol{F}_0$ is the information matrix under H_0 and $\mathscr{C}^o$ denotes the dual of the cone $\mathscr{C}$; see, for instance, Shapiro (1988), and Dardanoni and Forcina (1998). The survival function of $\bar{\chi}^2(\boldsymbol{\Sigma}, \mathscr{C})$ is given by

$$Pr(\bar{\chi}^2(\boldsymbol{\Sigma}, \mathscr{C}) \geqslant x) = \sum_{i=0}^{m} w_i(\boldsymbol{\Sigma}, \mathscr{C}) Pr(\chi_i^2 \geqslant x),$$

where χ_i^2 denotes a chi-squared random variable with i degrees of freedom, $w_i(\boldsymbol{\Sigma}, \mathscr{C})$, $i = 0, 1, \ldots, m$, are weights depending on $\boldsymbol{\Sigma}$ and $\mathscr{C}$, and m is the size of the squared matrix $\boldsymbol{\Sigma}$. The computation of the probability weights $w_i(\boldsymbol{\Sigma}, \mathscr{C})$ is a difficult numerical problem unless m is less than 4; however, accurate estimates can be easily obtained by Monte Carlo simulations (see Dardanoni and Forcina, 1998). Note that this asymptotic theory was, actually, developed with reference to the likelihood ratio statistic. However it can be used also for the Pearson's chi-squared statistic since these two discrepancy measures are asymptotically equivalent under the null hypothesis (see, for example Shapiro, 1985).

A referee pointed out that, to get rid of the nuisance parameters by conditioning on their observed values, these parameters should be asymptotically uncorrelated from those of interest. This always happens when the parameterization is based on local logits. Instead, when the parameterization is based on global logits, this happens only under the null hypothesis of independence (for a proof see the appendix in Bartolucci et al., 2001). However, even if independence does not hold, the asymptotic correlation between the two sets of parameters (marginal global logits and global log-odds ratio) is usually moderate. Consider, for example, a model in which the cell probabilities

correspond to the relative frequencies of the table we used in Section 7.2. Although this model is far from the independence one, as we have a strong association between the row and column variables, all the elements of the asymptotic correlation matrix between the two sets of parameters, computed on the basis of the expected information, are less than 0.25 in absolute value.

6. Extension to multiway tables

The framework presented here may be easily extended to deal with bivariate tables stratified according to one or more discrete explanatory variables. Assume, without loss of generality, that there is only one of such variables, say C, with K categories. Then, any stratified table corresponding to a level of C, say the kth, has multivariate hypergeometric distribution with canonical parameter vector denoted by $\boldsymbol{\theta}_k$; let also $\boldsymbol{\eta}_k$ be the vector of log-odds ratios for such a table. It is straightforward to see that several interesting hypotheses may be again formulated through the linear constraints $\boldsymbol{C}\boldsymbol{\eta}=\boldsymbol{0}$ and $\boldsymbol{D}\boldsymbol{\eta} \geqslant \boldsymbol{0}$, where, in this case, $\boldsymbol{\eta}$ is the vector obtained by stacking one below the other the vectors $\boldsymbol{\eta}_k$ for all k. For instance, when C is binary, the constraint of equal log-odds ratios (i.e. same association) may be formulated by letting $\boldsymbol{C}=(-\boldsymbol{I} \ \boldsymbol{I})$, whereas the constraint that all log-odds ratios in the second stratum are greater or equal to the corresponding log-odds ratios in the first stratum may be formulated as $\boldsymbol{D}=(-\boldsymbol{I} \ \boldsymbol{I})$. To fit models of this type, we can still use the constrained Fisher-scoring algorithm of Section 4. It is enough to replace $\boldsymbol{s}$ with the vector obtained by staking, one below the other, the vectors $\boldsymbol{s}_k$ and $\boldsymbol{F}$ with the diagonal matrix with blocks $\boldsymbol{F}_k$, where $\boldsymbol{s}_k$ and $\boldsymbol{F}_k$ are, respectively, the score vector and the information matrix corresponding to $\boldsymbol{\eta}_k$. Notice that, in this case, we have to generate MCMC tables independently for each of the strata. Also the hypothesis testing approach outlined in Section 5 may be exploited with minor adjustments.

