Bayesian Model Comparison for the Order Restricted RC Association Model

G. Iliopoulos, M. Kateri & I. Ntzoufras

(e-mails: geh@unipi.gr, mkateri@unipi.gr, ntzoufras@aueb.gr)

Dep. of Statistics & Insurance Science, University of Piraeus, Piraeus, GREECE

Department of Statistics, Athens University of Economics & Business, Athens, GREECE

First Draft: 09/07/2007

ABSTRACT

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 parametrization and we infer for the equality of subsequent scores using the Bayesian approach. In order 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 illustrated using two 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, 1985) play a predominant role in the relevant literature, 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 fixed, prespecified 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 additional to 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) effect 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 not linear (but multiplicative) in its parameters while the corresponding component in U, R and C models are linear 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 natural to expect the scores for an ordinal classification variable, to be monotonic. Estimation procedures subject to order constraints for the parametric scores have been developed initially for the R (or C) model by Agresti et al. (1987) and for the RC model by Ritov and Gilula (1991). Recently, alternative estimation methods have been proposed and compared by Galindo et al. (2002) and Galindo and Vermunt (2004, 2005).

In the Bayesian framework, the estimation of the RC association model has been considered by Chuang (1982) and Evans et al. (1993). The Bayesian analysis of the RC model can also be achieved through the procedure proposed in Kateri et al. (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 et al.(2006) used methodology adopted from product partition models to infer on the clustering of scores in the row effect models. None of these procedures considered order constraints for the parametric scores. Galindo and Vermunt (2005) facilitated posterior p-values to compare alternative association models. A first attempt for full Bayesian inference concerning the order constrained association models has been provided by Iliopoulos et al. (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. Nevertheless, this approach can not be considered as a formal Bayesian evaluation in favor or against merging specific scores.

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. Variations of the corresponding algorithm can be used for the simpler R and C models.

The RC model is presented in Section 2 under a new parametrization convenient for the needs of our procedure. This new parameterization exhibits also interpretational advantages, which are highlighted. Section 3 deals with models' Bayesian 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. Two illustrative examples are presented in Section 5 while the final section summarizes results and discusses related issues.

2 Modelling Details

Let $y = (y_{ij})$, $i = 1, \ldots, I$ and $j = 1, \ldots, J$, be a frequency $I \times J$ contingency table, produced by the cross-classification of two *ordinal* variables X (rows) and Y (columns). We assume

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

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

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

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 $\mu = (\mu_1, \mu_2, \ldots \mu_I)$ and $\nu = (\nu_1, \nu_2, \ldots \nu_J)$ 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, 1985) is achieved

$$\log \pi_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + \phi\mu_i\nu_j, \quad i = 1, \ldots, I,\ j = 1, \ldots, J. \tag{2.2}$$

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. The physical interpretation of the parameter ϕ is straightforward. It reflects the odds ratio of successive categories with score distances equal to one, since

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

In order to ensure identifiability of the parameters, certain constraints have to be imposed on the above model parameters. 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) constraints

$$\sum_{i=1}^{I}\mu_i = \sum_{j=1}^{J}\nu_j = 0 \quad \text{and} \quad \sum_{i=1}^{I}\mu_i^2 = \sum_{j=1}^{J}\nu_j^2 = 1. \tag{2.3}$$

In the context of Bayesian analysis of association models, the SSTO constraints complicate the structure of the posterior distribution, compared to the usual generalized linear models case.

This paper is restricted to the case of ordinal X and Y and consequently focus on the order restricted version of the RC association model. Thus, we assume an ordinal structure for the corresponding scores:

$$\mu_1 \leqslant \mu_2 \leqslant \cdots \leqslant \mu_I \quad \text{and} \quad \nu_1 \leqslant \nu_2 \leqslant \cdots \leqslant \nu_J \tag{2.4}$$

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 a new parameterization of the scores (subject to a different set of constraints). This parameterization is suitable for building a reversible jump MCMC algorithm (RJMCMC) for assessing the equality of successive row and column scores and simplifies further the MCMC scheme as well.

2.1 The proposed parametrization

In order to avoid the complicated structure of the posterior, we propose, instead the standard (2.3) constraints, to fix the minimum and maximum score for each classification variable. Thus, for pre-specified constants $\mu_{\min}$, $\mu_{\max}$, $\nu_{\min}$ and $\nu_{\max}$, with $\mu_{\min} < \mu_{\max}$ and $\nu_{\min} < \nu_{\max}$, we set

$$\mu_1 = \mu_{\min},\ \mu_I = \mu_{\max}, \qquad \nu_1 = \nu_{\min} \ \text{ and } \ \nu_J = \nu_{\max}. \tag{2.5}$$

Use of 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 $[0, 1]$.

The imposed constraints (2.5) on the extreme categories arise in a more natural way than the standard ones in (2.3) for the order resticted version of the model. For instance, the zero-one interval induces a proportion-like interpretation to the scores. Moreover, under this parametrization, the association parameter ϕ also has a straightforward interpretation 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.

3 Bayesian model formulation

In order to accommodate uncertainty concerning the equality of succesive scores, the model is further extented 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 $\gamma = (1, \gamma_2, \ldots, \gamma_I)$ and $\delta = (1, \delta_2, \ldots, \delta_J)$ 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. The above approach are analogous to the binary indicators used in Bayesian variable selection techniques (see George and McCulloch, 1993, Kuo and Mallick, 1998, Dellaportas et al., 2002).

