Journal of Multivariate Analysis 121 (2013) 109–126

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Bayesian estimation of order-restricted and unrestricted association models

Haydar Demirhan

Department of Statistics, Hacettepe University, Beytepe, 06800 Ankara, Turkey

ARTICLE INFO

Article history:
Received 19 June 2012
Available online 9 July 2013

AMS subject classifications:
62F15
62H20
62J12
62P15

Keywords:
Association models
Contingency table
Exchangeability assumption
Generalized multivariate log-gamma distribution
Gibbs sampling
Order statistics
Score
Uniform association

ABSTRACT

Association models include score parameters to multiplicatively represent the hierarchy between the levels of the considered ordinal factor. If order restrictions are placed on the scores, an estimation problem becomes a non-linear and restricted estimation, which is somewhat problematic with respect to the classical approaches. In this article, we consider the Bayesian estimation of the scores and other parameters of an association model both with and without order restrictions. We propose the use of a previously introduced multivariate prior in the unrestricted case and an order statistics approach in the order-restricted case. The advantages of using these prior structures are that we are able to consider the correlation patterns arising from the hierarchy between the levels of ordinal factors, there is no violation of the exchangeability assumption, the approaches are general for any size of contingency table, and the posterior inferences are easily derived. The proposed approaches are applied to both a previously analyzed popular two-way contingency table and a three-way contingency table. Smaller standard deviations than those of previous analyses are obtained, and a new best-fitting model is identified for the two-way table.

1. Introduction

Log-linear models are used to figure out the interaction structures of contingency tables and are widely applied in social and biomedical investigations. These models are characterized by the types of variable that constitute the contingency table of interest. A contingency table can be composed of nominal variables, ordinal variables, or both. A contingency table in which some of the variables are nominal and others are ordinal is called a mixed contingency table [19]. In addition to its association structure, a mixed contingency table contains information about the hierarchy between the levels of ordinal variables. Score values corresponding to the levels of each ordinal variable represent the hierarchy between the adjacent levels of a variable. Because results and inferences are directly affected by score values, the appropriate determination or estimation of the score values is the most vital part of an analysis. The hierarchy is handled in two ways. In the first, the structure of the hierarchy is taken as fixed and is assigned by the researcher. Thus, score values are taken as constants and are determined by the researcher before the analysis is performed. In the second, the score values are perceived as parameters and are estimated from the sample by using various statistical approaches. The complexity of analysis is higher in the latter approach, but this approach provides more reliable results than the former.

Various approaches have been used for the determination of score values in the classical setting. Goodman [17] uses the rank number of the relevant level as the corresponding score value. Graubard and Korn [18] use standardized rank numbers

E-mail address: haydarde@hacettepe.edu.tr.
URL: http://yunus.hacettepe.edu.tr/~haydarde.

0047-259X/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmva.2013.06.008

as the score values if the levels of the ordinal variable of interest are equally spaced. Agresti [2] states that half of the rank numbers of observations on the level of interest can be assigned as the corresponding score value. All approaches of this type are subjective.

In some models, scores are considered to be unknown parameters, with or without the order restrictions concerning their ordinality. The scores are estimated simultaneously with the parameters of the considered log-linear model. Log-linear models constructed for the analysis of mixed contingency tables are association, Goodman's RC, or RC(m) models. The RC and RC(m) models multiplicatively include scores and parameters. In this case, the model is no longer log-linear but rather log-multiplicative, and it is possible to have a convex likelihood function. Hence, a local maximum cannot be obtained. When the scores are treated as parameters, the model becomes less parsimonious, and the number of degrees of freedom increases; as a result, less powerful tests might be obtained [2]. Beh and Farver [6] review some of the problems associated with iterative estimation procedures and compare the performance of non-iterative approaches in the estimation of the linear-by-linear association parameter of ordinal log-linear models. The presence of order restrictions increases the complexity of the estimation problem. Agresti et al. [3] and Ritov and Gilula [27] give approaches for the R or C and RC models under order restrictions, respectively. Galindo-Garre and Vermunt [15] propose and compare alternative approaches for the RC models under order restrictions. Bartolucci and Forcina [5] introduce an extended RC model under order restrictions. These approaches are based on large-sample approximations. Therefore, these approaches are problematic for small samples or sparse tables [15]. Sin and Wong [31] give some estimation approaches for association models constructed over multidimensional contingency tables without order restrictions. However, in the estimation process, the convergence problems increase as the dimensionality of the considered table increases.

Bayesian approaches provide a solution to this issue. In this approach, it is possible to treat scores as random variables, assigning the scores a prior distribution. Then, posterior estimates of scores are used to draw inferences regarding the structures of both hierarchy and association. If researchers lack prior knowledge of score values, they are estimated using the information provided by the sample using a non-informative Bayesian analysis. In the Bayesian framework, inferences are based on the direct interpretation of posterior probabilities. The exact posterior distributions of the parameters producing posterior probabilities are estimated using Markov chain Monte Carlo (MCMC) methods. Additionally, convexity is not a problem in the Bayesian setting.

The pioneers of Bayesian estimation for the RC association model are Chuang [9] and Evans et al. [13]. Chuang [9] proposes an empirical Bayesian approach for two-way tables. Evans et al. [13] give an importance sampling approach for the Bayesian estimation of the RC model. Kateri et al. [23] propose a more general Bayesian approach to estimate scores simultaneously with the parameters of an RC(K) model. The approach proposed by Kateri et al. [23] requires complex transformations, which increase the complexity of the prior distributions. Therefore, Kateri et al. [23] do not derive the full conditional posterior distributions. Instead, they use complex MCMC algorithms that do not require full conditional distributions. Iliopoulos et al. [20] propose a new parameterization for order-restricted and unrestricted RC models, stating that their approach simplifies the approach of Kateri et al. [23]. Iliopoulos et al. [20] develop various constraints for employing the MCMC methods because their model is log-multiplicative when scores are treated as random variables. Their parameterization makes it possible to run Metropolis-within-Gibbs sampling. Iliopoulos et al. [21] propose an alternative parameterization and give an approach for the model selection of the order-restricted RC association model.

The exchangeability of the levels of variables and the correlations between scores should be considered in the analysis of a mixed table. Let us begin with an ordinal variable Education (EDU), with levels High School (HS), Undergraduate (UG), and Graduate (G). The place of each level among the levels of EDU is meaningful due to the hierarchy between the levels of EDU. This issue is related to the exchangeability assumption and should be regarded when determining the prior distribution of the scores of EDU. This issue is discussed in detail in Section 3.1. An ordinal variable incorporates the true hierarchy between its levels. Suppose that the true values of scores corresponding to HS, UG, and G are 1, 2, and 3, respectively. If the value of the score of UG is increased to 3, then, due to the interrelation between scores, the values of the scores for HS and G should increase to 2 and 4, respectively, to ensure the accurate representation of the true hierarchy. This interrelation causes the existence of a correlation structure between scores. If it is ignored and the values of the scores of both UG and G become 3, the location of UG in the hierarchy overlaps that of G. In this case, the structure of the hierarchy cannot be represented accurately.

In consideration of these issues, it is necessary to develop a prior structure that addresses the exchangeability assumption and the correlation patterns between scores in the Bayesian estimation of an unrestricted association model. In the presence of order restrictions, we should be concerned with the accurate reflection of the order restrictions within the analysis. Based on these motivations, we propose the use of previously introduced prior structures for the Bayesian estimation of scores and the remaining model parameters in both the presence and absence of order restrictions. The prior structure used for the unrestricted case is based on the multivariate distribution introduced by Demirhan and Hamurkaroglu [12]. For the restricted analysis, we derive our prior structure from the joint distribution of order statistics. In addition to the general advantages of Bayesian approaches, as mentioned above, for both types of analysis, the approaches based on prior structures are general for any size of contingency table. Further, the posterior inferences are easily drawn by using the MCMC methods; the possible correlations between the model parameters and scores are taken into consideration; the elicitation of the hyper-parameters of prior distribution is easier; the assumption of exchangeability is taken into account; we are able to define simultaneously a non-informative prior for the locations of scores and a slightly (or more) informative prior for the correlations between scores; it is possible to consider the scores of some ordinal variables as order restricted and the rest as unrestricted; and our approach is suitable for an informative Bayesian analysis, if subjective prior knowledge is available. Additionally, our prior structure is both useful and beneficial for model selection, if prior variances and correlations are important.

Notation and the likelihood function are given in Section 2. Our prior elicitation strategies are given in Section 3. Posterior inferences are mentioned in Section 4. Model comparison is outlined in Section 5. An illustrative application is taken into account in Section 6, and a discussion is given in Section 7. In addition, another application of our approach to a three-way contingency table is given in the electronic Appendix.

2. Notation and likelihood function

A contingency table or a log-linear model is called a nominal, ordinal, or mixed contingency table or log-linear model if all of the variables are either nominal, ordinal, or a mixture of nominal and ordinal, respectively. Although notation is given for two-way tables in this section, the approaches are general for any size of contingency table.

In our notation, each element of a log-linear model is denoted as follows: an expected count corresponding to the t and s levels of variables of an N-way contingency table is denoted by n_{ts}; the number of variables in a contingency table is denoted by N; the general effect term is denoted by u; a main effect of level t of variable i is denoted by $u_{i(t)}$; a two-way interaction between t and s levels of i and j nominal variables, respectively, is denoted by $u_{ij(ts)}$, which can easily be generalized for higher-way interactions; a row or column effect of level t of nominal variable i is denoted by $\tau_{i(t)}$; a score corresponding to the level t of ordinal variable i is denoted by $x_{i(t)}$; and an association parameter corresponding to the variables i and j is denoted by β_{ij}, where $i, j = 1, \ldots, N$. This notation is somewhat restrictive, but it is suitable for our purpose. More general notation for the definition of nominal log-linear models is given by King and Brooks [24]. For instance, the saturated log-linear model for a three-way nominal contingency table is written as follows:

$$\log n_{tsr} = u + u_{1(t)} + u_{2(s)} + u_{3(r)} + u_{12(ts)} + u_{13(tr)} + u_{23(sr)} + u_{123(tsr)}. \tag{1}$$

For a two-way contingency table, a multiplicative row and column effects (RC) model, in which both the row scores and the column scores are parameters, is written as follows:

$$\log n_{ts} = u + u_{1(t)} + u_{2(s)} + \beta_{12} x_{1(t)} x_{2(s)}. \tag{2}$$

A row effects (R) model, which has parameter row scores and fixed column scores, is written as follows:

$$\log n_{ts} = u + u_{1(t)} + u_{2(s)} + \tau_{1(t)} x_{2(s)}. \tag{3}$$

A column effects (C) model, which has parameter column scores and fixed row scores, is written as follows:

$$\log n_{ts} = u + u_{1(t)} + u_{2(s)} + \tau_{2(s)} x_{1(t)}. \tag{4}$$

In these models, $t = 1, \ldots, I$, $s = 1, \ldots, J$, and $r = 1, \ldots, K$. If both the column and the row scores are fixed, the model given in (2) becomes a linear-by-linear association (LL) model. For identifiability purposes, the following general sum-to-zero constraints are imposed, if the related parameters are present in the model:

$$\sum_t u_{i(t)} = 0, \quad \forall i; \qquad \sum_s u_{ij(st)} = \sum_t u_{ij(st)} = 0, \quad \forall (i,j); \qquad \sum_t \tau_{i(t)} = 0, \quad \forall i. \tag{5}$$

Inclusion of the parameter β_{12} in model (2) requires the following additional constraints:

$$\sum_t w_{i(t)} x_{i(t)} = 0, \qquad \sum_t w_{i(t)} x_{i(t)}^2 = 1, \quad \forall i, \tag{6}$$

where all $w_{i(j)}$ weights can be taken as 1 or the relevant marginal sum [20]. In order to simplify the analysis, β_{12} of model (2) can be set to 1. In this case, β_{12} is estimated by using the estimates of $x_{1(t)}$ and $x_{2(s)}$ as follows [20]:

$$\beta_{12} = \left(\sum_{t=1}^{I} x_{1(t)}^2 \right)^{0.5} \left(\sum_{s=1}^{J} x_{2(s)}^2 \right)^{0.5}, \tag{7}$$

and $x_{1(t)}$ and $x_{2(s)}$ are rescaled as follows:

$$\tilde{x}_{1(t)} = x_{1(t)} \left(\sum_{t=1}^{I} x_{1(t)}^2 \right)^{-0.5} \quad \text{and} \quad \tilde{x}_{2(s)} = x_{2(s)} \left(\sum_{s=1}^{J} x_{2(s)}^2 \right)^{-0.5}. \tag{8}$$

Under the multinomial sampling plan, the natural logarithm of the likelihood function for a three-way mixed contingency table is as follows:

$$\ell(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\tau}, \boldsymbol{\beta} | \boldsymbol{y}) = \sum_{tsr} y_{tsr} \log n_{tsr}, \tag{9}$$

where $\boldsymbol{x}$, $\boldsymbol{u}$, $\boldsymbol{\tau}$, and $\boldsymbol{\beta}$ are vectors of scores, main effect and interaction parameters, row or column effect parameters, and association parameters, respectively.

