Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics & Data Analysis 51 (2007) 4643–4655
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Bayesian estimation of unrestricted and order-restricted association models for a two-way contingency table

G. Iliopoulos[a], M. Kateri[a,*], I. Ntzoufras[b]

[a]*Department of Statistics and Insurance Science, University of Piraeus, 80 Karaoli & Dimitriou str., 18534 Piraeus, Greece*
[b]*Department of Statistics, Athens University of Economics and Business, 76 Patision str, 10434 Athens, Greece*

Received 29 March 2006; received in revised form 4 August 2006; accepted 4 August 2006
Available online 30 August 2006

Abstract

In two-way contingency tables analysis, a popular class of models for describing the structure of the association between the two categorical variables are the so-called "association" models. Such models assign scores to the classification variables which can be either fixed and prespecified or unknown parameters to be estimated. Under the row–column (RC) association model, both row and column scores are unknown parameters without any restriction concerning their ordinality. It is natural to impose order restrictions on the scores when the classification variables are ordinal. The Bayesian approach for the RC (unrestricted and restricted) model is adopted. MCMC methods are facilitated in order the parameters to be estimated. Furthermore, an alternative parametrization of the association models is proposed. This new parametrization simplifies computation in the MCMC procedure and leads to a natural parameter space for the order constrained model. The proposed methodology is illustrated via a popular dataset.

Keywords: Contingency tables; Association models; Ordinal classification variables; MCMC methods

1. Introduction

Let $\boldsymbol{y} = (y_{ij})$ be a frequency table, produced by the cross-classification of two categorical variables X and Y with I and J categories, respectively. The sampling scheme can be either multinomial or independent Poisson for each cell of the table. The total sample size of this table is denoted by $n = y_{\cdot\cdot} = \sum_{i=1}^{I}\sum_{j=1}^{J} y_{ij}$ while $\boldsymbol{\Pi} = (\pi_{ij})$ stands for the underlying probability table. Considering scores $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_I)$ and $\boldsymbol{v} = (v_1, v_2, \ldots, v_J)$ for the categories of the row and column classification variables, respectively, and replacing the interaction parameter λ_{ij}^{XY} of the saturated log-linear model (hereby denoted by S) by the product $\mu_i v_j$, the family of association models (Goodman, 1979, 1981) is generated. More formally, association models are expressed as

$$\log \pi_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + \mu_i v_j \quad \text{for } i = 1, \ldots, I, \quad j = 1, \ldots, J. \tag{1}$$

* Corresponding author. Tel.: +30 210 4142467; fax: +30 210 4142340.
E-mail address: mkateri@unipi.gr (M. Kateri).

0167-9473/$ - see front matter
doi:10.1016/j.csda.2006.08.013

For identifiability purposes, the sum-to-zero constraints are imposed on row and column main effects λ_i^X and λ_j^Y, i.e.,

$$\sum_{i=1}^{I} \lambda_i^X = \sum_{j=1}^{J} \lambda_j^Y = 0, \tag{2}$$

as well as on the row and column scores, as stated below

$$\sum_{i=1}^{I} \mu_i = \sum_{j=1}^{J} v_j = 0. \tag{3}$$

Moreover, an additional constraint is needed in order to identify the scores' estimates (due to the fact that (1) is not linear in its parameters). Association models were initially introduced by (1) but later on (cf. Goodman, 1985) they were defined through the equivalent expression

$$\log \pi_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + \phi \tilde{\mu}_i \tilde{v}_j \quad \text{for } i = 1, \ldots, I, \quad j = 1, \ldots, J, \tag{4}$$

which reveals the analogies of the RC model to the classical correspondence analysis (CA) or canonical correlation model. The parameter ϕ is an intrinsic association parameter and is redundant. Its introduction to the RC model definition serves interpretation purposes and results to the introduction of an extra constraint on the scores' parameters $\tilde{\mu}_i$ and $\tilde{v}_j$. This leads to the symmetric treatment of the row and column scores. In this case, the usual imposed constraints on the scores' parameters of model (4) are

$$\sum_{i=1}^{I} w_{1i} \tilde{\mu}_i = \sum_{j=1}^{J} w_{2j} \tilde{v}_j = 0 \quad \text{and} \quad \sum_{i=1}^{I} w_{1i} \tilde{\mu}_i^2 = \sum_{j=1}^{J} w_{2j} \tilde{v}_j^2 = 1, \tag{5}$$

where w_{1i}, w_{2j} are some positive weights. Common values of the weights are $w_{1i} = w_{2j} = 1$, for all i, j or $w_{1i} = \pi_{i\cdot}$ and $w_{2j} = \pi_{\cdot j}$, $i = 1, \ldots, I$, $j = 1, \ldots, J$. Arguments for the appropriate choice of the weights can be found in Becker and Clogg (1989). The use of marginal weights reveals analogies of association to canonical correlation models.

According to the assumptions made concerning the row and column scores, alternative association models are obtained. Hence, if both row and column scores are known and fixed then model (4) degenerates to the linear by linear (LL) association model which has one parameter (namely ϕ) additional to the independence model (denoted by $\mathscr{I}$). When only the row or the column scores are parametric then we end up with the row (R) or the column (C) model, respectively. Finally, when both row and column scores are parameters under estimation, then the multiplicative RC association model is achieved. The models LL, R and C are special cases of log-linear models while RC is not.

The association models (1) can equivalently be expressed in terms of the $(I-1)(J-1)$ odds ratios of the adjacent categories as

$$\log\left(\frac{\pi_{ij}\pi_{i+1,j+1}}{\pi_{i,j+1}\pi_{i+1,j}}\right) = (\mu_{i+1} - \mu_i)(v_{j+1} - v_j) \quad \text{for } i = 1, \ldots, I-1, \quad j = 1, \ldots, J-1, \tag{6}$$

offering an insight concerning the meaning of the association parameters. To be more specific, this expression clarifies that the difference between scores of successive categories is important for the association structure rather than their values themselves. When the scores are known and fixed, a natural and usual choice is to consider them equidistant for successive categories, unless other information is available. Hence, known scores are monotonic, which is not necessarily the case for parametric ones. Consequently, in the case of an ordinal classification variable, the estimated parametric scores may violate the natural underlying ordering. Such cases rise the necessity of setting up a model using an order-restricted score estimation.

Association models can be implemented through standard frequentists' approach either by the Newton's unidimensional algorithm (Goodman, 1979; Becker, 1990) or the Newton–Raphson algorithm (Haberman, 1995; Ait-Sidi-Allal et al., 2004). The algorithms proposed by these authors is for the general association model of order K, denoted as RC(K), and defined by

$$\log \pi_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + \sum_{k=1}^{K} \phi_k \mu_{ik} v_{jk},$$

for $i=1,\dots,I$, $j=1,\dots,J$, and $1\leqslant K\leqslant M=\min(I,J)-1$. For $K=1$ it is reduced to the RC model while for $K=M$, the RC(M) model is an equivalent expression of the saturated model. Order-restricted estimation of the C model (and consequently of the R) is provided by Agresti et al. (1987) through the pooled adjacent violation algorithm while for the general RC model the order-restricted maximum likelihood estimation is due to Ritov and Gilula (1991, 1993). They also provided the order-restricted maximum likelihood estimation for the correlation model (cf. Goodman, 1981).

