Inference for Positive Dependence and Marginal Modelling

Francesco Bartolucci
Dipartimento di Scienze Statistiche
Università di Perugia
bart@stat.unipg.it

Valentino Dardanoni
Istituto di Scienze Finanziarie
Università di Palermo
vdardanounipa.it

Antonio Forcina
Dipartimento Scienze Statistiche
Università di Perugia
forcinastat.unipg.it

Riassunto: Per un insieme di tabelle doppie con soggetti classificati rispetto alla stessa coppia di mutabili ordinali, viene proposto un approccio generale di inferenza basata sulla verosimiglianza per considerare simultaneamente ipotesi specificate da modelli marginali e vincoli di disuguaglianza sui parametri implicati da varie forme di concordanza. La metodologia si basa sulla parametrizzazione delle distribuzioni bivariate con logit marginali e logaritmi di rapporti degli odds appropriati. Per tali parametri viene tratteggiato un approccio generale per fare inferenza su ipotesi definite da vincoli di uguaglianza e disuguaglianza.

Keywords: Chi-bar square distribution; Conditional Inference; Fisher-scoring Algorithm; Positive Dependence.

1. Introduction

This paper is a revised and shorter version of Bartolucci, Forcina and Dardanoni (2001). Results are only stated; refer to the main paper for details and proofs.

In many applications involving bivariate distributions of ordered categorical variables, say, $A$ and $B$, it is natural to expect that larger values of $B$ are associated with larger values of $A$. A substantial literature now exists which formalizes this intuitive concept providing several notions of positive dependence which have found interesting applications in reliability theory, operational research, economics, finance and in many other fields. The book by Shaked and Shanthikumar (1994) reviews most of the theoretical results in this literature and presents many useful applications.

The condition of Positive Quadrant Dependence (PQD for short) was introduced by Lehmann (1966); it is also the least stringent notion of positive dependence within the class considered by Douglas et al. (1991). The well known property that PQD holds if and only if the covariance between any two arbitrary sets of increasing scores assigned respectively to the categories of $A$ and to those of $B$ is non negative (e.g. Karlin, 1983) provides an appealing motivation for PQD when the observed table may be thought of

---

1The research work for this paper has been funded by Palermo and Perugia University. Reproduced with permission from the Journal of the American Statistical Association, copyright 2001 by the American Statistical Association. All rights reserved.
as a discretized version of an unknown underlying bivariate distribution. Moreover, the fact that PQD holds for a given table if and only if all the global log-odds ratios are non-negative provides an interesting connection between the analysis of positive dependence and marginal modelling as considered, for instance, by Dale (1986), Molenberghs and Lesaffre (1994), Glonek and McCullagh (1995) or Bergsma and Rudas (2001). By parameterizing a bivariate distribution with the global logits on each marginal distribution and the global log-odds ratios, a variety of models for the marginal distributions may be combined with restrictions on the association. The novelty of the present approach consists in the possibility of dealing simultaneously with models defined by linear equality and inequality constraints. As instances of models concerning the marginal distributions we can consider: (i) linearity of global logits within one or both marginal distributions; (ii) with homologous response variables (in the terminology of McCullagh, 2000), that one marginal distribution is stochastically larger or has greater dispersion than the other; (iii) the proportional odds model and the model of marginal homogeneity for relating the row (or column) margins of several tables. On the other hand, restrictions on the global odds ratios are useful to model the PQD condition, the Plackett distribution (uniform global odds ratios), symmetry or one directional asymmetry (above or below the diagonal) of global odds-ratios in square tables and the hypothesis of stronger positive dependence when comparing several tables.

In this paper we propose a general approach for likelihood inference on an $r \times c$ table where subjects are cross classified according to ordinal categorical variables $A$ and $B$. We examine in detail three different sampling schemes: (i) each $r \times c$ table is a single sample; (ii) the $A$ margins are fixed; (iii) both $A$ and $B$ margins are fixed. We use this strategy both because conditioning may be used to reduce (or remove) the effect of nuisance parameters on the distribution of the likelihood ratio test.