Let us further define $\varGamma_i = \sum_{k=1}^{i} \gamma_k$, $i = 1, \ldots, I$, and $\varDelta_j = \sum_{k=1}^{j} \delta_k$, $j = 1, \ldots, J$ and denote by μ_γ the $\varGamma_I$-dimensional vector corresponding to the distinct μ-scores. Similarly we denote by ν_δ the $\varDelta_J$-dimensional vector corresponding to the distinct ν-scores. In order to avoid multiple indices, we denote by $\mu_\gamma(i)$ and $\nu_\delta(j)$ the i-th and j-th components of μ_γ and ν_δ respectively. In this setting, the original scores can be obtained as

$$\mu_i = \mu_\gamma(\Gamma_i) \quad \text{and} \quad \nu_j = \nu_\delta(\Delta_j). \tag{3.6}$$

Note that always $\mu_\gamma(1) = \mu_{\min}$, $\mu_\gamma(\Gamma_I) = \mu_{\max}$ ($\nu_\delta(1) = \nu_{\min}$, $\nu_\delta(\Delta_J) = \nu_{\max}$). Thus, there exist $\Gamma_I - 2$ and $\Delta_J - 2$ free parameters respectively.

3.1 Prior distributions

It is more convenient to set a prior distribution to the *positive* score differences rather than on the scores themselfs. When no prior information is available, a convenient choice for the prior is the Dirichlet distribution $\mathcal{D}(\underbrace{1,\dots,1}_{\Gamma_I-2})$ for the row scores and $\mathcal{D}(\underbrace{1,\dots,1}_{\Delta_J-2})$ for the column scores when they are rescaled to the $[0,1]$ interval. Equivalently, the (distinct) scores are a priori distributed as ordered iid uniform random variables with joint density

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

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

Normal prior distributions of vague variances are assigned to the main effects for rows and columns. 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 and 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.

3.2 Posterior Inference and Model Comparison

Since interest lies on the interaction terms of our model, focus will be given on the posterior distributions of the row and column scores as well as on the intrinsic association parameter ϕ. Moreover, since the latent indicators γ, δ specify the interaction and the model structure, the posterior distribution of $f(\gamma, \delta|y)$ and the marginal distributions $f(\gamma_i|y)$ and $f(\delta_j|y)$ for $i = 1, \dots, I$, $j = 1, \dots, J$ are also of special interest. The above posterior probabilities will identify the best models and indicate ties among the scores. Further on, the posterior distribution of the parameters conditional on a specific model structure will be also considered. Hence, we will focus on $f(\phi|\gamma,\delta)$, $f(\mu|\gamma,\delta)$ and $f(\nu|\gamma,\delta)$ as well, for those (γ,δ)'s having the highest posterior probabilities.

The above quantities will be estimated using reversible jump Markov chain Monte Carlo techniques introduced by Green (1995). Details concerning the implementation of this algorithm follow in the next section.

4 RJMCMC for row and column score merging

We proceed by describing the general MCMC algorithm and details about the matching functions and the proposal distributions are povided. In what follows, notation is similar to the one used in variable selection (see Dellaportas et al. 2002).

4.1 The MCMC Algorithm

The RJMCMC scheme can be summarized by the following steps

1. Update model structure by sampling γ and δ using successive RJMCMC moves for each component:

   - For $i = 2,\dots,I$, propose γ' such that $\gamma'_i = 1-\gamma_i$, $\gamma'_k = \gamma_k$ for $k \neq i$.
     - **Splitting two scores ($\gamma_i = 0 \rightarrow \gamma'_i = 1$)**: if $\gamma_i = 0$ then we propose to split μ_{i-1} and μ_i scores (i.e. $\mu'_{i-1} < \mu'_i$).
       (a) Generate a scalar u from a specified proposal density $q(u|\mu,\gamma,\gamma')$ which is used to equalize the dimensions of the compared models (see Green, 1995 for details).
       (b) Set $\mu'_{\gamma'} = g(\mu_\gamma, u)$, where $g(\mu_\gamma, u)$ can be any invertible function that matches the two models.
       (c) Obtain μ' from $\mu'_{\gamma'}$ using (3.6).
     - **Merging two scores ($\gamma_i = 1 \rightarrow \gamma'_i = 0$)**: if $\gamma_i = 1$ then we propose to move from $\mu_{i-1} < \mu_i$ to $\mu'_{i-1} = \mu'_i$.
       (a) Set $(\mu'_{\gamma'}, u) = g^{-1}(\mu_\gamma)$.
       (b) Obtain μ' from $\mu'_{\gamma'}$ using (3.6).

– The probability of acceptance of the proposed move $(\gamma, \mu) \rightarrow (\gamma', \mu')$ in each RJMCMC step equals $\alpha = \min(1, A)$, where

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

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

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

- For $j = 2, \ldots, J$, propose δ' such that $\delta'_j = 1 - \delta_j$, $\delta'_k = \delta_k$ for $k \neq j$.
  Similarly to the row scores, if $\delta_j = 1 \rightarrow \delta'_j = 0$ then we merge the j and $j-1$ successive column scores while if the move $\delta_j = 0 \rightarrow \delta'_j = 1$ is proposed then we split two successive column scores (i.e., set $\nu'_{j-1} < \nu'_j$). The sampling scheme will be analogous to the one described above for the row scores.

2. Update the model parameters $(\lambda^X, \lambda^Y, \phi, \mu, \nu)$, given the model structure (γ, δ), using the following steps:

   - Sample row and column effects (as in Iliopoulos et al. 2007).
   - Sample ϕ using a simple random walk Metropolis.
   - Use random walk for the logits of column and row scores' differences.

More details concerning the matching function and the proposal function follow. Technicalities concerning the domain of the proposed parameter and the Jacobian are given in Appendices A and B. A more detailed description of the split and merge moves is provided in Appendix C.

