Bayesian Model Comparison for the Order Restricted RC Association Model

George Iliopoulos, Maria Kateri & Ioannis Ntzoufras

Department of Statistics and Insurance Science
University of Piraeus
Piraeus, Greece;
e-mails: {geh; mkateri}@unipi.gr

Department of Statistics
Athens University of Economics and Business
Athens, Greece;
e-mail: ntzoufras@aueb.gr

6th June 2008, 1st Athens-Pavia Meeting on Statistics, Mati, Attica, Greece
The presentation and the paper are available at: http://stat-athens.aueb.gr/~jbn/papers/paper18.htm.

Synopsis

1 Introduction

- Let $\boldsymbol{y} = (y_{ij})$ be the frequencies and
- $\boldsymbol{\Pi} = (\pi_{ij})$ be the probabilities

of an $I \times J$ contingency table of two *ordinal* variables X and Y with I and J levels respectively.

Saturated log-linear model:

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

$$\Downarrow$$

$$\log \pi_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + \boxed{\phi \mu_i \nu_j} \qquad \text{(Goodman, 1979, 1985)} \qquad (1)$$

where $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots \mu_I)$ and $\boldsymbol{\nu} = (\nu_1, \nu_2, \ldots \nu_J)$ be the scores assigned to the levels of X (rows) and Y (columns) respectively.

Interpretation of ϕ

- ϕ is an intrinsic association parameter.
- The above formulation reveals the analogies to the classical correspondence analysis (CA) or canonical correlation model.
- **Interpretation of** ϕ: Log odds ratio of successive categories if the score distances are equal to one since $\log\left(\frac{\pi_{ij}\pi_{i+1,j+1}}{\pi_{i,j+1}\pi_{i+1,j}}\right) = \phi(\mu_{i+1} - \mu_i)(\nu_{j+1} - \nu_j)$.

USUAL CONSTRAINTS

- Sum-to-zero constraints on row and column main effects (λ_i^X and λ_j^Y).
- Sum-to-zero constraints on row and column scores (μ_i and ν_j).
- Two additional constraints on the row and column scores are needed in order to achieve the identifiability of the model (this due to the fact that (1) is multiplicative and not linear to its parameters).

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

Aim of this work

- Work with the order restricted RC model.
- Use the Bayesian approach to identify which scores μ_i, μ_{i+1} and ν_j, ν_{j+1} can be merged.
- Use Reversible jump MCMC to estimate posterior model probabilities (and odds) of each model
- Implement Bayesian model averaging

Why Use the Bayesian Approach in this Problem?

- They are not approximate and can be implemented even for samples with small size or with sparse contingency tables.
- Score merging in classical methods can be done using stepwise like methods and sequential implementation of significance tests (significance level is higher than the specified one, different model may selected if different starting points are selected).
- Using RJMCMC (or other varying dimension MCMC method) we automatically search the model space and estimate posterior model probabilities.
- Bayesian model averaging can be used in straightforward manner.

2 Modeling Details

- We focus on the order restricted version of the RC association model.
- X and Y ordinal $\Rightarrow$ natural to assume that the ordinal structure for scores

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

- **Which successive scores (μ_i, μ_{i+1}) and (ν_j, ν_{j+1}) are equal?**
- In all models we assume that at least two row and two column scores are different.

Proposed Constraints

- We propose to use an alternative set of constraints:

$$\mu_1 = \mu_{\min} < \mu_I = \mu_{\max} \text{ and } \nu_1 = \nu_{\min} < \nu_J = \nu_{\max}$$

- Row and column scores take values in the intervals $[\mu_{\min}, \mu_{\max}]$ and $[\nu_{\min}, \nu_{\max}]$ respectively.
- Sensible choices:
  - $\mu_{\min} = \nu_{\min} = -1$ and $\mu_{\max} = \nu_{\max} = 1$ [range similar to the parameters under constraints (2)]
  - We use: $\mu_{\min} = \nu_{\min} = 0$ and $\mu_{\max} = \nu_{\max} = 1$
    - ∗ simplifies computations
    - ∗ $\phi = \log\left(\frac{\pi_{11}\pi_{IJ}}{\pi_{1J}\pi_{I1}}\right)$
- Posterior distributions of scores under (2) can be obtained by transforming MCMC output of the proposed parametrization.

Model Formulation

- We introduce latent binary indicators

$$\boldsymbol{\gamma} = (1, \gamma_2, \ldots, \gamma_I) \text{ and } \boldsymbol{\delta} = (1, \delta_2, \ldots, \delta_J) \text{ and}$$

which are equal to

$\gamma_i = 1$ when $\mu_i > \mu_{i-1}$ (or $\delta_j = 1$ when $\nu_j > \nu_{j-1}$)

$\gamma_i = 0$ when $\mu_i = \mu_{i-1}$ (or $\delta_j = 0$ when $\nu_j = \nu_{j-1}$)

- The vectors $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$:
  - specify which scores are equal
  - are used instead of the usual model indicator m
- Estimate posterior model probabilities $f(\boldsymbol{\gamma}, \boldsymbol{\delta}|\boldsymbol{y})$.