3. Elicitation of prior distributions

3.1. Exchangeability assumption and determination of hyper-priors

Let $X_1, \ldots, X_r$ be random variables. If the subscripts contain no information, then $X_1, \ldots, X_r$ are exchangeable. In the categorical data analysis, the assumption of exchangeability is valid, if all of the variables are nominal [1]. The subscripts of the levels of an ordinal variable contain information because of the ordering between the levels. Generally, two hyper-parameters of interest are used to indicate the mean vector and covariance matrix. The mean vector reflects the subjective prior information, and the covariance matrix reflects the prior information on the correlations between the parameters of interest and the degree of prior belief in them. A researcher can take into account the information contained in the subscripts while determining the elements of a mean vector. However, this process is not easily done when determining the elements of a covariance matrix. If a common precision parameter is multiplied by the identity matrix to obtain a prior covariance matrix and then used for all of the scores, the information contained in the subscripts of the scores is ignored; hence, the exchangeability assumption is violated. One may use a precision parameter for each score. However, if there are more than three scores, the determination of a precision parameter for each score is not a convenient procedure. One cannot guarantee the positive definiteness of the relevant covariance matrix, which is required for posterior inferences. Therefore, in the Bayesian approaches to association models, the representation of the degree of belief in the induced prior information of the scores is related to the exchangeability concept. It is appropriate to decompose the covariance matrix and represent the degree of prior belief using the decomposed elements. Daniels and Pourahmadi [10] and Chen and Dunson [7] give various approaches for this decomposition. Daniels and Pourahmadi [10] use Cholesky decomposition to derive conditionally conjugate priors for covariance matrices. Chen and Dunson [7] give a reparameterization that is based on the Cholesky decomposition of the covariance matrix of parameters corresponding to the random part of a linear model. Demirhan and Hamurkaroglu [11] propose an approach for the representation of the degree of belief in prior knowledge concerning the log odds ratios using the decomposed elements.

We use the approach given by Chen and Dunson [7] to decompose the covariance matrix, and reflect the degree of belief in prior knowledge using the decomposed elements by the approach of Demirhan and Hamurkaroglu [11]. The approach of Demirhan and Hamurkaroglu [11] is outlined for log odds ratios. We reformulate the latter approach for scores and model parameters. Using our approach, it is possible to make individual representations of the degree of prior belief in each score. Therefore, the exchangeability assumption holds.

Let $\mathbf{\Sigma}_i$ be the covariance matrix of the scores corresponding to the variable i of the considered contingency table, and let I be the number of levels of the variable i. Note that we assume the mutual independency of the scores corresponding to the different variables.

The Cholesky decomposition of $\mathbf{\Sigma}_i$ is $\mathbf{\Phi}_i\mathbf{\Gamma}_i\mathbf{\Gamma}_i^T\mathbf{\Phi}_i^T$, where $\mathbf{\Phi}_i = \mathrm{diag}(\varphi_{i(t)})$, $\mathrm{diag}(\cdot)$ denotes a diagonal matrix, and $\mathbf{\Gamma}_i = (\gamma_{i(ts)})$, where $\gamma_{i(tt)} = 1$ and $\gamma_{i(ts)} = 0$ for $t = 1, \ldots, I$ and $s = t+1, \ldots, I$. Let $\sigma_{i(ts)}$ denote the (t, s)th element of $\mathbf{\Sigma}_i$. Then, $\sigma_{i(tt)}$ and $\sigma_{i(ts)}$ are obtained as follows [10,7,11]:

$$\sigma_{i(tt)} = \varphi_{i(t)}^2\left[1 + \sum_{s=1}^{t-1}\gamma_{i(ts)}^2\right] \quad \text{and} \quad \sigma_{i(st)} = \varphi_{i(t)}\varphi_{i(s)}\left[\gamma_{i(st)} + \sum_{r=1}^{t-1}\gamma_{i(tr)}\gamma_{i(sr)}\right].$$

In this decomposition, the parameter $\gamma_{i(ts)}$ reflects the dependency between scores, corresponding to levels t and s of ordinal variable i. We use the Pearson correlation coefficient to represent the prior knowledge regarding the degree of dependency between $x_{i(t)}$ and $x_{i(s)}$. The Pearson correlation coefficient, $\rho_{i(ts)}$, corresponds to the following transformation of $\gamma_{i(ts)}$:

$$\rho_{i(st)} = \left[\gamma_{i(st)} + \sum_{r=1}^{t-1}\gamma_{i(tr)}\gamma_{i(sr)}\right]\left[\left(1 + \sum_{r=1}^{t-1}\gamma_{i(tr)}^2\right)\left(1 + \sum_{r=1}^{s-1}\gamma_{i(sr)}^2\right)\right]^{-1/2}, \tag{10}$$

where we assume that the $\gamma_{i(rq)}$ values are given for $r = 1, \ldots, t-1$, $q = 1, \ldots, s-1$, and $r \neq q$. In simpler notation, Eq. (10) is rewritten as $\rho_{i(ts)} = (\gamma_{i(ts)} + A_{i(ts)})/B_{i(ts)}$, where A and B are constants. Note that $\rho_{i(ts)}$ includes neither $\varphi_{i(t)}$ nor $\varphi_{i(s)}$. Therefore, we treat $\varphi_{i(t)}$ parameters as given constants. Setting the values of all $\varphi_{i(t)}$ parameters to arbitrary numbers has no effect on the results. Demirhan and Hamurkaroglu [11] induce the following uniform prior on each $\rho_{i(ts)}$:

$$p(\rho_{i(ts)}) = (\mathrm{g}_{i(ts)})^{-1}, \qquad (1-d)\mathrm{g}_{i(ts)} < \rho_{i(ts)} < d\mathrm{g}_{i(ts)}, \tag{11}$$

where d is equal to 0 when $\mathrm{g}_{i(ts)} < 0$, and 1 when $\mathrm{g}_{i(ts)} > 0$; then the prior distribution of $\gamma_{i(ts)}$ given $\gamma_{i(rq)}$ for $r = 1, \ldots, t-1$, $q = 1, \ldots, s-1$, and $r \neq q$ is obtained by the inverse transformation as follows:

$$p(\gamma_{i(ts)}|\gamma_{i(rq)}) = B\mathrm{g}_{i(ts)}^{-1}, \qquad B(1-d)\mathrm{g}_{i(ts)} - A < \gamma_{i(ts)} < Bd\mathrm{g}_{i(ts)} - A. \tag{12}$$

In this way, the prior information on the degree of correlation between two scores is transformed into information on the relevant $\gamma_{i(ts)}$ and, hence, on the covariance between scores. We consider each level of an ordinal variable individually. Thus, exchangeability assumption is not violated. Here, $\mathrm{g}_{i(ts)}$ is a user-specified constant where the user directly expresses the prior knowledge of a correlation by determining a value for it [11]. One can determine a value for $\mathrm{g}_{i(ts)}$ by equating the quantified prior knowledge to $g_{i(ts)}/2$, which is the expected value of $\rho_{i(ts)}$ obtained over (11). For example, let us have such a prior knowledge that there is a negative and moderate correlation between the two scores of the first variable. Then, the value is quantified as −0.5. To reflect this information in the prior covariance matrix of the scores, we make the assignment that $g_{1(ts)}/2 = -0.5$; hence, we obtain $g_{1(ts)} = -1$. The prior distribution of $\rho_{1(ts)}$ is obtained as Uniform $(-1, 0)$, and that of $\gamma_{1(ts)}$ given the rest of the parameters is obtained from Eq. (12) as follows:

$$p(\gamma_{1(ts)}|\gamma_{1(rq)}) = B, \qquad -B - A < \gamma_{1(ts)} < -A.$$

Here, we should discuss the elicitation of the prior information of the correlations from experts, if available. Demirhan and Hamurkaroglu [11] give an approach for the elicitation. For an ordinal variable, the direction of the relevant correlation between levels is determined by the direction of the hierarchy, and the degree of association is determined by the expert according to the importance of passing from one level to the next one or the similarity between the levels considered. Considering the presence of correlations between adjacent levels of an ordinal variable and the absence of correlations between non-adjacent levels is reasonable. Therefore, we elicit prior information concerning the correlations between adjacent levels. For example, let the ordinal variable "Education" be categorized as "Middle School", "High School", "Undergraduate", and "Graduate". Here, we induce *a priori* positive correlations between the adjacent levels because an increasing hierarchy exists between the levels of education. The degree of correlation between the levels "Middle School" and "High School" would be elicited as moderate due to the similarity between these levels. Because the similarity between "High School" and "Undergraduate" is less than that of "Middle School" and "High School", we induce a low correlation between "High School" and "Undergraduate". Additionally, due to the higher similarity between "Undergraduate" and "Graduate", we induce a moderate correlation between these categories.

In summary, we decompose the covariance matrix of scores using the Cholesky decomposition, write the Pearson correlation coefficient in terms of the decomposed elements, and induce a uniform prior for each correlation coefficient. Accordingly, we are not only able to specify the prior covariances irrespective of the information contained in the subscripts of the levels of ordinal variables but also we can include the information derived from the hierarchy between the levels of an ordinal variable.

3.2. The generalized multivariate log-gamma distribution

We use the generalized multivariate log-gamma (G-MVLG) distribution introduced by Demirhan and Hamurkaroglu [12] as the prior distribution of scores and model parameters. If $\boldsymbol{Y} \sim$ G-MVLG$(\delta, \nu, \boldsymbol{\lambda}, \boldsymbol{\mu})$, the joint probability density function (pdf) of $\boldsymbol{Y} = (Y_1, \ldots, Y_k)$ is given as follows [12]:

$$p(\boldsymbol{y}|\boldsymbol{g}) \propto \delta^{\nu} \sum_{n=0}^{\infty} \frac{(1-\delta)^n \prod_{j=1}^{k} \mu_j \lambda_j^{-\nu-n}}{[\Gamma(\nu+n)]^{k-1}\Gamma(\nu)n!} \exp\left\{(\nu+n)\sum_{j=1}^{k}\mu_j y_j - \sum_{j=1}^{k}\frac{1}{\lambda_j}\exp\{\mu_j y_j\}\right\},$$

where $\boldsymbol{y} \in \mathbb{R}^k$, $\nu > 0$, $\lambda_j > 0$, $\mu_j > 0$, $\boldsymbol{\mu} = (\mu_j)$, $\boldsymbol{\lambda} = (\lambda_j)$, for $j = 1, \ldots, k$, $\delta = \det(\boldsymbol{\Omega})^{\frac{1}{k-1}}$,

$$\boldsymbol{\Omega} = \begin{pmatrix} 1 & \sqrt{\mathrm{abs}(\rho_{12})} & \cdots & \sqrt{\mathrm{abs}(\rho_{1k})} \\ \sqrt{\mathrm{abs}(\rho_{12})} & 1 & \cdots & \sqrt{\mathrm{abs}(\rho_{2k})} \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{\mathrm{abs}(\rho_{1k})} & \sqrt{\mathrm{abs}(\rho_{2k})} & \cdots & 1 \end{pmatrix}, \tag{13}$$

where $\det(\cdot)$ and abs $(\cdot)$ denote the determinant and absolute value of the inner expression, respectively, and $\boldsymbol{g} = (\delta, \nu, \boldsymbol{\lambda}^T, \boldsymbol{\mu}^T)$ includes parameters of the distribution. In fact, ρ_{ij} represents $\mathrm{Corr}(\exp(Y_i), \exp(Y_j))$. Demirhan and Hamurkaroglu [12] theoretically derive the exact expression of $\mathrm{Corr}(Y_i, Y_j)$, but this is a rather complex approach. Instead, they conduct a simulation study and demonstrate that $\rho_{ij} = \mathrm{Corr}(\exp(Y_i), \exp(Y_j)) \approx \mathrm{Corr}(Y_i, Y_j)$. Therefore, it is appropriate to use ρ_{ij} to represent $\mathrm{Corr}(Y_i, Y_j)$ when the G-MVLG distribution is used as the prior.

The marginal expected value and variance for elements of $\boldsymbol{Y}$ are given as follows:

$$E(Y_j) = \frac{1}{\mu_j}\left[\ln(\lambda_j/\delta) + \psi(\nu)\right] \quad \text{and} \quad V(Y_j) = \psi^{[1]}(\nu)/(\mu_j)^2, \tag{14}$$

where $\psi(\cdot)$ and $\psi^{[1]}(\cdot)$ are digamma and trigamma functions, respectively. Effects of μ_j, λ_j, δ, and ν over the marginal expectation and variance are discussed in detail by Demirhan and Hamurkaroglu [12].