2. A general framework for positive quadrant dependence

Let $A$ and $B$ be two ordered categorical variables having respectively $r$ and $c$ categories and $P$ be the $r \times c$ matrix of joint probabilities. We recall the original definition given by Lehmann (1966):

**Definition 1** A discrete bivariate distribution is said to be positive quadrant dependent if

$$Pr(A \leq a_h, B \leq b_k) \geq Pr(A \leq a_h)Pr(B \leq b_k),$$

where $h < r, k < c$. Essentially (1) says that within each of the $(r - 1)(c - 1)$ $2 \times 2$ tables obtained by partitioning the $r \times c$ joint probabilities matrix $P$, the first entry must be not less than the corresponding one under independence. This is clearly equivalent to the condition that the log-odds ratio within each of these tables, known as global log-odds ratio, is not negative. The global log odds ratios may be seen as the natural bivariate analog of the marginal global logits and together they provide a reparameterisation of $P$. Though the subject has already been discussed by various authors (e.g. Dale, 1986, Molenberghs and Lesaffre, 1994 and Glonek and McCullagh, 1995), it is useful to reformulate it in a more compact matrix notation. Given a vector of probabilities $\mathbf{\hat{f}} = (f_1 \cdots f_h)'$ summing up to 1, denote by $\mathbf{f}$ the vector containing the first $h - 1$ entries of $\mathbf{\hat{f}}$. Let $T_h$ be a
lower triangular matrix of ones of size $n$, and let $E_h$ be the $2(h-1) \times h$ matrix below

$$E_h = \begin{pmatrix} T_{h-1} & 0_{h-1} \\ 0_{h-1} & T'_{h-1} \end{pmatrix},$$

where $0_h$ is the $h$-dimensional column vector of zeros; notice that $E_h \hat{f} = (\gamma'_f - (1 - \gamma_f)'')$, where $\gamma_f = T_{h-1}f$ is the cumulative distribution of $f$.

Now let $\hat{\gamma} = P1_c$ and $\hat{\gamma} = P'1_r$ denote respectively the marginal distributions of $A$ and $B$ and $\hat{\gamma} = \text{vec}(P)$, where vec is the row vec operator. Letting $K_h = (-I_{h-1} I_{h-1})$, we are now ready to define the marginal global logits of $B$, the marginal global logits of $A$ and the global log-odds ratios respectively as

$$\eta_c = K_c \log(E_c \hat{c}), \quad \eta_r = K_r \log(E_r \hat{r}), \quad \eta_{rc} = (K_r \otimes K_c) \log([E_r \otimes E_c] \hat{p}),$$

where $\otimes$ denotes the Kronecker product. The PQD condition may be expressed as $\eta_{rc} \geq 0$; additional restrictions on the marginal logits $\eta_r$ and $\eta_c$ may also be of interest.

To deal with different sampling schemes, we introduce a general notation whose meaning will depend on the specific context. Let the data consist of a sample of $n$ subjects cross classified according to $A$ and $B$. The sampling scheme will determine the subset of the observed frequencies that constitute the response vector $y$ whose distribution will always be in the exponential family with an appropriate vector of canonical parameters $\theta$. Finally let $\eta$ be the vector of the parameters of interest: which, among $\eta_c$, $\eta_r$ and $\eta_{rc}$, will be the components of $\eta$ will also depend on the sampling scheme.

Hypotheses of interest are those that can be expressed by linear equality and/or inequality constraints on the vector $\eta$ and require a pair of matrices $C$ and $D$ such that the model of interest is defined by the space $\mathcal{H} = \{\eta \mid C\eta = 0, D\eta \geq 0\}$. For instance, when the inequality constraints are just those implied by PQD and the sampling scheme is single multinomial, $D = (0_{a,m} I_a)$ where $a = (r-1)(c-1)$ is the dimension of $\eta_{rc}$ and $m$ is the number of marginal parameters. On the whole this may be seen as an extension of the GLIM framework (Nelder and Wedderburn, 1972) in the sense that, though $\eta$ will be an invertible transformation of $\mu = E(y)$, this is not element wise.