Let

$$\Gamma_i = \sum_{k=1}^{i} \gamma_k \text{ and } \Delta_j = \sum_{k=1}^{j} \delta_k$$

be the distinct scores under estimation until row i or column j respectively.

Moreover the actual distinct unequal row and column scores will be denoted by the vectors $\boldsymbol{\mu}_\gamma$ and $\boldsymbol{\nu}_\delta$ of dimension Γ_I and Δ_J respectively given by

$$\boldsymbol{\mu}_\gamma = \Big(\{\mu_i : \gamma_i = 1; i = 1, 2, \ldots, I\}\Big) = \Big(\mu_\gamma(1), \mu_\gamma(2), \ldots, \mu_\gamma(\Gamma_I)\Big)^T$$

and

$$\boldsymbol{\nu}_\delta = \Big(\{\nu_j : \delta_j = 1; j = 1, 2, \ldots, J\}\Big) = \Big(\nu_\delta(1), \nu_\delta(2), \ldots, \nu_\delta(\Delta_J)\Big)^T.$$

Then the original scores are given by

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

Example of the Notation

i	1, 2,	3, 4,	5
μ_i	$\mu_1 = \mu_2 = 0$	$\mu_3 = \mu_4 = 0.6$	$\mu_5 = 1$
γ_i	1, 0,	1, 0,	1
Γ_i	1, 1,	2, 2,	3
$\boldsymbol{\mu}_\gamma(\ell)$	0	0.6	1

μ_i $\quad \mu_\gamma(\Gamma_1) = \mu_\gamma(1) = 0,\quad \mu_\gamma(\Gamma_3) = \mu_\gamma(2) = 0.6,\quad \mu_\gamma(\Gamma_5) = \mu_\gamma(3) = 1$

$\mu_\gamma(\Gamma_2) = \mu_\gamma(1) = 0,\quad \mu_\gamma(\Gamma_4) = \mu_\gamma(2) = 0.6,$

Differences and Variable Selection Representation

Consider the row and column score differences

$$D_{\mu_i} = \mu_i - \mu_{i-1} \text{ and } D_{\nu_j} = \nu_j - \nu_{j-1}$$

instead or the original parameters. Then

$$\mu_i = \sum_{k=1}^{i} \gamma_k D_{\mu_k} \text{ and } \nu_j = \sum_{k=1}^{j} \delta_k D_{\nu_k};\quad i = 1, \ldots, I,\quad j = 1, \ldots, J\ .$$

For scores of range one ($R_\mu = \mu_{\max} - \mu_{\min} = 1$) $\Rightarrow \sum_{i=2}^{I} \gamma_i D_{\mu_i} = 1 \Rightarrow$ we may use

$$\boldsymbol{D}_\gamma = \Big(\{D_{\mu_i} : \gamma_i = 1\}\Big) \sim \mathcal{D}(\mathbf{1}_{\Gamma_I - 1})$$

(Dirichlet prior of dimension $\Gamma_I - 1$ with all parameters equal to one)

as non informative prior for row score differences.

Similarly, for column scores $\rightarrow \boldsymbol{D}_\delta = \Big(\{D_{\nu_j} : \delta_j = 0\}\Big) \sim \mathcal{D}(\mathbf{1}_{\Delta_J - 1})$.

Prior Distributions on Scores

Equivalently, the scores are a priori distributed as ordered iid uniform random variables

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

Similarly, for the column scores

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

Prior Distributions on the rest of parameters

Normal with large variances for the rest of the parameters.

Bernoulli for γ_i and δ_j with prior probabilities equal to 1/2.

3 RJMCMC algorithm

1. Update model structure: Sample $(\boldsymbol{\gamma}, \boldsymbol{\delta})$ using successive RJMCMC moves:

   For $i = 2, \ldots, I$, propose $\boldsymbol{\gamma}'$: $\gamma'_i = 1 - \gamma_i$, $\gamma'_k = \gamma_k$ for $k \neq i$.

Split: $(\gamma_i = 0) \rightarrow (\gamma'_i = 1)$	**Merge**: $(\gamma_i = 1) \rightarrow (\gamma'_i = 0)$
(a) Propose $(\mu_{i-1} = \mu_i) \rightarrow (\mu'_{i-1} < \mu'_i)$.	(a) Propose $(\mu_{i-1} < \mu_i) \rightarrow (\mu'_{i-1} = \mu'_i)$.
(b) Generate u from $q(u|\boldsymbol{\mu}, \boldsymbol{\gamma}, \boldsymbol{\gamma}')$.	(b) (No generation is needed).
(c) Set $\boldsymbol{\mu}'_{\gamma'} = g(\boldsymbol{\mu}_\gamma, u)$.	(c) Set $(\boldsymbol{\mu}'_{\gamma'}, u) = g^{-1}(\boldsymbol{\mu}_\gamma)$.