One can simply use multivariate normal (MVN), multivariate *t* (MVt), or multivariate skew normal (MVSN) (see Azzalini [4] for the details of MVSN distribution) distributions as the multivariate prior distribution along with the multinomial likelihood, instead of G-MVLG. However, Demirhan and Hamurkaroglu [12] theoretically show that there are some problems in the coherency between the MVN, MVt, and MVSN prior distributions and the multinomial likelihood under the informative and non-informative settings, and these problems are not observed for the G-MVLG distribution.

3.3. Prior distributions for unrestricted and order-restricted cases

When there is no restriction on scores, we consider the dependence structures between the scores corresponding to each variable. Then, we assume the mutual independency of the scores corresponding to the separate variables, while assuming that the rest of the parameters of the log-linear model are mutually independent. Therefore, we present our approach only for the variable i.

We follow a two-stage approach for the determination of the prior distribution of the scores. In the first stage, we assign values of $g_{i(ts)}$, and we generate correlations from the uniform distribution in Eq. (12). Then, we determine the covariance matrix that does not violate the exchangeability assumption with the approach given in Section 3.1. The correlation matrix is obtained as $\boldsymbol{R}_i = \boldsymbol{D}_i^{-1/2}\boldsymbol{\Sigma}_i\boldsymbol{D}_i^{-1/2}$, where $\boldsymbol{D}_i$ is a diagonal matrix including the covariances. Then, we calculate $\boldsymbol{\Omega}_i$ by using (13), and obtain a proposed δ_i from $\boldsymbol{\Omega}_i$. We accept this move according to the full conditional distribution of δ, which is presented in the next section. This approach is an independence sampler. At the second stage, a G-MVLG prior is induced on the scores for a given model m. For the variable i, we define $\boldsymbol{x}_i = (x_{i(j)})$, $\boldsymbol{\gamma}_i = (\gamma_{i(ts)})$, $\boldsymbol{\lambda}_i = (\lambda_{i(j)})$, $\boldsymbol{\mu}_i = (\mu_{i(j)})$ for $j = 1, \ldots, I$, $t = 1, \ldots, I$, $s = t + 1, \ldots, I$, and $\boldsymbol{h}_i = \left(\boldsymbol{\varphi}, \boldsymbol{\gamma}_i, \boldsymbol{\lambda}_i, \boldsymbol{\mu}_i, \nu_i, \delta_i\right)$. Then the joint prior distribution of the scores is as follows:

$$\boldsymbol{x}_i|\boldsymbol{h}_i \sim \text{G-MVLG}(\delta_i, \nu_i, \boldsymbol{\lambda}_i, \boldsymbol{\mu}_i). \tag{15}$$

We set the value of ν_i to a suitable value that makes the values of both $\psi(\nu_i)$ and $\psi^{[1]}(\nu_i)$ appropriate. For example, when ν_i is set to 1, we obtain $\psi(1) = -0.577$ and $\psi^{[1]}(1) = 1.645$. We use elements of $\boldsymbol{\mu}_i$ to reflect the degree of belief in prior knowledge on each score. The hyper-parameter δ_i includes the prior information on the correlations between scores. A value of δ_i is obtained at the first stage. After determining the values of ν_i, δ_i and the elements of $\boldsymbol{\mu}_i$, the prior knowledge regarding the scores is reflected by using the expected value in Eq. (14). Assume that our prior knowledge on $x_{i(j)}$ is quantified as $\boldsymbol{B}_i = (B_{i(j)})$. Then, the corresponding value of $\lambda_{i(j)}$ is obtained from (14) as $\lambda_{i(j)} = \delta_i \exp\{B_{i(j)}\mu_{i(j)} - \psi(\nu_{i(j)})\}$. As seen here, values of $\mu_{i(j)}$ that are close to zero minimize the effect of $B_{i(j)}$, reflecting the weak prior belief in the subjective prior information for each score. In contrast, we should choose finite and sufficiently large values for $\mu_{i(j)}$ to make the prior belief as strong as desired. Assigning a large value to $\mu_{i(j)}$ seems to make the marginal prior mean zero. However, we place our quantified prior knowledge on $B_{i(j)}$ and derive $\lambda_{i(j)}$ from (14) for a given value of $\mu_{i(j)}$. Accordingly, the value of $\lambda_{i(j)}$ changes correspondingly with the value of $\mu_{i(j)}$; hence, the value of $\mu_{i(j)}$ is effectual only on the prior variance.

The use of a G-MVLG prior distribution for either the main effect, association, row or column effect parameters increases the complexity of the joint posterior distributions and, consequently, the amount of effort required for the posterior computations. Instead, assuming the independence of the parameters other than the scores, a Log-Gamma (θ, η) distribution is induced on each parameter, which is denoted by β for a general representation, as follows:

$$\beta \sim \text{Log-Gamma}(\theta, \eta). \tag{16}$$

To determine the values of the hyper-parameters of the prior distributions given in (16), we consider $E(\beta) = \log(\eta) + \psi(\theta)$ and $V(\beta) = \psi^{[1]}(\theta)$. The degree of belief in the prior information is reflected by the variance of the log-gamma distribution for the model parameters. The trigamma function, $\psi^{[1]}(\theta)$, decreases for $\theta > 0$. Thus, the values of $V(\beta)$ are large for values of θ that are close to zero; hence, small values of θ reflect a weak prior belief in the subjective prior information. In contrast, for larger values of θ, values for $V(\beta)$ are obtained close to zero; hence, we should choose sufficiently large values for θ to make the prior belief as strong as desired. Let the quantified prior knowledge for β be B. The knowledge is assigned to the expected value, and the value of η is derived such that $\eta = \exp(B - \psi(\theta))$.

When an order restriction is placed on scores, the restriction is inherently a strong dependence structure. Therefore, we assume the independence of scores while imposing order restrictions on scores, following the work of Iliopoulos et al. [20,21]. We use order statistics to represent order restrictions.

Suppose that the levels of the ordinal variable, upon which we impose order restrictions, are ordered according to the restrictions of interest. The ordered scores for the levels of variable i are represented with order statistics as $x_{i[1]} < x_{i[2]} < \cdots < x_{i[I]}$. We assume the independence of the ordered scores and induce a Log-Gamma $(\theta_{i[t]}, \eta_{i[t]})$ distribution on each score as follows:

$$x_{i[t]}|\theta_{i[t]}, \eta_{i[t]} \sim \text{Log-Gamma}(\theta_{i[t]}, \eta_{i[t]}). \tag{17}$$

The joint prior pdf of ordered scores is obtained from the general form of joint pdf of order statistics and (17) as follows:

$$p(x_{i[1]}, \ldots, x_{i[I]}|\boldsymbol{\theta}_i, \boldsymbol{\eta}_i) = I! \prod_{t=1}^{I} (\eta_{i[t]})^{\theta_{i[t]}} / \Gamma(\theta_{i[t]}) \exp\left\{ \sum_{t=1}^{I} \left(\theta_{i[t]} x_{i[t]} - (1/\eta_{i[t]}) \exp(x_{i[t]}) \right) \right\},$$
$$x_{i[1]} < x_{i[2]} < \cdots < x_{i[I]}, \tag{18}$$

where $\boldsymbol{\theta}_i = (\theta_{i[t]})$, $\boldsymbol{\eta}_i = (\eta_{i[t]})$ and $t = 1, \ldots, I$.

The prior distributions for other parameter sets are taken as described in Section 3.3. The degree of belief in the prior knowledge is reflected by $\theta_{i[t]}$. Following the discussion in Section 3.3, small values of $\theta_{i[t]}$ reflect a weak prior belief in the prior knowledge. Larger values of $\theta_{i[t]}$ make the prior belief as strong as desired. The quantified prior knowledge on $x_{i[t]}$, say $B_{i[t]}$, is assigned to the expected value, and $\eta_{i[t]}$ is solved as $\eta_{i[t]} = \exp(B_{i[t]} - \psi(\theta_{i[t]}))$. The discussion regarding the determination of the hyper-parameters, as given in Section 3.3, also applies to the rest of the model parameters.

The order-restricted version of the log-gamma distribution is preferable to the uniform prior used by Iliopoulos et al. [21], because we are able to define simultaneously an informative prior for some of the scores and a non-informative prior for the others in the ordered log-gamma prior setting. Additionally, the posterior computations in the ordered log-gamma prior setting are as easy to perform as those in the ordered uniform prior setting.

Note that, although the notation is given for two-way tables, the approaches given here are applicable to any size of contingency table.

4. Posterior inferences

We develop posterior inferences over an RC model constructed for a two-way contingency table; hence, $N = 2$. Posterior inferences for other models, in which $N > 2$, are straightforward following the context given here. The model of interest is as follows:

$$\log n_{ts} = u + u_{1(t)} + u_{2(s)} + x_{1(t)}x_{2(s)}, \quad t = 1, \ldots, I;\ s = 1, \ldots, J. \tag{19}$$

We obtain the posterior estimates of the following parameters: $\boldsymbol{u}_1 = (u_{1(2)}, \ldots, u_{1(I)})$, $\boldsymbol{u}_2 = (u_{2(2)}, \ldots, u_{2(J)})$, $\boldsymbol{\gamma}_1$, $\boldsymbol{\gamma}_2$, $\boldsymbol{x}_1 = (x_{1(2)}, \ldots, x_{2(I)})$, and $\boldsymbol{x}_2 = (x_{2(3)}, \ldots, x_{2(J)})$. The rest of the parameters are redundant and are obtained as a function of the given set of parameters. We set $x_{2(2)}$ to 1 for identifiability purposes, as done by Iliopoulos et al. [20]. The constant term u is obtained as follows:

$$u = -\log\left[\sum_{t=1}^{I}\sum_{s=1}^{J}\exp\{u_{1(t)} + u_{2(s)} + x_{1(t)}x_{2(s)}\}\right]. \tag{20}$$

If the prior distributions of scores and the main effect parameters are taken as given in (15) and (16), respectively, then the joint posterior distribution of the scores and model parameters is obtained as

$$p\big(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 | \boldsymbol{h}_1, \boldsymbol{h}_2, \boldsymbol{y}\big) \propto p\big(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 | \boldsymbol{h}_1, \boldsymbol{h}_2\big)\ell(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{u}_1, \boldsymbol{u}_2 | \boldsymbol{y}), \tag{21}$$

where

$$\ell(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{u}_1, \boldsymbol{u}_2 | \boldsymbol{y}) = \sum_{t=1}^{I}\sum_{s=1}^{J} y_{ts}(u + u_{1(t)} + u_{2(s)} + x_{1(t)}x_{2(s)}). \tag{22}$$

The joint posterior distribution obtained by (21) is a proper density. Because our G-MVLG and log-gamma distributions are proper, and we use a likelihood function based on a probability model, the propriety of the posterior follows from Bayes' theorem [12].

For the full conditional posterior distribution of δ_i, we use the relationship between δ_i and $\boldsymbol{\gamma}_i$, as explained in Section 3.1. Because δ_i is a transformation of $\gamma_{i(ts)}$ such that $\delta_i = f(\boldsymbol{\gamma}_i)$, we are able to obtain a candidate value for δ_i over the randomly generated values of $\gamma_{i(ts)}$ via $f(\boldsymbol{\gamma}_i)$. After several straightforward algebraic manipulations, the full conditional posterior distribution of δ_i given $\boldsymbol{g} = (\boldsymbol{x}_i, \boldsymbol{h}_1, \boldsymbol{y})$ is obtained as follows:

$$p(\delta_i | \boldsymbol{g}) \propto \delta_i^{\nu_i} \sum_{n=0}^{\infty} \frac{(1-\delta_i)^n \exp\left\{n \sum_{r=1}^{K} \mu_{i(r)} x_{i(r)}\right\} \prod_{r=1}^{K} \lambda_{i(r)}^{-n}}{[\Gamma(\nu_i + n)]^{K-1} n!}, \tag{23}$$

where $i = 1, 2$, $K = I$ for $i = 1$ and $K = J$ for $i = 2$. To simplify (23), we use the approach proposed by Demirhan and Hamurkaroglu [12] for $\Gamma(\nu_i + n)$ such that $\Gamma(\nu_i + n) \approx \exp(-7.24663 + 2.07728n + 1.9922\nu_i)$. The accuracy of this approach is evaluated via simulation in Demirhan and Hamurkaroglu [12]. As a result, when n is an integer, the accuracy of the approach is found to be better than that of the current approach given for the gamma function by Schmelzer and Trefethen [30]. When the approximation is applied to (23), the Taylor expansion of the exponential function at zero is obtained, and Eq. (23) is simplified as

$$p(\delta_i | \boldsymbol{g}) \propto \delta_i^{\nu_i} \exp\left\{-\delta_i \exp\left\{n \sum_{r=1}^{K} \mu_{i(r)} x_{i(r)}\right\} \prod_{r=1}^{K} \lambda_{i(r)}^{-n}\right\}, \quad \delta_i > 0. \tag{24}$$

The full conditional distribution in (24) is written in terms of $\gamma_{i(ts)}$ such that

$$\begin{gathered} p(f(\gamma_{i(ts)}) | \boldsymbol{g}) \propto f(\gamma_{i(ts)})^{\nu_i} \exp\left\{-f(\gamma_{i(ts)}) \exp\left\{n \sum_{r=1}^{K} \mu_{i(r)} x_{i(r)}\right\} \prod_{r=1}^{K} \lambda_{i(r)}^{-n}\right\}, \\ B_{i(ts)}(1 - d_{i(ts)})g_{i(ts)} - A_{i(ts)} < \gamma_{i(ts)} < B_{i(ts)} d_{i(ts)} g_{i(ts)} - A_{i(ts)}, \end{gathered} \tag{25}$$

where $A_{i(ts)}$, $B_{i(ts)}$, and $d_{i(ts)}$ are defined in Section 3.1.