**Parameter estimation**

The Fisher-scoring algorithm is a widely applicable method for computing maximum likelihood estimates under linear equality constraints. Recently, Dardanoni and Forcina (1998) have shown how the same algorithm can be used for fitting linear inequality constraints. Their results may be extended to allow for equality as well as inequality constraints. Because $\mu$ can be a rather complicated function of $P$, it is convenient to link $\theta$ directly to $\eta$. To this purpose let

$$\frac{\partial \theta}{\partial \eta'} = H, \quad \frac{\partial \mu}{\partial \theta'} = V;$$

by the chain rule, the score vector and the average Fisher information matrix may be written as

$$s = H'(y - \mu), \quad F = H'(V/n)H.$$
Maximum likelihood estimates of $\eta$ subject to $\mathcal{H}$ may be computed by an iterative algorithm based on maximizing, at each step, a quadratic form obtained by a second order Taylor series expansion of the log-likelihood around the updated estimate. More precisely, at step $h+1$, the algorithm performs the following operations:

1. compute the working dependent variable $e_h = \eta_h + H_h^{-1}V_h^{-1}(y - \mu_h)$;
2. set $\eta_{h+1}$ equal to the value of $\eta$ which maximizes

$$Q(\eta, \eta_h) = -\frac{n}{2}(e_h - \eta)'F_h(e_h - \eta),$$

subject to $\eta \in \mathcal{H}$.

The starting value, $\eta_0$, can be set equal to the unrestricted estimate of $\eta$.

Hypothesis testing

The problem of testing $\eta \in \mathcal{H}$ is an instance of order-restricted inference. A basic reference on the subject is the book by Robertson et al. (1988) which extends the earlier seminal contribution by Barlow et al. (1972). A concise introduction is in Shapiro (1988), and an application in much the same spirit as this paper is Dardanoni and Forcina (1998).

Let $S$ denote the saturated model, $t$ be the dimension of $y$ and $X$ be a $t \times l$ full-rank design matrix such that the regression model $\eta = X\beta$ is equivalent to the constraint $C\eta = 0$, with rank($C$) = $t - l$. Let also $\mathcal{D}$ be the convex cone $\{ \beta \mid DX\beta \geq 0 \}$ with rank($DX$) = $s$ and $B = (C' D')'$. Define as $\mathcal{H}_0 = \{ \eta \mid Be_n = 0 \}$ the model where all constraints hold as strict equalities. For example, when $\mathcal{H}$ defines the PQD condition, $\mathcal{H}_0$ is the independence model and when $\mathcal{H}$ states that one marginal distribution is stochastically larger than the other, $\mathcal{H}_0$ is the model of marginal homogeneity. Finally let $L(\hat{\eta} \mid \mathcal{M})$ denote the maximized log-likelihood for model $\mathcal{M}$ under the appropriate sampling scheme so that $\rho = 2[L(\hat{\eta} \mid \mathcal{S}) - L(\hat{\eta} \mid \mathcal{H})]$ and $\rho_0 = 2[L(\hat{\eta} \mid \mathcal{H}) - L(\hat{\eta} \mid \mathcal{H}_0)]$ are the log-likelihood ratio statistics for testing respectively $\mathcal{H}$ against the unrestricted model and $\mathcal{H}_0$ against $\mathcal{H}$.

An interesting result, which follows from the proof of Theorem 1 below, is that asymptotically $\rho_0$ and $\rho$ are distributed respectively like the squared norms, say $Q_0$ and $Q$, of the projection onto the convex cone defined by $\mathcal{H}$ and onto its dual of a vector having distribution $N(\eta, F^{-1})$, with $\eta \in \mathcal{H}_0$. The cumulative joint distribution of $Q_0$ and $Q$ takes the form (Raubertas et al., 1986, Theorem 3.4):

$$P_r(Q_0 \leq x, Q \leq y \mid \mathcal{H}_0) = \sum_{h=l-s}^l w_h(F^{-1}, \mathcal{H}) P_r(\chi_{h-(l-s)}^2 \leq x) P_r(\chi_{l-h}^2 \leq y)$$