   (d) Calculate $\boldsymbol{\mu}'$ by $\mu_i = \mu_\gamma(\Gamma_i)$

   (e) Accept/reject the proposed move.

   $\boldsymbol{\delta}$ is updated similarly.

2. Update model parameters $(\boldsymbol{\lambda}^X, \boldsymbol{\lambda}^Y, \phi, \boldsymbol{\mu}, \boldsymbol{\nu})$, given the model structure $(\boldsymbol{\gamma}, \boldsymbol{\delta})$ using a metropolis within Gibbs scheme.

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

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

$|J|$ is the absolute value of the RJMCMC Jacobian used in the split move and is given by

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

Remains to specify ...

- the linking function $g(\boldsymbol{\mu}_\gamma, u)$
- the proposal density $q(u|\ )$

Merge Central Scores
$(\gamma_i = 1 \rightarrow \gamma'_i = 0,\ \ i : 2 < \Gamma_i = \ell < \Gamma_I)$

$$\Big(\ldots \le \mu_\gamma(\ell-2) < \underbrace{\mu_\gamma(\ell-1) \;{\color{red}<}\; \mu_\gamma(\ell)} < \mu_\gamma(\ell+1) \le \ldots\Big)$$

$$\Downarrow \qquad\qquad \Downarrow \qquad\qquad \Downarrow$$

$$\Big(\ldots \le \mu'_{\gamma'}(\ell-2) < \mu'_{\gamma'}(\ell-1) < \mu'_{\gamma'}(\ell) \le \ldots\Big)$$

$$\Downarrow$$

Usual transformation: $\mu'_{\gamma'}(\ell-1) = \frac{\mu_\gamma(\ell-1)+\mu_\gamma(\ell)}{2}$

and leave the rest of the scores unchanged

$$\mu'_{\gamma'}(k) = \begin{cases} \mu_\gamma(k) & \text{for } k \ < \ \ell-1 \\ \mu_\gamma(k+1) & \text{for } k \ > \ \ell-1 \end{cases}$$

Split Central Scores (inverse move)
$(\gamma_i = 0 \rightarrow \gamma'_i = 1,\ \ i : 2 \le \Gamma_i = \ell < \Gamma_I)$

$$\Big(\ldots \le \mu_\gamma(\ell-1) < \mu_\gamma(\ell) < \mu_\gamma(\ell+1) \le \ldots\Big)$$

$$\Downarrow \qquad\qquad \Downarrow \qquad\qquad \Downarrow$$

$$\Big(\ldots \le \mu'_{\gamma'}(\ell-1) < \overbrace{\mu'_{\gamma'}(\ell) \;{\color{red}<}\; \mu'_{\gamma'}(\ell+1)} < \mu'_{\gamma'}(\ell+2) \le \ldots\Big)$$

$$\downarrow \qquad\qquad \downarrow$$

$$\mu_\gamma(\ell) - u \qquad \mu_\gamma(\ell) + u$$

- Generate $u \in \Big(0, \min\big\{\ \mu_\gamma(\ell)-\mu_\gamma(\ell-1),\ \mu_\gamma(\ell+1)-\mu_\gamma(\ell)\ \big\}\Big)$

- Set $\mu'_{\gamma'}(\ell) = \mu_\gamma(\ell) - u$ and $\mu'_{\gamma'}(\ell+1) = \mu_\gamma(\ell) + u$.
- Leave the rest of the scores unchanged, i.e. set

$$\mu'_{\gamma'}(k) = \begin{cases} \mu_\gamma(k) & \text{for } k \ < \ \ell \\ \mu_\gamma(k-1) & \text{for } k \ > \ \ell+1 \end{cases}$$

From the above we have

- In Split Move: $|J| = 2$ and $u = \frac{\mu'_{\gamma'}(\ell+1)-\mu'_{\gamma'}(\ell)}{2}$
- Hence in Merge Move $\rightarrow |J| = \frac{1}{2}$ and $u = \frac{1}{2}\Big\{\mu_\gamma(\ell) - \mu_\gamma(\ell-1)\Big\}$.

PROBLEM

The above transformation cannot be applied for merging/spliting the **lowest** or the **highest** scores.