4.2 Specification of the Matching Function.

The split and merge moves are directly defined by a three-armed function $g()$, each arm corresponding to a different type of move. 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; Iliopoulos et al., 2005).

Direct implementation of the above matching function on the first and last category scores leads to violation of the model constraints (2.5). 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.

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 μ_γ. On the other hand, in merge, we define as central scores the third to the last but one components of μ_γ. Hence, central scores are parameters $\mu_\gamma(\ell)$ such that $s \leq \ell = \Gamma_i \leq \Gamma_I - 1$; 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 such that $s \leq \Gamma_i \leq \Gamma_I - 1$.

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

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

The above merge move implies the following split move: when splitting $\mu_\gamma(i)$, the new vector of row scores $\mu'_{\gamma'}$ of dimension $\Gamma_I + 1$, is derived as

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

From (4.2) it follows that the pseudo-parameter u, which is used to generate the additional parameters in the split move, is given by

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

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 $\mu_\gamma(1)$ and $\mu_\gamma(2)$ or splitting the lowest score $\mu_\gamma(1)$. Imposing the above "central" scores transformation when merging the lowest two scores will result to $\mu'_{\gamma'}(1) = \frac{1}{2}\{\mu_{\min} + \mu_\gamma(2)\} \neq \mu_{\min}$. In order to obtain a set of parameters that satisfy (2.5) we rescale the above parameters in the $[0, 1]$ interval and then rescale them again to the desired $[\mu_{\min}, \mu_{\max}]$ interval. So, we subtract from all parameters the quantity $c = \frac{1}{2}\{\mu_{\min} + \mu_\gamma(2)\}$, multiply them by $(\mu_{\max} - \mu_{\min})/(\mu_{\max} - c)$ and finally add $\mu_{\min}$.

The above procedure results to the matching function

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

The above move is proposed for scores μ_i with $\Gamma_i = 2$. Moreover, for the additional pseudo-parameter u we set $u = \mu_\gamma(2)$ (see below in the split move for details).

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

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

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, $\mu_\gamma(\mathit{\Gamma}_I - 1)$ and $\mu_\gamma(\mathit{\Gamma}_I)$ or split the last one, $\mu_\gamma(\mathit{\Gamma}_I)$. Following similar arguments as above, when merging, the new highest score is firstly set equal to $\{\mu_\gamma(\mathit{\Gamma}_I - 1) + \mu_{\max}\}/2$ and then rescaled appropriately. All other scores are modified accordingly. Hence the new vector $\mu'_{\gamma'}$ of dimension $\mathit{\Gamma}'_I = \mathit{\Gamma}_I - 1$ is given by

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

For the additional parameter u, we set $u = \mu_{\max} - \mu_\gamma(\mathit{\Gamma}_I - 1)$.

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

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

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 Jacobian for the split moves is given by

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

while for the merge moves by

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

Details concerning their computation are provided in Appendix A.

4.3 Proposal Distribution of 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 an illustration, we refer to our illustrative examples, 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 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 1.

	Disturbance (from low to high)				
Age Group	1	2	3	4	Total
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

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

In all 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. Thus, they concluded to the correlation model (CA$_{RG}$) 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, posterior probabilities of the models under consideration are calculated and summarized in Table 2. The marginal posterior probabilities for the γ and δ indicators, given in Table 3, 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 2, 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 $\nu_1 < \nu_2 = \nu_3 = \nu_4$. This model is more parsimonious than the ones suggested by athors previously analyzed the same data.

The estimated posterior probability for the second highest model equals 0.154 and is very close to the first one. The corresponding posterior model odds between the two highest probability models is just 1.05 which suggests no clear separation between them. The second highest probability model is the one identified and proposed by Iliopoulos et al. (2007) as well as by Ritov and Gilula (1993). Further on, Table 2 indicates that eight models were found with posterior probability higher than 5%. All these models, 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 $f(\gamma_i|y)$ and $f(\delta_j|y)$ given in Table 3 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 are higher than 0.95 . Similarly, posterior model probability $f(\delta_2|y) > 0.99$ suggests that the first and the second categories of 'dreams disturbance' are different. For the rest of the row and column scores, we observe an increased 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 4-5 (highest levels of dreams' disturbance) posterior probabilities indicate mild evidence in favor of their equality with values 0.606 and 0.516 respectively.

By analyzing the posterior distribution over all visited models (Bayesian model averaging, BMA,

k	**Model (score structure)**	**Posterior Probability**	PO_{1k}	Mean	St.Dev.	Percentiles (2.5%, 97.5%)	Odds Ratio ($e^{\text{mean of } \phi}$)
				Posterior Summaries of ϕ			
1	$\mu_1 = \mu_2 < \mu_3 = \mu_4 < \mu_5$ $\nu_1 < \nu_2 = \nu_3 = \nu_4$	0.1620	1.00	-2.06	0.44	(-2.94, -1.21)	0.127
2	$\mu_1 = \mu_2 < \mu_3 = \mu_4 < \mu_5$ $\nu_1 < \nu_2 = \nu_3 < \nu_4$	0.1540	1.05	-2.55	0.61	(-3.87, -1.46)	0.078
3	$\mu_1 = \mu_2 < \mu_3 < \mu_4 < \mu_5$ $\nu_1 < \nu_2 = \nu_3 = \nu_4$	0.0877	1.85	-1.98	0.45	(-2.86, -1.12)	0.138
4	$\mu_1 = \mu_2 < \mu_3 < \mu_4 < \mu_5$ $\nu_1 < \nu_2 = \nu_3 < \nu_4$	0.0725	2.23	-2.42	0.60	(-3.73, -1.33)	0.089
5	$\mu_1 = \mu_2 < \mu_3 = \mu_4 < \mu_5$ $\nu_1 < \nu_2 < \nu_3 < \nu_4$	0.0609	2.66	-2.52	0.61	(-3.85, -1.44)	0.081
6	$\mu_1 = \mu_2 < \mu_3 = \mu_4 < \mu_5$ $\nu_1 < \nu_2 < \nu_3 = \nu_4$	0.0579	2.80	-2.16	0.48	(-3.12, -1.25)	0.115
7	$\mu_1 < \mu_2 < \mu_3 = \mu_4 < \mu_5$ $\nu_1 < \nu_2 = \nu_3 < \nu_4$	0.0541	2.99	-2.71	0.66	(-4.15, -1.52)	0.067
8	$\mu_1 < \mu_2 < \mu_3 = \mu_4 < \mu_5$ $\nu_1 < \nu_2 = \nu_3 = \nu_4$	0.0522	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