where $w_h(F^{-1}, \mathcal{H})$ are nonnegative weights which sum to 1. The cumulative marginal distributions of $Q_0$ and $Q$ are easily derived from this equation and are an average of cumulative $\chi^2$ distributions and because of this such distributions are called chi-bar squared. In particular we write $Q_0 \sim \bar{\chi}^2(F^{-1}, \mathcal{H})$ to denote that $Q_0$ has a chi-bar squared distribution based on the covariance matrix $F^{-1}$ and the convex cone $\mathcal{H}$. The fact that $Q_0$ and $Q$ are determined by the same set of weights used in reverse order is one aspect of the so called duality relating the two random variables. General properties of such random
variables are given among others by Shapiro (1988) and Wolak (1989); see also Dardanoni and Forcina (1998) for a concise exposition. The main result of this section is stated precisely in the following theorem where, for two independent random variables with distribution $F$ and $G$, by $F + G$ we denote the distribution of their sum.

**Theorem 1** Under $H_0$ and the assumption that the elements of $P$ are strictly positive, the asymptotic distributions of $\rho$ and $\rho_0$ are

$$\rho \sim \chi^2(t-1) + \chi^2(\Sigma, D^\prime)$$

and

$$\rho_0 \sim \chi^2(\Sigma, D),$$

where $\Sigma = (X'FX)^{-1}$ and $D^\prime$ is the dual of $D$.

### 3. Unconditional inference

When the data consist of a sample of $n$ subjects, the observed frequencies may be treated as the outcome of a multinomial distribution with vector of parameters $p$ which, following the general dot convention established in section 2, is the vector of the joint probabilities without the last entry and so is of dimension $t = rc - 1$. To apply our general framework, let the response vector $y$ contain the observed frequencies arranged in the same way as in $p$ and the vector of canonical parameters $\theta$ be equal to $\log(p/p_c)$. Because of (2) and (3), the relation between $p$ and the vector $\eta$ may be written in the form $\eta = R \log(Mp)$, $R = \text{diag}(K_r, K_r \otimes K_r, M = (1_r \otimes E'_r, E'_r \otimes 1_c, E'_r \otimes E'_c)'$. This is an instance of a saturated generalized log-linear model as defined by Lang (1996).

To maximize the log-likelihood $L(\eta) = y \theta - n \log[1 + 1' \exp(\theta)]$ subject to $\eta \in H$, we need to compute the derivative of $\theta$ with respect to $\eta$. Though an explicit expression for the $H$ matrix is not available, its inverse may be computed by the chain rule as

$$H^{-1} = \frac{\partial \eta}{\partial \hat{p}} \frac{\partial \hat{p}}{\partial p} \frac{\partial p}{\partial \theta} = R \text{diag}(Mp)^{-1} M \tilde{K}'_{t+1, \Omega p}$$

where $\tilde{K}_h = (I_{h-1} - 1_{h-1})$ so that $\hat{p} = \text{constant} + \tilde{K}'_{t+1, p}$ and, for any vector of probabilities $f$ without the last element, $\Omega_f = \text{diag}(f) - ff'$ is the kernel of the multinomial variance. Moreover, because $\mu = np$, from general properties of the exponential family, $V/n = \Omega_p$, the expected information matrix has the form $F = H \Omega_p H$.

**Theorem 2** Under independence of $A$ and $B$, the asymptotic distribution of the log-likelihood ratio statistics for testing PQD against the saturated model and independence against PQD are

$$\rho \sim \tilde{\chi}^2(S_r \otimes S_c, \Omega^o_a)$$

and

$$\rho_0 \sim \tilde{\chi}^2(S_r \otimes S_c, \Omega_a)$$

where $a = (r-1)(c-1)$, $\Omega_a$ is the positive orthant, $\Omega^o_a$ its dual and, for any vector of probabilities $f$, $S_f = T_{h-1,1} \Omega_f T_{h-1,1}'$.