The full conditional posterior distribution of score j of variable i given $\boldsymbol{g} = (\boldsymbol{x}_{-i(j)}, \boldsymbol{h}_1, \boldsymbol{h}_2, \boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{y})$, where $\boldsymbol{x}_{-i(j)}$ includes scores of both variables except $x_{i(j)}$, is obtained after some straightforward algebraic manipulations as follows:

$$p(x_{i(t)}|\boldsymbol{g}) \propto \frac{\exp\left\{nu + (\nu_i\mu_{i(t)} + S_{i(t)})x_{i(t)} - e^{\mu_{i(t)}x_{i(t)}}/\lambda_{i(t)}\right\}}{\Gamma(\nu_i)} \times \sum_{n=0}^{\infty} \frac{(1-\delta_i)^n \exp\left\{n \sum_{r=1,r\neq t}^{K} \mu_{i(r)}x_{i(r)}\right\} \prod_{r=1}^{K} \lambda_{i(r)}^{-n}}{[\Gamma(\nu_i+n)]^{K-1}n!} \exp\{n\mu_{i(t)}x_{i(t)}\}, \tag{26}$$

where $i = 1, 2$, $K = I$ for $i = 1$ and $K = J$ for $i = 2$. When we use the approach for the $\Gamma(\nu_i + n)$ function, the following is obtained:

$$\begin{aligned} p(x_{i(t)}|\boldsymbol{g}) &\propto \exp\left\{nu + (\nu_i\mu_{i(t)} + S_{i(t)})x_{i(t)} - e^{\mu_{i(t)}x_{i(t)}}/\lambda_{i(t)}\right\} \\ &\quad \times \exp\left\{\frac{(1-\delta_i)\exp\left\{\mu_{i(t)}x_{i(t)} + \sum_{r=1,r\neq t}^{K} \mu_{i(r)}x_{i(r)}\right\} \prod_{r=1}^{K} \lambda_{i(r)}^{-1}}{(\exp(2.07728))^{K-1}}\right\} \\ &\propto \exp\left\{nu + (\nu_i\mu_{i(t)} + S_{i(t)})x_{i(t)} - c_{i(t)}e^{\mu_{i(t)}x_{i(t)}}\right\}, \quad x_{i(t)} \in \mathbb{R}, \end{aligned} \tag{27}$$

where u is defined in (20), $S_{1(t)} = \sum_{s=1}^{J}(y_{ts} - y_{1s})x_{2(s)}$, $S_{2(s)} = \sum_{t=1}^{I}(y_{ts} - y_{t1})x_{1(t)}$, and

$$c_{i(t)} = (\lambda_{i(t)})^{-1} - (1-\delta_i)\exp\left\{2.07728(1-K) + \sum_{r=1,r\neq t}^{K} \mu_{i(r)}x_{i(r)}\right\} \prod_{r=1}^{K} \lambda_{i(r)}^{-1}.$$

Assuming that all of the main effect parameters are mutually independent, and that the prior mentioned in (16) is induced on each main effect parameter, the full conditional posterior distribution of parameter $u_{i(j)}$, given $\boldsymbol{g} = (\boldsymbol{u}_{-i(j)}, \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{h}_1, \boldsymbol{h}_2, \boldsymbol{y})$, where $\boldsymbol{u}_{-i(j)}$ includes the main effect parameters of both variables except $u_{i(j)}$, is straightforwardly obtained as follows:

$$p(u_{i(t)}|\boldsymbol{g}) \propto \exp\left\{nu + (\theta_{i(t)} + T_{i(t)})u_{i(t)} - (1/\eta_{i(t)})e^{u_{i(t)}}\right\}, \quad u_{i(t)} \in \mathbb{R}, \tag{28}$$

where u is defined in (20), $T_{1(t)} = (y_{t+} - y_{1+})$, and $T_{2(s)} = (y_{+s} - y_{+1})$. Here, the subscript over which a sum is calculated is replaced by a "+".

For the order-restricted case, the model of interest is the same as that of the unrestricted case. We obtain posterior estimates of $\boldsymbol{u}_1 = (u_{1(2)}, \ldots, u_{1(I)})$, $\boldsymbol{u}_2 = (u_{2(2)}, \ldots, u_{2(J)})$, $\boldsymbol{x}_{[1]} = (x_{1[2]}, \ldots, x_{2[I]})$, and $\boldsymbol{x}_{[2]} = (x_{2[3]}, \ldots, x_{2[J]})$. The rest of the parameters are redundant and are obtained as a function of the given set of parameters. We set $x_{2(2)}$ to 1 for identifiability purposes. The constant term u is obtained as in Eq. (20) using the ordered scores.

If the joint prior distribution of the scores is taken as given in (18), and the prior distributions of the main effect parameters are taken as given in (16), then the joint posterior distribution of scores and model parameters is obtained as follows:

$$p\big(\boldsymbol{x}_{[1]}, \boldsymbol{x}_{[2]}, \boldsymbol{u}_1, \boldsymbol{u}_2|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \boldsymbol{y}\big) \propto p\big(\boldsymbol{x}_{[1]}, \boldsymbol{x}_{[2]}, \boldsymbol{u}_1, \boldsymbol{u}_2|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\eta}_1, \boldsymbol{\eta}_2\big)\ell(\boldsymbol{x}_{[1]}, \boldsymbol{x}_{[2]}, \boldsymbol{u}_1, \boldsymbol{u}_2|\boldsymbol{y}), \tag{29}$$

where $\ell(\boldsymbol{x}_{[1]}, \boldsymbol{x}_{[2]}, \boldsymbol{u}_1, \boldsymbol{u}_2|\boldsymbol{y})$ is as given in Eq. (22) with the ordered scores in place of unordered ones. Propriety of the joint posterior distribution in (29) follows from the Bayes theorem.

The full conditional posterior distribution of score j of variable i given $\boldsymbol{g} = (\boldsymbol{x}_{-i[t]}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \boldsymbol{y})$, where $\boldsymbol{x}_{-i[t]}$ includes the scores of both variables except $x_{i[t]}$, is obtained from Eq. (18) and the corresponding likelihood function as follows:

$$p(x_{i[t]}|\boldsymbol{g}) \propto \exp\left\{nu + (\theta_{i[t]} + S_{i[t]})x_{i[t]} - (1/\eta_{i[t]})\exp(x_{i[t]})\right\}, \quad x_{i[t-1]} < x_{i[t]} < x_{i[t+1]}, \tag{30}$$

where u is as defined in (20), $S_{1[t]} = \sum_{s=1}^{J}(y_{ts} - y_{1s})x_{2[s]}$, and $S_{2[s]} = \sum_{t=1}^{I}(y_{ts} - y_{t1})x_{1[t]}$.

In the order-restricted case, the full conditional posterior distributions of the main effect parameters are the same as in the unrestricted case, but the unordered scores are replaced by ordered scores.

In our approach, due to the independency between score groups corresponding to ordinal variables, it is possible to impose order restrictions on the scores of some ordinal variables, leaving the scores of others unrestricted. In this case, we call the corresponding model a mixed association model. For a mixed model, the full conditional posterior distributions of the main effect parameters and unrestricted scores are the same as in the unrestricted case, and those of the order-restricted scores are taken as in (30).

Because we cannot directly sample from the full conditional distributions derived above, we employ a random-walk-Metropolis step for this purpose. As a result, we construct a Metropolis-within-Gibbs sampling algorithm by using the full conditional posterior distributions given in (27) and (28) for the unrestricted case, and those given in (28) and (30) for the restricted case. Posterior computations are implemented over the algorithm given in Appendix A.

5. Model comparison

Chib and Jeliazkov [8] introduce a practical approach for the calculation of marginal likelihood that can be used to obtain the Bayes factors for the model comparison. In their approach, the output of the Metropolis–Hastings algorithm is used directly, and the required Bayes factors are easily obtained at the end of each run of the algorithm. Additionally, Mira and Nicholls [26] show that the multi-block estimators of Chib and Jeliazkov [8] are bridge sampling estimators. Let $\boldsymbol{\beta}$ include all main effect parameters, δ_i, hyper-parameters, and scores; then the acceptance probability of a move to the point $\boldsymbol{\beta}'$ in our Metropolis-within-Gibbs sampling algorithm is defined as follows:

$$\alpha(\boldsymbol{\beta}, \boldsymbol{\beta}') = \min\left\{1, \frac{p(\boldsymbol{y}|\boldsymbol{\beta}')p(\boldsymbol{\beta}')q(\boldsymbol{\beta}', \boldsymbol{\beta}|\boldsymbol{y})}{p(\boldsymbol{y}|\boldsymbol{\beta})p(\boldsymbol{\beta})q(\boldsymbol{\beta}, \boldsymbol{\beta}'|\boldsymbol{y})}\right\}, \tag{31}$$

where $q(\boldsymbol{\beta}, \boldsymbol{\beta}'|\boldsymbol{y})$ is the proposal density of the move from $\boldsymbol{\beta}$ to $\boldsymbol{\beta}'$, $p(\boldsymbol{\beta})$ is the prior distribution, and $p(\boldsymbol{y}|\boldsymbol{\beta})$ is the likelihood. In our case, the parameters are grouped in $B = (I-1)+(J-1)+2+(I-1)+(J-2) = 2I+2J-3$ blocks. According to the approach of Chib and Jeliazkov [8], the reduced posterior ordinate obtained for the main effect parameter $u^*_{i(t)}$ for $t = 2, \dots, I$ for $i = 1$ and $t = 2, \dots, J$ for $i = 2$ is given as follows:

$$\hat{p}(u^*_{i(t)}|\boldsymbol{y}, u^*_{i(1)}, \dots, u^*_{i(j-1)}) = \frac{M\sum_{g=1}^{M}\alpha(u^{(g)}_{i(t)}, u^*_{i(t)}|\boldsymbol{y}, \boldsymbol{\psi}^*_{t-1}, \boldsymbol{\psi}^{t+1,(g)})q(u^{(g)}_{i(t)}, u^*_{i(t)}|\boldsymbol{y}, \boldsymbol{\psi}^*_{t-1}, \boldsymbol{\psi}^{t+1,(g)})}{R\sum_{r=1}^{R}\alpha(u^*_{i(t)}, u^{(r)}_{i(t)}|\boldsymbol{y}, \boldsymbol{\psi}^*_{t-1}, \boldsymbol{\psi}^{t+1,(r)})}, \tag{32}$$

where $\boldsymbol{\psi}_{t-1} = (u_{i(1)}, \dots, u_{i(t-1)})$, $\boldsymbol{\psi}^{t+1} = (u_{i(t+1)}, \dots, u_{i(K)})$, $K = I$ for $i = 1$ and $K = J$ for $i = 2$, $\boldsymbol{y}$ is the observed sample, $\{u^{(g)}_{i(t)}\}$ are M draws sampled from the posterior distribution, and $\{u^{(r)}_{i(t)}\}$ are R draws sampled from the proposal distribution given $\boldsymbol{u}^*_i$. The algorithm given in Appendix A is used for additional M and R draws. The reduced posterior ordinate for δ_i hyper-parameter, $\hat{p}(\delta^*_i|\boldsymbol{y}, \boldsymbol{x}^*_i)$ for $i = 1, 2$, is obtained by replacing $u_{i(t)}$ by δ_i in Eq. (32). The reduced posterior ordinate for the score $x^*_{i(t)}$, $\hat{p}(x^*_{i(t)}|\boldsymbol{y}, x^*_{i(1)}, \dots, x^*_{i(t-1)})$, for $t = 2, \dots, I$ for $i = 1$ and $t = 3, \dots, J$ for $i = 2$, is obtained by replacing $u_{i(t)}$ by $x_{i(t)}$ in Eq. (32).