From the Bayesian point of view, an empirical Bayes analysis of the RC model has been considered by Chuang (1982) and Evans et al. (1993). They set a prior distribution for the parameters of the saturated log-linear model and estimated the posterior distribution of the parameters of the RC through minimization of the Euclidean squared distance between the interaction terms of the two models. Further on, a smoothing of the cell probabilities based on the LL model has been suggested by Agresti and Chuang (1989), while Albert (1997) worked also the Bayesian estimation of the LL model. However, the above approaches do not allow for the implementation of the full Bayesian inference, which is provided for the general RC(K) model by Kateri et al. (2005). The algorithm provided by the latter work can be directly applied for the simple RC model by setting $K=1$. In the following we propose an alternative parametrization for the RC model which considerably simplifies the parametric space and helps us to construct simpler MCMC samplers. Moreover, the Bayesian inference concerning the order-restricted RC model has not been treated yet. This constitutes the main contribution of our paper.

The parametrization used in Kateri et al. (2005) serves the needs of the more complex RC(K) model and since it could be simplified in the context of the simple RC model, we first propose in Section 2 an alternative approach for the Bayesian estimation of the RC model. Of course, variations of this algorithm can be used for the simpler R and C models. Next, we proceed in Section 3 with the estimation of the RC model with order constraints applied on one or both set of scores. The evaluation of the model adequacy through information based criteria (AIC, BIC and DIC) is provided in Section 4. In Section 5, the proposed algorithms and evaluation procedure are illustrated on a classical data set. Results are summarized and topics for further research are introduced in the final section.

2. Bayesian analysis of the RC model

2.1. Scores' parametrization and prior distributions

Although the usual form of the RC model (4) and its parametrization is convenient for interpretation, it complicates the corresponding Bayesian analysis. Constraints (5) need to be obeyed by the prior distributions of the parameters, a fact that leads to complicated posterior distributions. As a result, sophisticated and complex MCMC algorithms need to be constructed in order to estimate such posterior distributions. For details on this procedure in the context of the RC(K) model, see in Kateri et al. (2005).

Here, we shall adopt the alternative parametrization of the RC model in (1). Hence the constraints (3) lead to

$$\mu_1=-\sum_{i=2}^{I}\mu_i \quad \text{and} \quad \nu_1=-\sum_{j=2}^{J}\nu_j. \tag{7}$$

As already argued in Section 1, an additional constraint is needed for identifiability purposes. Without loss of generality we set

$$\nu_2=1. \tag{8}$$

Alternatively, we could set a similar constraint on μ_2 or any other row or column category score. This reparametrization is feasible for $I\geqslant 3$ (for the row scores) and for $J\geqslant 3$ (for the column scores). In different case ($I=2$ or $J=2$) the RC model is saturated and equivalent to the C or R model, respectively, and will be fitted through the corresponding procedures (presented later on).

In this framework, the resulting posterior distributions (and the corresponding MCMC scheme) are simpler and easier to work, since we avoid the unit constraints on the I- and J-dimensional spheres. Moreover, it can be used directly for the ordered RC model which is the main objective of this paper. Since the cells probabilities' estimates under an association model are not affected by linear transformations of the scores, one could find the sample posterior means of the scores and rescale them in order to obey the desired constraints. Hence, a sample from the standard parametrization

(5) can be easily obtained by transforming the generated values of the MCMC output in each iteration (see Viele and Srinivasan, 2000, for similar practice in the analysis of variance framework).

Concerning the prior structure of the parameters, we facilitate the usual normal priors for $\boldsymbol{\lambda}^X$ and $\boldsymbol{\lambda}^Y$ given by

$$\lambda_i^X \sim \text{Normal}(0, \sigma^2_{\lambda_i^X}), \quad i = 2, \ldots, I,$$

and

$$\lambda_j^Y \sim \text{Normal}(0, \sigma^2_{\lambda_j^Y}), \quad j = 2, \ldots, J.$$

No prior is needed for λ_1^X and λ_1^Y since they are simply functions of the rest of the parameters ($\lambda_1^X = -\sum_{i=2}^I \lambda_i^X$ and $\lambda_1^Y = -\sum_{j=2}^J \lambda_j^Y$). Similarly, since μ_1, v_1 and v_2 are specified via (7) and (8), we need to assign prior distributions only for the remaining row and column scores. Moreover, these constraints simplify the parameter space and allow the use of normal prior distributions of the form

$$\mu_i \sim \text{Normal}(0, \sigma^2_{\mu_i}), \quad i = 2, \ldots, I,$$

and

$$v_j \sim \text{Normal}(0, \sigma^2_{v_j}), \quad j = 3, \ldots, J.$$

When no information is available we use large prior variances $\sigma^2_{\lambda_i^X}$, $\sigma^2_{\lambda_j^Y}$, $\sigma^2_{\mu_i}$ and $\sigma^2_{v_j}$ (for example 10^3) to represent prior ignorance.

2.2. MCMC

The vector of the non-redundant model parameters involved in the posterior distribution are

$$\begin{aligned}\boldsymbol{\theta} &= (\boldsymbol{\lambda}^X_{\backslash\{1\}}, \boldsymbol{\lambda}^Y_{\backslash\{1\}}, \boldsymbol{\mu}_{\backslash\{1\}}, \boldsymbol{v}_{\backslash\{1,2\}}) \\ &= (\lambda_2^X, \ldots, \lambda_I^X, \lambda_2^Y, \ldots, \lambda_J^Y, \mu_2, \ldots, \mu_I, v_3 \ldots, v_J).\end{aligned}$$

The remaining parameters are functions of $\boldsymbol{\theta}$ given by constraints (2), (7) and (8). Similarly, the constant term λ is evaluated by

$$\lambda = \lambda(\boldsymbol{\theta}) = -\log \sum_{i=1}^I \sum_{j=1}^J \exp\{\lambda_i^X + \lambda_j^Y + \mu_i v_j\}.$$

The likelihood is given by

$$f(\boldsymbol{y}|\boldsymbol{\theta}) = \frac{n!}{\prod_{i,j} y_{ij}!} \exp\left\{ n\lambda(\boldsymbol{\theta}) + \sum_{i=1}^I y_{i.}\lambda_i^X + \sum_{j=1}^J y_{.j}\lambda_j^Y + \sum_{i=1}^I \sum_{j=1}^J y_{ij}\mu_i v_j \right\},$$

while the posterior distribution

$$f(\boldsymbol{\theta}|\boldsymbol{y}) \propto f(\boldsymbol{y}|\boldsymbol{\theta}) \prod_{i=2}^I f(\lambda_i^X) \prod_{j=2}^J f(\lambda_j^Y) \prod_{i=2}^J f(\mu_i) \prod_{j=3}^J f(v_i)$$

is estimated using MCMC. The conditional posterior distributions used in the Gibbs sampler are given by

$$f(\lambda_i^X | \boldsymbol{\theta}_{\backslash \lambda_i^X}, \boldsymbol{y}) \propto \exp\{n\lambda(\boldsymbol{\theta}) + (y_{i.} - y_{1.})\lambda_i^X\} f(\lambda_i^X) \quad \text{for } i = 2, \ldots, I, \tag{9}$$