Merge the Lowest Scores $\mu_\gamma(1)$ and $\mu_\gamma(2)$
$(\gamma_i = 1 \rightarrow \gamma'_i = 0,\ \ i : \Gamma_i = 2)$

$$\begin{array}{llccc}
 & \underbrace{\mu_{\min} = \mu_\gamma(1) \;{\color{red}<}\; \mu_\gamma(2)} & < & \mu_\gamma(3) & < \ldots \\
 & \Downarrow & & \Downarrow & \\
 & \mu_{\min} = \mu'_{\gamma'}(1) & < & \mu'_{\gamma'}(2) & < \ldots \\
 & \Downarrow & & \Downarrow & \\
\textbf{Usual Transformation} & \frac{\mu_{\min}+\mu_\gamma(2)}{2} < & & \mu_\gamma(3) & < \ldots \\
\textbf{Not Valid Since} & \neq \mu_{\min} & & &
\end{array}$$

(VIOLATES THE CONSTRAINT $\mu'_{\gamma'}(1) = \mu_{\min}$)

Using similar logic we apply the following transformations

$$\begin{array}{ccccc}
\underbrace{\mu_{\min} = \mu_\gamma(1) \;{\color{red}<}\; \mu_\gamma(2)} & < & \mu_\gamma(3) & < \ldots < & \mu_\gamma(\Gamma_I) = \mu_{\max} \\
\Downarrow & & \Downarrow & & \Downarrow \\
\frac{\mu_{\min}+\mu_\gamma(2)}{2} & < & \mu_\gamma(3) & < \ldots < & \mu_\gamma(\Gamma_I) = \mu_{\max} \\
\Downarrow & & \Downarrow & & \Downarrow \\
0 & < & \mu_\gamma(3) - \frac{\mu_{\min}+\mu_\gamma(2)}{2} & < \ldots < & \mu_{\max} - \frac{\mu_{\min}+\mu_\gamma(2)}{2} \\
\Downarrow & & \Downarrow & & \Downarrow \\
0 & < & \frac{\mu_\gamma(3) - \frac{\mu_{\min}+\mu_\gamma(2)}{2}}{\mu_{\max} - \frac{\mu_{\min}+\mu_\gamma(2)}{2}} & < \ldots < & 1 \\
\Downarrow & & \Downarrow & & \Downarrow \\
\mu_{\min} & < & \mu_{\min} + \frac{2\mu_\gamma(3)-\mu_{\min}-\mu_\gamma(2)}{2\mu_{\max}-\mu_{\min}-\mu_\gamma(2)}(\mu_{\max}-\mu_{\min}) & < \ldots < & \mu_{\max} \\
\downarrow & & \downarrow & & \downarrow \\
\mu'_{\gamma'}(1) & < & \mu'_{\gamma'}(2) & < \ldots < & \mu'_{\gamma'}(\Gamma'_I)
\end{array}$$

Merge the Lowest Scores $\mu_\gamma(1)$ and $\mu_\gamma(2)$
$(\gamma_i = 1 \rightarrow \gamma'_i = 0,\ \ i : \Gamma_i = 2)$

Final transformation

$$\mu'_{\gamma'}(k) = \begin{cases} \mu_{\min}, & k = 1, \\ \mu_{\min} + (\mu_{\max}-\mu_{\min})\dfrac{2\mu_\gamma(k+1)-\mu_{\min}-\mu_\gamma(2)}{2\mu_{\max}-\mu_{\min}-\mu_\gamma(2)}, & k > 1. \end{cases} \qquad (3)$$

Split the Lowest Score $\mu_\gamma(1)$ (reverse move)
$(\gamma_i = 0 \rightarrow \gamma'_i = 1,\ \ i : \Gamma_i = 1)$

$$\Big(\mu_{\min} = \mu_\gamma(1) \quad < \quad \mu_\gamma(2) \quad < \ldots\Big)$$

$$\Downarrow \qquad\qquad \Downarrow$$

$$\Big(\overbrace{\mu_{\min} = \mu'_{\gamma'}(1) \;{\color{red}<}\; \mu'_{\gamma'}(2)} < \mu'_{\gamma'}(3) \quad < \ldots\Big)$$

- Set $\mu'_{\gamma'}(2) = u$.
- Generate u in the interval

$$u \in \left(\mu_{\min},\ \mu_\gamma(2) + \frac{(\mu_\gamma(2)-\mu_{min})[\mu_{max}-\mu_\gamma(2)]}{\mu_\gamma(2)+\mu_{max}-2\mu_{min}}\right).$$

Finally obtain the new proposed scores by

$$\mu'_{\gamma'}(k) = \begin{cases} \mu_{\min}, & k = 1, \\ u, & k = 2, \\ \dfrac{1}{2}\Big\{\mu_{\min} + u + (2\mu_{\max}-\mu_{\min}-u)\dfrac{\mu_\gamma(k-1)-\mu_{\min}}{\mu_{\max}-\mu_{\min}}\Big\}, & k > 2. \end{cases} \qquad (4)$$