Once the reduced posterior ordinates are obtained from (32), the logarithm of the marginal likelihood, $\log(\hat{m}(\boldsymbol{y}))$, for a model M_l is obtained as

$$\log(\hat{m}(\boldsymbol{y})|M_l, \boldsymbol{\beta}^*) = \log(p(\boldsymbol{y}|\boldsymbol{\beta}^*)) + \log(p(\boldsymbol{\beta}^*)) - \sum_{i=1}^{B}\log(\hat{p}(\boldsymbol{\beta}^*_i|\boldsymbol{y}, \boldsymbol{\beta}^*_1, \dots, \boldsymbol{\beta}^*_{i-1})). \tag{33}$$

Implementation details of the approach, determination of $\boldsymbol{\beta}^*$, and various examples are given by Chib and Jeliazkov [8].

The logarithm of the required Bayes factors for model comparison is obtained as $\log(B_{10}) = \log(\hat{m}(\boldsymbol{y})|M_1, \boldsymbol{\beta}^*) - \log(\hat{m}(\boldsymbol{y})|M_0, \boldsymbol{\beta}^*)$. Interpretations of the Bayes factor are summarized by Kass and Raftery [22].

6. An illustrative example

We illustrate our approach by revisiting the dreams data set introduced by Maxwell [25], as well as the breathing test results data set presented by Forthofer and Lehnen [14, p. 21]. The breathing test results data set, which includes a three-way contingency table, is analyzed to show that our approaches can be implemented in tables of higher dimensionality. The analysis is presented in Appendix B. We present a non-informative analysis for the dreams data set to make our approaches comparable to those of Iliopoulos et al. [20,21] and the other authors, and conduct a slightly informative analysis for the breathing test results data set to illustrate the use of our approach in an informative setting.

The dreams data set was analyzed by Agresti et al. [3] and Ritov and Gilula [28] from the classical perspective, and by Iliopoulos et al. [20] using the Bayesian approach. The data set consists of a 5×4 contingency table. The variables of interest are age (with categories 5–7, 8–9, 10–11, 12–13, 14–15) and the severity of the dream disturbances (with categories 1–4 from low to high). Agresti et al. [3] fit an order-restricted C model, and Ritov and Gilula [28] fit a correlation model to this data set. Iliopoulos et al. [20] fit an order-restricted RC model with various equality restrictions on the scores of the age and the disturbance. All of the authors detect a negative association between age and the severity of disturbance. We apply our approach to this data set to illustrate the use of our approaches and compare our results with the previous results obtained using the Bayesian and classical perspectives.

In all posterior computations of this section, the algorithm given in Appendix A, which is called Algorithm 1 throughout the rest of the manuscript, is run over five independent chains for 10 000 iterations. The first 1000 iterations are taken as the

Table 1
Values of log(ML), their standard errors, and posterior model probabilities for the fitted models.

Model	log(ML)	SE	PMP
RC	−76.99	2.42	2.29E−35
RC1	−79.30	1.29	2.27E−36
RC4	−80.04	3.86	1.08E−36
RC5	−81.12	0.28	3.67E−37
R	−81.43	3.15	2.71E−37
RC_{ord}	−81.78	0.42	1.91E−37
RC2	−82.97	0.40	5.82E−38
R2	−83.11	1.38	5.03E−38
C1	−83.86	0.49	2.38E−38
C	−84.20	2.09	1.69E−38
RC3	−88.76	0.20	1.77E−40
R3	−92.28	0.07	5.25E−42
R1	−92.75	0.14	3.28E−42
R_{ord}	−94.39	0.10	6.32E−43
C_{ord}	−94.93	0.90	3.7E−43
IDP	−96.10	0.12	1.14E−43
LL	−101.81	0.25	3.81E−46

SE: standard error; PMP: posterior model probability.

burn-in period, and the Gibbs sequence is thinned by recording every tenth point after the burn-in to reduce autocorrelation. The potential scale reduction factor ($\hat{R}$) given by Gelman [16] is used to evaluate the convergence for each parameter. All $\hat{R}$ values are close to 1 and less than 1.1, implying that convergence has been achieved. At steps 3, 10, and 20 of Algorithm 1, the following are taken: $\sigma^2_{u_{i(t)}} = 100$, $\sigma^2_{x_{i(t)}} = 100$, $\boldsymbol{\beta}_1 = (2, 7, 5, 2, 0.3)$ and $\boldsymbol{\beta}_2 = (2, 2, 5, 0.45)$, respectively. These values are chosen to obtain the overall acceptance rate for each parameter and δ parameters of approximately 0.35.

The hyper-parameters of the prior distributions of the main effect parameters are determined to obtain the prior expected value and variance of each parameter equal to 0 and 100, respectively. For the unrestricted analyses, we should determine the hyper-parameters of the MVLG distribution in (15). We have no direct knowledge regarding the correlation between the adjacent scores of the age and the disturbance. However, there are naturally increasing sequences of the levels of both age and the severity of disturbance; thus, it is appropriate to induce positive correlations on the corresponding adjacent scores. Because the correlation concept is a measure of similarity, we can evaluate the similarity between adjacent levels to assign the prior correlations. Because there is low similarity between the age levels, we assign 0.1 to the prior correlations of the adjacent scores of age. We set to 0.001 all $g_{1(ts)}$ for $t, s = 2, \ldots, 5$ and $t \neq s$ but $g_{1(21)} = g_{1(32)} = g_{1(43)} = g_{1(54)} = 0.2$. The similarity between the levels of disturbance is also low, although slightly greater. Therefore, we use 0.15 as the prior correlations of the adjacent scores of disturbance, and we set $g_{2(34)} = 0.3$. We set to 1.42 both ν_1 and ν_2. Because prior information on the values of the scores is unavailable, all elements of $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ are set to 0, and those of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are set to 0.3. In this case, the prior variance of each score is obtained from Eq. (14) as $\psi^{[1]}(1.42)/(0.3)^2 = 11.1$, which is a relatively large value for prior variances. For the order-restricted analyses, all elements of $\boldsymbol{B}_{[1]}$ and $\boldsymbol{B}_{[2]}$ are set to 0, and all elements of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are set to 0.3. Thus, the prior variance of each ordered score is 12.3.

To identify the best-fitting model, we fit a set of candidate models that consists of unrestricted and order-restricted models, and the models with order and equality restrictions by using the approach of Chib and Jeliazkov [8]. These models also have been fitted by Iliopoulos et al. [20]. The parameter vectors $\boldsymbol{u}_1^*$, $\boldsymbol{u}_2^*$, $\boldsymbol{x}_1^*$ and $\boldsymbol{x}_2^*$, which are used for all unrestricted models, are taken as $\boldsymbol{u}_1^* = (-0.7146, 0.1436, 0.1779, 0.3411, 0.0521)$, $\boldsymbol{u}_2^* = (0.6762, -0.2035, -0.2183, -0.2544)$, $\boldsymbol{x}_1^* = (-0.6325, -0.3162, 0, 0.3162, 0.6325)$, $\boldsymbol{x}_2^* = (-0.6708, 1, 0.2236, 0.6708)$, $\delta_1 = 0.6718$, and $\delta_2 = 0.4912$. Those used for all of the order-restricted models are taken as $\boldsymbol{u}_1^* = (-0.6375, 0.1826, 0.1913, 0.3171, -0.0534)$, $\boldsymbol{u}_2^* = (0.5568, -0.1672, -0.1779, -0.2118)$, $\boldsymbol{x}_1^* = (-0.2349, 0.0097, 0.0334, 0.0654, 0.1265)$ and $\boldsymbol{x}_2^* = (-3.3819, 1, 1.1056, 1.2763)$. These values are obtained by using the script prepared by Iliopoulos et al. [20] [follow the link given in page 4651]. As the other authors analyzed this data set, we also identified that the models with positive association ($\phi = 1$) give worse fits than all of the models with negative association. The log-marginal likelihood values, their standard errors, and model probabilities are presented in Table 1. Let B_{10} be the Bayes factor for a model M_1 against a model M_0. Using the Bayes factors corresponding to the independence, LL, and saturated models and the models with negative association, $2\log(B_{10})$ values are calculated and listed in Table 2. In Tables 1 and 2, IDP is the independence model; R_{ord}, C_{ord}, and RC_{ord} are the order-restricted R, C, and RC models, respectively; and R1 : $x_{1[1]} = x_{1[2]}$, R2 : $x_{1[3]} = x_{1[4]}$, R3 : $x_{1[1]} = x_{1[2]} = x_{1[3]} = x_{1[4]}$, C1 : $x_{2[2]} = x_{2[3]}$, RC1 : $x_{1[1]} = x_{1[2]}$, RC2 : $x_{2[2]} = x_{2[3]}$, RC3 : $x_{1[1]} = x_{1[2]}; x_{2[2]} = x_{2[3]}$, RC4 : $x_{1[1]} = x_{1[2]}; x_{1[3]} = x_{1[4]}; x_{2[2]} = x_{2[3]}$, and RC5 : $x_{1[1]} = x_{1[2]}; x_{1[3]} = x_{1[4]}; x_{2[2]} = x_{2[3]} = x_{2[4]}$ are equality-restricted models with the given restrictions.

According to Table 2, all restricted or unrestricted models give better fits than either of the LL or IDP models. The unrestricted RC model gives the best fit among the unrestricted models, and the unrestricted R model is better than the unrestricted C model. The restricted RC model gives the best fit among the restricted models without equality restrictions. Among the order-restricted models with equality restrictions, the RC1 and RC4 models give satisfactory fits. However, the standard errors of the log(ML) estimates for these models are higher than those of the other models. Therefore, the RC5

Table 2
Values of $2\log(B_{10})$ for the fitted models.

Model in M_1	M_0															
	RC	RC1	RC4	RC5	R	RC_{ord}	RC2	R2	C1	C	RC3	R3	R1	R_{ord}	C_{ord}	IDP
RC																
RC1	−4.6															
RC4	−6.1	−1.5														
RC5	−8.3	−3.6	−2.2													
R	−8.9	−4.2	−2.8	−0.6												
RC_{ord}	−9.6	−5.0	−3.5	−1.3	−0.7											
RC2	−11.9	−7.3	−5.8	−3.7	−3.1	−2.4										
R2	−12.2	−7.6	−6.1	−4.0	−3.4	−2.7	−0.3									
C1	−13.7	−9.1	−7.6	−5.5	−4.9	−4.2	−1.8	−1.5								
C	−14.4	−9.8	−8.3	−6.2	−5.6	−4.8	−2.5	−2.2	−0.7							
RC3	−23.5	−18.9	−17.4	−15.3	−14.7	−14.0	−11.6	−11.3	−9.8	−9.1						
R3	−30.6	−26.0	−24.5	−22.3	−21.7	−21.0	−18.6	−18.3	−16.8	−16.2	−7.0					
R1	−31.5	−26.9	−25.4	−23.3	−22.6	−21.9	−19.6	−19.3	−17.8	−17.1	−8.0	−0.9				
R_{ord}	−34.8	−30.2	−28.7	−26.5	−25.9	−25.2	−22.9	−22.6	−21.1	−20.4	−11.3	−4.2	−3.3			
C_{ord}	−35.9	−31.3	−29.8	−27.6	−27.0	−26.3	−23.9	−23.6	−22.1	−21.5	−12.3	−5.3	−4.4	−1.1		
IDP	−38.2	−33.6	−32.1	−30.0	−29.4	−28.7	−26.3	−26.0	−24.5	−23.8	−14.7	−7.7	−6.7	−3.4	−2.3	
LL	−49.6	−45.0	−43.5	−41.4	−40.8	−40.1	−37.7	−37.4	−35.9	−35.2	−26.1	−19.1	−18.1	−14.8	−13.8	−11.4

model, the corresponding standard error of which is small, would be considered as giving a satisfactory fit. According to the RC1 model, the 5–7 and 8–9 age categories are located at the first level of the hierarchy between age groups. It can be inferred from the RC4 model that the 5–7 and 8–9 age categories are at the first level of hierarchy, while the 10–11 and 12–13 age categories are at the second level of hierarchy between the age groups. The 2 and 3 disturbance categories are at the second levels of hierarchy between the disturbance levels. Additionally, the last level of the disturbance is homogeneous with respect to the second and third levels in the RC5 model. These inferences are in accordance with those given by Iliopoulos et al. [20,21] and maximum likelihood estimates (MLEs) given by the other authors mentioned. However, we identify the unrestricted RC model as the best one among all of the models, whereas Iliopoulos et al. [20] identified the RC4 model as the best based on a criterion-based comparison. Instead of a criterion-based comparison, Iliopoulos et al. [21] present a model comparison by using the reversible jump Markov chain Monte Carlo (RJMCMC) method. They identify the RC5 model as the best and confirm their inferences about the RC4 model. A reason for this situation would be that the amount of included information in our analysis is more than that in the analyses of other authors, because we include possible correlations between scores corresponding to the adjacent levels of ordinal variables. Additionally, for model comparison, Iliopoulos et al. [20] use information criteria that take the number of effective parameters as a penalization factor. Because of the equality restrictions, the number of effective parameters in models including equality restrictions is less than that in the restricted or unrestricted models. This makes not only the RC4 model but also the models with equality restrictions more favorable than their restricted or unrestricted counterparts in a criterion-based comparison. The RC5 model is the simplest among the compared models. Equality restrictions simplify the model by incorporating a certain amount information. Thus, the underlying model becomes more parsimonious and less uncertain. This affects the model selection or comparison algorithms. We would like to know the implementation details of the RJMCMC algorithm of Iliopoulos et al. [21] to comment on whether their RJMCMC algorithm tends to move simpler models or not. Nonetheless, placing equality restrictions on scores is inappropriate unless one has a very strong prior belief about the imposed equality restriction. We directly use the Bayes factors, which do not penalize models with the number of effective parameters. Therefore, the results obtained by the use of Bayes factors for the comparison of models, as in our situation, seem more appropriate and reliable.