7. Some applications

7.1. Classification of employees by salary and experience

In a study about probability estimation in sparse contingency tables, Simonoff (1987) reports the data presented in Table 1, taken from a survey conducted by the Department of Energy. The sample consists of 147 nonsupervisory female employees having the Bachelors (but not higher) degree, who were practicing mathematics or statistics in 1981. The sample is cross-classified by monthly salary and years since degree. Note that the table is quite sparse, with an average cell frequency equal to 2.70; moreover, only 20% of the cells have a frequency larger or equal to 5 and 28% of the cells have null frequency.

On these data we fitted two models of positive association, the first one based on continuation log-odds ratios (cc) and the second one on local log-odds ratios (ll). The values of the Pearson's chi-squared statistic are shown in Table 2 together with the

Table 1
Salary and years since degree for a sample of 147 female employees

Salary	Years since degree								
	0–2	3–5	6–8	9–11	12–14	15–17	18–23	24–29	30+
950–1350	7	1	1	0	0	0	0	2	0
1351–1750	10	6	5	3	0	1	1	1	0
1751–2150	12	14	7	1	4	2	2	1	2
2151–2550	0	1	8	3	3	3	5	0	4
2551–2950	0	0	3	2	0	6	5	2	7
2951–3750	1	0	1	0	1	1	6	0	2

Table 2
Pearson's chi-squared statistics for the model of positive association (X_1^2) and for the independence model (X_2^2)

Type of log-odds ratios	X_1^2	Monte Carlo p-value	Asym. p-Value	X_2^2	Monte Carlo p-value	Asym. p-Value
cc	17.415	0.6240	0.6000	94.920	0.0000	0.0000
ll	68.542	0.0020	0.0007	43.792	0.0000	0.0000

corresponding asymptotic and simulated p-values; the latter should be more reliable due to the sparseness of the data.

First of all note that the data conform to the notion of positive association expressed through continuation log-odds ratios; in this case, in fact, the Pearson's chi-squared statistic between the saturated model and the model under this constraint is $X_1^2 = 17.415$ with a corresponding simulated p-value of 0.6240. This implies that also PQD holds for these data. Instead, TP$_2$ has to be rejected since we have $X_1^2 = 68.542$ with a simulated p-value equal to 0.0020. Moreover, the independence model has to be rejected since we have a p-value for X_2^2 less than 10^{-4}, regardless of the type of positive association considered. In summary, we can state that, as suspected by Simonoff (1987), a certain degree of positive association exists between salary and years since degree, in the sense that females with more years since degree have a better chance of getting higher salaries. However, since the data show evidence against TP$_2$, the association between the two variables is not so strong as one may expect. This is a more precise conclusion than Simonoff's who simply recognized the existence of a positive association between salary and years of experience, without specifying the strength of this association.

As a final remark, it can be noticed that the simulated p-values are always very close to the asymptotic ones. The asymptotic theory seems, therefore, to provide a reasonable approximation even for sparse tables. This is a considerable advantage in terms of computing time. In this case, for instance, computing the asymptotic p-values took approximately 4.5 min (on a Pentium III, 700 MHz), while computing the p-values through simulation took 6 h.

Table 3
Father (x_i) and son (y_j) occupational status for a sample of 3488 British males

	y_1	y_2	y_3	y_4	y_5	y_6
x_1	125	60	26	49	14	5
x_2	47	65	66	123	23	21
x_3	31	58	110	223	64	32
x_4	50	114	185	715	258	189
x_5	6	19	40	179	143	71
x_6	3	14	32	141	91	106

7.2. Classification of men by social class and social class of their fathers

Dardanoni and Forcina (1998) analyzed a data set, considered also by Goodman (1991), in which 3488 British men are classified according to their own and their father's occupational status (see Table 3). In particular, they found that, under the product multinomial sampling schemes (i.e. conditioning only on the row margin), the difference in the deviance between the saturated model and the model of positive association, when log-odds ratios local in the row variable and global in the column variable are used, is null and so the data conform exactly to this constrained model. Therefore, data also conform exactly to less restrictive stochastic orderings such as PQD. They also concluded in favor of the positive association model based on continuation log-odds ratios in the column variable and local log-odds ratios in the row variable. Hence sons coming from a better family have a better chance of success not only in general, but also conditional to having already had a certain amount of success. However, due to the presence of nuisance parameters, represented by the column margins, it is not possible to reach a definitive conclusion about TP_2 as we can do by conditioning on both row and column totals.

