PSYCHOMETRIKA—VOL. 74, NO. 4, 561–587
DECEMBER 2009
DOI: 10.1007/S11336-009-9117-0

BAYESIAN MODEL COMPARISON FOR THE ORDER RESTRICTED RC ASSOCIATION MODEL

G. ILIOPOULOS AND M. KATERI

UNIVERSITY OF PIRAEUS

I. NTZOUFRAS

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

Association models constitute an attractive alternative to the usual log-linear models for modeling the dependence between classification variables. They impose special structure on the underlying association by assigning scores on the levels of each classification variable, which can be fixed or parametric. Under the general row-column (RC) association model, both row and column scores are unknown parameters without any restriction concerning their ordinality. However, when the classification variables are ordinal, order restrictions on the scores arise naturally. Under such restrictions, we adopt an alternative parameterization and draw inferences about the equality of adjacent scores using the Bayesian approach. To achieve that, we have constructed a reversible jump Markov chain Monte Carlo algorithm for moving across models of different dimension and estimate accurately the posterior model probabilities which can be used either for model comparison or for model averaging. The proposed methodology is evaluated through a simulation study and illustrated using actual datasets.

Key words: contingency tables, ordinal variables, reversible jump MCMC algorithm, equality of odds, Bayesian model averaging.

1. Introduction

In the context of two-way contingency tables, the association models, mainly developed by Goodman (cf. Goodman, 1979, 1986) play a predominant role in social sciences and biomedical applications, especially when the classification variables are ordinal. In the framework of these models, scores are assigned to the categories of both classification variables of the table. According to the assumptions made about these scores (i.e., whether they are prespecified, that is, fixed or parameters to be estimated), different possible association models occur. A natural choice for a set of fixed scores is a sequence of increasing scores, usually equidistant for successive categories. This particular choice results to the so-called uniform (U) association model which is the simplest association model having just one parameter more than the independence model. When the scores of just the column (row) classification variable are fixed, the derived model is the row (R) effect (column effect, C) association model. Finally, when both sets of scores are parametric, then the Row-Column (RC) association model arises. The RC model is the most computationally involved in terms of its estimation and fit, due to the fact that its systematic component is multiplicative in its parameters while the corresponding components in U, R, and C models are additive in their parameters.

In case of parametric scores, their monotonicity is not ensured by the standard estimation procedures. Since monotonicity of the scores is related to stochastic ordering of the corresponding classification variable (Goodman, 1981), it is usually natural to expect the scores for ordinal classification variables to be monotonic. Estimation procedures subject to order constraints for

Requests for reprints should be sent to M. Kateri, Department of Statistics and Insurance Science, University of Piraeus, Piraeus, Greece. E-mail: mkateri@unipi.gr

the parametric scores have been developed initially for the R (or C) model by Agresti, Chuang, and Kezouh (1987) and for the RC model by Ritov and Gilula (1991). Recently, alternative algorithms have been proposed and compared by Galindo-Garre and Vermunt (2004). Order restrictions yield also an extended RC model, introduced by Bartolucci and Forcina (2002).

Ordinary or order-restricted inferences for these models rely on large-sample asymptotic methods. As it is stated in Galindo-Garre and Vermunt (2004), these methods do not work well for sparse tables or small sample sizes, common in social and biomedical applications, where the usual asymptotic chi-squared p-values are known to be inaccurate. In contrast, in the Bayesian approach we draw inferences based on the posterior model probabilities which have a direct probability interpretation. These probabilities as well as the exact posterior distributions of the parameters can be estimated using Markov Chain Monte Carlo (MCMC) methods while the sampling error due to simulation (i.e., Monte Carlo error) is easily eliminated by running the algorithm for a sufficiently large number of iterations. Thus, the Bayesian approach, which has been also suggested by Galindo-Garre and Vermunt (2004) as "an interesting approach to find solutions", is a promising and attractive alternative in spite its computational complexity.

Furthermore, the model selection procedure is benefited in the Bayesian framework. MCMC methods (and the proposed algorithm in this article) enable an efficient search of the model space even if it is large. Such methods are automatic in the sense that no manual forward or backward moves are needed. For the models visited, the algorithm provides the posterior model probabilities, a powerful tool for models' evaluation and estimation of their uncertainty. Model uncertainty might be high in small samples or in existence of more than one models of similar performance. High model uncertainty can be incorporated in the Bayesian statistical inference. These issues are also discussed in Section 3.2.

In the Bayesian framework, the estimation of the RC association model has been considered by Chuang (1982) and Evans, Gilula, and Guttman (1993). The Bayesian analysis of the RC model can also be achieved through the procedure proposed in Kateri, Nicolaou, and Ntzoufras (2005), who dealt with the Bayesian estimation of the more general RC(K) association model, since the RC model is the RC(K) with $K = 1$. Recently, Tarantola, Consonni, and Dellaportas (2008) used methodology adopted from product partition models to make inferences about the clustering of scores in the row effect model. None of these procedures considered order constraints for the parametric scores. A first attempt for full Bayesian inference concerning the order constrained association models has been provided by Iliopoulos, Kateri, and Ntzoufras (2007). Their approach for identifying possible score equalities was based on calculating the posterior probabilities of possible order violations for successive categories in the unrestricted model. These probabilities were used in an isotonic regression type logic, indicating which scores should be merged. Furthermore, the deviance information criterion (DIC, Spiegelhalter, Best, Carlin, & van der Linde, 2002) was applied to identify the most appropriate model in terms of goodness of fit. Nevertheless, this approach can not be considered as a formal Bayesian evaluation in favor or against merging specific scores, since it is not based on the posterior model odds and probabilities (see Kass & Raftery, 1995, for details).

In the present work, we focus on the estimation of posterior model probabilities of the RC order constrained model, in a formal way, by allowing for ties in the prior distribution level. A trans-dimensional MCMC algorithm (reversible jump MCMC, Green, 1995) is constructed for assessing the equality of successive row and column scores.

The RC model is presented in Section 2 under an alternative parametrization convenient for the needs of our procedure. This alternative parameterization also exhibits interpretational advantages, which are highlighted. Section 3 deals with the Bayesian model formulation including the description of additional latent indicators used in the proposed algorithm to identify equal scores. A reversible jump MCMC algorithm for the estimation of the order restricted row or/and column scores, subject also to possible ties, is introduced in Section 4. A simulation study along with illustrative examples are presented in Section 5, and the final section summarizes results and discusses related issues.

2. Modeling Details

Let $\boldsymbol{y} = (y_{ij})$, $i = 1, \dots, I$ and $j = 1, \dots, J$, be an $I \times J$ cross-classification of two *ordinal* variables X (rows) and Y (columns). We assume that

$$\mathrm{vec}(\boldsymbol{y}) \sim \mathcal{M}ultinomial\left(\sum_{i=1}^{I} \sum_{j=1}^{J} y_{ij}, \mathrm{vec}(\boldsymbol{\Pi}) \right),$$

where $\boldsymbol{\Pi} = (\pi_{ij})$ is the underlying probability table. Under the saturated log-linear model,

$$\log \pi_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}, \quad i = 1, \dots, I, \; j = 1, \dots, J.$$

For identifiability purposes, the sum-to-zero constraints are imposed on the parameters, i.e.,

$$\sum_{i=1}^{I} \lambda_i^X = \sum_{j=1}^{J} \lambda_j^Y = \sum_{i=1}^{I} \lambda_{ij}^{XY} = \sum_{j=1}^{J} \lambda_{ij}^{XY} = 0. \tag{2.1}$$

Assigning parametric scores to the categories of X and Y, denoted by $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_I)$ and $\boldsymbol{\nu} = (\nu_1, \nu_2, \dots, \nu_J)$, respectively, and substituting the interaction terms λ_{ij}^{XY} by the product $\phi \mu_i \nu_j$, the multiplicative row-column association (RC) model (Goodman, 1979) is achieved:

$$\log \pi_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + \phi \mu_i \nu_j, \quad i = 1, \dots, I, \; j = 1, \dots, J.$$

The parameter ϕ is a global measure of association. The above formulation reveals the analogies to the classical correspondence analysis (CA) or canonical correlation model (Goodman, 1986). The physical interpretation of the parameter ϕ is straightforward. It reflects the odds ratio of successive categories with score distances equal to one, since

$$\log \left(\frac{\pi_{ij} \pi_{i+1,j+1}}{\pi_{i,j+1} \pi_{i+1,j}} \right) = \phi (\mu_{i+1} - \mu_i)(\nu_{j+1} - \nu_j).$$

Certain constraints have to be imposed on the above model to ensure its identifiability. Additional to the natural constraint $\sum_{i=1}^{I} \sum_{j=1}^{J} \pi_{ij} = 1$ and the sum-to-zero (STZ) constraints on the row and column main effects in (2.1), four additional constraints are imposed on the row and column scores (μ_i and ν_j). Usually, these restrictions are the STZ and the sum of squares equal to one (SSTO) weighted constraints

$$\sum_{i=1}^{I} w_{1i} \mu_i = \sum_{j=1}^{J} w_{2j} \nu_j = 0 \quad \text{and} \quad \sum_{i=1}^{I} w_{1i} \mu_i^2 = \sum_{j=1}^{J} w_{2j} \nu_j^2 = 1, \tag{2.2}$$

fixing the location and scale of the scores, respectively. The usual choices for the weights are either the uniform ($w_{1i} = w_{2j} = 1$, $i = 1, \dots, I$, $j = 1, \dots, J$) or the marginal ($w_{1i} = \pi_{i\cdot}$, $i = 1, \dots, I$; $w_{2j} = \pi_{\cdot j}$, $j = 1, \dots, J$) ones. Within the context of Bayesian analysis of association models, the SSTO constraints complicate the structure of the posterior distribution; see Kateri et al. (2005) for more details.

When both classification variables, X and Y, are ordinal, the focus lies on the order restricted version of the RC association model. Thus, an ordinal structure for the corresponding scores is assumed:

$$\mu_1 \leqslant \mu_2 \leqslant \cdots \leqslant \mu_I \quad \text{and} \quad \nu_1 \leqslant \nu_2 \leqslant \cdots \leqslant \nu_J$$

with $\mu_1 < \mu_I$ and $\nu_1 < \nu_J$ (i.e., at least two distinct and unequal row and column scores are assumed). The aim of the current work is the Bayesian estimation of the order restricted RC model with simultaneous identification of possible score equalities over successive scores $\mu_i,\ \mu_{i+1}$ and $\nu_j,\ \nu_{j+1}$.

For the needs of our procedure, we introduce an alternative parameterization of the scores (subject to a different set of constraints). This parameterization is suitable for building a reversible jump MCMC (RJMCMC) algorithm for assessing the equality of successive row and column scores and simplifies further the MCMC scheme as well.

2.1. *The Alternative Parameterization*

Under the standard SSTO constraints, the posterior for the row scores turns out to be an I-dimensional distribution constrained on the I-dimensional unit hypersphere for uniform weights or on an I-dimensional ellipsoid for any other weights. Analogously, the posterior distribution of the column scores will be constrained on the J-dimensional unit hypersphere or, more generally, on a J-dimensional ellipsoid. Sampling on this constrained parametric space becomes cumbersome since the redundant score-parameters are expressed as functions of the remaining scores. Kateri et al. (2005) proposed a complicated algorithm to sample from the parameters of the RC(K) model under this parametrization. Note that here the posterior is further complicated due to the additional ordering constraints. We propose, instead of using the standard constraints (2.2), to fix the minimum and maximum score for each classification variable. This way, the redundant score-parameters are set to be fixed and so, derivation of posterior distribution for them is avoided. Thus, we set

$$\begin{aligned} &\mu_1 = \mu_{\min}, \quad \mu_I = \mu_{\max}, \\ &\nu_1 = \nu_{\min} \quad \text{and} \quad \nu_J = \nu_{\max}, \end{aligned} \tag{2.3}$$

with $\mu_{\min} < \mu_{\max}$ and $\nu_{\min} < \nu_{\max}$, for prespecified constants $\mu_{\min}$, $\mu_{\max}$, $\nu_{\min}$, and $\nu_{\max}$. This approach results to row and column scores taking values in the intervals $[\mu_{\min}, \mu_{\max}]$ and $[\nu_{\min}, \nu_{\max}]$, respectively. Typical choices for these intervals could be $[-1, 1]$ or $[0, 1]$.

For the order restricted RC model, due to the ordering of the scores, the constraints (2.3) imposed on the extreme categories arise in a more natural way than the standard ones (2.2). For instance, the association parameter ϕ has a straightforward interpretation for scores in the zero-one interval, since

$$\phi = \log\left(\frac{\pi_{11}\pi_{IJ}}{\pi_{I1}\pi_{1J}}\right),$$

i.e., it is equal to the log-odds ratio that results from the extreme categories of the contingency table. Throughout this paper, we shall adopt the (2.3) constraints in the zero-one interval.

Usually, the main focus of the RC model lies on modeling the association and providing sensitive interpretations regarding score distances among categories of the classification variables. In this context, the main effect terms have no direct interpretation and are treated as nuisance parameters. However, they become important in the IRT interpretation of the RC model (see Anderson & Yu, 2007). Its role is also crucial in the distance perspective point of view of the association model (De Rooij, 2007). Note that, under our setup, the main effects have a straightforward interpretation as well. More specifically, it holds

$$\log\left\{\frac{P(X=i_1|Y=1)}{P(X=i_2|Y=1)}\right\}=\lambda_{i_1}^X-\lambda_{i_2}^X \quad \text{and} \quad \log\left\{\frac{P(Y=j_1|X=1)}{P(Y=j_2|X=1)}\right\}=\lambda_{j_1}^Y-\lambda_{j_2}^Y,$$

i.e., their differences are equal to the logarithms of the corresponding probability ratios, conditional on being in the first column or row categories, respectively.