(Inverse transformation of equation (3) - given in the corresponding merge move)

- In Split Move $\rightarrow |J| = \left(1 - \frac{1}{2}\frac{u-\mu_{\min}}{\mu_{\max}-\mu_{\min}}\right)^{\Gamma_I - 2}$
- In Merge Move $\rightarrow u = \mu_\gamma(2)$ and

$$|J| = \left[\left(1 - \frac{1}{2}\frac{u-\mu_{\min}}{\mu_{\max}-\mu_{\min}}\right)^{\Gamma_I'-2}\right]^{-1} = \left(1 - \frac{1}{2}\frac{\mu_\gamma(2)-\mu_{\min}}{\mu_{\max}-\mu_{\min}}\right)^{3-\Gamma_I}.$$

Reminder:

- Γ_I is the number of scores of the current model (In split "smaller", In merge: "larger" model)
- Γ_I' is the number of scores of the proposed model (In split "larger", In merge: "smaller" model)

Merge the Highest Scores $\mu_\gamma(\Gamma_I - 1)$ and $\mu_\gamma(\Gamma_I)$ $(\gamma_i = 1 \rightarrow \gamma_i' = 0,\ i : \Gamma_i = \Gamma_I\ )$

$$\begin{array}{ccccc} \mu_{\min} = \mu_\gamma(1) & < \ldots < & \mu_\gamma(\Gamma_I - 2) < & \underbrace{\mu_\gamma(\Gamma_I - 1) \color{red}{<} \mu_\gamma(\Gamma_I) = \mu_{\max}} \\ \Downarrow & & \Downarrow & \Downarrow \\ \mu_{\min} = \mu'_{\gamma'}(1) & < \ldots < & \mu'_{\gamma'}(\Gamma_I - 2) < & \mu'_{\gamma'}(\Gamma_I - 1) = \mu_{\max} \end{array}$$

Note: $\Gamma_I' = \Gamma_I - 1$ since we merge two scores into one.

$$\begin{array}{cccc} \mu_{\min} = \mu_\gamma(1) & < \ldots < & \mu_\gamma(\Gamma_I - 2) < & \underbrace{\mu_\gamma(\Gamma_I - 1) \color{red}{<} \mu_\gamma(\Gamma_I) = \mu_{\max}} \\ \Downarrow & & \Downarrow & \Downarrow \\ \mu_{\min} = \mu_\gamma(1) & < \ldots < & \mu_\gamma(\Gamma_I - 2) < & \frac{\mu_\gamma(\Gamma_I)+\mu_\gamma(\Gamma_I-1)}{2} \\ \Downarrow & & \Downarrow & \Downarrow \\ 0 & < \ldots < & \mu_\gamma(\Gamma_I - 2) - \mu_{\min} < & \frac{\mu_\gamma(\Gamma_I)+\mu_\gamma(\Gamma_I-1)}{2} - \mu_{\min} \\ \Downarrow & & \Downarrow & \Downarrow \\ 0 & < \ldots < & \frac{\mu_\gamma(\Gamma_I-2)-\mu_{\min}}{\frac{\mu_\gamma(\Gamma_I)+\mu_\gamma(\Gamma_I-1)}{2}-\mu_{\min}} < & 1 \\ \Downarrow & & \Downarrow & \Downarrow \\ \mu_{\min} & < \ldots < & \mu_{\min} + 2\frac{(\mu_{\max}-\mu_{\min})(\mu_\gamma(\Gamma_I-2)-\mu_{\min})}{\mu_\gamma(\Gamma_I)+\mu_\gamma(\Gamma_I-1)-2\mu_{\min}} < & \mu_{\max} \end{array}$$

Merge the Highest Scores $\mu_\gamma(\Gamma_I - 1)$ and $\mu_\gamma(\Gamma_I)$ $(\gamma_i = 1 \rightarrow \gamma_i' = 0,\ i : \Gamma_i = \Gamma_I\ )$

Final transformation

$$\mu'_{\gamma'}(k) = \begin{cases} \mu_{\min} + 2(\mu_{\max} - \mu_{\min})\dfrac{\mu_\gamma(k) - \mu_{\min}}{\mu_\gamma(\Gamma_I - 1) + \mu_{\max} - 2\mu_{\min}}, & k \leqslant \Gamma_I' - 1 = \Gamma_I - 2, \\ \mu_{\max}, & k = \Gamma_I' = \Gamma_I - 1. \end{cases} \quad (5)$$