In fact, for the data at issue we have that the Pearson's chi-squared statistic between the saturated model and the model under TP_2 is $X_1^2 = 9.012$, with a *p*-value equal to 0.1426. Therefore, TP_2 should not be rejected; hence, we can state that sons coming from a better family have a better chance of success also conditional on remaining within any given subset of neighboring classes. Independence is instead a rather implausible model in this context: a *p*-value less than 10^{-4} is obtained when testing independence against the TP_2 model.

For the data at issue we also consider some equality constraints on the log-odds ratios. In particular, we are interested in two types of models, linear-by-linear with equispaced score and quasi-symmetry, already illustrated in Section 2. The first model does not seem to fit adequately; for instance, with local log-odds ratios the deviance with respect to the saturated model is 104.15 with 24 degrees of freedom. Instead the symmetry model fits adequately for all types of log-odds ratios. For instance, with the type local we have a deviance equal to 7.6671 with 10 degrees of freedom (*p*-value 0.6643). Now, if we add the constraint of positive association, the deviance increases by 6.1083 with a *p*-value equal to 0.3105 which confirms that TP_2 may not be rejected.

Table 4
Eye testing for a sample of 10,719 employees

Right eye grade	Female left eye grade				Male left eye grade			
	4	3	2	1	4	3	2	1
4	1520	266	124	66	821	112	85	35
3	234	1512	432	78	116	494	145	27
2	117	362	1772	205	72	151	583	87
1	36	82	179	492	43	34	106	331

Table 5
Pearson's chi-squared statistic for the model of equal log-odds ratios (d.o.f. = 9)

Log-odds ratios	X^2	p-value	Log-odds ratios	X^2	p-value
gg	35.268	0.0000	*rr*	14.664	0.1006
cc	33.395	0.0001	*ll*	28.951	0.0007

7.3. Classification of employees by left eye grade and right eye grade

Table 4, based on case records of the eye-testing of employees in Royal Ordinance factories in 1943–1946, has been analyzed by Stuart (1953). The 10,719 people in the sample are cross-classified according to gender and left and right eye grade.

First of all we fitted the model of equal log-odds ratios between the female and male groups; using the notation in Section 6, this constraint may be simply expressed as $C\eta = 0$, where $C = (I_9 \quad -I_9)$ and $\eta = (\eta_1' \quad \eta_2')'$. The Pearson's chi-squared statistic between the saturated model and the model under this constraint is shown, for different type of log-odds ratio, in Table 5.

Note that the constraint at issue seems to hold only when reverse continuation (*rr*) log-odds ratios are used. In fact, the deviance from the saturated model is equal to $X^2 = 14.664$ with a p-value slightly greater than 0.1. So, we base the analysis of the data at issue on log-odds ratios of this type. Note that, since the row and column categories are in reverse order, using these log-odds ratios is natural since they correspond to the continuation log-odds ratios when the categories are ordered in the usual way. Moreover, we have that the lowest of these log-odds ratio is equal to 0.0533. So, we have surely positive association between left and right eye grade. In summary using reverse continuation odds ratios we reach the same conclusion that Stuart (1953) reached using a modified form of Kendall's rank correlation coefficient: there is positive association between the right and the left eye grade in both female and male groups and the strength of the association is the same in the two groups.

Acknowledgements

Appendix A.