Typically, the scores of the RC model are estimated so that ϕ is positive due to the fact that RC is a special case of RC(K) model where the parameters ϕ_k's, $k=1,\ldots,K$, are actually the singular values of a particular matrix (and so they satisfy $\phi_1\geqslant\cdots\geqslant\phi_K\geqslant 0$, cf. Goodman, 1986). However, it is clear that the positivity constraint is not necessary for the ϕ of the RC model. Furthermore, this constraint must be completely removed in our set-up, since under the proposed parametrization both row and column scores, are forced to be of increasing order. Thus, the sign of ϕ implies the direction of the underlying association between the two classification variables.

Speaking of the general RC(K) model, we did not consider order restricted inference for the case $K\geqslant 2$ for technical and interpretation reasons. Under the RC(K) model, the row scores are the I-dimensional eigenvectors $\boldsymbol{\mu}_k=(\mu_{1k},\ldots,\mu_{Ik})$, $k=1,\ldots,K$, of a particular matrix and so, they are mutually orthogonal to each other. The column scores are defined analogously. Technically, it is not possible for two (or more) mutually orthogonal vectors to stay in the isotonic region. According to our knowledge, no order restricted maximum likelihood estimation has been developed for $K\geqslant 2$, and furthermore, we can not think of any meaningful physical interpretation of order restricted scores in models of higher dimensional association.

2.2. *Comparison with the Standard Parametrization*

The RC model under the alternative proposed parameterization is equivalent to the RC under the standard constraints, leading to the same expected cell frequencies. The proposed parametrization affects only the values of the parameters and consequently the reference basis for their physical interpretation. There is a one-to-one correspondence between the two sets of parameters. For instance, if we denote the parameters under the STZ and SSTO constraints with uniform weights by $\tilde{\lambda}$, $\tilde{\lambda}_i^X$, $\tilde{\lambda}_j^Y$, $\tilde{\phi}$, $\tilde{\mu}_i$, $\tilde{\nu}_j$, then they can be obtained from the parameters of our parametrization through the expressions

$$\tilde{\lambda}=\lambda+\phi\bar{\mu}\bar{\nu}, \qquad \tilde{\lambda}_i^X=\lambda_i^X+\phi\bar{\nu}(\mu_i-\bar{\mu}), \qquad \tilde{\lambda}_j^Y=\lambda_j^Y+\phi\bar{\mu}(\nu_j-\bar{\nu}), \tag{2.4}$$

$$\tilde{\phi}=\phi s_\mu s_\nu, \qquad \tilde{\mu}_i=(\mu_i-\bar{\mu})/s_\mu, \qquad \tilde{\nu}_j=(\nu_j-\bar{\nu})/s_\nu, \tag{2.5}$$

where $\bar{\mu}=\sum_{i=1}^I\mu_i/I$, $s_\mu^2=\sum_{i=1}^I(\mu_i-\bar{\mu})^2$, $\bar{\nu}=\sum_{j=1}^J\nu_j/J$, $s_\nu^2=\sum_{j=1}^J(\nu_J-\bar{\nu})^2$. On the other hand, our parameters with the scores in the interval [0, 1] are given in terms of the original ones (with uniform weights constraints) as

$$\lambda=\tilde{\lambda}-\tilde{\phi}\tilde{\mu}_1\tilde{\nu}_1, \qquad \lambda_i^X=\tilde{\lambda}_i^X+\tilde{\phi}\tilde{\nu}_1\tilde{\mu}_i, \qquad \lambda_j^Y=\tilde{\lambda}_j^Y+\tilde{\phi}\tilde{\mu}_1\tilde{\nu}_j,$$

$$\phi=\tilde{\phi}R_{\tilde{\mu}}R_{\tilde{\nu}}, \qquad \mu_i=(\tilde{\mu}_i-\tilde{\mu}_1)/R_{\tilde{\mu}}, \qquad \nu_j=(\tilde{\nu}_j-\tilde{\nu}_1)/R_{\tilde{\nu}},$$

where $R_{\tilde{\mu}}=\tilde{\mu}_I-\tilde{\mu}_1$, $R_{\tilde{\nu}}=\tilde{\nu}_J-\tilde{\nu}_1$. Similar (although more cumbersome) expressions hold for any choice of the weights and our parametrization's scores intervals as well.

3. Bayesian Model Formulation

To accommodate uncertainty concerning the equality of successive scores, the model is further extended by introducing the latent binary indicators γ_i (for $i = 2, \ldots, I$) and δ_j (for $j = 2, \ldots, J$), which are equal to 1 when $\mu_i - \mu_{i-1} > 0$ (or $\nu_j - \nu_{j-1} > 0$ respectively) and 0 when $\mu_i - \mu_{i-1} = 0$ (or $\nu_j - \nu_{j-1} = 0$, respectively). The indicator vectors $\boldsymbol{\gamma} = (1, \gamma_2, \ldots, \gamma_I)$ and $\boldsymbol{\delta} = (1, \delta_2, \ldots, \delta_J)$ specify which successive scores are equal and can be used as a model indicator. Without loss of generality, we set $\gamma_1 = \delta_1 = 1$ in order to retain the same dimension between the indicator and the corresponding score vectors. These indicators are analogous to the binary indicators used in Bayesian variable selection techniques (see George & McCulloch, 1993; Kuo & Mallick, 1998, and Dellaportas, Forster, & Ntzoufras, 2002).

Let us define $\boldsymbol{\mu}_{\boldsymbol{\gamma}}$ as the vector of the *distinct* row scores $\boldsymbol{\mu}$, which is indexed by $\Gamma_i = \sum_{k=1}^{i} \gamma_k$, for $i = 1, \ldots, I$. Similarly, we define $\boldsymbol{\nu}_{\boldsymbol{\delta}}$ as the vector of the distinct row scores $\boldsymbol{\nu}$, which is indexed by $\Delta_j = \sum_{k=1}^{j} \delta_k$, $j = 1, \ldots, J$. Hence, Γ_i stands for the number of distinct scores up to the ith row of the table while the interpretation of Δ_j is analogous. Thus, the total number of distinct $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ scores (and hence the lengths of $\boldsymbol{\mu}_{\boldsymbol{\gamma}}$ and $\boldsymbol{\nu}_{\boldsymbol{\delta}}$) are equal to Γ_I and Δ_J, respectively. For notational convenience, the ith element of the $\boldsymbol{\mu}_{\boldsymbol{\gamma}}$ vector, μ_{γ_i}, will be denoted by $\boldsymbol{\mu}_{\boldsymbol{\gamma}}(i)$. Similarly, $\boldsymbol{\nu}_{\boldsymbol{\delta}}(j)$ stands for ν_{δ_j}, the jth element of the $\boldsymbol{\nu}_{\boldsymbol{\delta}}$ vector. Notice that there is a one-to-one correspondence between $\boldsymbol{\mu}$ and $(\boldsymbol{\mu}_{\boldsymbol{\gamma}}, \boldsymbol{\gamma})$ and between $\boldsymbol{\nu}$ and $(\boldsymbol{\nu}_{\boldsymbol{\delta}}, \boldsymbol{\delta})$ given by the expressions

$$\mu_i = \boldsymbol{\mu}_{\boldsymbol{\gamma}}(\Gamma_i) \quad \text{and} \quad \nu_j = \boldsymbol{\nu}_{\boldsymbol{\delta}}(\Delta_j), \tag{3.1}$$

respectively. Note that, it always holds $\boldsymbol{\mu}_{\boldsymbol{\gamma}}(1) = \mu_{\min}$, $\boldsymbol{\mu}_{\boldsymbol{\gamma}}(\Gamma_I) = \mu_{\max}$, $\boldsymbol{\nu}_{\boldsymbol{\delta}}(1) = \nu_{\min}$, $\boldsymbol{\nu}_{\boldsymbol{\delta}}(\Delta_J) = \nu_{\max}$. Thus, there exist $\Gamma_I - 2$ and $\Delta_J - 2$ free row and column score parameters, respectively.

3.1. *Prior Distributions*

We assume that the (distinct) scores are a priori distributed as ordered iid uniform random variables with joint densities

$$f(\boldsymbol{\mu}_{\boldsymbol{\gamma}}|\boldsymbol{\gamma}) = \frac{(\Gamma_I - 2)!}{(\mu_{\max} - \mu_{\min})^{\Gamma_I - 2}} \mathcal{I}\big(\mu_{\min} < \text{ordered distinct } \mu\text{'s} < \mu_{\max}\big),$$

$$f(\boldsymbol{\nu}_{\boldsymbol{\delta}}|\boldsymbol{\delta}) = \frac{(\Delta_J - 2)!}{(\nu_{\max} - \nu_{\min})^{\Delta_J - 2}} \mathcal{I}\big(\nu_{\min} < \text{ordered distinct } \nu\text{'s} < \nu_{\max}\big),$$

respectively. Normal prior distributions with vague variances are assigned to the row and column main effects. A normal prior distribution with large prior variance is also considered for the association parameter ϕ. The effect of this choice on the posterior model probabilities will be minimal due to the presence of these parameters in all compared models (see, for details, Kass & Raftery, 1995, Section 5.3). To complete the prior specification, Bernoulli priors with success probabilities equal to 1/2 are assigned to the γ_i's and δ_j's.

Remark 3.1. Alternatively, instead of imposing an ordered uniform prior on the distinct row and column scores, one could consider a prior on the differences of successive scores $D_{\mu_i} = \mu_i - \mu_{i-1}$, for $i = 2, \ldots, I$, with $D_{\mu_1} = \mu_1 = \mu_{\min}$, and $D_{\nu_j} = \nu_j - \nu_{j-1}$, for $j = 2, \ldots, J$, with $D_{\nu_1} = \nu_1 = \nu_{\min}$. Thus, $\sum_{i=2}^{I} \gamma_i D_{\mu_i} = \mu_{\max} - \mu_{\min}$ and in particular, $\sum_{i=2}^{I} \gamma_i D_{\mu_i} = 1$ for range of scores equal to one. In the latter case, when no prior information is available, a convenient choice for the prior on the row scores differences is the Dirichlet distribution $\mathcal{D}(\underbrace{1, \ldots, 1}_{\Gamma_I - 1})$. Similarly, a noninformative prior for the column scores differences is $\mathcal{D}(\underbrace{1, \ldots, 1}_{\Delta_J - 1})$.

It can be shown that the two prior considerations (the ordered uniform on the scores and the Dirichlet with unit parameters) are equivalent. In what follows, we work with the first representation. The reasons we have also presented the second one is clarity as well as the fact that it is used in the `R` function we constructed to implement the method.

3.2. Posterior Inference and Model Comparison

Since interest lies on the interaction terms of our model, focus is placed on the posterior distributions of the row and column scores as well as of the intrinsic association parameter ϕ. Moreover, since the latent indicators $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ specify the interaction and the model structure, the posterior distribution of $f(\boldsymbol{\gamma}, \boldsymbol{\delta}|\boldsymbol{y})$ and the marginal distributions of the elements of $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$, $f(\gamma_i|\boldsymbol{y})$, $i = 1, \ldots, I$, and $f(\delta_j|\boldsymbol{y})$, $j = 1, \ldots, J$, respectively, are also of interest. These posterior probabilities identify the best models and indicate ties among the scores. The posterior distribution of the parameters conditional on a specific model structure will be also considered. Hence, we will focus on $f(\phi|\boldsymbol{\gamma}, \boldsymbol{\delta})$, $f(\boldsymbol{\mu}|\boldsymbol{\gamma}, \boldsymbol{\delta})$ and $f(\boldsymbol{\nu}|\boldsymbol{\gamma}, \boldsymbol{\delta})$ as well, for those $(\boldsymbol{\gamma}, \boldsymbol{\delta})$'s having the highest posterior probabilities. Further on, we shall realize that an important issue in model comparison is not just the selection of a single "best model" but also the estimation of the degree of uncertainty for the selected model as well as the detection of models with close predictive behavior. In the classical approach, the second issue has not been sufficiently addressed. The confrontation through the use of p-values provides a partial (questionable by many) answer. They are usually based on asymptotic distributions which demand adequate sample sizes. Moreover, difficulties can appear for large samples as well, where the p-values tend to reject the null hypothesis even if it is in fact true. For a detailed related discussion, we refer to Raftery (1995). On the other hand, the Bayesian approach facilitates posterior probabilities of the models under consideration which have direct interpretation and constitute an important tool for model judgment. Under this approach, a single "best model" is not selected whenever model uncertainty is high (i.e., a large number of models have close posterior probabilities). In such cases, the uncertainty is incorporated in the inference through Bayesian model averaging techniques (cf. Hoeting, Madigan, Raftery, & Volinski, 1999).

The aforementioned quantities will be estimated using reversible jump Markov chain Monte Carlo (RJMCMC) techniques introduced by Green (1995). This allows us to explore the model space and simultaneously appraise both the posterior model probabilities and the posterior distribution of model parameters. Accurate estimates can be obtained only for the parameters of high posterior probability models since a sufficient number of iterations of the algorithm will be available for them. If we wish to estimate parameters with increased precision (i.e., reduce the Monte Carlo error), we can run the algorithm for more iterations or run an MCMC algorithm for the single model under consideration. No further actions are needed since the algorithm provides samples (and hence results) directly from the correct posterior distribution. Note that from the output of the RJMCMC algorithm we can also obtain a posterior sample for the parameters satisfying the original STZ/SSTO parametrization (2.2). This can be achieved by simply implementing (2.4) and (2.5) on the set of values obtained in each iteration of the imposed MCMC algorithm.

Details concerning the implementation of the proposed RJMCMC algorithm follow in the next section.

4. RJMCMC for Row and Column Score Merging

Reversible jump MCMC (RJMCMC, Green, 1995) is a generalization of the usual MCMC algorithms. While the latter are used to estimate the posterior distribution of a given model with a parameter vector of fixed dimension, RJMCMC can effectively move between models of different dimensions. For this reason, it can be used to estimate both the posterior distributions of the parameters within each model and the posterior probabilities of the models under consideration. Furthermore, when the number of competing models is large, it can be effectively applied as a model search algorithm, i.e., an algorithm which identifies the models with high posterior probabilities. In terms of implementation, when proposing to jump to models of higher parametric dimension, additional proposal parameters need to be generated so that the two compared models will be of the same dimensionality.