Split the Highest Score $\mu_\gamma(\Gamma_I)$ (reverse move) $(\gamma_i = 0 \rightarrow \gamma_i' = 1,\ i : \Gamma_i = \Gamma_I\ )$

$$\begin{array}{cccc} \mu_{\min} = \mu_\gamma(1) & < \ldots < & \mu_\gamma(\Gamma_I - 1) < & \mu_\gamma(\Gamma_I) = \mu_{\max} \\ \Downarrow & & \Downarrow & \Downarrow \\ \mu_{\min} = \mu'_{\gamma'}(1) & < \ldots < & \mu'_{\gamma'}(\Gamma_I - 1) < & \overbrace{\mu'_{\gamma'}(\Gamma_I) \color{red}{<} \mu'_{\gamma'}(\Gamma_I + 1) = \mu_{\max}} \end{array}$$

- Generate u in the interval

$$u \in \left(0, 2\frac{\Big(\mu_{\max} - \mu_{\min}\Big)\Big(\mu_{\max} - \mu_\gamma(\Gamma_I - 1)\Big)}{\Big(\mu_{\max} - \mu_{\min}\Big) + \Big(\mu_{\max} - \mu_\gamma(\Gamma_I - 1)\Big)}\right)$$

- and set $\mu'_{\gamma'}(\Gamma_I' - 1) = \mu'_{\gamma'}(\Gamma_I) = \mu_{\max} - u$.

Split the Highest Score $\mu_\gamma(\Gamma_I)$ (reverse move) $(\gamma_i = 0 \rightarrow \gamma_i' = 1,\ i : \Gamma_i = \Gamma_I\ )$

Final transformation

$$\mu'_{\gamma'}(k) = \begin{cases} \mu_\gamma(k) - \frac{u}{2}\frac{\mu_\gamma(k)-\mu_{\min}}{\mu_{\max}-\mu_{\min}}, & k \leqslant \Gamma_I' - 2 = \Gamma_I - 1 \\ \mu_{\max} - u, & k = \Gamma_I' - 1 = \Gamma_I \\ \mu_{\max}, & k = \Gamma_I' = \Gamma_I + 1. \end{cases} \quad (6)$$

- In Split move
  - Determinant of the Jacobian: $|J| = \left(1 - \frac{1}{2}\frac{u}{\mu_{\max}-\mu_{\min}}\right)^{\Gamma_I - 2}$
  - Γ_I is the number of scores in the smaller (current) model.
- In the Merge move
  - $u = \mu_{\max} - \mu_\gamma(\Gamma_I - 1)$ and
  - Det. of Jacobian: $|J| = \left(1 - \frac{1}{2}\frac{u}{\mu_{\max}-\mu_{\min}}\right)^{2-\Gamma_I'} = \left(1 - \frac{1}{2}\frac{\mu_{\max}-\mu_\gamma(\Gamma_I-1)}{\mu_{\max}-\mu_{\min}}\right)^{3-\Gamma_I}$
  - Here:
    - Γ_I is the number of scores in the "bigger" (current) model.
    - Γ_I' is the number of scores in the "smaller" (proposed) model.

Additional Details

- In practice we have used $\mu_{\min} = \nu_{\min} = 0$ and $\mu_{\max} = \nu_{\max} = 1$.
- When $\Gamma_I = 2$ then only two scores are different and set equal to $\mu_{\min}$ and $\mu_{\max}$. No further splitting is allowed. Similar is the case for column scores ν_j.
- Rescaled Beta proposals can be used for proposing values for u.
- In practice we have used **Uniform** proposal which has been proved sufficient for datasets we have implemented the methodology.
- Further investigation is needed in order to construct proposals leading to more efficient RJMCMC schemes.

4 Illustrative Examples.

4.1 Simulated data.

- Monte Carlo study following Galindo-Garre and Vermunt (2004, Psychometrika).
- 1000 simulated datasets for a 5×3 contingency table with $\pi_{ij} = \exp(\phi^* \mu_i^* \nu_j^*)$.

  For the three models we have

  1. Model m_1: Different but equidistant Row + Column scores.
  2. Model m_2: $\mu_1^* = \mu_2^*$; rest of the scores are equidistant.
  3. Model m_3: $\mu_1^* = \mu_2^*$ and $\nu_2^* = \nu_3^*$; rest of the scores are equidistant.