Notice that the distributions of $\rho$ and $\rho_0$ depend on the marginal distributions of $A$ and $B$; to get rid of the corresponding nuisance parameters we search for distributions that are least favorable to the null hypothesis. Though such distributions assign to the nuisance parameters values that are often very different from any realistic estimate, they
provide easily computed limits for the significance level. One could also condition on the observed margins, as we do in the following sections. Below, for any pair of random variables $X_1$ and $X_2$ let $X_1 \leq_s X_2$ denote the fact that $X_1$ is stochastically smaller than $X_2$.

**Theorem 3** Under independence and for any possible pair of marginal distributions, the asymptotic distributions of $\rho$ and $\rho_0$ satisfy

$$
\bar{\chi}^2(1, O_1) \leq_s \rho \leq_s \bar{\chi}^2(I_a, O_a) \\
\bar{\chi}^2(I_a, O_a) \leq_s \rho_0 \leq_s \chi^2_{a-1} + \bar{\chi}^2(1, O_1)
$$

where $a = (r - 1)(c - 1)$.

**Conditioning on the row totals**

When the marginal probabilities are nuisance parameters, we may at least reduce their number by conditioning on the row totals $(n_{1a} \ldots n_{ra})$. By doing so, the frequencies of the $i$-th row may be treated as the outcome of a multinomial distribution with parameter $q_i$, let $\mathbf{q} = (q_i^1 \ldots q_i^c)'$, $y_i$ contain the frequencies of the $i$-th row arranged as in $q_i$. $\mathbf{y} = (y_i^1 \ldots y_i^c)'$ and the vector of canonical parameters contain, within each row, the logits relative to the last column so that $\theta = \text{log} \{[\text{diag}(q_1, \ldots, q_c)^{-1} \otimes I_{c-1}] \mathbf{q}\}$. When constructing the vector of parameters for the saturated model drop the block corresponding to the marginal logits of $A$ so that $\mathbf{q} = R \log(M \hat{\mathbf{p}})$ where $R = \text{diag}(\mathbf{K}_c, \mathbf{K}_r \otimes \mathbf{K}_c)$ and $M' = (1_r \otimes \mathbf{E}_c' \mathbf{E}_c' \otimes \mathbf{E}_c')$.

The crucial step for maximizing the log-likelihood $L(\eta) = \mathbf{y}' \theta - \sum_i n_{ia} \log [1 + 1' \exp(\theta_i)]$ concerns again the relation between $\eta$ and $\theta$. About this notice that we may write $\hat{\mathbf{p}} = \text{constant} + [\text{diag}(\hat{\mathbf{r}}) \otimes \hat{\mathbf{K}}_c'] \mathbf{q}$. Moreover, because $E(y_i) = n_{ia} q_i = n_{ia} \exp(\theta_i)/[1 + 1' \exp(\theta_i)]$,

$$H^{-1} = \frac{\partial \eta}{\partial \hat{\mathbf{p}}} \frac{\partial \mathbf{q}}{\partial \hat{\mathbf{q}}} \frac{\partial \hat{\mathbf{q}}}{\partial \theta'} = R \text{diag}(M \hat{\mathbf{p}})^{-1} M [\text{diag}(\hat{\mathbf{r}}) \otimes \hat{\mathbf{K}}_c'] \Omega(q),$$

where $\Omega(q)$ is block diagonal with elements equal to $\Omega_{q_i}$; the variance matrix $\mathbf{V}$ is also block diagonal, the $i$th block being $n_{ia} \Omega_{q_i}$. In the proof of Theorem 2 we show that the information matrix may be simply obtained from the information matrix of the unconditional approach by deleting the block corresponding to the marginal distribution of $A$ and replacing $\hat{\mathbf{r}}$ with its estimate. It follows that the asymptotic distribution of $\rho$ and $\rho_0$ for testing PQD are, under independence, the same as in the unconditional case. However, while the vector $\hat{\mathbf{r}}$ is now obtained from the observed frequencies, the same does not hold for $\hat{\mathbf{e}}$ and so $\eta_c$ is still a vector of nuisance parameters. The following Theorem may be useful when these are unknown:

**Theorem 4** Under independence and for any possible marginal distribution of $B$, the asymptotic distributions of $\rho$ and $\rho_0$ satisfy

$$
\bar{\chi}^2(S_r^{-1}, O_{r-1}) \leq_s \rho \leq_s \sum_{i=1}^{c-1} \bar{\chi}^2(S_r^{-1}, O_{r-1}) \\
\sum_{i=1}^{c-1} \bar{\chi}^2(S_r, O_{r-1}) \leq_s \rho_0 \leq_s \chi^2_{(r-1)(c-2)} + \bar{\chi}^2(S_r, O_{r-1}),
$$

$- 118 -$
where $\sum \chi_i^2(S^{-1}, O_{r-1})$ denotes convolution of i.i.d. chi-bar squared variables.

**Conditioning on both row and column totals**

In principle, one could get rid of all nuisance parameters by conditioning on both row and column totals. However, this conditional distribution, known as the multivariate generalized hypergeometric, is almost intractable whenever $n$ is moderately large or $r$ and $c$ are greater than 3. Because this distribution belongs to the exponential family, our version of the constrained Fisher-scoring algorithm is applicable and usually converges in very few steps. However, it still requires computation of the likelihood, expectation and variance again at each step. Thus its use is not feasible in many real applications and the reader is referred to the extended version of this paper.

For larger tables we suggest replacing the exact procedure with the following which is based on a similar approach used by Agresti and Coull (1998). Take the product multinomial, obtained by conditioning only to the row totals, and maximize under the additional constraint that the column totals must be equal to those of the observed table. This constraint, once expressed into the notations of the previous section, has the form $C\eta = K_\sigma \log(E_c)\hat{c}$ where $\hat{c}$ are the observed column proportions and $C = (I_{c-1} \ 0_{c-1,a})$.

To assess the performance of this product multinomial approximation, we generated 500 samples of size 30 from the hypergeometric with uniform row and column margins under independence and compared the actual distributions of $\rho_0$ and $\rho$ produced by the exact procedure with those produced by this method and those produced by a normal approximation of the hypergeometric. The correlation coefficient between the corresponding statistics computed exactly and with the product multinomial approximation is always greater than .999 and the set of samples in the rejection area at 10, 5 and 1% are identical. The method based on conditioning on row and column totals after approximating the full multinomial with the normal distribution does slightly worse: the correlation coefficient is approximately 0.97, with approximately 2% disagreement relative to the exact rejection regions.

**4. Stronger notions of positive dependence**

The parameterization adopted in this paper can be easily extended to deal with stronger notions of positive dependence by simply redefining the $E_r$ and $E_c$ matrices in an appropriate way. For instance, the three notions of positive dependence (simple, uniform, likelihood ratio) considered by Dardanoni and Forcina (1998) are equivalent to the condition that the local-global, local-continuation and local-local log-odds ratios respectively are non negative, and are also known as regression dependence, hazard rate and TP2 orderings in the literature. Now if we define

$$E^L_h = \begin{pmatrix} I_{h-1} & 0_{h-1} \\ 0_{h-1} & I_{h-1} \end{pmatrix} \quad \text{and} \quad E^C_h = \begin{pmatrix} I_{h-1} & 0_{h-1} \\ 0_{h-1} & T_{h-1} \end{pmatrix}$$

and replace $E_r$ with $E^L_r$ in equations (2) and (3), $\eta_{rc}$ becomes the local-global log-odds ratios. The local-continuation and the local-local parameterizations can be obtained by using $E^L_r$ and $E^C_r$ and $E^L_c$ and $E^C_c$ respectively. In this way the approach of Dardanoni and Forcina (1998) can be extended to deal with models involving equality constraints, additional inequality constraints apart from those
5. Application

A mobility table. Consider a dataset taken from Hand et al. (1994) where 847 Hertfordshire men are classified according to social class and social class of their fathers. The data were arranged in a $9 \times 9$ table and we collapsed the last 3 categories into 1 to reduce sparseness obtaining the following $7 \times 7$ table.