In this section, we proceed by describing the RJMCMC algorithm adapted for our problem. Details concerning the matching functions and the proposal distributions are provided. In what follows, notation is similar to the one used in variable selection (see Dellaportas et al., 2002).

4.1. *The Algorithm*

The RJMCMC scheme can be summarized by the following steps:

1. For $i = 2, \dots, I$, update $(\boldsymbol{\mu}_{\boldsymbol{\gamma}}, \boldsymbol{\gamma})$ using the following steps:
    - (a) Propose $\boldsymbol{\gamma}'$ such that $\gamma'_i = 1 - \gamma_i$ and $\gamma'_k = \gamma_k$ for $k \neq i$.
    - (b_1) If $\gamma_i = 0 \rightarrow \gamma'_i = 1$ then *split scores* $\mu_{i-1} = \mu_i$ to $\mu'_{i-1} < \mu'_i$.
        - i. Generate a scalar u from a specified proposal density $q(u|\boldsymbol{\mu}, \boldsymbol{\gamma}, \boldsymbol{\gamma}')$ which is used to equalize the dimensions of the compared models (see Green, 1995, for details).
        - ii. Set $\boldsymbol{\mu}'_{\boldsymbol{\gamma}'} = g(\boldsymbol{\mu}_{\boldsymbol{\gamma}}, u)$, where $g(\boldsymbol{\mu}_{\boldsymbol{\gamma}}, u)$ can be any invertible function that matches the two models.
    - (b_2) If $\gamma_i = 1 \rightarrow \gamma'_i = 0$ then *merge scores* $\mu_{i-1} < \mu_i$ to $\mu'_{i-1} = \mu'_i$.
        - i–ii. Calculate simultaneously u and the proposed parameter vector $\boldsymbol{\mu}'_{\boldsymbol{\gamma}'}$ using the expression $(\boldsymbol{\mu}'_{\boldsymbol{\gamma}'}, u) = g^{-1}(\boldsymbol{\mu}_{\boldsymbol{\gamma}})$.
    - (c) Obtain $\boldsymbol{\mu}'$ from $\boldsymbol{\mu}'_{\boldsymbol{\gamma}'}$ using (3.1).
    - (d) Accept the proposed move $(\boldsymbol{\gamma}, \boldsymbol{\mu}) \rightarrow (\boldsymbol{\gamma}', \boldsymbol{\mu}')$ with probability $\alpha = \min(1, A)$, where

$$A = \frac{f(y|\lambda^X, \lambda^Y, \phi, \boldsymbol{\mu}', \boldsymbol{\nu})}{f(y|\lambda^X, \lambda^Y, \phi, \boldsymbol{\mu}, \boldsymbol{\nu})} \frac{f(\boldsymbol{\mu}'_{\boldsymbol{\gamma}'}|\boldsymbol{\gamma}')f(\boldsymbol{\gamma}')}{f(\boldsymbol{\mu}_{\boldsymbol{\gamma}}|\boldsymbol{\gamma})f(\boldsymbol{\gamma})} \frac{q(u|\boldsymbol{\mu}'_{\boldsymbol{\gamma}'}, \boldsymbol{\gamma}', \boldsymbol{\gamma})^{\gamma_i}}{q(u|\boldsymbol{\mu}_{\boldsymbol{\gamma}}, \boldsymbol{\gamma}, \boldsymbol{\gamma}')^{1-\gamma_i}} |\boldsymbol{J}|^{1-2\gamma_i}.$$

      Here, $|\boldsymbol{J}|$ is the absolute value of the determinant of the Jacobian matrix for the matching function $g(\boldsymbol{\mu}_{\boldsymbol{\gamma}}, u)$, used in the split scores move, given by

$$\boldsymbol{J} = \left| \frac{\partial g(\boldsymbol{\mu}_{\boldsymbol{\gamma}}, u)}{\partial(\boldsymbol{\mu}_{\boldsymbol{\gamma}}, u)} \right|.$$

2. For $j = 2, \dots, J$, update $(\boldsymbol{\nu}_{\boldsymbol{\delta}}, \boldsymbol{\delta})$ using the same procedure as in step 1 above.
3. Given the model structure $(\boldsymbol{\gamma}, \boldsymbol{\delta})$, update the model parameters $(\lambda^X, \lambda^Y, \phi, \boldsymbol{\mu}, \boldsymbol{\nu})$ using the following steps:
    - Sample row and column effects.
    - Sample ϕ using a simple random walk Metropolis.
    - Use random walk Metropolis for the logits of column and row scores' differences.

For updating the row and column scores, we have used a random walk Metropolis on the logits of their differences in order to ensure that the ordering constraints will be satisfied for all sampled values. A simpler alternative, as pointed out by an anonymous referee, would be rejection sampling, i.e., to sample repeatedly from the unconstrained posterior distribution and accept the first set of values that satisfy the inequality constraints. However, this method may be quite inefficient especially when violations of the ordering constraints are implied by the unconstrained posterior distribution. In such cases, the corresponding algorithm will "stick" since the probability of sampling acceptable values will be extremely low.

More details concerning the matching and proposal functions follow. Technicalities concerning the Jacobians as well as the domain of the dimensions equalizing parameter u are given in Appendices A and B, respectively. A more detailed description of the split and merge moves is provided in Appendix C. Finally, additional details concerning the updating of the parameters given the model structure (step 3) are available in Iliopoulos et al. (2007).

Variations of the corresponding algorithm can be used for the simpler R and C models. In particular, one needs just to skip the `for` loop of the proposed algorithm that corresponds to the vector of known scores.

4.2. *Specification of the Matching Function*

A common function used to specify a merge move can be obtained by simply considering the mean (arithmetic or geometric) of the successive scores that we wish to merge and leave the rest of the parameters unchanged (see, for example, in Richardson and Green, 1997). Direct implementation of the above matching function on the first and last category scores leads to violation of the model constraints (2.3). Thus, when merging the two lowest or the two highest scores, we use a rescaled version of the above considered matching function, so that the required constraints are satisfied. The same comments are also true for the inverse move (split of two scores) since they are directly defined by the inverse of the function for the merge moves. Hence, the split and merge moves are directly defined by a three-armed function g, each arm corresponding to a different type of move.

In the following, we provide details concerning the split and merge moves only for the row scores, since the corresponding moves for the column scores are derived analogously.

4.2.1. Split and Merge Moves for the "Central" Scores The definition of central scores depends on the type of the proposed move. When splitting is proposed, we define as central scores the second to the last but one components of $\boldsymbol{\mu}_\gamma$. On the other hand, when merging, we define as central scores the third to the last but one components of $\boldsymbol{\mu}_\gamma$. Hence, central scores are the $\boldsymbol{\mu}_\gamma(\ell)$'s with $s \leqslant \ell = \Gamma_i \le \Gamma_I - 1$, where $s = 3$ for merging moves and $s = 2$ for splitting moves. Concerning the vector of the original row scores μ_i, we define as central scores the ones with $s \leqslant \Gamma_i \leqslant \Gamma_I - 1$.

Merging of $\boldsymbol{\mu}_\gamma(\ell - 1) < \boldsymbol{\mu}_\gamma(\ell)$ leads to a new parameter vector $\boldsymbol{\mu}'_{\gamma'}$ of dimension $\Gamma'_I = \Gamma_I - 1$ through the following transformation:

$$\boldsymbol{\mu}'_{\gamma'}(k) = \begin{cases} \boldsymbol{\mu}_\gamma(k), & k < \ell - 1, \\ \frac{1}{2}\{\boldsymbol{\mu}_\gamma(\ell - 1) + \boldsymbol{\mu}_\gamma(\ell)\}, & k = \ell - 1, \\ \boldsymbol{\mu}_\gamma(k+1), & k > \ell - 1, \end{cases} \tag{4.1}$$

for $k = 1, \dots, \Gamma_I - 1$ and $\ell = \Gamma_i$. This merge move implies the following split move: When splitting $\boldsymbol{\mu}_\gamma(i)$, the new vector of row scores $\boldsymbol{\mu}'_{\gamma'}$ of dimension $\Gamma_I + 1$, is derived as

$$\boldsymbol{\mu}'_{\gamma'}(k) = \begin{cases} \boldsymbol{\mu}_\gamma(k), & k < \ell, \\ \boldsymbol{\mu}_\gamma(\ell) - u, & k = \ell, \\ \boldsymbol{\mu}_\gamma(\ell) + u, & k = \ell + 1, \\ \boldsymbol{\mu}_\gamma(k-1), & k > \ell + 1, \end{cases} \tag{4.2}$$

for $k = 1, \dots, \Gamma_I + 1$ and $\ell = \Gamma_i$.

TABLE 1.
Stepwise implementation of the transformation used for merging the lowest scores.

$\underbrace{\mu_{\min} = \mu_\gamma(1) < \mu_\gamma(2)}$	$<$	$\mu_\gamma(3)$	$< \cdots <$	$\mu_\gamma(\Gamma_I) = \mu_{\max}$
$\Downarrow$		$\Downarrow$		$\Downarrow$
$\frac{\mu_{\min}+\mu_\gamma(2)}{2}$	$<$	$\mu_\gamma(3)$	$< \cdots <$	$\mu_\gamma(\Gamma_I) = \mu_{\max}$
$\Downarrow$		$\Downarrow$		$\Downarrow$
0	$<$	$\mu_\gamma(3) - \frac{\mu_{\min}+\mu_\gamma(2)}{2}$	$< \cdots <$	$\mu_{\max} - \frac{\mu_{\min}+\mu_\gamma(2)}{2}$
$\Downarrow$		$\Downarrow$		$\Downarrow$
0	$<$	$\frac{\mu_\gamma(3) - \frac{\mu_{\min}+\mu_\gamma(2)}{2}}{\mu_{\max} - \frac{\mu_{\min}+\mu_\gamma(2)}{2}}$	$< \cdots <$	1
$\Downarrow$		$\Downarrow$		$\Downarrow$
$\mu_{\min}$	$< \mu_{\min} +$	$\frac{2\mu_\gamma(3)-\mu_{\min}-\mu_\gamma(2)}{2\mu_{\max}-\mu_{\min}-\mu_\gamma(2)}(\mu_{\max}-\mu_{\min})$	$< \cdots <$	$\mu_{\max}$
$\downarrow$		$\downarrow$		$\downarrow$
$\mu'_{\gamma'}(1)$	$<$	$\mu'_{\gamma'}(2)$	$< \cdots <$	$\mu'_{\gamma'}(\Gamma'_I)$

In the split move, the pseudo-parameter u is generated from the proposal distribution $q(u|\boldsymbol{\mu}, \boldsymbol{\gamma}, \boldsymbol{\gamma}')$, while in the merge move it is calculated from the inverse function of (4.2) resulting in

$$u = \frac{\boldsymbol{\mu}'_{\boldsymbol{\gamma}'}(\ell) - \boldsymbol{\mu}'_{\boldsymbol{\gamma}'}(\ell - 1)}{2}.$$

4.2.2. Merging the Lowest Two Scores and/or Splitting the Lowest Score Here, we consider the case of merging the two lowest scores $\boldsymbol{\mu}_{\boldsymbol{\gamma}}(1)$ and $\boldsymbol{\mu}_{\boldsymbol{\gamma}}(2)$ or splitting the lowest score $\boldsymbol{\mu}_{\boldsymbol{\gamma}}(1)$. Imposing the above "central" scores transformation when merging the lowest two scores will result to $\boldsymbol{\mu}'_{\boldsymbol{\gamma}'}(1) = \frac{1}{2}\{\mu_{\min} + \boldsymbol{\mu}_{\boldsymbol{\gamma}}(2)\} \neq \mu_{\min}$. To obtain a set of parameters that satisfy (2.3), we rescale the above parameters in the [0, 1] interval or, more generally, to any particular $[\mu_{\min}, \mu_{\max}]$ interval we have chosen. So, we subtract from all parameters the quantity $c = \frac{1}{2}\{\mu_{\min} + \boldsymbol{\mu}_{\boldsymbol{\gamma}}(2)\}$, multiply them by $(\mu_{\max} - \mu_{\min})/(\mu_{\max} - c)$ and finally add $\mu_{\min}$; see Table 1 for a detailed description. This procedure results to the matching function

$$\boldsymbol{\mu}'_{\boldsymbol{\gamma}'}(k) = \begin{cases} \mu_{\min}, & k = 1, \\ \mu_{\min} + (\mu_{\max} - \mu_{\min}) \frac{2\boldsymbol{\mu}_{\boldsymbol{\gamma}}(k+1) - \mu_{\min} - \boldsymbol{\mu}_{\boldsymbol{\gamma}}(2)}{2\mu_{\max} - \mu_{\min} - \boldsymbol{\mu}_{\boldsymbol{\gamma}}(2)}, & k > 1, \end{cases} \tag{4.3}$$

for $k = 1, \ldots, \Gamma_I - 1$. The above move is proposed for scores μ_i with $\Gamma_i = 2$. Moreover, for the additional pseudo-parameter u, we set $u = \boldsymbol{\mu}_{\boldsymbol{\gamma}}(2)$ (this is implied by the split move below).

The corresponding split move is applicable for μ_i with $\Gamma_i = 1$ and the new vector $\boldsymbol{\mu}'_{\boldsymbol{\gamma}'}$ of dimension $\Gamma'_I = \Gamma_I + 1$ is given by

$$\boldsymbol{\mu}'_{\boldsymbol{\gamma}'}(k) = \begin{cases} \mu_{\min}, & k = 1, \\ u, & k = 2, \\ \frac{1}{2}\left\{\mu_{\min} + u + (2\mu_{\max} - \mu_{\min} - u) \frac{\boldsymbol{\mu}_{\boldsymbol{\gamma}}(k-1) - \mu_{\min}}{\mu_{\max} - \mu_{\min}}\right\}, & k > 2, \end{cases} \tag{4.4}$$

for $k = 1, \ldots, \Gamma_I + 1$.