  Furthermore we have

  - ϕ^*, μ_i^*, ν_j^* satisfy SSTO constraints.
  - Three different values of $\phi^* = 1, 2, 3$.
  - Two sample sizes $n = 100, 1000$

True Model (m_t)	ϕ	$n=1000$ Mean Rank	Rel. Freq. (%) $R_t=1$	$R_t \le 3$	$n=100$ Mean Rank	Rel. Freq. (%) $R_t=1$	$R_t \le 3$	Median $\log PO_{bt}$
m_1	1	2.04	0.442	0.866	10.50	0.025	0.092	0.955
All scores	2	1.08	0.930	1.000	5.17	0.107	0.378	0.722
Different	3	1.00	0.997	1.000	3.05	0.238	0.691	0.436
m_2	1	1.85	0.560	0.902	11.01	0.035	0.124	0.953
$\mu_1=\mu_2$	2	1.13	0.885	0.999	4.78	0.186	0.508	0.624
	3	1.09	0.910	1.000	2.90	0.341	0.732	0.329
m_3	1	2.09	0.519	0.859	10.16	0.053	0.242	0.777
$\mu_1=\mu_2$	2	1.20	0.847	0.989	4.37	0.184	0.566	0.543
$\nu_2=\nu_3$	3	1.11	0.897	0.999	2.86	0.329	0.766	0.345

m_t: True model

R_t: Ranking of posterior probability of model m_t in descending order,

True Model	n	ϕ	Monte Carlo Means of Posterior probabilities: $f(\gamma_i=1\|\boldsymbol{y})$ γ_2	γ_3	γ_4	γ_5	$f(\delta_j=1\|\boldsymbol{y})$ δ_2	δ_3
m_1	100	1	58	57	56	57	71	71
All scores		2	64	66	65	63	81	82
Different		3	68	73	73	68	92	92
	1000	1	74	79	78	73	97	97
		2	74	79	78	73	97	97
		3	99	99	99	99	100	100
m_2	100	1	48	56	59	61	71	71
$\mu_1=\mu_2$		2	42	65	72	68	81	83
		3	40	73	78	75	89	94
	1000	1	37	81	84	80	97	98
		2	28	99	98	98	100	100
		3	24	100	100	100	100	100
m_3	100	1	48	57	59	60	86	48
$\mu_1=\mu_2$		2	40	69	71	65	99	28
$\nu_2=\nu_3$		3	35	78	77	70	100	22
	1000	1	36	84	83	77	100	20
		2	25	99	98	92	100	13
		3	21	100	100	97	100	10

4.2 Application to Data

The method is also implemented in three datasets

1. Dreams Disturbance Data

   5×4 table; $n = 223$ children

   (Agresti et al., 1987, Ritov and Gilula, 1993).

2. Student Survey based Schizotypal Personality Questionnaire data

   7×6 table; 202 students.

3. Family size and happiness data

   5×4 table; $n = 1517$ families

   (see Clogg, 1982, Table 2, Galindo-Garre and Vermunt, 2004).

see for more details in `http://stat-athens.aueb.gr/~jbn/papers/paper18.htm`.

4.3 Dreams Disturbance Data.

Age Group	Disturbance (from low to high) 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

Results: Most frequently visited models

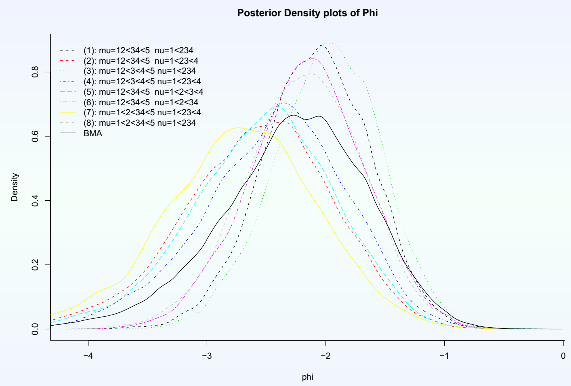
k	Model (scores)	Post. prob.	PO_{1k}	AIC	BIC	DIC	p_m	d_m
1	$\mu_1=\mu_2<\mu_3=\mu_4<\mu_5$ $\nu_1<\nu_2=\nu_3=\nu_4$	0.1620	1.00	1265.0	1295.7	1265.0	9.0	9
2	$\mu_1=\mu_2<\mu_3=\mu_4<\mu_5$ $\nu_1<\nu_2=\nu_3<\nu_4$	0.1540	1.05	1265.9	1300.0	1265.1	9.6	10
3	$\mu_1=\mu_2<\mu_3<\mu_4<\mu_5$ $\nu_1<\nu_2=\nu_3=\nu_4$	0.0877	1.85	1267.6	1301.6	1266.3	9.4	10
4	$\mu_1=\mu_2<\mu_3<\mu_4<\mu_5$ $\nu_1<\nu_2=\nu_3<\nu_4$	0.0725	2.23	1268.6	1306.1	1266.4	9.9	11
5	$\mu_1=\mu_2<\mu_3=\mu_4<\mu_5$ $\nu_1<\nu_2<\nu_3<\nu_4$	0.0609	2.66	1269.0	1306.5	1266.4	9.7	11
6	$\mu_1=\mu_2<\mu_3=\mu_4<\mu_5$ $\nu_1<\nu_2<\nu_3=\nu_4$	0.0579	2.80	1267.6	1301.7	1266.5	9.4	10
7	$\mu_1<\mu_2<\mu_3=\mu_4<\mu_5$ $\nu_1<\nu_2=\nu_3<\nu_4$	0.0541	2.99	1269.0	1306.5	1266.7	9.9	11
8	$\mu_1<\mu_2<\mu_3=\mu_4<\mu_5$ $\nu_1<\nu_2=\nu_3=\nu_4$	0.0522	3.10	1268.3	1302.4	1266.8	9.2	10

Single RJMCMC (R RESULTS): 100,000 iterations + additional burn-in of 10,000 iterations.