$$f(\lambda_j^Y | \boldsymbol{\theta}_{\backslash \lambda_j^Y}, \boldsymbol{y}) \propto \exp\{n\lambda(\boldsymbol{\theta}) + (y_{.j} - y_{.1})\lambda_j^Y\} f(\lambda_j^Y) \quad \text{for } j = 2, \ldots, J, \tag{10}$$

$$f(\mu_i | \boldsymbol{\theta}_{\backslash \mu_i}, \boldsymbol{y}) \propto \exp\left\{ n\lambda(\boldsymbol{\theta}) + \mu_i \sum_{j=1}^J (y_{ij} - y_{1j}) v_j \right\} f(\mu_i), \tag{11}$$

for $i = 2, \ldots, I$, and

$$f(v_j|\boldsymbol{\theta}_{\backslash v_j}, \mathbf{y}) \propto \exp\left\{n\lambda(\boldsymbol{\theta}) + v_j \sum_{i=1}^{I}(y_{ij} - y_{i1})\mu_i\right\} f(v_i), \tag{12}$$

for $j = 3, \ldots, J$. We use the above equations to construct a Metropolis within Gibbs sampler. All the parameters are generated using the random walk Metropolis approach.

In case we would like to have estimates for the parameters of the (4) version of the RC model, we could rescale the posterior mean estimates of the row and column scores by $\tilde{\mu}_i = \mu_i/\sqrt{\sum_{i=1}^{I}\mu_i^2}$ and $\tilde{v}_j = v_j/\sqrt{\sum_{j=1}^{J}v_j^2}$, respectively, in order to harmonize them completely with the constraint (5) with unit weights. The transformed scores are in agreement with model (4) with $\phi = \sqrt{\sum_{i=1}^{I}\mu_i^2}\sqrt{\sum_{j=1}^{J}v_j^2}$. If a different weighting system is used, the scores are transformed analogously.

3. The ordered RC model

3.1. Prior distributions and ordered constraints

As it is argued in Section 1, in case of estimating scores of an ordinal variable, the ordering of the categories is not necessarily reflected in the simple RC model, giving ground for the development of the order-restricted estimation of the scores. Thus, the order constrainted RC model requires for the row scores to have ascending ($\mu_1 < \mu_2 < \cdots < \mu_I$) or descending ($\mu_1 > \mu_2 > \cdots > \mu_I$) order and/or the same argument holds for the column scores $v_1, \ldots, v_J$. If this is the case, in order to simplify the estimation procedure, we always consider the ascending order for the scores ($\mu_1 < \mu_2 < \cdots < \mu_I$ and $v_1 < v_2 < \cdots < v_J$) and multiply their product term in model (1) by 1 or -1 according to whether the association is assumed positive or negative.

Since the crucial quantity is the difference between two successive scores and not the individual scores (see (6)), we shall focus on these differences. Hence, we define the differences between two subsequent row and column scores,

$$\delta_{\mu_1} = \mu_1, \quad \delta_{\mu_i} = \mu_i - \mu_{i-1}, \;\; i = 2, \ldots, I,$$

and

$$\delta_{v_1} = v_1, \quad \delta_{v_j} = v_j - v_{j-1}, \;\; j = 2, \ldots, J.$$

From the above we have that the row and column scores, μ_i and v_j, respectively, for $i \neq 1$ and $j \neq 1$ are provided by

$$\mu_i = \sum_{k=1}^{i}\delta_{\mu_k}, \;\; i = 2, \ldots, I, \quad v_j = \sum_{k=1}^{j}\delta_{v_k}, \;\; j = 3, \ldots, J,$$

while μ_1 and v_1 are given by

$$\mu_1 = \delta_{\mu_1} = -I^{-1}\sum_{k=2}^{I}(I - k + 1)\delta_{\mu_k} \tag{13}$$

and

$$v_1 = \delta_{v_1} = -\sum_{k=3}^{J}(J - k + 1)\delta_{v_k} - (J - 1), \tag{14}$$

respectively. Note that constraint (8) implies that $\delta_{v_2} = 1 + \sum_{k=3}^{J}(J - k + 1)\delta_{v_k} + (J - 1)$. Hence we need to specify priors directly on δ_{μ_i} for $i = 2, \ldots, I$ and δ_{v_j} for $j = 3, \ldots, J$. Additionally, the prior distributions for δ_{μ_i} and δ_{v_j} are defined in the range of positive numbers, so that the ascending nature of the row and column scores is assumed. In order to retain a prior similar to the normal one used in the simple (unordered) RC model, we may use a truncated normal prior, for example $\delta_{\mu_i} \sim N(0, \sigma^2_{\delta_{\mu_i}})I(\delta_{\mu_i} > 0)$. Another alternative is the lognormal distribution of the type $\delta_{\mu_i} \sim LN(0, \sigma^2_{\log \delta_{\mu_i}})$. The main disadvantage of both the above prior setups is that the resulting prior means $E(\delta_{\mu_i})$ will depend on the prior scale parameters $\sigma^2_{\delta_{\mu_i}}$ and $\sigma^2_{\log \delta_{\mu_i}}$, respectively. Thus, for the cases where we need to express low prior information using large values of $\sigma^2_{\delta_{\mu_i}}$ or $\sigma^2_{\log \delta_{\mu_i}}$ this will result to large prior means that may finally influence the posterior distributions or cause computational problems. For this reason, we prefer to adopt a gamma prior distribution for positive parameters, i.e.,

$$\delta_{\mu_i} \sim \text{Gamma}(a_{\mu_i}, b_{\mu_i}), \quad i = 2 \ldots I,$$

and

$$\delta_{\nu_j} \sim \text{Gamma}(a_{\nu_j}, b_{\nu_j}), \quad j = 3, \ldots, J.$$

We remind that δ_{μ_1}, δ_{ν_1} and δ_{ν_2} will be specified directly via Eqs. (13) and (14).

3.2. MCMC sampler

To facilitate notation, we denote the vector of all model parameters of the order-restricted model by

$$\begin{aligned}\boldsymbol{\vartheta} &= (\boldsymbol{\lambda}^X_{\setminus\{1\}}, \boldsymbol{\lambda}^Y_{\setminus\{1\}}, \boldsymbol{\delta}_{\mu\setminus\{1\}}, \boldsymbol{\delta}_{\nu\setminus\{1,2\}}) \\ &= (\lambda^X_2, \ldots, \lambda^X_I, \lambda^Y_2, \ldots, \lambda^Y_J, \delta_{\mu_2}, \ldots, \delta_{\mu_I}, \delta_{\nu_3}, \ldots, \delta_{\nu_J}).\end{aligned}$$

Note that for every set of the ordinal parameters $\boldsymbol{\vartheta}$ we can calculate a corresponding set of $\boldsymbol{\theta}$ parameters using the equations above (13) and (14). Hence, in the order-restricted model, the row and column parameters μ_i and ν_j are simply functions of their corresponding differences δ_{μ_i} and δ_{ν_j} denoted by $\mu_i(\boldsymbol{\delta}_\mu)$ and $\nu_j(\boldsymbol{\delta}_\nu)$. Moreover, the constant term λ is a function of the whole parameter vector $\boldsymbol{\vartheta}$ and is given in a similar way as in the simple RC model by

$$\lambda(\boldsymbol{\vartheta}) = -\log \sum_{i=1}^{I} \sum_{j=1}^{J} \exp\{\lambda^X_i + \lambda^Y_j + \mu_i(\boldsymbol{\delta}_\mu)\nu_j(\boldsymbol{\delta}_\nu)\}.$$