4.2.3. Merging the Highest Two Scores and/or Splitting the Highest Score The last move is performed when we wish to merge the two highest scores, $\boldsymbol{\mu}_{\boldsymbol{\gamma}}(\Gamma_I - 1)$ and $\boldsymbol{\mu}_{\boldsymbol{\gamma}}(\Gamma_I)$, or split the last one, $\boldsymbol{\mu}_{\boldsymbol{\gamma}}(\Gamma_I)$. Similarly to the case of lowest scores, when merging, the new highest score is firstly set equal to $\frac{1}{2}\{\boldsymbol{\mu}_{\boldsymbol{\gamma}}(\Gamma_I - 1) + \mu_{\max}\} \neq \mu_{\max}$ and then all scores are rescaled appropriately. Hence, the new vector $\boldsymbol{\mu}'_{\boldsymbol{\gamma}'}$ of dimension $\Gamma'_I = \Gamma_I - 1$ is given by

$$\boldsymbol{\mu}'_{\boldsymbol{\gamma}'}(k) = \begin{cases} \mu_{\min} + 2(\mu_{\max} - \mu_{\min}) \frac{\boldsymbol{\mu}_{\boldsymbol{\gamma}}(k) - \mu_{\min}}{\boldsymbol{\mu}_{\boldsymbol{\gamma}}(\Gamma_I - 1) + \mu_{\max} - 2\mu_{\min}}, & k \leqslant \Gamma'_I - 1 = \Gamma_I - 2, \\ \mu_{\max}, & k = \Gamma'_I = \Gamma_I - 1, \end{cases} \tag{4.5}$$

for $k = 1, \ldots, \Gamma_I - 1$. For the additional parameter u, we set $u = \mu_{\max} - \boldsymbol{\mu}_{\boldsymbol{\gamma}}(\Gamma_I - 1)$.

The reverse split move is given by the vector $\boldsymbol{\mu}'_{\boldsymbol{\gamma}'}$ of dimension $\Gamma'_I = \Gamma_I + 1$

$$\boldsymbol{\mu}'_{\boldsymbol{\gamma}'}(k) = \begin{cases} \boldsymbol{\mu}_{\boldsymbol{\gamma}}(k) - \frac{u}{2} \frac{\boldsymbol{\mu}_{\boldsymbol{\gamma}}(k) - \mu_{\min}}{\mu_{\max} - \mu_{\min}}, & k \leqslant \Gamma'_I - 2 = \Gamma_I - 1, \\ \mu_{\max} - u, & k = \Gamma'_I - 1 = \Gamma_I, \\ \mu_{\max}, & k = \Gamma'_I = \Gamma_I + 1, \end{cases} \tag{4.6}$$

for $k = 1, \ldots, \Gamma_I + 1$.

4.2.4. The Jacobian Matrix The Jacobian matrix in any case will be a $(\Gamma_I - 1) \times (\Gamma_I - 1)$ matrix corresponding to the $\Gamma_I - 2$ free score parameters and the proposed parameter u.

The determinant of the Jacobian for the split moves is given by

$$|\boldsymbol{J}| = \begin{cases} 2, & \text{for central scores,} \\ \left(1 - \frac{1}{2} \frac{u - \mu_{\min}}{\mu_{\max} - \mu_{\min}}\right)^{\Gamma_I - 2}, & \text{for the lowest score,} \\ \left(1 - \frac{1}{2} \frac{u}{\mu_{\max} - \mu_{\min}}\right)^{\Gamma_I - 2}, & \text{for the highest score,} \end{cases}$$

while for the merge moves by

$$|\boldsymbol{J}| = \begin{cases} 1/2, & \text{for central scores,} \\ \left(1 - \frac{1}{2} \frac{\boldsymbol{\mu}_{\boldsymbol{\gamma}}(2) - \mu_{\min}}{\mu_{\max} - \mu_{\min}}\right)^{3 - \Gamma_I}, & \text{for lowest scores,} \\ \left(1 - \frac{1}{2} \frac{\mu_{\max} - \boldsymbol{\mu}_{\boldsymbol{\gamma}}(\Gamma_I - 2)}{\mu_{\max} - \mu_{\min}}\right)^{3 - \Gamma_I}, & \text{for highest scores.} \end{cases}$$

Details concerning its computation are provided in Appendix A.

4.3. *Proposal Distribution of Pseudo-Parameter u*

In all three proposed moves, the pseudo-parameter u lies within an interval of the type (L, U). The values of L and U for each case as well as their derivation are provided in Appendix B. A natural choice for the proposal distribution is the uniform distribution $\mathcal{U}(L, U)$, which has the advantage of avoiding any additional parameter specification. For illustration, we refer to the examples below, where it performed satisfactorily.

Alternatively, we can use a rescaled Beta distribution. In this case, we sample $u^* \sim \mathcal{B}eta(a, b)$ and then set $u = L + u^*(U - L)$. The parameters a and b can be obtained by matching the mean and variance of this density with values taken from a pilot study (see, for example, in Dellaportas et al., 2002).

5. Illustrative Examples

5.1. *Simulated Data*

To demonstrate the performance of the proposed methodology, we have carried out a Monte Carlo study following the simulation pattern of Galindo-Garre and Vermunt (2004). Therefore, we generated 1,000 contingency tables of size 5 × 3, based on model m_1 with $\pi_{ij} = \exp(\tilde{\lambda} + \tilde{\phi}\tilde{\mu}_i\tilde{\nu}_j)$ and equidistant row and column scores (under uniform association). In this subsection, the results are presented in terms of $\tilde{\phi}$, $\tilde{\mu}_i$, and $\tilde{\nu}_j$, i.e., with respect to the STZ/SSTO constraints, in order to be comparable with those obtained by Galindo-Garre and Vermunt. Thus, although we implemented the algorithm of the previous section, we transformed the parameters through the procedure of Section 2.2. Four different set-ups were considered for $\tilde{\phi} \in \{0.3, 1, 2, 3\}$ and two different sample sizes $n = 100$ and $n = 1{,}000$, resulting to a total of eight Monte Carlo based comparisons. The cell probabilities lie in intervals of increasing range in $\tilde{\phi}$. In particular, for $\tilde{\phi} = 0.3$, 1, 2, and 3, the corresponding intervals are $[0.06, 0.08]$, $[0.04, 0.10]$, $[0.02, 0.14]$, and $[0.01, 0.19]$, respectively. Hence, even medium sample sizes could lead to sparse tables. Note, for example, that from our 1,000 generated contingency tables of $n = 100$, the proportion of tables having at least 3 out of all 15 cells (20%) of low frequency (<5) was 66.7%, 85.9%, 99.3% and 100% for $\tilde{\phi} = 0.3$, 1, 2, and 3, respectively. The corresponding proportions for having at least 6 cells (40%) of low frequency were 2.2%, 8%, 44.8%, and 88.3%, respectively. Furthermore, two additional models, m_2 and m_3, were used to generate samples, using the values of $\tilde{\phi}$ and n specified above. Under m_2, we generated samples as before but with the additional equality constraint on the first two row scores, i.e., with $\tilde{\mu}_1 = \tilde{\mu}_2$, while under m_3, the constraints imposed on the scores were $\tilde{\mu}_1 = \tilde{\mu}_2$ and $\tilde{\nu}_2 = \tilde{\nu}_3$.

Table 2 presents the mean ranks of the true model's posterior probability, calculated over 1,000 simulated datasets of size $n = 100$ and $n = 1{,}000$. The proportions of samples for which the true underlying model is detected (i) as the maximum a-posteriori model ($R_{\mathrm{t}} = 1$) and (ii) among the three highest a-posteriori models ($R_{\mathrm{t}} \leqslant 3$) are also provided in the same table. Finally, the median (over all generated samples) of the posterior model odds between the true and the highest a-posteriori model is given. From this table, it turns out that, when $n = 1{,}000$ and $\tilde{\phi} \in \{1, 2, 3\}$, the proposed methodology works efficiently and traces the correct model with a high relative frequency, ranging from 44% to 99.7%. Note that for $\tilde{\phi} = 1$, a higher degree of uncertainty is observed in the selection of the true model, since the true model is identified in around 50% of the generated samples. Nevertheless, the true model was traced among the three best models in more than 85% of the samples.

For $n = 1{,}000$ with $\tilde{\phi} = 0.3$ and for all cases with $n = 100$, results are more vague. This is not surprising, due to the sparseness that can occur when $n = 100$. On the other hand, $\tilde{\phi} = 0.3$ corresponds to very weak association, difficult to be identified. Let us mention that in this case the average (over 1,000 simulations) p-value for testing independence was 0.44 and 0.20 for $n = 100$ and $n = 1{,}000$, respectively. The cell counts are almost uniformly distributed with the observed relative frequencies ranging from 0.06 to 0.08. Nevertheless, the true model is usually among the best 11 models for $\tilde{\phi} = 0.3$ when $n = 1{,}000$ and among the best 22, 11, 5, and 3 models for $\tilde{\phi} = 0.3, 1, 2, 3$, respectively, when $n = 100$. Even though the true model is not indicated as the best one, the median (across generated datasets) of the log-posterior model odds between the true and the highest a-posteriori model is lower than one for all Monte Carlo set-ups. This indicates that for over 50% of the generated datasets the relative difference of the posterior probabilities between the true and the highest a-posteriori model does not provide evidence that is "worth more than a bare mention" according to Kass and Raftery (1995) evaluation scale of posterior model odds. Often sparseness or weak association are responsible for high model uncertainty, reflected in the corresponding posterior model probabilities. In such cases, selection of a "single" best

TABLE 2.
Summaries across 1,000 simulated datasets of the posterior probability and odds of the true model.

True model	$\tilde{\phi}$	$n = 100$				$n = 1{,}000$			
(m_{t})		Mean	Relative frequencies		Median	Mean	Relative frequencies		Median
		rank	$R_{\mathrm{t}} = 1$	$R_{\mathrm{t}} \leqslant 3$	$\log PO_{\mathrm{bt}}$	rank	$R_{\mathrm{t}} = 1$	$R_{\mathrm{t}} \leqslant 3$	$\log PO_{\mathrm{bt}}$
m_1	0.3	14.68	0.006	0.020	0.818	10.71	0.011	0.070	0.864
All scores	1	10.50	0.025	0.092	0.955	2.04	0.442	0.866	0.124
different	2	5.17	0.107	0.378	0.722	1.08	0.930	1.000	0.000
	3	3.05	0.238	0.691	0.436	1.00	0.997	1.000	0.000
m_2	0.3	18.25	0.002	0.013	0.887	11.15	0.039	0.138	0.835
$\mu_1 = \mu_2$	1	11.01	0.035	0.124	0.866	1.85	0.560	0.902	0.000
	2	4.78	0.186	0.508	0.624	1.13	0.885	0.999	0.000
	3	2.90	0.341	0.732	0.329	1.09	0.910	1.000	0.000
m_3	0.3	21.61	0.012	0.064	0.886	11.10	0.043	0.234	0.740
$\mu_1 = \mu_2$	1	10.16	0.053	0.242	0.777	2.09	0.519	0.859	0.000
$\nu_2 = \nu_3$	2	4.37	0.184	0.566	0.543	1.20	0.847	0.989	0.000
	3	2.86	0.329	0.766	0.345	1.11	0.897	0.999	0.000

R_{t}: Ranking of the posterior probability of the true model m_{t} in descending order.
$R_{\mathrm{t}} = 1$: The true model is the model of highest posterior probability.
$R_{\mathrm{t}} \leq 3$: The true model is included in the three highest posterior probability ones.
Mean rank: Mean of the ranking of the posterior probability of the true model m_{t}.
Median $\log PO_{\mathrm{bt}}$: Median over all simulated samples of the log posterior odds of the maximum a-posteriori (best) model versus the true model.
$\tilde{\phi}$ is the association parameter ϕ under the original parametrization with constraints (2.2).

TABLE 3.
Means of posterior marginal probabilities (%) $f(\gamma_i = 1|\boldsymbol{y})$, $i = 2, 3, 4, 5$, and $f(\delta_j = 1|\boldsymbol{y})$, $j = 2, 3$, across 1,000 simulated datasets.

True model	$\tilde{\phi}$	$n = 100$						$n = 1{,}000$					
(m_{t})		Posterior probabilities (%)						Posterior probabilities (%)					
		γ_2	γ_3	γ_4	γ_5	δ_2	δ_3	γ_2	γ_3	γ_4	γ_5	δ_2	δ_3
m_1	0.3	56	51	51	57	67	67	56	57	57	57	71	70
All scores	1	58	57	56	57	71	71	74	79	78	73	97	97
different	2	64	66	65	63	81	82	94	95	94	95	100	100
	3	68	73	73	68	92	92	99	99	99	99	100	100
m_2	0.3	55	51	52	57	68	67	50	56	60	60	71	70
$\mu_1 = \mu_2$	1	48	56	59	61	71	71	37	81	84	80	97	98
	2	42	65	72	68	81	83	28	99	98	98	100	100
	3	40	73	78	75	89	94	24	100	100	100	100	100
m_3	0.3	55	52	52	57	70	64	49	56	60	59	85	50
$\mu_1 = \mu_2$	1	48	57	59	60	86	48	36	84	83	77	100	20
$\nu_2 = \nu_3$	2	40	69	71	65	99	28	25	99	98	92	100	13
	3	35	78	77	70	100	22	21	100	100	97	100	10

model ignores model uncertainty and can lead to false interpretation of association and/or to underestimated standard errors. Instead, the Bayesian model average approach incorporates this model uncertainty in the inference.

Finally, in Table 3, the marginal probabilities for the equalities between successive row and column scores are provided. For $\tilde{\phi} \in \{1, 2, 3\}$, the method works efficiently and traces differences between row and column scores in most of the generated samples. Specifically, for $n = 1{,}000$, all

TABLE 4.
Cross–classification of 223 boys by severity of disturbances of dreams and age (Example 1: Dream's disturbance data).