Table 2: 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.

Row Scores	Posterior Probability	Column Scores	Posterior Probability
$f(\gamma_2 = 1\|y) =$	0.285	$f(\delta_2 = 1\|y) =$	0.996
$f(\gamma_3 = 1\|y) =$	0.940	$f(\delta_3 = 1\|y) =$	0.286
$f(\gamma_4 = 1\|y) =$	0.391	$f(\delta_4 = 1\|y) =$	0.484
$f(\gamma_5 = 1\|y) =$	0.964		

Table 3: 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.

estimate) provided in Table 2, 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 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 an 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. Plot of the posterior distribution of ϕ for the best eight models (given in Table 2) and the corresponding weighted distribution over these models is provided in Figure 1.

Concerning the row and column scores, graphical representation of their posterior distributions is provided in Figure 2. The posterior summaries of the related estimated odds ratios can be extracted directly by the 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 $i \geq 2$, $j \geq 2$ to the baseline cell $(i = 1, j = 1)$. For the highest probability model, the

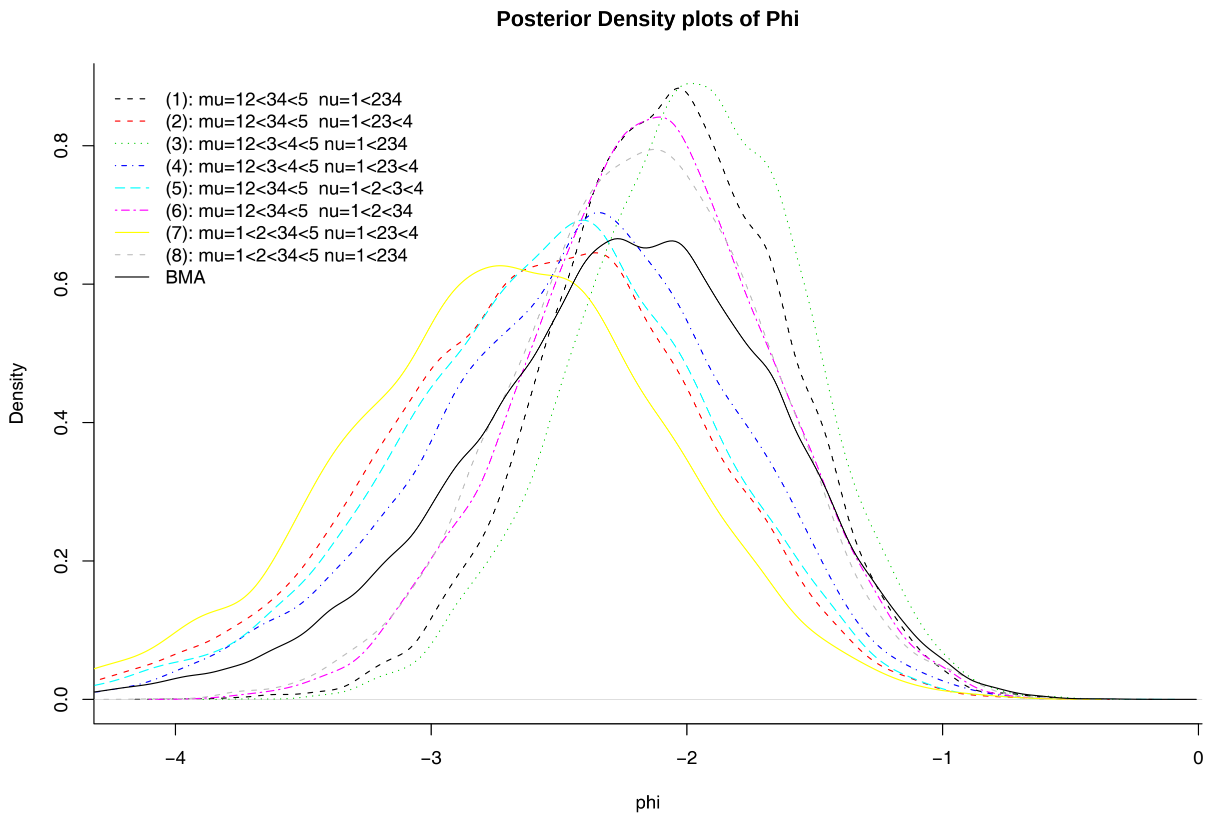

Figure 1: Posterior Distributions of ϕ over models with highest posterior probabilities for Example 1 (dreams' distrurbance data).

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 4. For the two highest probability proposed models, 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.

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	2	1.00	2	1.00	1.00
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	0.10

Table 4: Estimated posterior odds ratios (posterior geometric means) using model averaging for the two highest probability models for example 1 (dream's disturbance data).