A natural ordering exists between the levels of disturbance. Therefore, even if we find the unrestricted RC model to be the best, an analysis with unrestricted scores for disturbance seems to be unrealistic, or at least we expect to obtain ordered posterior estimates for the scores of the disturbance. If this is not the case, we should consider fitting mixed RC models, in which the scores of some factors are unrestricted while those of the others are restricted. In addition to the models considered in Table 2, we fit two mixed RC models. In the first (RCM model), the disturbance and age scores are taken as order restricted and unrestricted, respectively. Because the RC2 model gives a satisfactory fit, we added an equality restriction to the disturbance scores of levels 2 and 3 to the RCM model in the second analysis (RCM2). For the comparison of RCM, RCM2, and satisfactory models that are obtained over Table 2, $2\log(B_{10})$ values are given in Table 3. In addition, posterior model probabilities of RCM2 and RCM models are 4.45E−33 and 5.67E−34, respectively. The standard errors of log(ML) for those models are 0.42 and 1.36, respectively.

We have positive evidence in favor of using the RCM model instead of the RC model, as well as strong evidence in favor of using the RCM model over the RC_{ord}, RC1, and RC4 models. Additionally, we have positive evidence supporting the RCM2 model over the RCM model, in addition to strong evidence in favor of the RCM2 model instead of the RC, RC_{ord}, RC1, RC4, and RC5 models. As expected, mixed RCM models, for which the estimated standard errors of log(ML) are relatively small, give better fits than the rest of the models considered. In this case, we identify the RCM2 model as the best model for the dream disturbance data set.

Table 3
Values of $2\log(B_{10})$ for satisfactory models.

log(ML)	Model in M_1	M_0					
		RCM2	RCM	RC	RC1	RC4	RC5
−72.69	RCM2						
−74.76	RCM	−4.1					
−76.99	RC	−8.6	−4.5				
−79.30	RC1	−13.2	−9.1	−4.6			
−80.04	RC4	−14.7	−10.6	−6.1	−1.5		
−81.12	RC5	−16.9	−12.7	−8.3	−3.6	−2.2	
−81.78	RC_{ord}	−18.2	−14.0	−9.6	−5.0	−3.5	−1.3

Table 4
Posterior estimates of scores and association parameters.

Model	Prms	Actual					Rescaled
		Posterior		95% HPD int		Rescaled[a]	Posterior
		Mean	StDev	LB	UB		Mean
RC	$x_{1(1)}$	0.226	0.188	0.069	0.283	−0.273	−0.249
	$x_{1(2)}$	0.507	0.159	0.496	0.588	−0.612	−0.579
	$x_{1(3)}$	−0.032	0.140	−0.101	0.084	0.039	0.042
	$x_{1(4)}$	−0.094	0.120	−0.152	−0.070	0.114	0.109
	$x_{1(5)}$	−0.607	0.196	−0.685	−0.341	0.732	0.676
	$x_{2(1)}$	−1.530	0.357	−1.689	−1.226	−0.813	−0.773
	$x_{2(2)}$	1.000	0.000	1.000	1.000	0.531	0.537
	$x_{2(3)}$	0.088	0.275	−0.065	0.316	0.047	0.017
	$x_{2(4)}$	0.441	0.236	0.292	0.572	0.235	0.219
	ϕ	−1.000	0.000	–	–	−1.559	−0.945
RC_{ord}	$x_{1[1]}$	−0.205	0.082	−0.226	−0.177	−0.851	−0.841
	$x_{1[2]}$	0.006	0.010	0.003	0.009	0.037	0.036
	$x_{1[3]}$	0.030	0.022	0.019	0.054	0.126	0.121
	$x_{1[4]}$	0.057	0.031	0.046	0.060	0.237	0.231
	$x_{1[5]}$	0.109	0.042	0.103	0.111	0.450	0.454
	$x_{2[1]}$	−4.122	**1.739**	−5.954	−3.293	−0.829	−0.770
	$x_{2[2]}$	1.000	0.000	1.000	1.000	0.201	0.220
	$x_{2[3]}$	0.594	**1.428**	0.507	1.988	0.119	0.067
	$x_{2[4]}$	2.528	**1.065**	1.967	2.786	0.508	0.483
	ϕ	−1.000	0.000	–	–	−1.198	−1.172
RC4	$x_{1[1]}$	−0.074	0.001	−0.076	−0.070	−0.408	−0.408
	$x_{1[2]}$	−0.074	0.001	−0.076	−0.070	−0.408	−0.408
	$x_{1[3]}$	0.000	0.000	0.000	0.000	0.000	0.000
	$x_{1[4]}$	0.000	0.000	0.000	0.000	0.000	0.000
	$x_{1[5]}$	0.148	0.004	0.141	0.152	0.816	0.817
	$x_{2[1]}$	−3.673	**0.241**	−4.314	−3.564	−0.859	−0.858
	$x_{2[2]}$	1.000	0.000	1.000	1.000	0.234	0.238
	$x_{2[3]}$	1.000	0.000	1.000	1.000	0.234	0.238
	$x_{2[4]}$	1.673	**0.241**	1.314	1.956	0.391	0.382
	ϕ	−1.000	0.000	–	–	−0.776	−0.774
RCM2	$x_{1(1)}$	0.194	0.084	0.186	0.217	0.224	0.184
	$x_{1(2)}$	0.584	0.064	0.480	0.634	0.674	0.668
	$x_{1(3)}$	−0.188	0.070	−0.237	−0.131	−0.217	−0.201
	$x_{1(4)}$	−0.009	0.040	−0.082	0.005	−0.011	0.003
	$x_{1(5)}$	−0.581	0.091	−0.619	−0.473	−0.670	−0.654
	$x_{2[1]}$	−3.491	0.010	−3.497	−3.481	−0.862	−0.861
	$x_{2[2]}$	1.000	0.000	1.000	1.000	0.247	0.248
	$x_{2[3]}$	1.000	0.000	1.000	1.000	0.247	0.248
	$x_{2[4]}$	1.491	0.010	1.490	1.497	0.368	0.366
	ϕ	−1.000	0.000	–	–	−3.513	−3.513

Int: interval; LB: lower bound; UB: upper bound; Prms: parameters; StDev: standard deviation.
[a] Actual parameters are rescaled.

The posterior means of the actual and rescaled scores, the standard deviations, and 95% highest probability density (HPD) intervals of the actual scores and association parameters of the RC, RC_{ord}, RC4, and RCM2 models for negative association are given in Table 4. The rescaled scores and association parameters are obtained by using Eqs. (8) and (7), respectively. The error bars for the rescaled posterior estimates of the row and column scores of the RC and RC_{ord} models are given in Fig. 1, and those of the RC4 and RCM2 models are given in Fig. 2.

Our analyses are intended to be non-informative in both the unrestricted and order-restricted cases. When the MLEs of the parameters (scores and association parameter) of the RC model, which are provided by Iliopoulos et al. [20], are compared

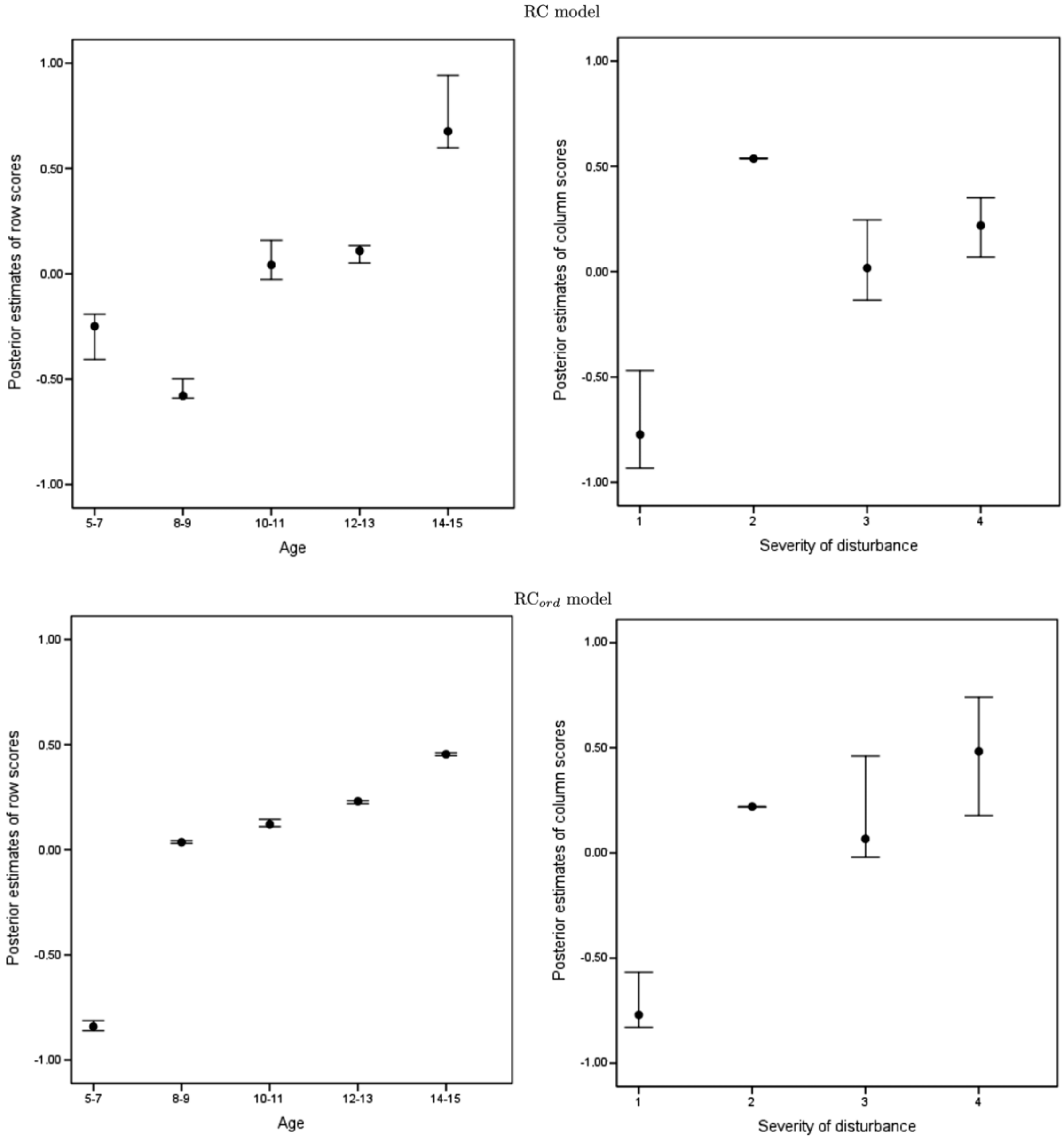

Fig. 1. Error bars for rescaled posterior estimates of row and column scores of RC and RC_{ord} models.

with our posterior estimates, our posterior estimates are observed to be close to the corresponding MLEs. This implies that we are able to conduct a nearly non-informative analysis with our MVLG prior, as intended. Our analysis is "nearly" non-informative because we induce some information on the correlations between the adjacent levels of the ordinal factors. That is why we obtain some deviation from the MLEs in our posterior estimates.