Age group	Disturbance (from low to high)				Total
	1	2	3	4	
5–7	7	4	3	7	21
8–9	10	15	11	13	49
10–11	23	9	11	7	50
12–13	28	9	12	10	59
14–15	32	5	4	3	44
Total	100	42	41	40	223

equalities between scores ($\gamma_i = 0$ or $\delta_j = 0$) and all distinct scores ($\gamma_i = 1$ or $\delta_j = 1$) are traced correctly with high-posterior probabilities. Even for $n = 100$, marginal posterior probabilities indicate the true patterns but with lower clarity due to the large uncertainty of the generated tables. On the other hand, for $\tilde{\phi} = 0.3$ the method fails to discriminate which row and column scores are different since the posterior distributions are close. The situation is slightly better for the larger datasets (with $n = 1{,}000$) but still more data are needed to have a clearer picture. Nevertheless, in this case, the majority of the means (over all simulated datasets) of posterior inclusion probabilities are close to 0.5 indicating high uncertainty and lack of data information to decide about the equality of the corresponding successive scores.

5.2. *Example 1: Dreams' Disturbance Data*

The classical dataset by Maxwell (1961) on the severity of dreams' disturbance for boys aged 5 to 15, has been used to illustrate the order restricted maximum likelihood estimation of association and correlation models by Agresti et al. (1987) and Ritov and Gilula (1993), respectively. A first Bayesian ordered restricted estimation of the association model for the same data set has been provided in Iliopoulos et al. (2007). The data are listed in Table 4.

In all of the above publications, negative association between age and severity of disturbances was identified. In particular, Agresti et al. (1987) proposed an order restricted C model under which $\hat{\nu}_1 < \hat{\nu}_2 = \hat{\nu}_3 < \hat{\nu}_4$ while Ritov and Gilula (1993) suggested order restriction on the age classification variable as well. They concluded to the correlation model (CA$_{RC}$) with $\hat{\nu}_1 < \hat{\nu}_2 = \hat{\nu}_3 < \hat{\nu}_4$ and $\hat{\mu}_1 = \hat{\mu}_2 < \hat{\mu}_3 = \hat{\mu}_4 < \hat{\mu}_5$.

After implementing the proposed methodology using $\mu_{\min} = \nu_{\min} = 0$ and $\mu_{\max} = \nu_{\max} = 1$, eight models were found with posterior model probabilities higher than 5%. Results for these models are summarized in Table 5. The marginal posterior probabilities for the γ and δ indicators, given in Table 6, provide direct information about merging or not specific row or/and column scores according to the differences of the corresponding categories in terms of the underlying association structure. According to Table 5, the highest probability model (16.2%) sets equal the scores of rows 1–2 and 3–4, indicating that the age categories 5–7 and 8–9 (as well as ages 10–11 and 12–13) do not differentiate in terms of association to severity of dream disturbance. Concerning the dreams disturbance level, the highest probability model suggests that only the first column category of lowest disturbance level distinguishes from the remaining categories since it sets $\nu_1 < \nu_2 = \nu_3 = \nu_4$.

The maximum a posteriori model of our proposed method is more parsimonious than that suggested by Ritov and Gilula (1993) who analyzed the same data. Their model is the one identified and proposed also by Iliopoulos et al. (2007) and is the second highest model of our approach. It is actually very close to the first one, with an estimated posterior probability equal to 0.154. The corresponding posterior model odds (PO_{1k} in Table 5) between the two highest probability models is just 1.05, which suggests no clear separation between them.

TABLE 5.
Estimated posterior model probabilities for most frequently visited models and posterior summaries of ϕ for Example 1 (dreams' disturbance data); Single RJMCMC using R: 100,000 iterations and additional burn-in of 10,000 iterations.

k	Model (score structure)	Posterior probability (%)	PO_{1k}^{**}	Posterior summaries of ϕ			
				Mean	St.Dev.	Percentiles (2.5%, 97.5%)	Odds ratio: $\exp\{\bar{\phi}\}$
1	$\mu_1 = \mu_2 < \mu_3 = \mu_4 < \mu_5$ $\nu_1 < \nu_2 = \nu_3 = \nu_4$	16.20	1.00	−2.06	0.44	(−2.94, −1.21)	0.127
2	$\mu_1 = \mu_2 < \mu_3 = \mu_4 < \mu_5$ $\nu_1 < \nu_2 = \nu_3 < \nu_4$	15.40	1.05	−2.55	0.61	(−3.87, −1.46)	0.078
3	$\mu_1 = \mu_2 < \mu_3 < \mu_4 < \mu_5$ $\nu_1 < \nu_2 = \nu_3 = \nu_4$	8.77	1.85	−1.98	0.45	(−2.86, −1.12)	0.138
4	$\mu_1 = \mu_2 < \mu_3 < \mu_4 < \mu_5$ $\nu_1 < \nu_2 = \nu_3 < \nu_4$	7.25	2.23	−2.42	0.60	(−3.73, −1.33)	0.089
5	$\mu_1 = \mu_2 < \mu_3 = \mu_4 < \mu_5$ $\nu_1 < \nu_2 < \nu_3 < \nu_4$	6.09	2.66	−2.52	0.61	(−3.85, −1.44)	0.081
6	$\mu_1 = \mu_2 < \mu_3 = \mu_4 < \mu_5$ $\nu_1 < \nu_2 < \nu_3 = \nu_4$	5.79	2.80	−2.16	0.48	(−3.12, −1.25)	0.115
7	$\mu_1 < \mu_2 < \mu_3 = \mu_4 < \mu_5$ $\nu_1 < \nu_2 = \nu_3 < \nu_4$	5.41	2.99	−2.71	0.66	(−4.15, −1.52)	0.067
8	$\mu_1 < \mu_2 < \mu_3 = \mu_4 < \mu_5$ $\nu_1 < \nu_2 = \nu_3 = \nu_4$	5.22	3.10	−2.17	0.48	(−3.13, −1.23)	0.114
	Weighted Estimate using BMA			−2.26	0.62	(−3.62, −1.16)	0.104

[**] The posterior model odds of the highest versus the kth highest a-posteriori probable model.

TABLE 6.
Estimated posterior probabilities for γ and δ "split" indicators for Example 1 (dreams disturbance data); single RJMCMC Using R: 100,000 iterations and additional burn-in of 10,000 iterations.

Row scores	Posterior probability	Column scores	Posterior probability
$f(\gamma_2 = 1\|\boldsymbol{y})$	0.285	$f(\delta_2 = 1\|\boldsymbol{y})$	0.996
$f(\gamma_3 = 1\|\boldsymbol{y})$	0.940	$f(\delta_3 = 1\|\boldsymbol{y})$	0.286
$f(\gamma_4 = 1\|\boldsymbol{y})$	0.391	$f(\delta_4 = 1\|\boldsymbol{y})$	0.484
$f(\gamma_5 = 1\|\boldsymbol{y})$	0.964		

All models with posterior probability higher than 5% (provided in Table 5), when compared with the model of highest probability, provide "not worth than a bare mention" evidence in favor of the latter, according to Kass and Raftery (1995) evaluation table for Bayes factors.

The marginal posterior probabilities $f(\gamma_i|\boldsymbol{y})$ and $f(\delta_j|\boldsymbol{y})$ given in Table 6 provide a detailed insight to the structure of the supported model. Hence, there is a clear evidence that the scores for the row categories 2 and 3 (ages 8–9 and 10–11) as well as 4 and 5 (ages 12–13 and 14–15) are different, since their marginal posterior probabilities equal 0.94 and 0.96, respectively. Similarly, posterior model probability $f(\delta_2|\boldsymbol{y}) > 0.99$ suggests that the first and the second categories of "dreams disturbance" are different as well. For the rest of the row and column scores, we observe a high uncertainty concerning their equality. Indeed, their marginal probabilities provide evidence in favor of merging the subsequent scores. Thus, merging of row scores 1–2 and column scores 2–3 are supported with posterior probability of about 0.715. Finally, for row scores 3–4

(10–11, 12–13 years old) and column scores 3–4 (highest levels of dreams' disturbance) posterior probabilities indicate mild evidence in favor of their equality with values 0.609 and 0.516, respectively.

By analyzing the posterior distribution over all visited models (Bayesian model averaging, BMA, estimate) provided in the bottom row of Table 5, we observe a clear negative association between age and severity of dreams' distrurbance ($\phi < 0$) with posterior mean and median equal to -2.26 and -2.21, respectively. The BMA estimate is valuable for this data set, since it provides the posterior distribution of the association parameter ϕ by taking into account also the uncertainty concerning the final structure of the row and column scores. The above values correspond to odds ratio equal to 0.10 having a direct interpretation using our proposed parametrization. Thus, we can argue that the odds of severe dreams' disturbance versus lower disturbances for older children (aged between 14 and 15 years old) is 90% lower than the corresponding odds for younger children. The posterior standard deviation was found equal to 0.62 while 95% and 99% intervals lie between -3.62 and -1.16, -4.17, and -0.88, respectively, supporting the strong negative association assumption for this table (corresponding intervals for odds ratios are given by $(0.027, 0.314)$ and $(0.015, 0.416)$, respectively).

The posterior distribution of ϕ for the a-posteriori highest probable model indicates that we expect under this model the odds of high versus low dreams disturbances to be 7.87 times higher for children aged 14–15 than for children aged 5–9. On the other hand, under the a posteriori second highest probable model this odds is 12.8 times higher for the older children compared to the younger.

Concerning the row and column scores, graphical representations of their posterior distributions are provided in Fig. 1. The posterior summaries of the parameters can be extracted directly from the MCMC output for models with sufficient number of generated posterior values. Table 7 presents posterior means and standard deviations for the parameters of the two models with the highest posterior probabilities. Posterior summaries for the original STZ/SSTO parameterization are also estimated by implementing (2.4) and (2.5) on each iteration of our MCMC output.

In the context of odds ratios, note that in case of no equality restrictions among the scores, there exist $(I-1)(J-1)$ odds ratios comparing each cell of the table with $i \geqslant 2$, $j \geqslant 2$ to the baseline cell ($i = j = 1$). For the highest probability model, the equalities on row and column scores impose restrictions on the odds ratios, and hence their number is reduced just to two (including ϕ itself commented above) while for the second highest probable model we need to calculate four odds ratios. Posterior geometric means using model averaging for the two highest probability models are provided in Table 8. For the two models with the highest probability, the odds ratios for the second row are equal to one, since the corresponding first and second row scores are equal. Note that the odds ratios for all models are decreasing in each column, indicating negative association between dreams' disturbance severity and age.

5.3. Example 2: Association of Schizotypal Personality Subscales in a Student Survey

The data analyzed here are part of a student survey in Greece which aimed to assess the association between schizotypal traits and impulsive and compulsive buying behavior of university students (Iliopoulou, 2004).

The cross-classification of 202 students of the survey according to "social anxiety" and "odd behavior" is given in Table 9. These variables refer to two of the nine specific characteristics of a "schizotypal personality," as they are defined in the DSM-III-R diagnostic and statistical manual of mental disorders, edited by the American Psychiatric Association (1987). Social anxiety refers to excessive stress, nervousness, or feeling extremely uncomfortable when being with other people which does not disappear with familiarity. Odd behavior is related to eccentric appearance, unusual habits, and peculiar actions that may not be acceptable in society.

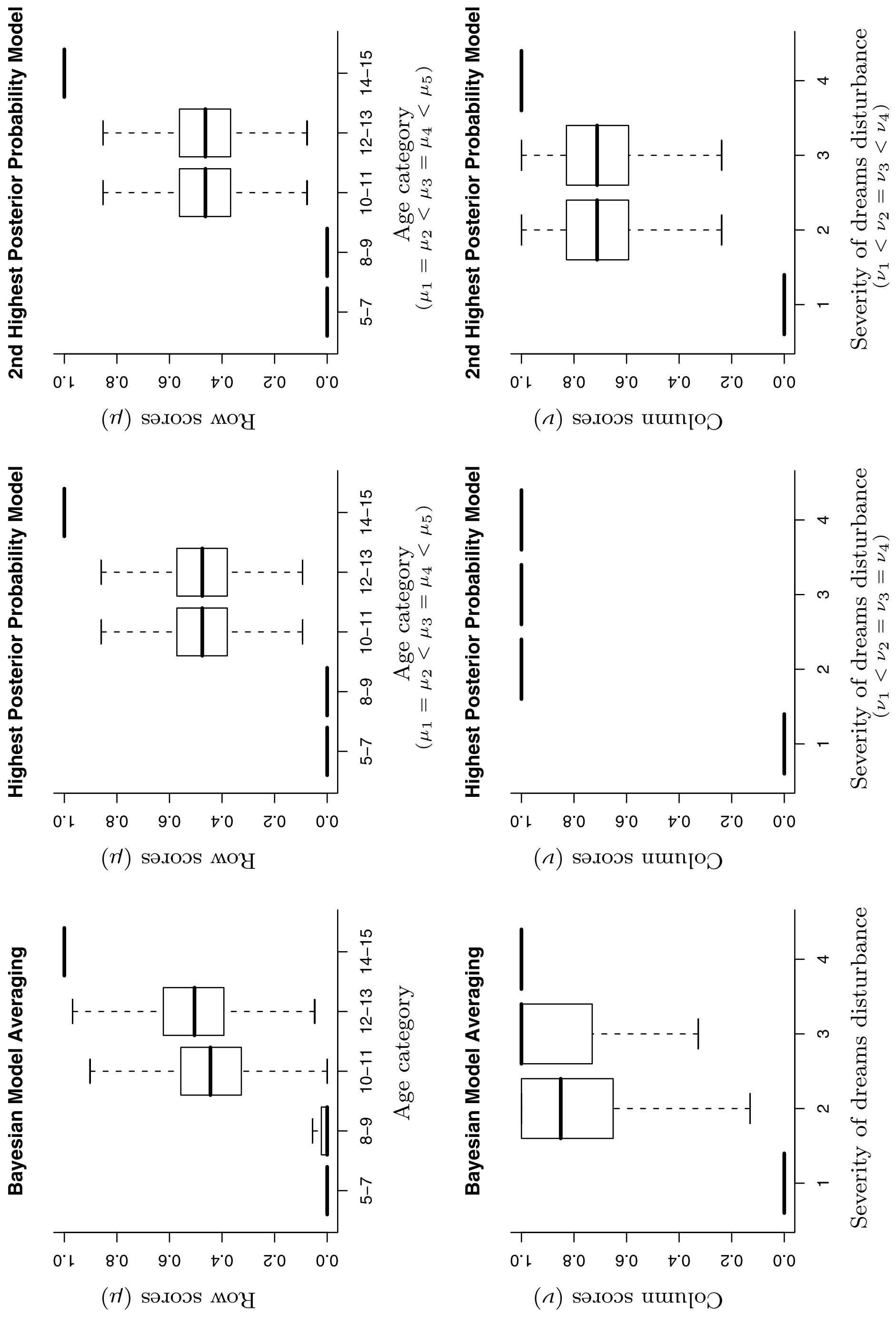

FIGURE 1.
Boxplots of the posterior distribution of the row and column scores for model averaging and the two a posteriori most probable models for Example 1 (dream's disturbance data).