The conditional posterior distributions of λ^X_i and λ^Y_j will remain the same as in the unordered case, hence $f(\lambda^X_i|\boldsymbol{\vartheta}_{\setminus\lambda^X_i})$ and $f(\lambda^Y_j|\boldsymbol{\vartheta}_{\setminus\lambda^Y_j})$ are given by (9) and (10). The row and column parameters are functions of $\boldsymbol{\delta}_\mu$ and $\boldsymbol{\delta}_\nu$ resulting in the following posterior distributions of row and column score differences:

$$f(\delta_{\mu_i}|\boldsymbol{\vartheta}_{\setminus\delta_{\mu_i}}, \boldsymbol{y}) \propto \exp\left\{n\lambda(\boldsymbol{\vartheta}) + \left(\sum_{j=3}^{J} \psi_{ij}\delta_{\nu_j}\right)\delta_{\mu_i}\right\} f(\delta_{\mu_i})$$

and

$$f(\delta_{\nu_j}|\boldsymbol{\vartheta}_{\setminus\delta_{\nu_j}}, \boldsymbol{y}) \propto \exp\left\{n\lambda(\boldsymbol{\vartheta}) + \left(\sum_{i=3}^{I} \psi_{ij}\delta_{\mu_i}\right)\delta_{\nu_j}\right\} f(\delta_{\nu_j}),$$

where

$$\psi_{ij} = (I - i + 1)(J - j + 1)y_{11} - (J - j + 1)Y_{i+,1} - (I - i + 1)Y_{1,j+} + Y_{i+,j+}$$

for $Y_{i+,j} = \sum_{k=i}^{I} y_{kj}$, $Y_{i,j+} = \sum_{\ell=j}^{J} y_{i\ell}$ and $Y_{i+,j+} = \sum_{k=i}^{I}\sum_{\ell=j}^{J} y_{k\ell}$. We construct a Metropolis within Gibbs algorithm. We use symmetric proposals for updating $\boldsymbol{\lambda}^X$ and $\boldsymbol{\lambda}^Y$ resulting to simple random walk Metropolis steps. For the differences $\boldsymbol{\delta}_{\mu\setminus\{1\}}$ and $\boldsymbol{\delta}_{\nu\setminus\{1,2\}}$ we use multiplicative random walks (symmetric normal proposals on their logarithms) in order to propose positive values. In all steps, the variances of the proposals have been tuned to achieve acceptance rate equal to 30–50%.

4. Model evaluation using AIC, BIC and DIC

In order to compare the RC unrestricted and ordered-restricted models proposed above and evaluate their performance, we facilitate the DIC (Spiegelhalter et al., 2002) and versions of the AIC/BIC (see Brooks, 2002) measures. Hence we consider models

$$m \in \mathscr{M} = \{\mathscr{I}, \mathrm{LL}, \mathrm{R}, \mathrm{C}, \mathrm{RC}, \mathrm{S}, \mathrm{R}_{\mathrm{ord}}, \mathrm{C}_{\mathrm{ord}}, \mathrm{RC}_{\mathrm{ord}}\},$$

with corresponding parameter vector $\boldsymbol{\theta}_m$, where $\mathrm{R}_{\mathrm{ord}}$, $\mathrm{C}_{\mathrm{ord}}$ and $\mathrm{RC}_{\mathrm{ord}}$ stand for the order constrained R, C and RC models, respectively.

DIC has been introduced by Spiegelhalter et al. (2002) as a measure of model adequacy and is given by

$$\mathrm{DIC}(m) = 2\overline{D(\boldsymbol{\theta}_m, m)} - D(\overline{\boldsymbol{\theta}}_m, m) = D(\overline{\boldsymbol{\theta}}_m, m) + 2p_m,$$

where $p_m = \overline{D(\boldsymbol{\theta}_m, m)} - D(\overline{\boldsymbol{\theta}}_m, m)$ can be interpreted as the number of "effective" parameters for model m, $\overline{\boldsymbol{\theta}}_m$ is the posterior mean of the parameters involved in model m, $D(\boldsymbol{\theta}_m, m) = -2\log f(\boldsymbol{y}|\boldsymbol{\theta}_m, m)$ is the deviance measure of a model m and $\overline{D(\boldsymbol{\theta}_m, m)}$ is its posterior mean. Note that assuming multinomial sampling, the above measure becomes

$$D(\boldsymbol{\theta}_m, m) = C - 2\sum_{i=1}^{I}\sum_{j=1}^{J} y_{ij}\log \pi_{ij},$$

where C is a constant common across all models that depends only on $\boldsymbol{y}$. To simplify computations, we use the unnormalized version of the likelihood in the deviance term.

An important problem with DIC is that, in some cases, p_m might take negative values hence inflating the log-likelihood by providing an additional positive bonus (instead of penalizing it) for some particular models. Moreover, in some models with complicated parameter structure the mean of the parameters may not provide a stable estimate of the minimum deviance and hence it may be more efficient to use the posterior means of $\overline{\boldsymbol{\pi}}_m$ for each cell probability (see for an example in Kateri et al., 2005). Hence, in our setting we propose to use

$$\mathrm{DIC}(m) = D(\overline{\boldsymbol{\pi}}_m, m) + 2p_m,$$

with

$$p_m = \overline{D(\boldsymbol{\theta}_m, m)} - D(\overline{\boldsymbol{\pi}}_m, m).$$

Furthermore, we use BIC which can be thought as a rough approximation of the Bayes factor (for details, see Raftery, 1995). In its original form it is defined as

$$\mathrm{BIC}(m) = D(\hat{\boldsymbol{\theta}}_m, m) + d_m \log n,$$

where $\hat{\boldsymbol{\theta}}_m$ denotes the maximum likelihood estimate of model's m parameter vector $\boldsymbol{\theta}_m$ and d_m is its dimension. Alternatively, in the calculation of BIC we may use the posterior mode or the posterior mean $\overline{\boldsymbol{\theta}}_m$ (Raftery, 1996) or even BICs whole posterior distribution (see Brooks, 2002). Following the same arguments as in DIC, we propose to use in the calculation of BIC the estimated posterior means of the multinomial probabilities which will provide a more accurate and stable estimate of the minimum deviance measure, especially when the parametric space is very complicated. Hence, we may estimate the minimum value of BIC,

$$\mathrm{BIC}(m) = D(\overline{\boldsymbol{\pi}}_m, m) + d_m \log n.$$

Similarly, we may also use the corresponding versions of AIC (Akaike, 1973) given by

$$\mathrm{AIC}(m) = D(\overline{\boldsymbol{\pi}}_m, m) + 2\, d_m.$$

When fixed effects models and strict equality constraints are considered, then d_m and p_m (used in DIC) are approximately equal resulting to similar AIC and DIC values.