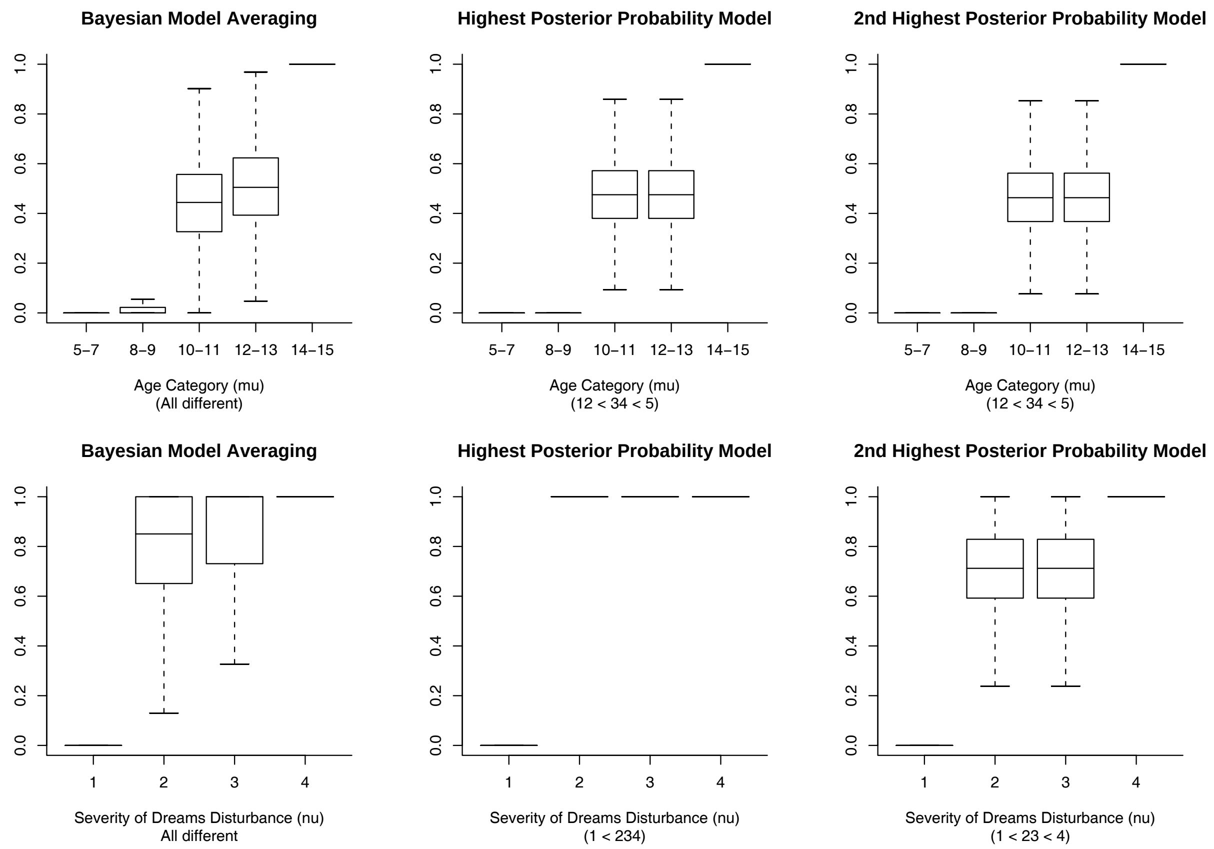

Figure 2: Boxplots of the posterior distribution of the row and column scores for model averaging and the two a-posteriori most probable models (dream's disturbance data).

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

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

Social Anxiety Score	i	Odd Behavior Score 0 $j=1$	1 $j=2$	2 $j=3$	3 $j=4$	4 $j=5$	5–7 $j=6$	Total
0	1	11	5	1	0	1	0	18
1	2	13	8	8	2	2	3	36
2	3	8	9	4	1	4	0	26
3	4	6	7	5	4	4	1	27
4	5	6	9	5	3	2	4	29
5	6	3	13	5	4	1	5	31
6–8	7	0	11	5	10	3	6	35
Total		47	62	33	24	17	19	202

Table 5: Cross–classification of 202 Students by Social Anxiety and Odd Behavior sub-scales (Example 2: Schizotypy data).

The cross-classification of 202 students of the survey according to 'social anxiety' and 'odd behavior' is given in Table 5. 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.

Note that a 'schizotype' suffers from minor episodes of pseudoneurotic problems. In general, the prevalence rate of schizotypy is about 10% in the general population. The importance of schizotypal personality in psychiatric research is prominent for two reasons: Firstly, shizotypal subjects have increased risk to develop schizophrenia during their life. Secondly, since they are healthy persons, they can participate in psychiatric/psychological research studies (by completing questionnaires - psychometric instruments) in which schizophrenia cases are unable to do. Although several scales have been proposed in the literature, Raine's (1991) SPQ scale, a 74-item self-administered questionnaire, is the most popular questionnaire used to measure the concepts of schizotypal personality. The questionnaire consists of binary zero-one (yes-no) items. It provides subscales for nine schizotypal features as well as an overall scale for schizotypy. Each subscale is calculated as the sum of the questionnaire items that refer to each schizotypal subscale.