TABLE 7.
Posterior summaries of parameters for the two highest probability models.

Parameter	Proposed parametrization[1]				STZ/SSTO parametrization[2]			
	Model m_1		Model m_2		Model m_1		Model m_2	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
λ	−2.61	0.14	−2.63	0.13	−3.21	0.05	−3.22	0.05
λ_1^X	−1.18	0.23	−1.16	0.23	−0.57	0.19	−0.57	0.19
λ_2^X	−0.32	0.19	−0.30	0.19	0.29	0.14	0.29	0.14
λ_3^X	0.34	0.17	0.33	0.17	0.21	0.14	0.22	0.14
λ_4^X	0.51	0.16	0.50	0.16	0.38	0.13	0.38	0.13
λ_5^X	0.64	0.18	0.62	0.18	−0.30	0.19	−0.32	0.19
λ_1^Y	0.00	0.21	0.03	0.20	0.60	0.11	0.61	0.11
λ_2^Y	0.03	0.14	−0.06	0.15	−0.17	0.13	−0.16	0.13
λ_3^Y	0.00	0.14	−0.08	0.15	−0.20	0.13	−0.18	0.13
λ_4^Y	−0.02	0.15	0.12	0.18	−0.23	0.14	−0.28	0.14
ϕ	−2.06	0.44	−2.55	0.61	−1.51	0.31	−1.61	0.35
μ_1	0.00	–	0.00	–	−0.46	0.05	−0.46	0.05
μ_2	0.00	–	0.00	–	−0.46	0.05	−0.46	0.05
μ_3	0.48	0.15	0.47	0.15	0.10	0.10	0.09	0.10
μ_4	0.48	0.15	0.47	0.15	0.10	0.10	0.09	0.10
μ_5	1.00	–	1.00	–	0.72	0.09	0.73	0.09
ν_1	0.00	–	0.00	–	−0.87	–	−0.80	0.07
ν_2	1.00	–	0.71	0.16	0.29	–	0.13	0.10
ν_3	1.00	–	0.71	0.16	0.29	–	0.13	0.10
ν_4	1.00	–	1.00	0.00	0.29	–	0.53	0.13

m_1: Highest probability model; m_2: 2nd highest probability model; see Table 5 for details.
Mean: posterior mean; S.D.: posterior standard deviation.

[1] Parametrization based on (2.1) and (2.3) with $\mu_{\min} = \nu_{\min} = 0$ and $\mu_{\max} = \nu_{\max} = 1$.

[2] Parametrization based on (2.1) and (2.2) with $w_{1i} = w_{2j} = 1$ for all i, j.

TABLE 8.
Estimated posterior odds ratios (posterior geometric means) using model averaging for the two highest probability models for Example 1 (dream's disturbance data).

Bayesian model averaging				Highest prob. model		2nd highest prob. model		
i	$j=2$	$j=3$	$j=4$	i	$j=2,3,4$	i	$j=2,3$	$j=4$
2	0.93	0.93	0.91					
3	0.46	0.43	0.37	3, 4	0.38	3, 4	0.47	0.35
4	0.40	0.38	0.32					
5	0.17	0.15	0.10	5	0.13	5	0.20	0.08

Since both classification variables of Table 9 are ordinal, we shall apply the order restricted RC model with $\mu_{\min} = \nu_{\min} = 0$ and $\mu_{\max} = \nu_{\max} = 1$. Details for the most probable models are summarized in Table 10. Note that, for this example, traditional asymptotic inference cannot be applied due to the small cell frequencies.

According to the information provided in Table 10, the most probable model (3%) is the one with just three distinguished values for the column and five distinguished values for the row scores. Namely, it sets $\mu_2 = \mu_3$, $\mu_4 = \mu_5$ for the row scores and $\nu_2 = \nu_3$, $\nu_4 = \nu_5 = \nu_6$

TABLE 9.
Cross–classification of 202 students by social anxiety and odd behavior sub-scales (Example 2: Schizotypy data).

Social anxiety score	Odd behavior score						
	0	1	2	3	4	5–7	Total
0	11	5	1	0	1	0	18
1	13	8	8	2	2	3	36
2	8	9	4	1	4	0	26
3	6	7	5	4	4	1	27
4	6	9	5	3	2	4	29
5	3	13	5	4	1	5	31
6–8	0	11	5	10	3	6	35
Total	47	62	33	24	17	19	202

TABLE 10.
Estimated posterior model probabilities for most frequently visited models for Example 2 (schizotypy data); single RJMCMC using R: 500,000 iterations and additional burn-in of 10,000 iterations.

k	Model (score structure)	Posterior probability (%)	PO_{1k}**
1	$\mu_1 < \mu_2 = \mu_3 < \mu_4 = \mu_5 < \mu_6 < \mu_7$ $\nu_1 < \nu_2 = \nu_3 < \nu_4 = \nu_5 = \nu_6$	3.01	1.00
2	$\mu_1 < \mu_2 < \mu_3 < \mu_4 = \mu_5 < \mu_6 < \mu_7$ $\nu_1 < \nu_2 = \nu_3 < \nu_4 = \nu_5 = \nu_6$	1.94	1.55
3	$\mu_1 < \mu_2 = \mu_3 < \mu_4 = \mu_5 = \mu_6 < \mu_7$ $\nu_1 < \nu_2 = \nu_3 < \nu_4 = \nu_5 = \nu_6$	1.91	1.57
4	$\mu_1 < \mu_2 = \mu_3 < \mu_4 = \mu_5 < \mu_6 < \mu_7$ $\nu_1 < \nu_2 = \nu_3 < \nu_4 = \nu_5 < \nu_6$	1.69	1.78
5	$\mu_1 < \mu_2 < \mu_3 = \mu_4 = \mu_5 < \mu_6 < \mu_7$ $\nu_1 < \nu_2 = \nu_3 < \nu_4 = \nu_5 = \nu_6$	1.68	1.79
6	$\mu_1 < \mu_2 = \mu_3 < \mu_4 < \mu_5 < \mu_6 < \mu_7$ $\nu_1 < \nu_2 = \nu_3 < \nu_4 = \nu_5 = \nu_6$	1.61	1.87

**The posterior model odds of the highest versus the k-th highest a-posteriori probable model.

for the column scores, with the analogous interpretation. The low probability of the maximum a posteriori model implies high uncertainty concerning the structure of the model for this dataset, which is mainly due to the large number of models under consideration (equal to $(2^5 - 1) \times (2^6 - 1) = 1953$) and the small sample size. The posterior odds ratios, presented in Table 11, are increasing with row and column levels, indicating an underlying positive association between the two classification variables. Moreover, values of last row and column scores are much higher than the rest ones indicating that the association between the two subscales is stronger at the higher levels of the variables under consideration. In this dataset, the model uncertainty is high as well and hence odds ratios estimated using BMA are of high importance.

5.4. *Example 3: Family Size and Happiness Data*

Finally, we briefly comment (without analytically presenting the results) on the data of Clogg (1982, Table 2) which were analyzed in the context of ordered RC models by Galindo-Garre and Vermunt (2004). The data refer to the cross-classification of 1,517 families according to the

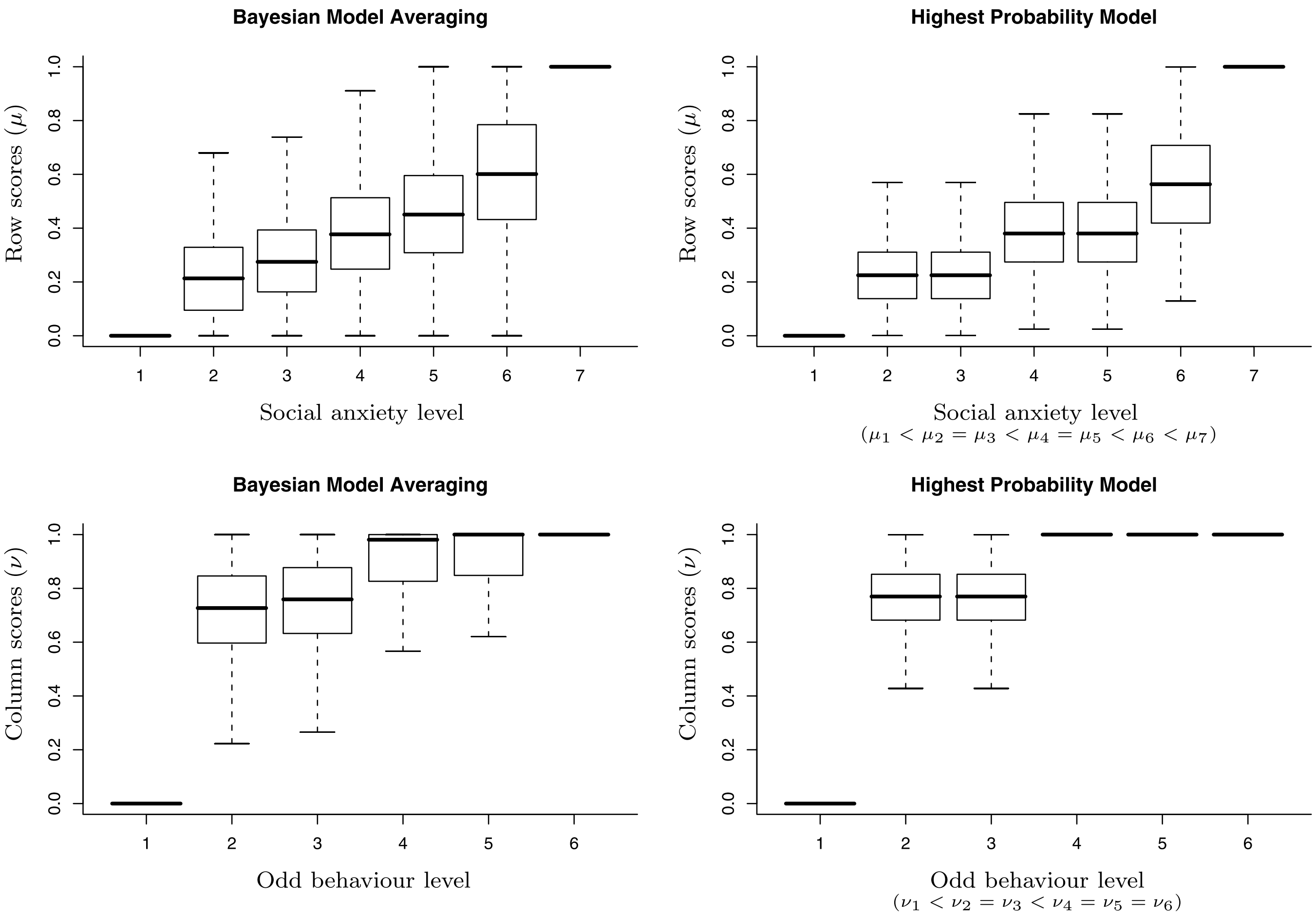

FIGURE 2.
Boxplots of the posteriors of the row and column scores for model averaging and the a posteriori most probable model for Example 2 (schizotypy data).

TABLE 11.
Estimated posterior odds ratios (posterior geometric means) for Example 2 (schizotypy data), using model averaging and the highest probability model.

	Bayesian model averaging					Highest prob. model		
i	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$		$j=2,3$	$j=4,5,6$
2	2.25	2.36	2.94	2.99	3.43	2, 3	2.72	3.95
3	2.79	2.95	3.87	3.95	4.66			
4	3.95	4.25	6.07	6.24	7.78	4, 5	5.16	9.38
5	5.05	5.49	8.22	8.49	11.14			
6	8.85	9.82	16.25	16.93	24.19	6	11.19	25.71
7	61.18	71.82	162.04	172.18	278.61	7	115.17	469.00

number of siblings (five levels) and the level of happiness (three levels). The model supported by the method of Galindo-Garre and Vermunt (2004) with $\nu_1 = \nu_2$ is the model with the highest posterior probability (0.16) while the second most probable model is the one with $\nu_1 = \nu_2$ and $\mu_4 = \mu_5$ with close posterior probability (0.149). Hence, the Bayesian solution here indicates that the data cannot discriminate clearly between the two a posteriori most probable models and, therefore, BMA may be used.

5.5. *Computational Details*

To implement the proposed algorithms the authors have developed `R` functions and stand alone `Fortran` software. Both were used for cross checking of results. Software is available from the authors upon request.

Results for Examples 1–3 illustrated in this paper have been generated using `R` functions. Presented results regarding the Monte Carlo study have been obtained using the `Fortran` code to speed up computations. For Example 1, results were generated using a sample of 100,000 iterations after discarding additional initial 10,000 iterations as burn-in period. Running time was approximately equal to 10 minutes, on a Pentium dual core 2.0 PC using version 2.0.1 of `R`. For Example 2, results were based on 500,000 iterations after discarding initial 10,000 iterations as burn-in period. Running time was approximately equal to 18 minutes per 100 thousand iterations on the same machine. Both examples were also run extensively using the corresponding `Fortran` program which was considerably faster. For both cases, samples of 500,000 iterations were generated and compared with the corresponding `R` results to ensure convergence as well as lack of programming bugs.