All the criteria described above have similar structure since they are composed by two components. The first one is the deviance term $D(\boldsymbol{\theta}_m, m)$ (evaluated in the posterior mean or mode) and can be thought as a measure of fit while the second one is a penalty term which accounts for the complexity of the model. A major disadvantage when using AIC and BIC for the comparison between the unrestricted and order-restricted versions of the same model is that the penalty term depends only on the dimension of the corresponding parameter space and hence it will not account for any inequality constraints. Therefore, when comparing a model to an order-restricted version of it, for example models RC and RC_{ord}, then the penalty of both AIC and BIC will be the same although the complexity of the order-restricted model will be lower (see for similar arguments Myung et al., 2005). Hence, these measures will always support the unrestricted model. This is not the case for the DIC which penalizes the deviance term with the number of the "effective" parameters p_m, accounting also for the imposed order-restriction constraints (see for details Spiegelhalter et al., 2002). In the limited case where the posterior distributions of the scoring parameters of RC have a strict hierarchy (which means that $P(\mu_1 < \cdots < \mu_I | \mathbf{y}) = 1$) then the order restrictions do not imply additional information (or constraints) for the fitted model. As a result the fit as well as the number of "effective" estimated parameters will be the same for the RC and RC_{ord} models. Consequently, viewing the problem from the reverse standpoint, equal or close DIC values for the RC and RC_{ord} models indicate that the data themselves provide information for the ordinal structure of the scoring parameters and therefore no additional order constraints are needed. When ordering violations exist then the inequality constraints will force the posterior distributions of the corresponding parameters to be close, indirectly reducing the parameter space by this way. Hence, the difference in the number of "effective" parameters between RC and RC_{ord} provide valuable information for the number of ordering violations. Similar arguments have been used by Myung et al. (2005) in order to implement DIC for the comparison of order-restricted axiomatic statements. However, AIC and BIC can still be used for comparing models of different dimension but with the same ordering structure.

5. Illustration

In this section we implement our proposed analysis on a classical dataset introduced by Maxwell (1961), concerning the severity of dreams' disturbance for boys aged 5 to 15. The data, provided in Table 1, form a 5×4 contingency table cross-tabulating two variables: the age of the child participating in the study (divided in five categories: 5–7, 8–9, 10–11, 12–13, 14–15) and the severity of his dreams disturbance (levels 1–4 from low to high). Among others, these data have been analyzed by Agresti et al. (1987) using the order-restricted column association model and by Ritov and Gilula (1993) applying maximum likelihood estimated correlation models. Both identified negative association between age and severity of disturbances. To be more specific, Agresti et al. (1987) proposed an order-restricted C model under which $\hat{v}_1 < \hat{v}_2 = \hat{v}_3 < \hat{v}_4$. Ritov and Gilula (1993) suggested order restriction on the age classification variable as well. Thus, they concluded to the correlation model with $\hat{v}_1 < \hat{v}_2 = \hat{v}_3 < \hat{v}_4$ and $\hat{\mu}_1 = \hat{\mu}_2 < \hat{\mu}_3 = \hat{\mu}_4 < \hat{\mu}_5$.

To illustrate our proposed methodology we have fitted a large variety of models including LL, R, C and RC and the corresponding order-restricted models. The independence ($\mathscr{I}$) and saturated/full (S) models are also provided for comparative purposes. In order to compare these models we have used the BIC, AIC and DIC, described in

Table 1
Cross-classification of 223 boys by severity of disturbances of dreams and age. Means of the predicted values for the frequencies y_{ij} under the models RC, RC_{ord} and RC_{ord} with $\delta_{\mu_2} = \delta_{\mu_4} = \delta_{v_3} = 0$, respectively, are provided in parentheses

Age group	Disturbance (from low to high)				Total
	1	2	3	4	
5–7	7	4	3	7	21
	(6.8/ 4.3/ 5.2)	(5.0/ 5.1/ 5.0)	(4.3/ 5.4/ 4.9)	(4.9/ 6.2/ 5.9)	
8– 9	10	15	11	13	49
	(10.9/ 12.8/ 12.0)	(14.2/ 11.6/ 11.8)	(10.5/ 11.9/ 11.5)	(13.3/ 12.9/ 13.8)	
10–11	23	9	11	7	50
	(23.3/ 21.3/ 23.4)	(8.8/ 10.0/ 9.4)	(9.4/ 9.6/ 9.2)	(8.4/ 9.0/ 8.0)	
12–13	28	9	12	10	59
	(28.0/ 29.6/ 27.6)	(10.2/ 10.6/ 11.1)	(11.1/ 10.0/ 10.8)	(9.9/ 8.8/ 9.4)	
14–15	32	5	4	3	44
	(30.9/ 32.0/ 31.8)	(3.7/ 4.8/ 4.7)	(5.6/ 4.2/ 4.6)	(3.7/ 3.1/ 2.8)	
Total	100	42	41	40	223

Table 2
Unnormalized versions of model selection criteria for association models

Model	d_m	p_m	DIC	AIC		BIC	
				AIC($\bar{\pi}_m$)	(2.5–97.5%)	BIC($\bar{\pi}_m$)	(2.5–97.5%)
Independence	7	7.0	1288.0	1288.0	1289.6–1303.8	1311.8	1313.5–1327.7
Saturated	19	19.5	1280.5	1279.5	1289.0–1313.0	1344.2	1353.0–1378.0
LL	8	8.2	1272.0	1271.5	1273.8–1289.5	1298.8	1301.0–1316.8
Row	11	11.3	1273.2	1272.5	1276.5–1295.1	1310.0	1314.0–1332.6
Column	10	10.1	1270.8	1270.6	1273.9–1291.5	1304.7	1308.0–1325.6
RC	13	12.8	1270.6	1271.0	1275.8–1295.9	1315.3	1320.1–1340.2
Negative association							
Ordered models							
Row	11	9.7	1271.5	1274.2	1276.8–1294.6	1311.7	1314.2–1332.0
Column	10	8.8	1269.7	1272.0	1274.2–1291.2	1306.1	1308.3–1325.3
RC	13	10.0	1268.1	1274.0	1276.6–1294.3	1318.3	1320.9–1338.6
Row $\delta_{\mu_2}=0$	10	9.4	1270.2	1273.4	1276.1–1293.2	1310.8	1313.5–1330.6
Row $\delta_{\mu_4}=0$	10	9.4	1270.4	1273.5	1276.2–1293.3	1310.9	1313.7–1370.8
Row $\delta_{\mu_2}=\delta_{\mu_4}=0$	9	9.1	1268.8	1272.6	1275.4–1291.7	1310.1	1312.9–1329.2
Column $\delta_{\nu_3}=0$	9	8.7	1268.4	1271.0	1273.5–1289.6	1305.0	1307.6–1323.7
RC $\delta_{\mu_2}=0$	12	9.8	1266.6	1270.9	1273.8–1290.9	1311.8	1314.6–1331.7
RC $\delta_{\nu_3}=0$	12	10.1	1267.9	1271.7	1274.3–1292.6	1312.6	1315.2–1333.5
RC $\delta_{\mu_2}=\delta_{\nu_3}=0$	11	9.8	1265.9	1268.4	1271.4–1288.8	1305.9	1308.9–1326.3
RC $\delta_{\mu_2}=\delta_{\mu_4}=\delta_{\nu_3}=0$	10	9.9	1265.7	1265.8	1269.0–1986.0	1299.9	1303.1–1320.1

Section 4. Detailed results of these measures are provided in Table 2. The association models presented in Table 2 assume negative association ($\phi=-1$). The corresponding positive association models are not listed since their criteria values indicated much worse fit. For example, the RC_{ord} for positive association gave DIC $= 1295.4$, which is even worse than the independence model.