It can be seen from Figs. 1 and 2 and their corresponding counterparts given by Iliopoulos et al. [20] that, while our rescaled posterior estimates for the scores of age over the RC, RC_{ord}, and RC4 models are similar to those of Iliopoulos et al. [20], our estimates have less variability. This implies that our prior structures are beneficial when used in a non-informative setting. The same situation also can be observed for most of the rescaled estimates of the column scores. For the RC_{ord} model, our posterior estimates of the age scores are similar to those of Iliopoulos et al. [20], while those of the disturbance are significantly different. The estimates of the posterior standard deviations of the disturbance scores are very high (8.06, 2.75, 5.71 for $x_{2[1]}$, $x_{2[3]}$ and $x_{2[4]}$, respectively) in the results given by Iliopoulos et al. [20]. We obtain smaller estimates, highlighted as 1.739, 1.428, and 1.065 for $x_{2[1]}$, $x_{2[3]}$, and $x_{2[4]}$, respectively. Our posterior estimates for the scores of disturbance are more stable and reliable than those of Iliopoulos et al. [20]. Note that this situation affects the results of model selection and would cause the difference between our model selection results and that of Iliopoulos. For the RC4

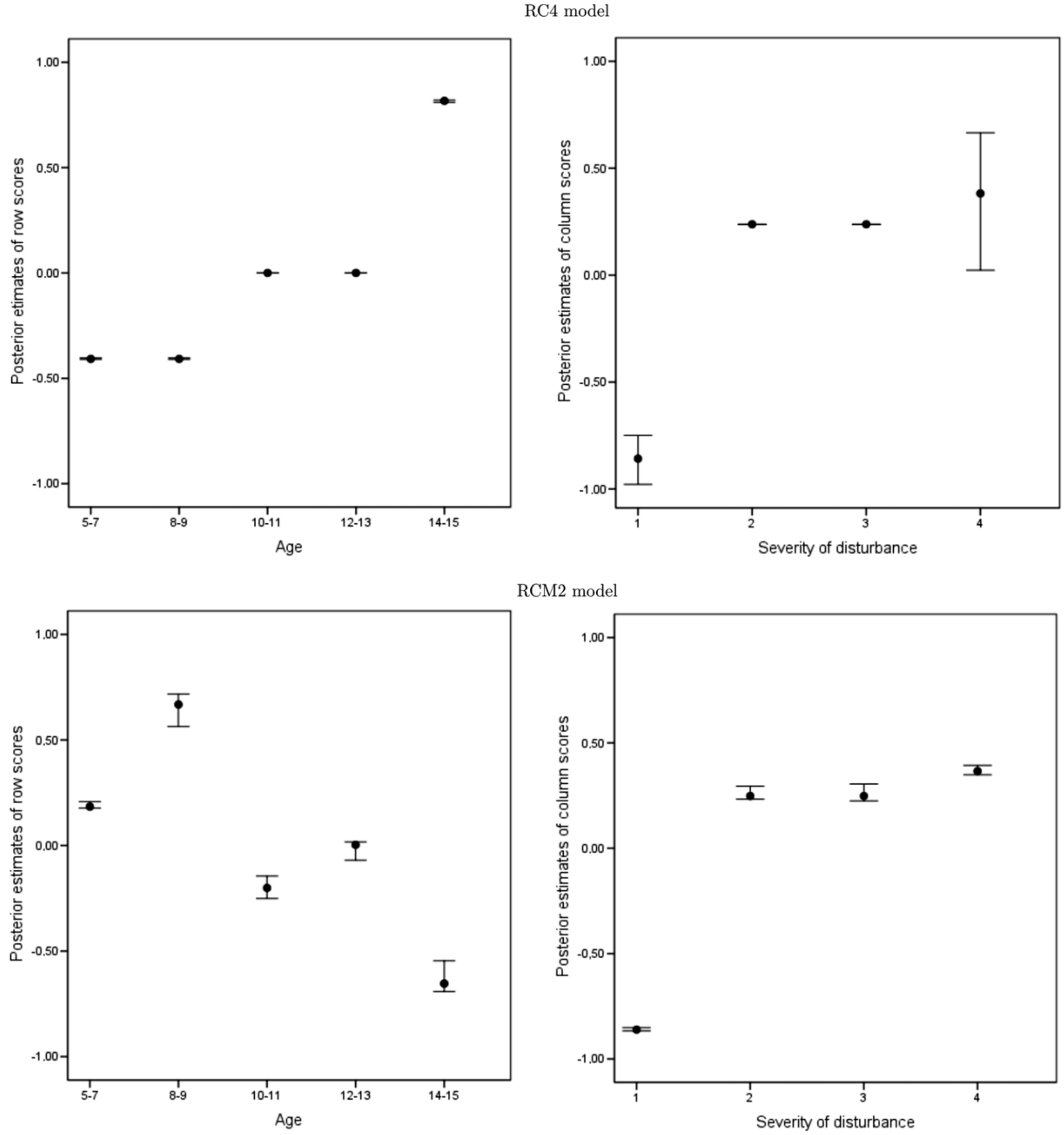

Fig. 2. Error bars for rescaled posterior estimates of row and column scores of RC4 and RCM2 models.

model, the same situation can be observed. The estimates of the posterior standard deviations of the disturbance scores provided by Iliopoulos et al. [20] are significantly higher (18.09 for both $x_{2[1]}$ and $x_{2[4]}$) in this case. Our corresponding estimates are 0.241 for both $x_{2[1]}$ and $x_{2[4]}$. Again, our posterior estimates of the scores of the disturbance are more reliable. The rest of our actual and rescaled estimates are in accordance with those of Iliopoulos et al. [20]. We obtain the smallest posterior standard deviations for the RCM2 model, in which the scores of age are unrestricted while those of disturbance are order restricted. The most stable rescaled posterior estimates can also be observed for the RMC2 model (Fig. 2). According to the RCM2 model, age level 8–9 is at the top, while level 14–15 is at the bottom of the hierarchy between age levels. The same inference is also valid for the RC model. Bearing in mind that the RC and RCM2 models are the best among the fitted models, it can be inferred that imposing order restrictions on the levels of age is inappropriate; 8–9 is the most important group with respect to the severity of dream disturbance. Because the RCM2 gives a better fit than the RC model, we infer that the order restriction on the levels of severity of disturbance is compatible with the hierarchy between the levels of disturbance.

To conduct a sensitivity analysis for our G-MVLG and ordered log-gamma prior structures, we use the measure proposed by Ruggeri and Sivaganesan [29]. We fit the RC and RC_{ord} models to evaluate the sensitivity of the G-MVLG and ordered

Table 5
Ranges of values of relative sensitivity measure over four prior settings.

RC			$\mathrm{RC}_{\mathrm{ord}}$	
Score	$r(\mathrm{RSM}_1)$	$r(\mathrm{RSM}_2)$	Score	$r(\mathrm{RSM}_{[1]})$
$x_{1(1)}$	0.756	0.714	$x_{1[1]}$	0.356
$x_{1(2)}$	0.399	2.053	$x_{1[2]}$	0.313
$x_{1(3)}$	0.519	2.272	$x_{1[3]}$	0.252
$x_{1(4)}$	0.675	0.904	$x_{1[4]}$	0.301
$x_{1(5)}$	0.645	1.081	$x_{1[5]}$	0.302
$x_{2(1)}$	0.689	1.797	$x_{2[1]}$	0.433
$x_{2(3)}$	0.532	2.366	$x_{2[3]}$	0.177
$x_{2(4)}$	0.721	1.472	$x_{2[4]}$	0.669
ϕ	0.507	0.777	ϕ	0.264

log-gamma prior structures. The relative sensitivity measure (RSM) of Ruggeri and Sivaganesan [29] is defined as follows:

$$\mathrm{RSM} = \frac{(E_\pi - E_0)^2}{V_\pi}, \tag{34}$$

where V_π is the posterior variance of the parameter considered with respect to the prior π, while E_0 and E_π are the posterior expectations of each parameter obtained for a prior π_0 and another plausible prior π, respectively. The prior π is a member of the class $\mathcal{C}$ of many plausible priors, including π_0. The RSM has a common interpretation irrespective of the problem context that a small (large) "range of $\sqrt{\mathrm{RSM}}$" ($r(\mathrm{RSM})$) values over $\mathcal{C}$ implies low (high) sensitivity. For the sensitivity of the G-MVLG prior, π_0 is taken as the G-MVLG prior with the hyper-parameters defined at the beginning of this section. We use four different π. We change only the elements of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, while the values of the rest of the hyper-parameters remain the same for all π. Algorithm 1 is run over the RC model. We set all elements of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ equal to 0.2, 0.25, 0.5, and 0.75 for π_1, π_3, π_3, and π_4, respectively. These values make the prior variance of each score 25, 16, 4, and 1.8, respectively. In this case, the ranges of the $\sqrt{\mathrm{RSM}}$ values obtained for the scores of the RC model over the four prior setups are shown in the $r(\mathrm{RSM}_1)$ column of Table 5. Then, we fix the values of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ to 0.3 and change the value of $g_{1(ts)} = g_{2(ts)}$ within the set $\mathcal{A} = \{0.1, 0.5, 0.9, 1.3, 1.7\}$ to evaluate the effect of including the possible correlations between adjacent scores. For this case, the ranges of the $\sqrt{\mathrm{RSM}}$ values obtained for the scores of the RC model over the four prior setups are shown in the $r(\mathrm{RSM}_2)$ column of Table 5. If inclusion of the correlations is effectual on the parameter estimates, we should obtain larger values for $r(\mathrm{RSM}_2)$. We follow the same procedure as that of the order-restricted case for the sensitivity analysis of log-gamma prior. We run Algorithm 1 over the $\mathrm{RC}_{\mathrm{ord}}$ model by setting all elements of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ equal to 0.2, 0.25, 0.5, and 0.75 for $\pi_{[1]}, \pi_{[2]}, \pi_{[3]}$, and $\pi_{[4]}$, respectively. This setting makes the prior variance of each score 26.3, 17.2, 4.9, and 2.5, respectively. For this case, the ranges of the $\sqrt{\mathrm{RSM}}$ values obtained for the scores of the $\mathrm{RC}_{\mathrm{ord}}$ model over the four prior setups are shown in the $r(\mathrm{RSM}_{[1]})$ column of Table 5. Scores $x_{2(3)}$ and $x_{2[3]}$ are not included in Table 5 because their values are fixed to 1 in the computations for identifiability purposes.

Considering the fact that our posterior variances are small, the $r(\mathrm{RSM}_1)$ and $r(\mathrm{RSM}_{[1]})$ values obtained for all scores are satisfactorily narrow, such that we may conclude the insensitivity of both of our G-MVLG and log-gamma prior structures with respect to the prior variance over the RC and $\mathrm{RC}_{\mathrm{ord}}$ models. It can be seen from this sensitivity analysis that even a value of prior variance near 2 produces a non-informative setting for both of the G-MVLG and log-gamma priors. One should use values smaller than 2 for the prior variances to obtain an informative analysis for this data set. As expected, the $r(\mathrm{RSM}_2)$ values obtained by changing the prior information on the correlations of the adjacent scores are larger than the $r(\mathrm{RSM}_1)$ values. Thus, our G-MVLG prior structure is sensitive to changes in the prior information on the correlations, and the incorporation of the correlations is effectual on the parameter estimates.

We compare the sensitivity of our approach with that of the approaches of Iliopoulos et al. [20] by using $r(\mathrm{RSM})$. The results of this sensitivity analysis are presented in Appendix C. Thus, we figure out that the sensitivity of our approach in the unrestricted case is significantly less than that of the corresponding approach of Iliopoulos et al. [20]. The lower insensitivity is another beneficial feature of our G-MVLG approach in the non-informative setting. In the order-restricted case, both our approach and that of Iliopoulos et al. [20] are satisfactorily insensitive according to the $r(\mathrm{RSM})$.

In conclusion, we can identify the negative association between the age and severity of dream disturbance, consistent with the results of other authors who have analyzed this data set. Unlike the other authors, we identify an order restriction on the scores of disturbance and no restriction over the scores of age. The second and third levels of disturbance are indistinguishable within the hierarchy of the levels of disturbance. Accordingly, we obtain that the second level of age is at the top of the hierarchy between the age levels; hence, the ages of 8 and 9 are the most important ages with respect to the severity of dream disturbance. The differences between our inferences and previous ones are due to the amount of information considered in the analysis. Because we include information regarding the possible correlations between the adjacent levels of ordinal variables, we are working with more information relevant to the subject of interest.

7. Discussion

In this article, we propose new prior structures for the Bayesian estimation of association models with and without order restrictions on scores. We consider the exchangeability assumption for prior distributions. This assumption should be taken into account for prior distributions related with ordinal categorical variables because the hierarchy between the levels of the factors composing the table of interest violates the exchangeability assumption. We propose a prior determination strategy for this issue.

When there is no order restriction on the scores, we induce a generalized multivariate log-gamma (G-MVLG) prior on scores and independent log-gamma priors on the main effect parameters of an association model. The G-MVLG prior is able to represent the possible correlation structures between the adjacent scores and score groups. Prior knowledge regarding the scores and model parameters and the degree of belief in this knowledge are easily represented using the G-MVLG prior. One should be careful when placing prior information on the correlations because that it should not conflict with the data. One can obtain such prior information by evaluating the direction of hierarchy and the characteristics of the variable of interest. Because it is possible to obtain a simple and tractable full conditional posterior distribution when the G-MVLG prior is used, the derivations of posterior inferences for scores and model parameters are very simple when using Metropolis steps within Gibbs sampling.