Since both classification variables of Table 5 are ordinal, we shall apply the order restricted RC model. Details for the most probable table are summarized in Table 6. Note that, for this example, traditional asymptotic inference cannot be applied due to the small cell entries of the resulted contingency table.

k	Model (score structure)	Posterior Probability (%)	PO_{1k}
1	$\mu_1 < \mu_2 = \mu_3 < \mu_4 = \mu_5 < \mu_6 < \mu_7$ $\nu_1 < \nu_2 = \nu_3 < \nu_4 = \nu_5 = \nu_6$	3.011	1.00
2	$\mu_1 < \mu_2 < \mu_3 < \mu_4 = \mu_5 < \mu_6 < \mu_7$ $\nu_1 < \nu_2 = \nu_3 < \nu_4 = \nu_5 = \nu_6$	1.937	1.55
3	$\mu_1 < \mu_2 = \mu_3 < \mu_4 = \mu_5 = \mu_6 < \mu_7$ $\nu_1 < \nu_2 = \nu_3 < \nu_4 = \nu_5 = \nu_6$	1.914	1.57
4	$\mu_1 < \mu_2 = \mu_3 < \mu_4 = \mu_5 < \mu_6 < \mu_7$ $\nu_1 < \nu_2 = \nu_3 < \nu_4 = \nu_5 < \nu_6$	1.688	1.78
5	$\mu_1 < \mu_2 < \mu_3 = \mu_4 = \mu_5 < \mu_6 < \mu_7$ $\nu_1 < \nu_2 = \nu_3 < \nu_4 = \nu_5 = \nu_6$	1.684	1.79
6	$\mu_1 < \mu_2 = \mu_3 < \mu_4 < \mu_5 < \mu_6 < \mu_7$ $\nu_1 < \nu_2 = \nu_3 < \nu_4 = \nu_5 = \nu_6$	1.614	1.87

Table 6: 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).

According to the information provided in Table 6, the most probable model (3%) is the one

Row Scores	Posterior Probability	Column Scores	Posterior Probability
$f(\gamma_2 = 1\|y) =$	0.8358	$f(\delta_2 = 1\|y) =$	0.9997
$f(\gamma_3 = 1\|y) =$	0.4734	$f(\delta_3 = 1\|y) =$	0.2746
$f(\gamma_4 = 1\|y) =$	0.6129	$f(\delta_4 = 1\|y) =$	0.6921
$f(\gamma_5 = 1\|y) =$	0.4878	$f(\delta_5 = 1\|y) =$	0.1653
$f(\gamma_6 = 1\|y) =$	0.6750	$f(\delta_6 = 1\|y) =$	0.4432
$f(\gamma_7 = 1\|y) =$	0.8930		

Table 7: Estimated Posterior Probabilities for γ and δ 'Split' Indicators for Example 2 (schizotypy data); single RJMCMC Using R (500,000 iterations and additional burn-in of 10,000 iterations).

with just three distinguished values for 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 $\nu_2 = \nu_3$, $\nu_4 = \nu_5 = \nu_6$ for the column scores, with the analoguous interpretation. The posterior odds ratios, presented in Table 8, are increasing with row and column levels, indicating an underlying positive association between the two clasiffication variables. Moreover, values of last row and column scores are much higher than the rest ones indicating that the association between the two sub-scales is stronger at the higher levels of the variables under consideration.

	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

Table 8: Estimated posterior odds ratios (posterior geometric means) for Example 2 (schizotypy data), using model averaging and the highest probability model .

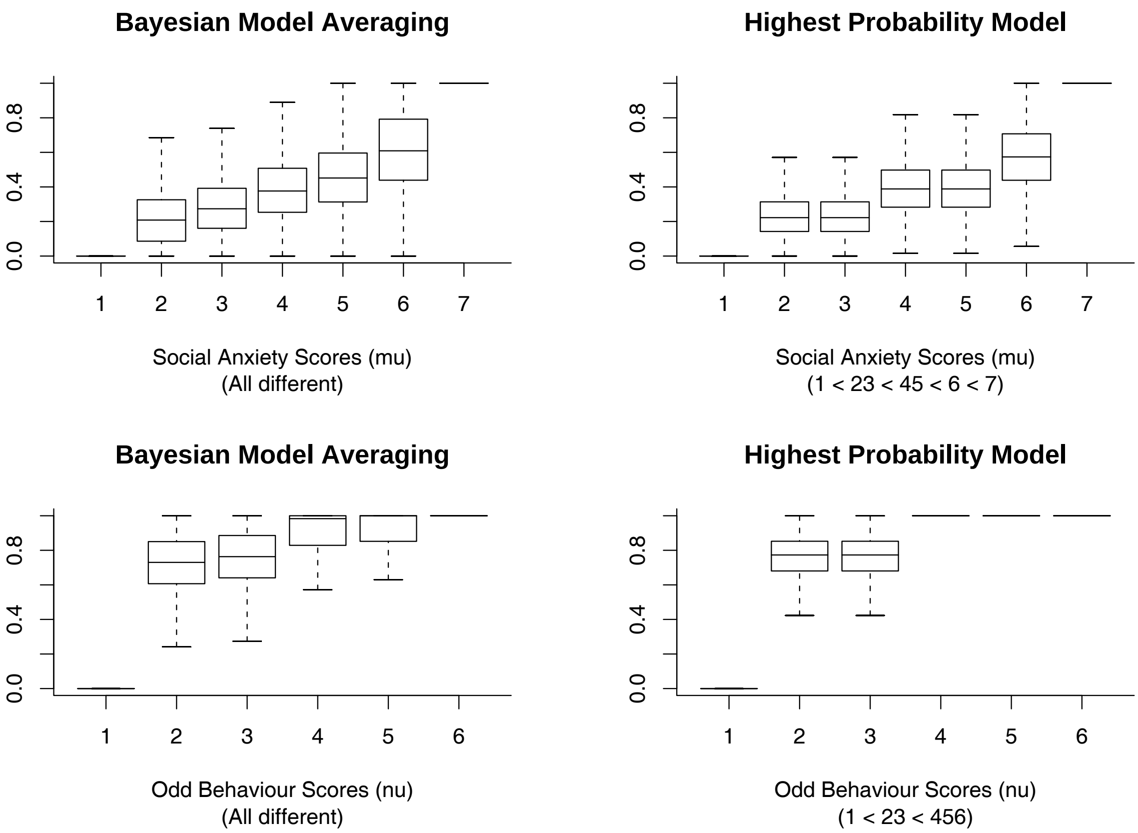

Figure 3: Boxplots of the posteriors of the row and column scores for model averaging and the a-posteriori most probable model for Example 2 (schizotypy data).