All models considered here are implemented in R and Fortran. The R routine is available at http://stat-athens.aueb.gr/jbn/papers/paper15.htm. In R we have used univariate random walks for all parameters with acceptance rates between 30% and 50% while in Fortran, multivariate random walks with similar acceptance rate. Results presented in this section have been reproduced using 10,000 iterations and additional burn-in period of 1000 iterations with the exception of the RC_{ord} model with $\delta_{\mu_2}=\delta_{\mu_4}=\delta_{\nu_3}=0$, for which the iterations were increased to 20,000 in order to reduce the Monte Carlo error. Convergence has been checked graphically (using ergodic means) and by CODA diagnostics (Best et al., 1995). Furthermore, we observed stability of the posterior summaries when increasing the number of iterations.

Starting from the unrestricted models, we observe for Table 2 that both DIC and AIC support C and RC models with identical values. BIC is more strict and supports the simpler uniform association model. Posterior summaries of row and column scores are given in Table 3 while Boxplots of the rescaled parameters are provided in Fig. 1.

In order to assess the ordinality of the row and column scores of the RC model and to be guided toward a certain order-restricted model (if there is evidence for it), we calculate the posterior probabilities of ordering violations for the scores of the unrestricted RC model. Thus, without loss of generality, let preassume the strict ordering $\mu_1 < \cdots < \mu_I$ and $\nu_1 < \cdots < \nu_J$ for the row and column scores, respectively. In this case, $P(\boldsymbol{\delta}_{\mu_i} < 0|\boldsymbol{y}) = P(\mu_i < \mu_{i-1}|\boldsymbol{y}), i=2,\ldots,I$, and $P(\boldsymbol{\delta}_{\nu_j} < 0|\boldsymbol{y}) = P(\nu_j < \nu_{j-1}|\boldsymbol{y}), j=2,\ldots,J$, are the posterior probabilities of order violations between successive categories. If all these probabilities corresponding to row (column) scores are low, then this is considered as evidence in favour of the assumed ordering for the corresponding classification variable. On the other hand, if all of them are high, then this is also evidence of ordering but of the inverse direction. Otherwise, when only some ordering violations occur with high posterior probabilities, then this is considered as indication of violation of the assumed ordering for the corresponding scores and an equality constraint for these scores in the order restricted model is imposed.

For our example, these probabilities are provided in Table 4 from which we observe high posterior probability (>0.5) for the ordering violation of μ_1, μ_2 and μ_3, μ_4 in model R. For the C model, there is evidence only for one violation (between ν_2 and ν_3) with posterior probability greater than 0.7. When considering the RC model, clear evidence of

Table 3
Posterior summaries for $RC^{(1)}$, $RC^{(1)}_{ord}$ and $RC^{(2)}_{ord}$ with $\delta_{\mu_2} = \delta_{\mu_4} = \delta_{\nu_3} = 0$ model parameters (burn-in = 1000 iterations; total iterations kept: (1) 10,000 iterations, (2) 20,000 iterations)

Parameters	Posterior summaries			Rescaled summaries[a]		Summaries of rescaled parameters[b]			MLE
	Mean	Median	St. dev.	Mean	Median	Mean	Median	St. dev.	
RC									
ϕ	1.00	1.00		−1.87	−1.43	−1.67	−1.66	0.34	−1.56
μ_1	0.18	0.15	0.16	−0.30	−0.28	−0.28	−0.30	0.20	−0.30
μ_2	0.36	0.34	0.19	−0.61	−0.60	−0.58	−0.58	0.13	−0.61
μ_3	−0.05	−0.04	0.11	0.08	0.07	0.08	0.08	0.15	−0.09
μ_4	−0.06	−0.05	0.10	0.10	0.09	0.10	0.10	0.15	0.11
μ_5	−0.43	−0.42	0.21	0.72	0.74	0.68	0.69	0.10	0.72
ν_1	−2.62	−2.09	1.62	−0.84	−0.82	−0.79	−0.81	0.07	−0.84
ν_2	1.00	1.00	0.00	0.32	0.40	0.38	0.38	0.15	−0.30
ν_3	0.26	0.09	0.88	0.08	0.04	0.04	0.04	0.20	0.09
ν_4	1.36	1.02	1.22	0.44	0.40	0.37	0.39	0.18	0.44
RC_{ord}									
ϕ	−1.00	−1.00		−2.37	−1.59	−1.72	−1.70	0.34	−1.61
μ_1	−0.13	−0.12	0.07	−0.55	−0.55	−0.53	−0.53	0.08	−0.47
μ_2	−0.09	−0.08	0.05	−0.38	−0.37	−0.37	−0.38	0.08	−0.47
μ_3	0.00	0.00	0.03	0.02	0.01	0.01	0.02	0.11	0.10
μ_4	0.04	0.04	0.04	0.19	0.17	0.18	0.18	0.11	0.10
μ_5	0.17	0.17	0.09	0.72	0.73	0.71	0.72	0.09	0.73
ν_1	−8.26	−5.92	8.06	−0.83	−0.84	−0.84	−0.84	0.02	−0.83
ν_2	1.00	1.00	0.00	0.10	0.14	0.14	0.14	0.06	0.16
ν_3	2.34	1.54	2.75	0.24	0.22	0.24	0.25	0.07	0.16
ν_4	4.92	3.28	5.71	0.49	0.47	0.45	0.45	0.08	0.51
RC_{ord} with $\delta_{\mu_2} = \delta_{\mu_4} = \delta_{\nu_3} = 0$									
ϕ	−1.00	−1.00	0.00	−3.30	−1.57	−1.67	−1.66	0.34	−1.61
μ_1	−0.11	−0.10	0.06	−0.47	−0.47	−0.46	−0.47	0.05	−0.47
μ_2	−0.11	−0.10	0.06	−0.47	−0.47	−0.46	−0.47	0.05	−0.47
μ_3	0.02	0.02	0.03	0.10	0.08	0.09	0.10	0.09	0.10
μ_4	0.02	0.02	0.03	0.10	0.08	0.09	0.10	0.09	0.10
μ_5	0.17	0.16	0.09	0.73	0.74	0.73	0.74	0.09	0.73
ν_1	−11.04	−5.82	18.09	−0.77	−0.82	−0.81	−0.82	0.04	−0.83
ν_2	1.00	1.00	0.00	0.07	0.14	0.14	0.14	0.08	0.16
ν_3	1.00	1.00	0.00	0.07	0.14	0.14	0.14	0.08	0.16
ν_4	9.04	3.82	18.09	0.63	0.54	0.53	0.54	0.11	0.51

[a]Posterior summaries of row and column scores are rescaled in order to satisfy the sum of squares equal to one constraint.
[b]Posterior summaries of the rescaled sample resulted by transforming the original sample so that (in each iteration) row and column scores satisfy the sum of squares equal to one constraint.