<table>
<thead>
<tr>
<th>Father ($x_h$) and Son ($y_h$) Social Class</th>
<th>$y_1$</th>
<th>$y_2$</th>
<th>$y_3$</th>
<th>$y_4$</th>
<th>$y_5$</th>
<th>$y_6$</th>
<th>$y_7$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>7</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>$x_2$</td>
<td>4</td>
<td>20</td>
<td>9</td>
<td>22</td>
<td>16</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>$x_3$</td>
<td>4</td>
<td>12</td>
<td>5</td>
<td>7</td>
<td>5</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>$x_4$</td>
<td>12</td>
<td>54</td>
<td>32</td>
<td>143</td>
<td>50</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>$x_5$</td>
<td>5</td>
<td>29</td>
<td>34</td>
<td>116</td>
<td>67</td>
<td>22</td>
<td>6</td>
</tr>
<tr>
<td>$x_6$</td>
<td>0</td>
<td>88</td>
<td>50</td>
<td>18</td>
<td>12</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>$x_7$</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>14</td>
<td>11</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

The fact that independence is a rather unplausible model in this context is supported by the large value of $\rho_0 = 84.019$ when testing independence against PQD. The model of quasi-independence implies 29 equality constraints on the local-local log-odds ratios and gives also a bad fit with $\rho = 67.165$.

When testing PQD against general alternatives, $\rho = 4.146$. To address the question as to whether observed violations of positive association are random, we report the results of the various methods for a comparison. The estimated local $p$-value (based on replacing the nuisance parameters with their maximum likelihood estimates) equal to 0.7799. The corresponding bounds derived from Theorem 3, which are equal to 0.0209 and 0.9977 respectively, are relatively useless in this case. The situation improves if we condition on the row totals giving $\rho = 4.149$, with $p$-value bounded between 0.1478 and 0.8318 (Theorem 4). Using the multinomial approximation to the hypergeometric, $\rho = 4.21$ and the Monte Carlo estimate of the $p$-value is 0.823.

Sommers and Conlisk (1978) have argued that observed mobility tables often tend to display a symmetric pattern. A weaker notion of symmetry in our context is symmetry of the GOR, that is the assumption that $\eta_{c}(i, j) = \eta_{c}(j, i)$, for all $i \neq j$, with the marginal distributions unconstrained. This model, which we may call QGS, is the analog of the model of quasi-symmetry (based on symmetry of the local-local odds-ratios). A one sided alternative to QGS is the hypothesis $\eta_{c}(i, j) \geq \eta_{c}(j, i)$, for $i > j$ which can be interpreted as more mobility from poor fathers to richer sons than viceversa. The fact that frequencies in table 2 are larger on the lower diagonal might suggest an alternative in the opposite direction; notice however that this is a by-product of the fact that the row marginal is stochastically larger than the column marginal. Both hypotheses are easily
implemented in our framework, and the statistics for testing QGS against asymmetry and asymmetry against no restrictions are respectively $\rho_0 = 17.132$ and $\rho = 7.3163$. Their estimated $p$-values are equal to 0.1401 and to 0.1026 respectively and seem to indicate that the model of QGS can be accepted.

If, in addition to PQD, we assume a proportional odds model for the marginals, $\rho = 14.7859$ with the unconditional approach, and the contribution due to the additional restrictions is 10.6396, which is to be compared against a $\chi^2_5$ and is non significant at the 5% level. The marginal logits exhibit almost a linear trend with the exception of a shift due to the third social class, for both fathers and sons. This structure is consistent with social classes being defined by equally spaced cut points (except for that between the third and 4th social class) on an underlying logistic distribution. Fitting this in addition to PQD implies 3 additional equality constraints on each marginal giving $\rho = 19.6247$ and the net contribution due to the equality constraints is 15.4784 with a $p$-value of 0.0168 computed under the $\chi^2_6$ distribution.

References


University Press.