5.3 Computational Details

In order to implement the proposed algorithms the authors have developed `R` functions and stand alone Fortran software. These two versions were used for crosschecking of results. Software is available by the authors upon request.

Results illustrated in this paper have been generated using `R` functions. 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, in a Pentium duo cure 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 for 100 thousand iteration was approximately equal to 18 minutes in the same machine.

Both examples were also run extensively using the corresponding Fortran program which was considerably faster. For both cases, samples of 500 thousand iteration were generated and compared with the corresponding `R` results to ensure convergence and the lack of programming bugs.

The algorithm was highly mobile in both examples. In example 1 the algorithm visited 69, 86 and all 105 models in 10, 100 iterations 400 thousand iterations respectively. In example 2 the algorithm visited 358, 730 and 894 models in 1, 10 and 100 thousand iterations respectively (out of 1953 models).

6 Discussion and Further Research

In this paper we dealt with the problem of the 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). In order to achieve that, we propose an alternative parametrization which is convenient in terms of interpretation and computation. We used order-restricted uniform prior for the model scores and constructed a flexible RJMCMC algorithm to explore the model space. 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), Bayesian or standard versions of AIC and BIC (Spiegelhalter et al. 2002, Brooks, 2002) or other Bayesian measures as the posterior p-values used by Galindo and Vermunt (2005) and the ones proposed for association models (in a more general context) by Kateri et al. (2005). For a relevant illustration in RC models, 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 U, R and C models. 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 et al. (2000) based on imaginary data which will express our prior beliefs. By this way, prior distributions that are compatible across models can be constructed in a straightforward manner.

Finally, future research on the area can be focused on the extension of such types of association models (with identification of sets of classification variables' categories with equal scores) to high-dimensional contingency tables.

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 $\mu_\gamma(\ell)$, we have

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

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

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

where I_k is the identity matrix of order k, $0_{k\times l}$ is the $k \times l$ matrix of zeros and $k_1 = \mathit{\Gamma}_\ell - 2$, $k_2 = \mathit{\Gamma}_I - \mathit{\Gamma}_\ell - 1$.

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

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

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

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

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

where $\frac{\partial \mu'_{\gamma'}}{\partial u}$ 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, $\mathit{\Gamma}_I$ by $\mu_\gamma(2)$, $\mathit{\Gamma}_I - 1$ respectively. Therefore, in this case the absolute value of the Jacobian determinant is

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

Similarly, for splitting the highest scores we obtain

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

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

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

whilst for the merge move of the highest scores it becomes

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

B. Domains of the Proposal Parameter u

In order 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 $\mu_{\gamma}(\ell)$ into the new scores $\mu'_{\gamma'}(\ell)$ and $\mu'_{\gamma'}(\ell+1)$, the proposed scores satisfy

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

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

$$
\mu_{\gamma}(\ell - 1) < \mu_{\gamma}(\ell) - u < \mu_{\gamma}(\ell) + u < \mu_{\gamma}(\ell + 1).
$$

It follows that

$$
L = 0 < u < \min\left\{\mu_{\gamma}(\ell) - \mu_{\gamma}(\ell - 1), \mu_{\gamma}(\ell + 1) - \mu_{\gamma}(\ell)\right\} = U.
$$

Analogously, for the lowest scores we have the constraint

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

resulting in

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

Finally, for the highest scores we have that

$$\mu_\gamma(\mathit{\Gamma}_I - 1) - \frac{\{\mu_\gamma(\mathit{\Gamma}_I - 1) - \mu_{\min}\}u}{2R_\mu} < \mu_{\max} - u < \mu_{\max}$$

leading to

$$L = 0 < u < \frac{2R_\mu\{\mu_{\max} - \mu_\gamma(\mathit{\Gamma}_I - 1)\}}{R_\mu + \mu_{\max} - \mu_\gamma(\mathit{\Gamma}_I - 1)} = U.$$

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 μ_γ by $\mu'_{\gamma'}$; see Table 9 for the corresponding expressions.

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

Table 9: Limits of the pseudo parameter u for the inverse merge move ($R_\mu = \mu_{\max} - \mu_{\min}$).

C. Summarizing the Split and Merge Moves

From the above we summarize the RJMCMC details, for $i = 2, \ldots, \Gamma_I$ 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

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

where μ' is the proposed value of μ.

1. If $\gamma_i = 0 \rightarrow \gamma'_i = 1$ (split move) then:

   (a) For $\mathit{\Gamma}_i = 1$ (lowest scores) we proceed with the following steps:

      i. Generate u from $\mathcal{U}\left(\mu_{\min}, \mu_\gamma(2) + \dfrac{(\mu_\gamma(2) - \mu_{\min})(\mu_{\max} - \mu_\gamma(2))}{R_\mu + \mu_\gamma(2) - \mu_{\min}}\right)$.

ii. Calculate $\mu'_{\gamma'}$ by (4.4).