ordering violation is observed between μ_1, μ_2 and ν_2, ν_3 (posterior probability > 0.8). Weaker evidence of ordering violation is provided for μ_3, μ_4 (posterior probability about 0.4). Additionally, the marginal posterior probabilities of the number of positive score differences in the unrestricted models are provided in Table 5. These probabilities suggest for the R model the consideration of four or three distinct ordered scores (i.e. one or two equality constraints) with posterior probabilities 0.5 and 0.33, respectively. Analogously, three distinct ordered scores (i.e. one equality constraint) with posterior probability 0.75 could be suggested for the column scores of the C model. Similar are the conclusions for the RC model based on the corresponding marginal probabilities. Note that also the joint posterior probabilities of scores' ordering violations (not shown) are higher for the model with three row and three column distinct ordered scores (equal to 0.44) followed by the model with four row and three column distinct ordered scores (with posterior probability equal to 0.31). Thus, possible score equalities are $\mu_1 = \mu_2$ ($\delta_{\mu_2} = 0$), $\mu_3 = \mu_4$ ($\delta_{\mu_4} = 0$) and $\nu_2 = \nu_3$ ($\delta_{\nu_3} = 0$). These order violations are in agreement with the maximum likelihood analysis of Agresti et al. (1987) and Ritov and Gilula (1993).

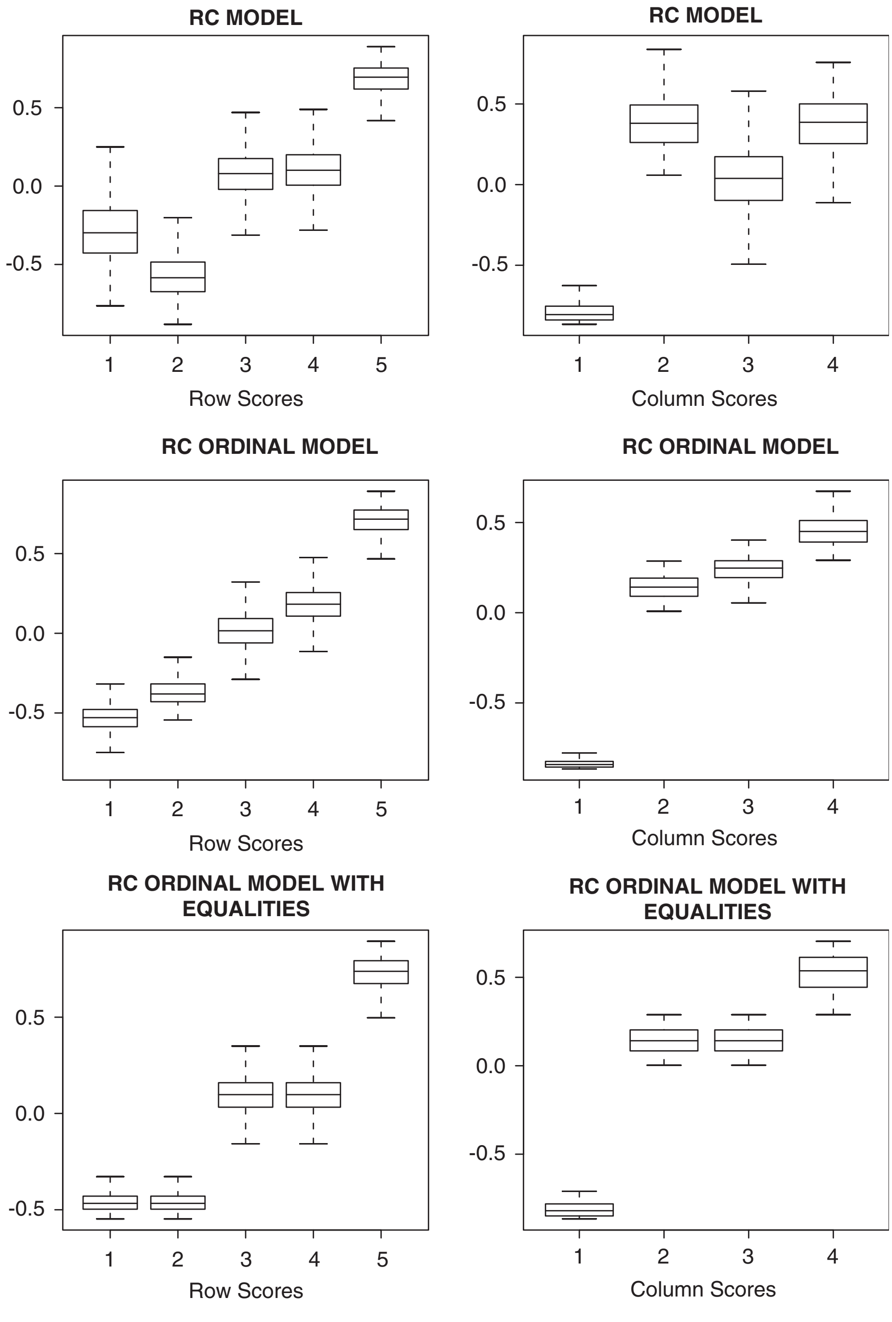

Fig. 1. Boxplots of the rescaled row and column scores for RC, RC_{ord} and RC_{ord} with $\delta_{\mu_2} = \delta_{\mu_4} = \delta_{v_3} = 0$ models.

So, we have proceeded by fitting the strictly ordered restricted versions of R, C and RC, denoted by R_{ord}, C_{ord} and RC_{ord}, respectively, as well as, models using row and column score equalities, as indicated by the above analysis. For the comparison of these models we base our inference mainly on DIC which also penalizes the deviance measure for the additional ordering constraints (see Section 4 for more details). The number of effective parameters of the RC and RC_{ord} models differ by almost three (12.8 vs. 10). This is an indication that three violations exist in the ordering of row and column scores and is in concordance with the posterior probabilities of ordering violation, which indicated two possible violations for row scores and one for the column scores. Similar are the conclusions when we compare the

Table 4
Posterior probabilities of ordering violations (negative score differences δ_{μ_i}, δ_{v_j}) in unrestricted models

Model	Row scores' differences				Column scores' differences		
	$P(\delta_{\mu_2}<0)$	$P(\delta_{\mu_3}<0)$	$P(\delta_{\mu_4}<0)$	$P(\delta_{\mu_5}<0)$	$P(\delta_{v_2}<0)$	$P(\delta_{v_3}<0)$	$P(\delta_{v_4}<0)$
R	0.5829	0.0170	0.5513	0.0025			
C					0.0001	0.7323	0.0974
RC	0.8190	0.0029	0.4427	0.0075	0.0000	0.8874	0.1928

Table 5
Marginal posterior probabilities of the number of positive score differences in the unrestricted models

Model	Distribution of $\{\#\delta_\mu>0\}$					Distribution of $\{\#\delta_v>0\}$			
	4	3	2	1	0	3	2	1	0
R	0.1748	0.4968	0.3282	0.0002	0.0000				
C						0.2063	0.7576	0.0361	0.0000
RC	0.1104	0.5501	0.3390	0.0004	0.0000	0.1309	0.7794	0.0897	0.0000

dimension and the number of "effective" parameters in the rest of the order constrained models. Moreover, DIC values indicate the RC_{ord} model with $\mu_1=\mu_2$ ($\delta_{\mu_2}=0$), $\mu_3=\mu_4$ ($\delta_{\mu_4}=0$) and $v_2=v_3$ ($\delta_{v_3}=0$) as the best fitted model, closed followed by the RC_{ord} model with $\mu_1=\mu_2$ ($\delta_{\mu_2}=0$) and $v_2=v_3$ ($\delta_{v_3}=0$). Note that the best DIC model is also supported by AIC and is the second best according to BIC. This model was indicated by Ritov and Gilula (1993) as well. They proceeded further by collapsing the categories of equal scores and concluded thus to a 3×3 table. In general, whether we adopt the collapsing of the corresponding categories is a matter of policy (Kateri and Iliopoulos, 2003). Posterior summaries of the association parameters for the finally selected model and the general RC_{ord} model are listed in Table 3. The means of the predicted cell frequencies under this model as well as under the RC and RC_{ord} are also provided in Table 1.