The algorithm was highly mobile in all our examples. In particular, in Example 1 the algorithm visited 69, 86 and eventually all 105 models in 10, 100 and 400 thousand iterations, respectively. On the other hand, in Example 2, the algorithm visited 358, 730, 894, and 1,083 (out of 1,953) models in 1, 10, 100, and 500 thousand iterations, respectively.

6. Discussion and Further Research

In this paper, we dealt with the problem of Bayesian score merging for the ordered restricted association models used in two-way contingency tables. We have focused on the comparison of scores using trans-dimensional methods (RJMCMC, Green, 1995). To achieve this, we proposed a nonstandard parametrization which is convenient in terms of interpretation and computation. For exploring the model space, we constructed a flexible RJMCMC algorithm using an order-restricted uniform prior for each model's scores. Our approach can easily handle sparse tables since it is not based on asymptotic results and avoids sequential pairwise testing (and stepwise procedures) making our method automatic in the sense that the best a posteriori models are directly available by ordering the estimated posterior model probabilities.

The approach presented in this paper assumes that a log-multiplicative structure between the ordinal variables exists. Comparison of the order-restricted association model with other standard models, such as the independence and the saturated model, can be done in a straightforward manner using the deviance information criterion (DIC, Spiegelhalter et al., 2002), Bayesian or standard versions of AIC and BIC (Brooks, 2002) or other Bayesian measures as posterior p-values and the ones proposed for association models (in a more general context) by Kateri et al. (2005). For such a comparison in RC models using the DIC, see also Iliopoulos et al. (2007).

An obvious extension of our proposed method is to embody to our algorithm other association models such as, for example, the R and C models. Special types of U models with possibility of equality of some successive scores and equidistant the distinct ones could also be of interest. This will considerably complicate the algorithm, due to the appearance of additional issues, such as, for example, the construction of a set of sensible and compatible priors across different models. We did not pursued this issue further within this paper, since this work was focused on inference concerning the merging of parametric scores for ordinal variables.

Other interesting issues for future research may include the prior elicitation and how we should incorporate prior information in such models. An interesting prior can be constructed by using the power prior distribution proposed by Chen, Ibrahim, and Shao (2000) based on imaginary data which will express our prior beliefs. In this way, prior distributions that are compatible across models can be constructed in a straightforward manner.

A challenging task is to extend this Bayesian approach to high-dimensional contingency tables (i.e., to multiway association models with order restricted scores). This will be of particular interest, since high-dimensional tables often face the problem of sparseness under which the standard asymptotic approach is improper. Goodness of fit testing and expected cell frequencies estimation in the high-dimensional contingency tables context is an area of increasing interest. Indicatively, we mention the pioneering work of Simonoff (1983), as well as Dong and Simonoff (1995), and Burman (2004). Maydeu-Olivares and Joe (2006) propose a family of goodness-of-fit statistics for testing composite null hypothesis in multidimensional contingency tables based on a limited information method (see also references cited therein). Our Bayesian approach could be developed and compared to this.

Finally, the method could potentially be extended to other families of models, such as in analyzing mortality rates by the “Lee-Carter model” (Lee and Carter, 1992), since it would be natural to impose order-restriction because of the expected monotonicity of mortality within age. Another option could be in a regression type context with multiple predictors using the stereotype model (Anderson, 1984).

Acknowledgements

The authors wish to thank the editor, the associate editor as well as the three anonymous referees for their detailed, precise, and productive comments which considerably improved the paper.

Appendix A. Calculation of the Jacobians

The Jacobian matrix is a $(\Gamma_I - 1) \times (\Gamma_I - 1)$ matrix with the derivatives of the new scores with respect to the old.

When splitting a "central" score $\boldsymbol{\mu}_{\boldsymbol{\gamma}}(\ell)$, we have

$$\frac{\partial \boldsymbol{\mu}'_{\boldsymbol{\gamma}'}(k)}{\partial \boldsymbol{\mu}_{\boldsymbol{\gamma}}(l)} = I(k \leqslant i)\delta_{k,l} + I(k > i)\delta_{k,l+1},$$
$$\frac{\partial \boldsymbol{\mu}'_{\boldsymbol{\gamma}'}(k)}{\partial u} = \delta_{k,i+1} - \delta_{k,i},$$

where $I(\cdot)$ is the indicator function and $\delta_{k,l} = I(k = l)$ is Kronecker's delta. The corresponding Jacobian determinant is given by

$$|\boldsymbol{J}| = \begin{vmatrix} \boldsymbol{I}_{k_1} & \mathbf{0}_{1\times 2} & \mathbf{0}_{1\times k_2} \\ \mathbf{0}_{1\times k_1} & (1,1) & \mathbf{0}_{1\times k_2} \\ \mathbf{0}_{k_2\times k_1} & \mathbf{0}_{k_2\times 2} & \boldsymbol{I}_{k_2} \\ \mathbf{0}_{1\times k_1} & (-1,1) & \mathbf{0}_{1\times k_2} \end{vmatrix} = \begin{vmatrix} 1 & 1 \\ -1 & 1 \end{vmatrix} = 2,$$

where $\boldsymbol{I}_k$ is the identity matrix of order k, $\mathbf{0}_{k\times l}$ is the $k \times l$ matrix of zeros and $k_1 = \Gamma_\ell - 2$, $k_2 = \Gamma_I - \Gamma_\ell - 1$.

When merging two central scores, the Jacobian is simply the inverse of the above quantity, i.e. $|\boldsymbol{J}| = 1/2$.

For the split move of lowest scores, the partial derivatives occurring in the Jacobian are

$$\frac{\partial \boldsymbol{\mu}'_{\boldsymbol{\gamma}'}(k)}{\partial \boldsymbol{\mu}_{\boldsymbol{\gamma}}(l)} = \left(1 - \frac{u - \mu_{\min}}{2R_{\boldsymbol{\mu}}}\right)\delta_{k,l+1} I(k > 2),$$
$$\frac{\partial \boldsymbol{\mu}'_{\boldsymbol{\gamma}'}(k)}{\partial u} = \frac{1}{2}\left(1 - \frac{\boldsymbol{\mu}_{\boldsymbol{\gamma}}(k-1) - \mu_{\min}}{R_{\boldsymbol{\mu}}}\right) I(k > 2) + I(k = 2),$$

where $R_{\boldsymbol{\mu}} = \mu_{\max} - \mu_{\min}$. Hence, the absolute value of the Jacobian determinant is given by

$$|\boldsymbol{J}| = \begin{vmatrix} \operatorname{diag}\left(\frac{\partial \boldsymbol{\mu}'_{\boldsymbol{\gamma}'}(k)}{\partial \boldsymbol{\mu}_{\boldsymbol{\gamma}}(k-1)}\right) & \frac{\partial \boldsymbol{\mu}'_{\boldsymbol{\gamma}'}}{\partial u} \\ \mathbf{0}_{1\times \Gamma_I - 2} & 1 \end{vmatrix} = \left(1 - \frac{u - \mu_{\min}}{2R_{\boldsymbol{\mu}}}\right)^{\Gamma_I - 2},$$

where $\frac{\partial \boldsymbol{\mu}'_{\boldsymbol{\gamma}'}}{\partial u}$ denotes the vector of the corresponding $\Gamma_I - 2$ partial derivatives with respect to u.

The inverse (merge) move of the lowest scores is obtained by inverting the above quantity and substituting u, Γ_I by $\boldsymbol{\mu}_{\boldsymbol{\gamma}}(2)$, $\Gamma_I - 1$, respectively. Therefore, in this case, the absolute value of the Jacobian determinant is

$$|\boldsymbol{J}| = \left(1 - \frac{\boldsymbol{\mu}_{\boldsymbol{\gamma}}(2) - \mu_{\min}}{2R_{\boldsymbol{\mu}}}\right)^{-(\Gamma_I - 1) + 2} = \left(1 - \frac{\boldsymbol{\mu}_{\boldsymbol{\gamma}}(2) - \mu_{\min}}{2R_{\boldsymbol{\mu}}}\right)^{3 - \Gamma_I}.$$

Similarly, for splitting the highest scores, we obtain

$$\frac{\partial \boldsymbol{\mu}'_{\boldsymbol{\gamma}'}(k)}{\partial \boldsymbol{\mu}_{\boldsymbol{\gamma}}(l)} = \left(1 - \frac{u}{2R_{\boldsymbol{\mu}}}\right)\delta_{k,l} I(k < \Gamma_I),$$
$$\frac{\partial \boldsymbol{\mu}'_{\boldsymbol{\gamma}'}(k)}{\partial u} = -\frac{\boldsymbol{\mu}_{\boldsymbol{\gamma}}(k) - \mu_{\min}}{2R_{\boldsymbol{\mu}}} I(k < \Gamma_I) - I(k = \Gamma_I).$$

Finally, the absolute value of the Jacobian determinant for the split move of the highest scores is given by

$$|\boldsymbol{J}| = \begin{vmatrix} \mathrm{diag}\left(\frac{\partial \boldsymbol{\mu}'_{\boldsymbol{\gamma}'}(k)}{\partial \boldsymbol{\mu}_{\boldsymbol{\gamma}}(k)}\right) & \frac{\partial \boldsymbol{\mu}'_{\boldsymbol{\gamma}'}}{\partial u} \\ \mathbf{0}_{1\times \Gamma_I - 2} & -1 \end{vmatrix} = \left(1 - \frac{u}{2R_{\mu}}\right)^{\Gamma_I - 2},$$

whilst for the merge move of the highest scores it becomes

$$|\boldsymbol{J}| = \left(1 - \frac{\mu_{\max} - \boldsymbol{\mu}_{\boldsymbol{\gamma}}(\Gamma_I - 2)}{2R_{\mu}}\right)^{3-\Gamma_I}.$$

Appendix B. Domains of the Proposal Parameter u

To identify appropriate proposal distributions we first need to identify the domain of the additional parameter u.

Let us first consider a split move for the "central" scores. In case we wish to split $\boldsymbol{\mu}_{\boldsymbol{\gamma}}(\ell)$ into the new scores $\boldsymbol{\mu}'_{\boldsymbol{\gamma}'}(\ell)$ and $\boldsymbol{\mu}'_{\boldsymbol{\gamma}'}(\ell+1)$, the proposed scores satisfy

$$\boldsymbol{\mu}'_{\boldsymbol{\gamma}'}(\ell-1) < \boldsymbol{\mu}'_{\boldsymbol{\gamma}'}(\ell) < \boldsymbol{\mu}'_{\boldsymbol{\gamma}'}(\ell+1) < \boldsymbol{\mu}'_{\boldsymbol{\gamma}'}(\ell+2)$$

for every ℓ such that $2 \leqslant \ell \leqslant \Gamma_I - 1$. Through (4.2), this leads to

$$\boldsymbol{\mu}_{\boldsymbol{\gamma}}(\ell-1) < \boldsymbol{\mu}_{\boldsymbol{\gamma}}(\ell) - u < \boldsymbol{\mu}_{\boldsymbol{\gamma}}(\ell) + u < \boldsymbol{\mu}_{\boldsymbol{\gamma}}(\ell+1).$$

It follows that

$$L = 0 < u < \min\bigl\{\boldsymbol{\mu}_{\boldsymbol{\gamma}}(\ell) - \boldsymbol{\mu}_{\boldsymbol{\gamma}}(\ell-1), \boldsymbol{\mu}_{\boldsymbol{\gamma}}(\ell+1) - \boldsymbol{\mu}_{\boldsymbol{\gamma}}(\ell)\bigr\} = U.$$

Analogously, for the lowest scores, we have the constraint

$$\mu_{\min} < u < \frac{1}{2}\left\{\mu_{\min} + u + (R_{\mu} + \mu_{\max} - u)\frac{\boldsymbol{\mu}_{\boldsymbol{\gamma}}(2) - \mu_{\min}}{R_{\mu}}\right\}$$

resulting in

$$L = \mu_{\min} < u < \boldsymbol{\mu}_{\boldsymbol{\gamma}}(2) + \frac{(\boldsymbol{\mu}_{\boldsymbol{\gamma}}(2) - \mu_{\min})\{\mu_{\max} - \boldsymbol{\mu}_{\boldsymbol{\gamma}}(2)\}}{R_{\mu} + \boldsymbol{\mu}_{\boldsymbol{\gamma}}(2) - \mu_{\min}} = U.$$

Finally, for the highest scores, we have that

$$\boldsymbol{\mu}_{\boldsymbol{\gamma}}(\Gamma_I - 1) - \frac{\{\boldsymbol{\mu}_{\boldsymbol{\gamma}}(\Gamma_I - 1) - \mu_{\min}\}u}{2R_{\mu}} < \mu_{\max} - u < \mu_{\max}$$

leading to

$$L = 0 < u < \frac{2R_{\mu}\{\mu_{\max} - \boldsymbol{\mu}_{\boldsymbol{\gamma}}(\Gamma_I - 1)\}}{R_{\mu} + \mu_{\max} - \boldsymbol{\mu}_{\boldsymbol{\gamma}}(\Gamma_I - 1)} = U.$$

The above values are needed also for the calculation of the acceptance probability α in the inverse merge moves of RJMCMC. The only difference is that we need to substitute $\boldsymbol{\mu}_{\boldsymbol{\gamma}}$ by $\boldsymbol{\mu}'_{\boldsymbol{\gamma}'}$; see Table 12 for the corresponding expressions.

TABLE 12.
Limits of the pseudo parameter u for the inverse merge move ($R_{\boldsymbol{\mu}} = \mu_{\max} - \mu_{\min}$).