A.1. Sampling from the multivariate hypergeometric distribution

To sample contingency tables with fixed margins from $\pi(y; \theta)$, we can use the Metropolis–Hastings algorithm described below. Suppose we have drawn a table, say y_t. To obtain the next table, y_{t+1}, we perform the following operations: (i) choose randomly a pair of rows, i_1 and i_2, and a pair of columns, j_1 and j_2; (ii) modify the entries (i_1, j_1), (i_1, j_2), (i_2, j_1) and (i_2, j_2) of the table (completed with the last row and column) corresponding to y_t according to the following rule:

$$
\begin{array}{cc}
+ & - \\
- & +
\end{array}
\quad \text{or} \quad
\begin{array}{cc}
- & + \\
+ & -
\end{array}
\quad \text{with probability } 1/2 \text{ each,}
$$

where $+$ means "add 1" and $-$ means "subtract 1"; let $y^\star$ be the table so obtained (proposed table); (iii) by the usual Metropolis procedure, let $y_{t+1} = y^\star$ or $y_{t+1} = y_t$ with probability α and $1 - \alpha$, respectively, where

$$
\alpha = \min\left\{ 1, \frac{\pi(y^\star; \theta)}{\pi(y_t; \theta)} \right\} = \min\left\{ 1, \frac{\exp\left[(y^\star)' \theta\right]}{\exp\left(y_t' \theta\right)} \right\}.
$$

Note that step (ii) may force negative entries; in this case, the table is discarded and a new pair of rows and columns is chosen. As an initial state of the chain, y_0, we can use the observed table; also a burn-in period is necessary. The convergence of the chain to its stationary distribution can be checked by plotting $m(y_t) \exp(y_t' \theta)$ against t. If the mixing is too slow, this can be improved by adding or subtracting c instead of 1 from each of the four entries selected, where c is a random integer from 1 to $c_{\max}$. Clearly, increasing $c_{\max}$ reduces the overall acceptance rate of the algorithm and, therefore, some preliminary simulations are required for a sensible choice.

A.2. Transformation from η to p and θ

Given η and the row and column margins, we can compute the corresponding vector of probabilities p, and consequently θ, by solving the system

$$
L \log(Np) = \phi
$$

through a Newton algorithm; in the expression above,

$$
L = \begin{pmatrix} K_J & 0 & 0 \\ 0 & K_I & 0 \\ 0 & 0 & K_I \otimes K_J \end{pmatrix}, \quad
N = \begin{pmatrix} 1_I' \otimes M_J(b) \\ M_I(a) \otimes 1_J' \\ M_I(a) \otimes M_J(b) \end{pmatrix}, \quad
\phi = \begin{pmatrix} \phi_B \\ \phi_A \\ \eta, \end{pmatrix},
$$

where ϕ_A and ϕ_B are the vectors of the marginal logits of A and B, respectively, that is $\phi_A = K_I \log[M_I(a)r]$ and $\phi_B = K_J \log[M_J(b)c]$, and $r = (r_1 \; \cdots \; r_I)'$ and $c = (c_1 \; \cdots \; c_J)'$ are the vectors corresponding to the row and column margins. Moreover, once that from

$\boldsymbol{\eta}$ we have obtained $\boldsymbol{p}$, we can simply compute $\boldsymbol{\theta}$ as indicated in Section 2. The Jacobian of the transformation from $\boldsymbol{\eta}$ to $\boldsymbol{\theta}$ may be expressed as

$$\boldsymbol{J}(\boldsymbol{\eta}) = \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\eta}'} = \{\boldsymbol{QL}\operatorname{diag}(\boldsymbol{Mp})^{-1}\boldsymbol{N}\operatorname{diag}(\boldsymbol{p})(\bar{\boldsymbol{K}}_I \otimes \bar{\boldsymbol{K}}_J)'\}^{-1},$$

where $\boldsymbol{Q}$ is obtained by removing the first $I + J - 2$ rows from an identity matrix of order $IJ - 1$.

References

Agresti, A., 1990. Categorical Data Analysis. Wiley, New York.

Agresti, A., 1992. A survey of exact inference for contingency tables. Statist. Sci. 7, 131–177.

Agresti, A., Coull, B.A., 1998. Order-restricted inference for monotone trend alternatives in contingency tables. Comput. Statist. Data Anal. 28, 139–155.