Summarizing, we conclude that there exist a negative association between age and severity of the dreams' disturbance. Further on, the first two age categories as well as the next two are indistinguishable in terms of their association to the disturbance severity. Also the two middle categories of the severity scale are indistinguishable with respect to their interaction with age. This means that provided we are restricted to the two first age categories, age and severity of dreams' disturbance seem to be independent (conditionally). The interpretation of the other two equality constraints on the scores is analogue. From the boxplots of the RC_{ord} model as well as the corresponding scores' estimates in Table 3, it is obvious that the distinct row scores are close to be equidistant while the column scores are not, with the first column category (low disturbance) being far apart from the rest of them. From the scores' estimates of the RC_{ord} model (see Table 3) and through (6) we could formulate conclusions of the form: the odds of having dreams' disturbance of middle instead of low severity is 4.78 times higher for boys aged 5–9 than aged 10–13.

6. Discussion

In this work, we have considered a different parametrization for the scores of the standard RC association model, which simplifies the Bayesian estimation procedure. Moreover, we imposed strictly order constraints on the scores. In order to allow scores of successive categories to be equal, we developed a procedure based on the order violation probabilities of the unrestricted models. Of course a more realistic choice would be to incorporate the non-strict ordering of the scores in the choice of priors. The consideration of such ordering restriction is the subject of our current research.

For the comparison of different model formulations, we have used the AIC, BIC and DIC, that can be directly computed via the MCMC output of each model. Within the above framework, an attractive approach involves building a single-chain transdimensional MCMC chain (see Sisson, 2005, for a review) to directly estimate the posterior model probabilities and the corresponding posterior model odds (Iliopoulos et al., 2006). Such an approach might be based directly on a RJMCMC sampler (Green, 1995) or on Gibbs based samplers (see for example Dellaportas et al., 2000).

References

Agresti, A., Chuang, C., 1989. Model-based Bayesian methods for estimating cell proportions in cross-classification tables having ordered categories. Comput. Statist. Data Anal. 7, 245–258.

Agresti, A., Chuang, C., Kezouh, A., 1987. Order-restricted score parameters in association models for contingency tables. J. Amer. Statist. Assoc. 82, 619–623.

Ait-Sidi-Allal, M.L., Baccini, A., Mondot, A.M., 2004. A new algorithm for estimating the parameters and their asymptotic covariance in correlation and association models. Comput. Statist. Data Anal. 45, 389–421.

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Second International Symposium on Information Theory.Academia Kiado, Budapest, pp. 267–281.

Albert, J.H., 1997. Bayesian testing and estimation of association in a two-way contingency table. J. Amer. Statist. Assoc. 92, 685–693.

Becker, M.P., 1990. Maximum likelihood estimation of the RC(M) association model. Appl. Statist. 39, 152–167.

Becker, M.P., Clogg, C.C., 1989. Analysis of sets of two-way contingency tables using association models. J. Amer. Statist. Assoc. 84, 142–151.

Best, N., Cowles, M.K., Vines, K., 1995. CODA: Convergence Diagnostics and Output Analysis Software for Gibbs Sampling Output. MRC Biostatistics Unit, Institute of Public Health, Cambridge, UK.

Brooks, S.P., 2002. Discussion of the paper by Spiegelhalter, Best, Carlin and van der Linde. J. Royal Statist. Soc. B 64, 616–618.

Chuang, C., 1982. Empirical Bayes methods for a two-way multiplicative interaction model. Comm. Statistics Theory Methods 11, 2977–2989.

Dellaportas, P., Forster, J.J., Ntzoufras, I., 2000. Bayesian variable selection using the Gibbs sampler. In: Dey, D.K., Ghosh, S., Mallick, B. (Eds.), Generalized Linear Models: A Bayesian Perspective. Marcel Dekker, New York, pp. 271–286.

Evans, M., Gilula, Z., Guttman, I., 1993. Computational issues in the Bayesian analysis of categorical data: log-linear and Goodman's RC model. Statist. Sin. 3, 391–406.

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

Goodman, L.A., 1981. Association models and canonical correlation in the analysis of cross-classifications having ordered categories. J. Amer. Statist. Assoc. 76, 320–334.

Goodman, L.A., 1985. The analysis of cross-classified data having ordered and/or unordered categories: association models correlation models and asymmetry models for contingency tables with or without missing entries. Ann. Statist. 13, 10–69.

Green, P., 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82, 711–732.

Haberman, S.J., 1995. Computation of maximum likelihood estimates in association models. J. Amer. Statist. Assoc. 90, 1438–1446.

Iliopoulos, G., Kateri, M., Ntzoufras, I., 2006. Bayesian model comparison for the order restricted RC association model, work in progress.

Kateri, M., Iliopoulos, G., 2003. On collapsing categories in two-way contingency tables. Statistics 37, 443–455.

Kateri, M., Nicolaou, A., Ntzoufras, I., 2005. Bayesian inference for the RC(M) association model. J. Comput. Graphical Statist. 14, 116–138.

Maxwell, A.E., 1961. Analyzing Qualitative Data. Methuen, London.

Myung, J.I., Karabatsos, G., Iverson, G.J., 2005. A Bayesian approach to testing decision making axioms. J. Math. Psychology 49, 205–225.

Raftery, A.E., 1995. Bayesian Model Selection in Social Research. In: Marsden, P.V. (Ed.), Sociological Methodology 1995. Blackwell, Oxford.

Raftery, A.E., 1996. Hypothesis testing and model selection. In: Gilks, W.R., Richardson, S., Spiegelhalter, D.J. (Eds.), Markov Chain Monte Carlo in Practice. Chapman & Hall, London, pp. 163–188.

Ritov, Y., Gilula, Z., 1991. The order-restricted RC model for ordered contingency tables: estimation and testing of fit. Ann. Statist. 19, 2090–2101.

Ritov, Y., Gilula, Z., 1993. Analysis of contingency tables by correspondence models subject to order constraints. J. Amer. Statist. Assoc. 88, 1380–1387.

Sisson, S.A., 2005. Transdimensional Markov Chains: a decade of progress and future perspectives. J. Amer. Statist. Assoc. 100, 1077–1089.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A., 2002. Bayesian measures of model complexity and fit (with discussion). J. Roy. Statist. Soc. B 64, 583–639.

Viele, K., Srinivasan, C., 2000. Parsimonious estimation of multiplicative interaction in analysis of variance using Kullback–Leibler Information. J. Statist. Plann. Inference 84, 201–219.