In the presence of order restrictions on scores, we utilize the joint probability density function (pdf) of order statistics and assume the independence of the scores. Independence is a reasonable assumption because an order restriction also represents strong knowledge regarding the dependency between the scores assigned to the levels of a factor. We place a log-gamma prior on each score and arrange the scores according to the relevant order restriction, handling them as order statistics. The joint pdf of the order statistics from a log-gamma-distributed population gives us the joint prior distribution of the scores under order restrictions. We derive the required full conditional posterior distributions to run the Gibbs sampling in a tractable form for posterior inferences in the order-restricted case. Therefore, our approaches are computationally efficient. We provide a detailed algorithm for the posterior computations in both the unrestricted and restricted cases.

We give a brief outline on the adaptation of the approach of Chib and Jeliazkov [8] to our approaches for the calculation of the Bayes factors for model comparison.

To illustrate our approaches and examine the impacts of order restrictions, we analyze a cross classification of age and the severity of dream disturbance, which includes a two-way contingency table composed of ordinal variables and a three-way breathing tests results classification. In accordance with the theory, our approach successfully imposes order restrictions in the application. When we compare the results of our restricted analysis with the unrestricted analysis, we see that order restrictions have significant impacts on inferences. Therefore, researchers should be careful when placing order restrictions on the scores. Researchers should, at least, verify the direction of hierarchy between the levels of the factor of interest by conducting an unrestricted and non-informative Bayesian analysis of the model intended for use under the order restrictions. For the dream disturbance data, our inferences are somewhat different from the inferences derived by other authors due to the inclusion of information regarding the correlations between adjacent scores. Our estimates of the posterior distributions are smaller, and we identify a best-fitting model that includes both the order-restricted and unrestricted scores.

As future work, we intend to present the reversible jump Markov chain Monte Carlo algorithms for selection and evaluation of order-restricted and unrestricted association models over our prior structures.

Acknowledgments

The author would like to thank the anonymous referee for constructive and helpful comments, which considerably enriched the content and improved the clarity and quality of the manuscript. The author would also like to thank Dr. Zeynep Kalaylıoğlu for valuable discussions on the topic and for carefully reading the manuscript.

Appendix A. Algorithm for posterior inferences

0. Initialize $\boldsymbol{u}_i^{(0)}, \boldsymbol{x}_i^{(0)}, \boldsymbol{\gamma}_i^{(0)}$, calculate $\delta_i^{(0)}$ by using $\boldsymbol{\gamma}_i^{(0)}$ for $i = 1, 2$.

1.1. If unrestricted analysis is being undertaken, read parameters of prior distributions of main effects, $g_{i(ts)}$ for $t \neq s, t, s = 2, \ldots, I$ when $i = 1$, and $t, s = 3, \ldots, J$ when $i = 2$, ν_i, $\boldsymbol{\mu}_i$ and $\boldsymbol{B}_i$, and calculate $\boldsymbol{\lambda}_i$ as described in Section 3.3 for $i = 1, 2$.

1.2. If order-restricted analysis is being undertaken, read parameters of prior distributions of main effects, $\boldsymbol{\theta}_i$ and $\boldsymbol{B}_i$, and calculate $\boldsymbol{\eta}_i$ as described in Section 3.3 for $i = 1, 2$.

2. Set $r = 1, i = 1$ and $\boldsymbol{u}_i^{(r)} = \boldsymbol{u}_i^{(r-1)}$.

3. Draw $u_{i(t)}^{(r)}$ from $N\left(u_{i(t)}^{(r-1)}, \sigma^2_{u_{i(t)}}\right)$ for $t = 2, \ldots, K$.

4. Calculate $u_{i(1)}^{(r)}$ by using $u_{i(t)}^{r}, \boldsymbol{u}_{-i(t)}^{(r-1)}, \boldsymbol{x}^{(r-1)}$ over the relevant constraint given in (5).

5. Calculate u by using $u_{i(t)}^{(r)}, \boldsymbol{u}_{-i(t)}^{(r-1)}, \boldsymbol{x}^{(r-1)}$ over (20).

6. Accept the proposed move with the following acceptance probability:

$$\alpha\left(u_{i(t)}^{(r-1)}, u_{i(t)}^{(r)}\right) = \min\left\{1, \frac{p\left(u_{i(t)}^{(r)}|\mathbf{g}\right)}{p\left(u_{i(t)}^{(r-1)}|\mathbf{g}\right)}\right\}.$$

7. Set $i = i + 1$ and go to step 2 while $i \le 2$ (all main effects are updated).
8. Set $i = 0$, and if unrestricted case is precessing go to step 9 else if order-restricted case is processing go to step 19.
9. Set $i = i + 1$ and $\boldsymbol{x}_i^{(r)} = \boldsymbol{x}_i^{(r-1)}$.
10. Draw $x_{i(t)}^{(r)}$ from $N\left(x_{i(t)}^{(r-1)}, \sigma_{x_{i(t)}}^2\right)$ for $t = 2, \ldots, I$ if $i = 1$ and $t = 3, \ldots, J$ if $i = 2$.
11. Calculate $x_{i(1)}^{(r)}$ by using $x_{i(t)}^{r}, \boldsymbol{x}_{-i(t)}^{(r-1)}$ over the relevant constraint given in (6).
12. Calculate u by using $x_{i(t)}^{(r)}, \boldsymbol{x}_{-i(t)}^{(r-1)}, \boldsymbol{u}^{(r)}$ over (20).
13. Draw all elements of $\boldsymbol{\gamma}_i$ from the distribution mentioned in (12) by using the predetermined $g_{i(ts)}$ values.
14. Calculate $\rho_{i(ts)}$ values by using the generated values of $\gamma_{i(ts)}$ at step 13 over Eq. (10).
15. Construct the matrix $\boldsymbol{\Omega}$ as given in (13), and calculate the value of $\delta_i^{(r)}$ as mentioned in Section 3.2.
16. Accept the proposed move for δ_i with the following acceptance probability:

$$\alpha\left(\delta_i^{(r-1)}, \delta_i^{(r)}\right) = \min\left\{1, \frac{p\left(\delta_i^{(r)}|\mathbf{g}\right)}{p\left(\delta_i^{(r-1)}|\mathbf{g}\right)}\right\}.$$

17. Accept the proposed move for each score with the following acceptance probability:

$$\alpha\left(x_{i(t)}^{(r-1)}, x_{i(t)}^{(r)}\right) = \min\left\{1, \frac{p\left(x_{i(t)}^{(r)}|\mathbf{g}\right)}{p\left(x_{i(t)}^{(r-1)}|\mathbf{g}\right)}\right\}.$$

18. If $i \le 2$, go to step 8, else go to step 24.
19. Set $i = i + 1$ and $\boldsymbol{x}_i^{(r)} = \boldsymbol{x}_i^{(r-1)}$.
20. Draw $x_{i[t]}^{(r)}$ from $Truncated\text{–}Gamma\left(1, \beta_{i[t]}\right)$ which is bounded below and above with $x_{i[t-1]}$ and $x_{i[t+1]}$, respectively.
21. Calculate $x_{i[1]}^{(r)}$ by using $x_{i[t]}^{r}, \boldsymbol{x}_{-i[t]}^{(r-1)}$ over the relevant constraint given in (6).
22. Calculate u by using $x_{i[t]}^{(r)}, \boldsymbol{x}_{-i[t]}^{(r-1)}, \boldsymbol{u}^{(r)}$ over (20).
23. Accept the proposed move for each score with the following acceptance probability:

$$\alpha\left(x_{i[t]}^{(r-1)}, x_{i[t]}^{(r)}\right) = \min\left\{1, \frac{p\left(x_{i[t]}^{(r)}|\mathbf{g}\right)}{p\left(x_{i[t]}^{(r-1)}|\mathbf{g}\right)}\right\}.$$

24. If $i \le 2$, go to step 18, else go to step 24.
25. Set $r = r + 1$ and go to step 2 until the total number of iterations is accomplished.

Appendix B. Analysis of the breathing test results data set

Analysis of the breathing test results data set is presented in supplementary material that can be found online at http://dx.doi.org/10.1016/j.jmva.2013.06.008.

Appendix C. Comparison of sensitivities of prior distributions

Results of sensitivity analysis of the approaches of Iliopoulos et al. [20] are presented in the supplementary material that can be found online at http://dx.doi.org/10.1016/j.jmva.2013.06.008.

References

[1] A. Agresti, Analysis of Ordinal Categorical Data, Wiley, New York, 1984.
[2] A. Agresti, Categorical Data Analysis, second ed., Wiley, New York, 2002.
[3] A. Agresti, C. Chuang, A. Kezouh, Order-restricted score parameters in association models for contingency tables, Journal of the American Statistical Association 82 (1987) 619–623.
[4] A. Azzalini, D. Valle, The multivariate skew normal distribution, Biometrika 83 (1996) 715–726.
[5] F. Bartolucci, A. Forcina, Extended RC association models allowing for order restrictions and marginal modeling, Journal of the American Statistical Association 97 (2002) 1192–1199.

[6] E.J. Beh, T.B. Farver, An evaluation of non-iterative methods for estimating the linear-by-linear parameter of ordinal log-linear models, Australian and New Zealand Journal of Statistics 335–352 (2009) 1–37.
[7] C. Chen, D. Dunson, Random effects selection in linear mixed models, Biometrics 59 (2003) 762–769.
[8] S. Chib, I. Jeliazkov, Marginal likelihood from the Metropolis–Hastings output, Journal of the American Statistical Association 96 (2001) 270–281.
[9] C. Chuang, Empirical Bayes methods for a two-way multiplicative-interaction model, Communications in Statistics—Theory and Methods 11 (1982) 2977–2989.
[10] M. Daniels, M. Pourahmadi, Bayesian analysis of covariance matrices and dynamic models for longitudinal data, Biometrika 89 (2002) 553–566.
[11] H. Demirhan, C. Hamurkaroglu, Bayesian estimation of log odds ratios from $r \times c$ and $2 \times 2 \times k$ contingency tables, Statistica Neerlandica 62 (2008) 405–512.
[12] H. Demirhan, C. Hamurkaroglu, On a multivariate log-gamma distribution and the use of the distribution in the Bayesian analysis, Journal of Statistical Planning and Inference 141 (2011) 1141–1152.
[13] M. Evans, Z. Gilula, I. Guttman, Computational issues in the Bayesian analysis of categorical data: log-linear and Goodman's RC model, Statistica Sinica 3 (1993) 391–406.
[14] R. Forthofer, R. Lehnen, Public Program Analysis, Life Time Learning, Belmont, 1981.
[15] F. Galindo-Garre, J. Vermunt, The order-restricted association model: two estimation algorithms and issues in testing, Psychometrika 69 (2004) 641–654.
[16] A. Gelman, Inference and monitoring convergence, in: Markov Chain Monte Carlo in Practice, Chapman and Hall/CRC, 1996.
[17] L. Goodman, Simple models for analysis of association in cross-classifications having ordered categories, Journal of the American Statistical Association 74 (1979) 537–552.
[18] B. Graubard, E. Korn, Choice of column scores testing independence in ordered $2 \times k$ contingency tables, Biometrics 43 (1987) 471–476.
[19] L. Hubert, A note on Freeman's measure of association for relating an ordered to an unordered factor, Psychometrika 39 (1974) 517–520.
[20] G. Iliopoulos, M. Kateri, I. Ntzoufras, Bayesian estimation of unrestricted and order-restricted association models for a two-way contingency table, Computational Statistics and Data Analysis 51 (2007) 4643–4655.
[21] G. Iliopoulos, M. Kateri, I. Ntzoufras, Bayesian model comparison for the order restricted rc association model, Psychometrika 74 (2009) 561–587.
[22] R. Kass, A. Raftery, Bayes factors, Journal of the American Statistical Association 90 (1995) 773–795.
[23] M. Kateri, A. Nicolaou, I. Ntzoufras, Bayesian inference for the RC(m) association model, Journal of Computational and Graphical Statistics 14 (2005) 116–138.
[24] R. King, S. Brooks, Prior induction in log-linear models for general contingency table analysis, The Annals of Statistics 29 (2001) 715–747.
[25] A. Maxwell, Analyzing Qualitative Data, Methuen, London, 1961.
[26] A. Mira, G. Nicholls, Birdge estimation of the probability density at a point, Statistica Sinica 14 (2004) 603–612.
[27] Y. Ritov, Z. Gilula, The order-restricted rc model for ordered contingency tables: estimation and testing of fit, The Annals of Statistics 19 (1991) 2090–2101.
[28] Y. Ritov, Z. Gilula, Analysis of contingency tables by correspondence models subject to order constraints, Journal of the American Statistical Association 88 (1993) 1380–1387.
[29] F. Ruggeri, S. Sivaganesan, On a global sensitivity measure for Bayesian inference, Sankhyā: The Indian Journal of Statistics 62 (2000) 110–127.
[30] T. Schmelzer, L. Trefethen, Computing the gamma function using contour integrals and rational approximations, SIAM Journal on Numerical Analysis 45 (2007) 558–571.
[31] R. Sin, K. Wong, Multidimensional association models: a multilinear approach, Sociological Methods and Research 30 (2001) 197–240.