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

$$A = LR \times \frac{2(\mathit{\Gamma}_I - 1)(\mu_\gamma(2) - \mu_{\min})}{R_\mu + \mu_\gamma(2) - \mu_{\min}} \times \left(1 - \frac{u - \mu_{\min}}{2R_\mu}\right)^{\mathit{\Gamma}_I - 2}.$$

(b) For $2 \leqslant \mathit{\Gamma}_i \leqslant \mathit{\Gamma}_I - 1$ (central scores):

i. Set $\ell = \mathit{\Gamma}_i$.

ii. Generate u from $\mathcal{U}(0, \max\{\mu_\gamma(\ell) - \mu_\gamma(\ell - 1), \mu_\gamma(\ell + 1) - \mu_\gamma(\ell)\})$.

iii. Calculate $\mu'_{\gamma'}$ by (4.2).

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

$$A = LR \times \frac{\mathit{\Gamma}_I - 1}{R_\mu} \times \max\{\mu_\gamma(\ell) - \mu_\gamma(\ell - 1), \mu_\gamma(\ell + 1) - \mu_\gamma(\ell)\} \times 2.$$

(c) For $\mathit{\Gamma}_i = \mathit{\Gamma}_I$ (highest scores):

i. Generate u from $\mathcal{U}\left(0, \dfrac{2R_\mu\{\mu_{\max} - \mu_\gamma(\mathit{\Gamma}_I - 1)\}}{R_\mu + \mu_{\max} - \mu_\gamma(\mathit{\Gamma}_I - 1)}\right)$.

ii. Calculate $\mu'_{\gamma'}$ by (4.6) .

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

$$A = LR \times \frac{2(\mathit{\Gamma}_I - 1)\{\mu_{\max} - \mu_\gamma(\mathit{\Gamma}_I - 1)\}}{R_\mu + \mu_{\max} - \mu_\gamma(\mathit{\Gamma}_I - 1)} \times \left(1 - \frac{u}{2R_\mu}\right)^{\mathit{\Gamma}_I - 2}.$$

2. If $\gamma_i = 1 \to \gamma'_i = 0$ (merge move) then:

(a) For $\mathit{\Gamma}_i = 2$ (lowest scores) we proceed with the following steps:

i. Set $u = \mu_\gamma(2)$.

ii. Calculate $\mu'_{\gamma'}$ by (4.3).

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

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

(b) For $3 \leqslant \mathit{\Gamma}_i \leqslant \mathit{\Gamma}_I - 1$ (central scores):

i. Set $\ell = \mathit{\Gamma}_i$.

ii. Set $u = \{\mu_\gamma(\ell) - \mu_\gamma(\ell - 1)\}/2$.

iii. Calculate $\mu'_{\gamma'}$ by (4.2).

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

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

(c) For $\Gamma_i = \Gamma_I$ (highest scores):

i. Set $u = \mu_{\max} - \mu_\gamma(\Gamma_I - 1)$.

ii. Calculate $\mu'_{\gamma'}$ by (4.6).

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

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

References

Agresti, A., Chuang, C. and 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. revised). Washington, DC: APA.

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.

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.

Dellaportas, P., Forster, J.J. and Ntzoufras, I. (2002). On Bayesian Model and Variable Selection Using MCMC. *Statistics and Computing*, **12**, 27–36.

Evans, M., Gilula, Z. and 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, F., and Vermunt, J.K. (2004). The order-restricted association model: Two estimation algorithms and issues in testing. *Psychometrika*, **69**, 641–654.

Galindo, F., and Vermunt, J.K. (2005). Testing log-linear models with inequality constraints: a comparison of asymptotic, bootstrap, and posterior predictive p-values. *Statistica Neerlandica*, **59**, 82-94

Galindo, F., J. K. Vermunt and M. A. Croon (2002). Likelihood-ratio tests for order restricted log-linear models: a comparison of asymptotic and bootstrap methods. *Metodología de las Ciencias del Comportamiento*, **4**, 325-337.

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

Green, P. (1995). Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. *Biometrika*, **82**, 711–732.

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. (1985). The analysis of cross-classified data having ordered and/or unordered categories: Association models, correlation models and asymmetry models for contingency tables with or without missing entries. *Annals of Statistics*, **13**, 10–69.

Iliopoulos, G., Karlis, D. and Ntzoufras, I. (2005) Bayesian estimation in Kibble's bivariate gamma distribution. *Canadian Journal of Statistics*, **33**, 571-589 .

Iliopoulos, G., Kateri, M. and 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. and Raftery, A.E. (1995). Bayes Factors. *Journal of the American Statistical Association* **90**, 773–795.

Kateri, M., Nicolaou, A. and Ntzoufras, I. (2005). Bayesian Inference for the RC(m) Association Model. *Journal of Computational and Graphical Statistics*, **14**, 116–138.

Kuo, L. and Mallick, B. (1998). Variable Selection for Regression Models. *Sankhyã B*, **60**, 65–81.

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

Raine, A. (1991). The SPQ: A Scale for the assessment of schizotypal personality based on DSM-III-R criteria. *Schizophrenia Bulletin*, **17**, 555-564.

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

Ritov, Y. and Gilula, Z. (1993). Analysis of Contingency Tables by Correspondence Models Subject to Order Constraints. *Journal of the American Statistical Association*, **88**, 1380–1387.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and 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., and Dellaportas, P. (2006). Bayesian clustering of row effects models (submitted). *Technical Report*, University of Pavia, Italy.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and van der Linde, A. (2002). Bayesian Measures of Model Complexity and Fit (with discussion). *Journal of the Royal Statistical Society B*, **64**, 583–639.