Scores	L	U
Lowest	0	$\boldsymbol{\mu}'_{\boldsymbol{\gamma}'}(2) + \{\boldsymbol{\mu}'_{\boldsymbol{\gamma}'}(2) - \mu_{\min}\}\{\mu_{\max} - \boldsymbol{\mu}'_{\boldsymbol{\gamma}'}(2)\}/\{R_{\boldsymbol{\mu}} + \boldsymbol{\mu}'_{\boldsymbol{\gamma}'}(2) - \mu_{\min}\}$
Central	$\mu_{\min}$	$\min\{\boldsymbol{\mu}'_{\boldsymbol{\gamma}'}(\ell - 1) - \boldsymbol{\mu}'_{\boldsymbol{\gamma}'}(\ell - 2),\ \boldsymbol{\mu}'_{\boldsymbol{\gamma}'}(\ell + 1) - \boldsymbol{\mu}'_{\boldsymbol{\gamma}'}(\ell)\}$
Highest	0	$2R_{\boldsymbol{\mu}}\{\mu_{\max} - \boldsymbol{\mu}'_{\boldsymbol{\gamma}'}(\Gamma'_I - 1)\}/\{R_{\boldsymbol{\mu}} + \mu_{\max} - \boldsymbol{\mu}'_{\boldsymbol{\gamma}'}(\Gamma'_I - 1)\}$

Appendix C. Summarizing the Split and Merge Moves

From the above, we summarize the RJMCMC details, for $i = 2, \dots, \Gamma_I$ using the following steps. In what follows, we denote by LR the likelihood ratio between the likelihood of the proposed and the current state of the chain. Hence,

$$\mathrm{LR} = \frac{f(\boldsymbol{y}|\boldsymbol{\lambda}^X, \boldsymbol{\lambda}^Y, \phi, \boldsymbol{\mu}', \boldsymbol{\nu})}{f(\boldsymbol{y}|\boldsymbol{\lambda}^X, \boldsymbol{\lambda}^Y, \phi, \boldsymbol{\mu}, \boldsymbol{\nu})},$$

where $\boldsymbol{\mu}'$ is the proposed value of $\boldsymbol{\mu}$.

1. If $\gamma_i = 0 \to \gamma'_i = 1$ (split move), then:
   (a) For $\Gamma_i = 1$ (lowest scores) we proceed with the following steps:
      i. Generate u from $\mathcal{U}(\mu_{\min}, \boldsymbol{\mu}_{\boldsymbol{\gamma}}(2) + \frac{(\boldsymbol{\mu}_{\boldsymbol{\gamma}}(2)-\mu_{\min})(\mu_{\max}-\boldsymbol{\mu}_{\boldsymbol{\gamma}}(2))}{R_{\boldsymbol{\mu}}+\boldsymbol{\mu}_{\boldsymbol{\gamma}}(2)-\mu_{\min}})$.
      ii. Calculate $\mu'_{\gamma'}$ by (4.4).
      iii. Accept the proposed move with probability $\alpha = \min\{1, A\}$ with A given by

$$A = \mathrm{LR} \times \frac{2(\Gamma_I - 1)(\boldsymbol{\mu}_{\boldsymbol{\gamma}}(2) - \mu_{\min})}{R_{\boldsymbol{\mu}} + \boldsymbol{\mu}_{\boldsymbol{\gamma}}(2) - \mu_{\min}} \times \left(1 - \frac{u - \mu_{\min}}{2R_{\boldsymbol{\mu}}}\right)^{\Gamma_I - 2}.$$

   (b) For $2 \leqslant \Gamma_i \leqslant \Gamma_I - 1$ (central scores):
      i. Set $\ell = \Gamma_i$.
      ii. Generate u from $\mathcal{U}(0, \max\{\boldsymbol{\mu}_{\boldsymbol{\gamma}}(\ell) - \boldsymbol{\mu}_{\boldsymbol{\gamma}}(\ell - 1), \boldsymbol{\mu}_{\boldsymbol{\gamma}}(\ell + 1) - \boldsymbol{\mu}_{\boldsymbol{\gamma}}(\ell)\})$.
      iii. Calculate $\boldsymbol{\mu}'_{\boldsymbol{\gamma}'}$ by (4.2).
      iv. Accept the proposed move with probability $\alpha = \min\{1, A\}$ with A given by

$$A = \mathrm{LR} \times \frac{\Gamma_I - 1}{R_{\boldsymbol{\mu}}} \times \max\{\boldsymbol{\mu}_{\boldsymbol{\gamma}}(\ell) - \boldsymbol{\mu}_{\boldsymbol{\gamma}}(\ell - 1), \boldsymbol{\mu}_{\boldsymbol{\gamma}}(\ell + 1) - \boldsymbol{\mu}_{\boldsymbol{\gamma}}(\ell)\} \times 2.$$

   (c) For $\Gamma_i = \Gamma_I$ (highest scores):
      i. Generate u from $\mathcal{U}(0, \frac{2R_{\boldsymbol{\mu}}\{\mu_{\max}-\boldsymbol{\mu}_{\boldsymbol{\gamma}}(\Gamma_I-1)\}}{R_{\boldsymbol{\mu}}+\mu_{\max}-\boldsymbol{\mu}_{\boldsymbol{\gamma}}(\Gamma_I-1)})$.
      ii. Calculate $\boldsymbol{\mu}'_{\boldsymbol{\gamma}'}$ by (4.6).
      iii. Accept the proposed move with probability $\alpha = \min\{1, A\}$ with A given by

$$A = \mathrm{LR} \times \frac{2(\Gamma_I - 1)\{\mu_{\max} - \boldsymbol{\mu}_{\boldsymbol{\gamma}}(\Gamma_I - 1)\}}{R_{\boldsymbol{\mu}} + \mu_{\max} - \boldsymbol{\mu}_{\boldsymbol{\gamma}}(\Gamma_I - 1)} \times \left(1 - \frac{u}{2R_{\boldsymbol{\mu}}}\right)^{\Gamma_I - 2}.$$

2. If $\gamma_i = 1 \to \gamma'_i = 0$ (merge move) then:
   (a) For $\Gamma_i = 2$ (lowest scores), we proceed with the following steps:
      i. Set $u = \boldsymbol{\mu}_{\boldsymbol{\gamma}}(2)$.
      ii. Calculate $\boldsymbol{\mu}'_{\boldsymbol{\gamma}'}$ by (4.3).

iii. Accept the proposed move with probability $\alpha = \min\{1, A\}$ with A given by

$$A = \mathrm{LR} \times \frac{R_{\boldsymbol{\mu}} + \boldsymbol{\mu}'_{\boldsymbol{\gamma}'}(2) - \mu_{\min}}{2(\Gamma_I - 2)(\boldsymbol{\mu}'_{\boldsymbol{\gamma}'}(2) - \mu_{\min})} \times \left(1 - \frac{\boldsymbol{\mu}_{\boldsymbol{\gamma}}(2) - \mu_{\min}}{2R_{\boldsymbol{\mu}}}\right)^{3-\Gamma_I}.$$

(b) For $3 \leqslant \Gamma_i \leqslant \Gamma_I - 1$ (central scores):
   i. Set $\ell = \Gamma_i$.
   ii. Set $u = \frac{\boldsymbol{\mu}_{\boldsymbol{\gamma}}(\ell) - \boldsymbol{\mu}_{\boldsymbol{\gamma}}(\ell-1)}{2}$.
   iii. Calculate $\boldsymbol{\mu}'_{\boldsymbol{\gamma}'}$ by (4.2).
   iv. Accept the proposed move with probability $\alpha = \min\{1, A\}$ with A given by

$$A = \mathrm{LR} \times \frac{R_{\boldsymbol{\mu}}}{2(\Gamma_I - 2)\min\{\boldsymbol{\mu}'_{\boldsymbol{\gamma}'}(\ell-1) - \boldsymbol{\mu}'_{\boldsymbol{\gamma}'}(\ell-2), \boldsymbol{\mu}'_{\boldsymbol{\gamma}'}(\ell+1) - \boldsymbol{\mu}'_{\boldsymbol{\gamma}'}(\ell)\}}.$$

(c) For $\Gamma_i = \Gamma_I$ (highest scores):
   i. Set $u = \mu_{\max} - \boldsymbol{\mu}_{\boldsymbol{\gamma}}(\Gamma_I - 1)$.
   ii. Calculate $\boldsymbol{\mu}'_{\boldsymbol{\gamma}'}$ by (4.6).
   iii. Accept the proposed move with probability $\alpha = \min\{1, A\}$ with A given by

$$A = \mathrm{LR} \times \frac{R_{\boldsymbol{\mu}} + \mu_{\max} - \boldsymbol{\mu}'_{\boldsymbol{\gamma}'}(\Gamma_I - 2)}{2(\Gamma_I - 2)\{\mu_{\max} - \boldsymbol{\mu}'_{\boldsymbol{\gamma}'}(\Gamma_I - 2)\}} \times \left(1 - \frac{\mu_{\max} - \boldsymbol{\mu}_{\boldsymbol{\gamma}}(\Gamma_I - 2)}{2R_{\boldsymbol{\mu}}}\right)^{3-\Gamma_I}.$$

References

Agresti, A., Chuang, C., & Kezouh, A. (1987). Order-restricted score parameters in association models for contingency tables. *Journal of the American Statistical Association*, *82*, 619–623.

American Psychiatric Association (1987). *DSM-III-R: Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington: APA.

Anderson, C.J., & Yu, H.T. (2007). Log-multiplicative association models as item response models. *Psychometrika*, *72*, 5–23.

Anderson, J.A. (1984). Regression and ordered categorical variables. *Journal of the Royal Statistical Society B*, *46*, 1–30.

Bartolucci, F., & Forcina, A. (2002). Extended RC association models allowing for order restrictions and marginal modeling. *Journal of the American Statistical Association*, *97*, 1192–1199.

Brooks, S.P. (2002). Discussion of the paper by Spiegelhalter, Best, Carlin and van der Linde. *Journal of the Royal Statistical Society B*, *64*, 616–618.

Burman, P. (2004). On some testing problems for sparse contingency tables. *Journal of Multivariate Analysis*, *88*, 1–18.

Chen, M.H., Ibrahim, J.G., & Shao, Q.M. (2000). Power prior distributions for generalized linear models. *Journal of Statistical Planning and Inference*, *84*, 121–137.

Chuang, C. (1982). Empirical Bayes methods for a two-way multiplicative-interaction model. *Communications in Statistics: Theory and Methods*, *11*, 2977–2989.

Clogg, C.C. (1982). Some models for the analysis of association in multi-way cross-classifications having ordered categories. *Journal of the American Statistical Association*, *77*, 803–815.

De Rooij, M. (2007). The Distance Perspective of Generalized Biadditive Models: Scalings and Transformations. *Journal of Computational and Graphical Statistics*, *16*, 210–227.

Dellaportas, P., Forster, J.J., & Ntzoufras, I. (2002). On Bayesian model and variable selection using MCMC. *Statistics and Computing*, *12*, 27–36.

Dong, J., & Simonoff, J.S. (1995). A geometric combination estimator for d-dimensional ordinal sparse contingency tables. *The Annals of Statistics*, *23*, 1143–1159.

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

Galindo-Garre, F., & Vermunt, J.K. (2004). The order-restricted association model: two estimation algorithms and issues in testing. *Psychometrika*, *69*, 641–654.

George, E.I., & McCulloch, R.E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, *88*, 881–889.

Goodman, L.A. (1979). Simple models for the analysis of association in cross-classifications having ordered categories. *Journal of the American Statistical Association*, *74*, 537–552.

Goodman, L.A. (1981). Association models and canonical correlation in the analysis of cross-classifications having ordered categories. *Journal of the American Statistical Association*, *76*, 320–334.

Goodman, L.A. (1986). Some useful extensions of the usual correspondence analysis approach and the usual log-linear models approach in the analysis of contingency tables (with a discussion and reply by the author). *International Statistical Review*, *54*, 243–309.

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

Hoeting, J.A., Madigan, D., Raftery, A.E., & Volinski, C.T. (1999). Bayesian model averaging: a tutorial. *Statistical Science*, *14*, 382–417.

Iliopoulos, G., Kateri, M., & Ntzoufras, I. (2007). Bayesian estimation of unrestricted and order-restricted association models for a two-way contingency table. *Computational Statistics and Data Analysis*, *51*, 4643–4655.

Iliopoulou, K. (2004). Schizotypy and consumer behavior (in Greek). M.Sc. Thesis, Department of Business Administration. University of the Aegean, Chios, Greece, http://stat-athens.aueb.gr/~jbn/courses/diplomatikes/business/Iliopoulou(2004).pdf.

Kass, R.E., & Raftery, A.E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*, 773–795.

Kateri, M., Nicolaou, A., & Ntzoufras, I. (2005). Bayesian inference for the RC(m) association model. *Journal of Computational and Graphical Statistics*, *14*, 116–138.

Kuo, L., & Mallick, B. (1998). Variable selection for regression models. *Sankhyã B*, *60*, 65–81.

Lee, R.D., & Carter, L.R. (1992). Modeling and forecasting US mortality. *Journal of the American Statistical Association*, *87*, 659–671.

Maxwell, A.E. (1961). *Analyzing Qualitative Data*. London: Methuen.

Maydeu-Olivares, A., & Joe, H. (2006). Limited information goodness-of-fit testing in multidimensional contingency tables. *Psychometrika*, *71*, 713–732.

Raftery, A.E. (1995). Bayesian model selection in social research. *Sociological Methodology*, *25*, 111–163.

Richardson, S., & Green, P.J. (1997). On Bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society B*, *59*, 731–792.

Ritov, Y., & Gilula, Z. (1991). The order-restricted RC model for ordered contingency tables: estimation and testing of fit. *The Annals of Statistics*, *19*, 2090–2101.

Ritov, Y., & Gilula, Z. (1993). Analysis of contingency tables by correspondence models subject to order constraints. *Journal of the American Statistical Association*, *88*, 1380–1387.

Simonoff, J.S. (1983). A penalty function approach to smoothing large sparse contingency tables. *The Annals of Statistics*, *11*, 208–218.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society B*, *64*, 583–639.

Tarantola, C., Consonni, G., & Dellaportas, P. (2008). Bayesian clustering of row effects models. *Journal of Statistical Planning and Inference*, *138*, 2223–2235.

Manuscript Received: 21 JUL 2007
Final Version Received: 20 FEB 2009