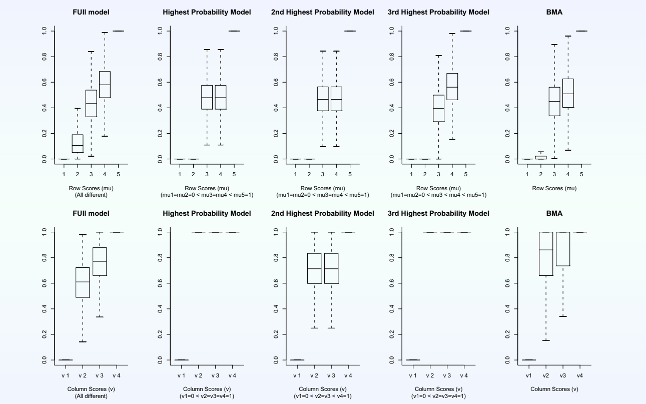
Results: Marginal Probabilities $f(\gamma_i = 1|\boldsymbol{y})$ and $f(\delta_j = 1|\boldsymbol{y})$

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

Single RJMCMC (R RESULTS): 100,000 iterations + additional burn-in of 10,000 iterations.



Posterior Distributions of ϕ over models with highest posterior probabilities .



Boxplots of the row and column scores over three models with highest posterior probabilties.

Some Comments on the Results

- Negative association between age and severity of dreams' distrurbance ($\phi < 0$).
- **Age**:
  - Categories 2-3 (8–9, 10–11 years old) and 4-5 (12–13, 14–15 years old) $\Rightarrow$ **different** in terms of the association (marginal post.prob. = 0.94 and 0.96 respectively).
  - Categories 1-2 (5–7, 8–9 years old) and 3-4 (10–11, 12–13years old) $\Rightarrow$ **indistinguishable** concerning the association (mild evidence with marginal post.probab.= 0.715 and 0.609 respectively).
- **Severity of dreams' disturbance**: More uncertainty is involved:
  - ⋄ Clear evidence that the first category differs than the rest [$f(\delta_2 = 1|\boldsymbol{y}) = 0.996$].
  - ⋄ Model with the highest posterior probability $\Rightarrow$ all the other three scores equal ($\nu_2 = \nu_3 = \nu_4$).
  - ⋄ Model with the 2nd highest posterior probability $\Rightarrow \nu_2 = \nu_3 < \nu_4$.
- The algorithm was highly mobile visiting 69, 86 and all 105 models in 10, 100 iterations 400 thousand iterations respectively.

Comparison to Previous Results

- RJMCMC indicated a more parsimonious model (according to highest posterior probability) than the one (2nd in rank) indicated by our previous analysis (see Iliopoulos *et al*. 2007).
- Agresti et al. (1987) proposed an order restricted C model under which $\hat{\nu}_1 < \hat{\nu}_2 = \hat{\nu}_3 < \hat{\nu}_4$.
- Ritov and Gilula (1993) suggested an order restriction model with $\hat{\nu}_1 < \hat{\nu}_2 = \hat{\nu}_3 < \hat{\nu}_4$ and $\hat{\mu}_1 = \hat{\mu}_2 < \hat{\mu}_3 = \hat{\mu}_4 < \hat{\mu}_5$ which is the second highest probability according to our method

5 Work in progress and future work

1. Comparison of the above models with the Uniform association, Independence and Saturated models [use different prior for ϕ].
2. Incorporate selection between unrestricted RC, Row, Column association models (can we use similar parametrization?)
3. Use similar approach in unrestricted RC model for merging/grouping scores
4. Expand methodology to high dimensional tables
5. Use different priors for scores; for example power prior and imaginary data.

Publications by the same Group

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.

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.

Iliopoulos, G., Kateri, M. and Ntzoufras, I. (2007). Bayesian Model Comparison for the Order Restricted RC Association Model. *Technical Report*, Dep. of Statistics, Athens University of Economics and Business. 1st Draft: 23/6/2007.

Related Work

Tarantola, C., Consonni, G. and Dellaportas, P. (2007) Bayesian clustering for row effects models. *Technical Report*, University of Pavia (submitted).

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.

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. (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.

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

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.