Bartolucci, F., Forcina, A., Dardanoni, V., 2001. Positive quadrant dependence and marginal modelling in two-way tables with ordered margins. J. Amer. Statist. Assoc. 96, 1497–1505.

Caussinus, H., 1966. Contribution à l'analyse statistique des tableaux de corrélation. Ann. Faculté Sci. Univ. Toulouse 29 (année 1965), 77–183.

Dale, J.R., 1986. Global cross-ratio models for bivariate, discrete, ordered responses. Biometrics 42, 909–917.

Dardanoni, V., Forcina, A., 1998. A unified approach to likelihood inference on stochastic orderings in a nonparametric context. J. Amer. Statist. Assoc. 93, 1112–1123.

Diaconis, P., Sturmfels, B., 1998. Algebraic algorithms for sampling from conditional distributions. Ann. Statist. 98, 363–397.

Douglas, R., Fienberg, S.E., Lee, M.T., Sampson, A.R., Whitaker, L.R., 1991. Positive dependence concepts for ordinal contingency tables. In: Block, H.W., Sampson, A.R., Savits, T.H. (Eds.), Topics in Statistical Dependence, IMS Lecture/Monograph Series, Hayward, CA, pp. 189–202.

Dykstra, R.L., Kochar, S., Robertson, T., 1991. Statistical inference for uniform stochastic ordering in several populations. Ann. Statist. 19, 870–888.

Dykstra, R.L., Kochar, S., Robertson, T., 1995. Inference for likelihood ratio ordering in the two-sample problem. J. Amer. Statist. Assoc. 90, 1034–1040.

Fisher, R.A., 1934. Statistical Methods for Research Workers (originally published 1925, 14th Edition, 1970), Oliver and Boyd, Edinburgh.

Fisher, R.A., 1935. The logic of inductive inference (with discussion). J. Roy. Statist. Soc. 98, 39–82.

Gelman, A., Meng, X.L., 1998. Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. Statist. Sci. 13, 163–185.

Geyer, C.J., 1991. Markov chain Monte Carlo maximum likelihood. In: Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface. Keramidas, E.M. (Ed.), Interface Foundation of North America, Fairfax Station, VA, pp. 156–163.

Geyer, C.J., 1996. Contrained Monte Carlo maximum likelihood for dependent data (with Discussion). J. Roy. Statist. Soc. B 54, 657–699.

Geyer, C.J., Thompson, E.A., 1992. Contrained Monte Carlo maximum likelihood for dependent data (with Discussion). J. Roy. Statist. Soc. B 54, 657–699.

Goodman, L., 1979. Simple models for the analysis of association in cross-classifications having ordered categories. J. Amer. Statist. Assoc. 74, 537–552.

Goodman, L., 1991. Measures, models and graphical displays in the analysis of cross-classified data. J. Amer. Statist. Assoc. 86, 1085–1111.

Hammersley, J., Handscomb, D., 1964. Monte Carlo Methods. Methuen, London.

Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57, 97–109.

Karlin, S., 1968. Total Positivity, Vol. 1. Stanford University Press, Standford, CA.

Lehmann, E.L., 1966. Some concepts of dependence. Ann. Math. Statist. 37, 1137–1153.

McCullagh, P., 2000. Invariance and Factorial Models (with discussion). J. Roy. Statist. Soc. B 62, 209–256.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equations of state calculations by fast computing machines. J. Chem. Phys. 21, 1087–1091.

Plackett, R.L., 1965. A class of bivariate distributions. J. Amer. Statist. Assoc. 60, 516–522.

Robertson, T., Wright, F.T., Dykstra, R., 1988. Order Restricted Statistical Inference. Wiley, New York.

Shaked, M., Shanthikumar, J.G., 1994. Stochastic Orders and their Applications. Academic Press, San Diego, CA.

Shapiro, A., 1985. Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints. Biometrika 72, 133–144.

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

Simonoff, J.S., 1987. Probability estimation via smoothing in sparse contingency tables with ordered categories. Statist. Probab. Lett. 5, 55–63.

Stuart, A., 1953. The estimation and comparison of strength of association in contingency tables. Biometrika 40, 105–110.