Probability Based Independence Sampler for Bayesian Quantitative Learning in Graphical Log-Linear Marginal Models

Ioannis Ntzoufras[1], **Claudia Tarantola**[2] & **Monia Lupparelli**[3]

[1] Department of Statistics, Athens University of Economics and Business, Greece;
[2] Department of Economics and Management, University of Pavia, Italy;
[3] Department of Statistical Sciences, University of Bologna, Italy

1. Motivation

Statistical models which impose restrictions on marginal distributions of categorical data have received considerable attention especially in social and economic science.
AIM: To develop a fully automatic, efficient MCMC strategy for quantitative learning for graphical log-linear marginal models.

WHY:
(a) Not many Bayesian methods;
(b) No conjugate analysis is available.
(c) The likelihood cannot be analytically expressed as a function of the marginal log-linear interactions.
⇒ Difficulties on the implementation of MCMC
⇒ At each MCMC iteration, an iterative procedure is applied to calculate the cell probabilities
(d) Construct algorithm which generates parameter values with compatible marginals.

Extension of our previous related work served as basis

- Ntzoufras and Tarantola (2013) → implemented in probability based parameters

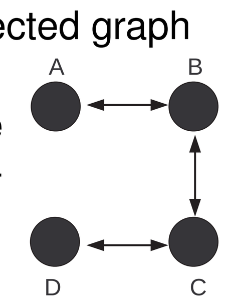
2. General framework

We consider the log-linear marginal models introduced by Bergsma and Rudas (2002):

$$\boldsymbol{\lambda} = \boldsymbol{C} \log\left(\boldsymbol{M}\boldsymbol{P}\right) \text{ with } \boldsymbol{P} = \text{vec}(\boldsymbol{p})$$

where $\boldsymbol{p}$ is the table of joint probabilities and $\boldsymbol{P}$ is the vectorized version of $\boldsymbol{p}$.

- Estimation using the frequencies of appropriate marginal contingency tables, and expressed in terms of log-odds ratios.
- Important in cases where information is available for specific marginal associations via odds ratios or when partial information (i.e. marginals) is available.
- It is defined by imposing zero constraints on specific log-linear interactions (Lupparelli *et al.*, 2009).
- The model can be represented by a bi-directed graph like the one besides; where a missing edge indicates that the corresponding variables are marginal independent.



3. Main Problems in Bayesian Analysis of Graphical Log-Linear Marginal Models

- Graphical log-linear marginal models belong to curved exponential families that are difficult to handle from a Bayesian perspective.
- The likelihood cannot be analytically expressed as a function of the marginal log-linear interactions.
- Posterior distributions cannot be directly obtained, and MCMC methods are needed.
- A well-defined model requires parameter values that lead to compatible marginal probabilities.

4. A novel MCMC strategy

- New fully automatic and efficient MCMC strategy.
- It handles the problems previously discussed.
- Prior: is expressed in terms of the marginal log-linear interactions
- Proposal: is defined on the probability parameter space.

Advantages

- The joint distribution factorises under certain conditional independence models, and the likelihood can be directly expressed in terms of probability parameters.
- Efficient proposal distributions: by exploiting a conditional conjugate approach of Ntzoufras and Tarantola (2013).
- Working on the Probability space ⇒ Always compatible marginals & contraints on marginal log-odds (i.e. interactions) are imposed automatically.

5. Prior Specification

Let $\vec{\boldsymbol{\lambda}}$ be the set of elements of $\boldsymbol{\lambda}$ not restricted to zero by the graphical structure.

1. Approach 1: Assign relatively flat normal priors on each element of $\vec{\boldsymbol{\lambda}}$.
2. Approach 2: Bases on the Dellaportas and Forster (1999) prior for standard log-linear models.
   - We work separately on each single set $\boldsymbol{\lambda}^{M_m}$ obtained from marginal M_m
   - $\boldsymbol{\lambda}_S^{M_m}$: parameter vector for the saturated model S that can be estimated from marginal M_m. By construction, it coincides with the parameter vector of the saturated standard log-linear model obtained from this marginal.
   - We implement the DF prior on the saturated model of each marginal, ending up to the prior

$$\boldsymbol{\lambda}^{M_m} \sim N\Big(\boldsymbol{\theta} - \log(N)\, \mathbf{X}_{M_m}^{-1}\mathbf{1},\;\; 2|\mathcal{I}_{M_m}|\left(\boldsymbol{X}_{M_m}^T \boldsymbol{X}_{M_m}\right)^{-1}\Big)$$

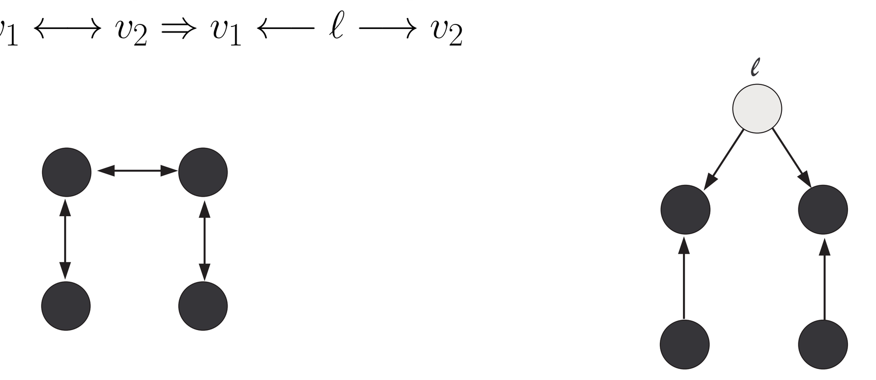
where $\boldsymbol{\theta} = (\log \overline{\boldsymbol{n}}, 0, \ldots, 0)^T$ is the prior mean of DF; $\boldsymbol{X}_M$= sum to zero design matrix for the marginal table M, and N is the total sample size (sum of all frequencies).

6. Augmented DAG Representation

Graphical log-linear marginal models ⇒ compatible (in terms of independencies), with a certain augmented DAG; see e.g. Cox and Wermuth (1993).

Construction of Augmented DAG:

- We consider the skeleton $\overline{G}$ of G
- Constructing the sink orientation of G: Assign arrows $v_i \longrightarrow v_j \longleftarrow v_k$ to each ∨ configuration of $\overline{G}$.
- For every bi-directed edge, we introduce a latent node ℓ: $v_1 \longleftrightarrow v_2 \Rightarrow v_1 \longleftarrow \ell \longrightarrow v_2$



(a) Bi-directed graph (b) Augmented DAG representation

- Augmented DAG ⇒ standard factorisation of conditional probability parameters $\boldsymbol{\Pi}$ ⇒ (Conditional) conjugate Bayesian approach.

7. The Proposed General MCMC Algorithm

For $t = 1, \ldots, T$, repeat the following steps:

1. Propose a new vector $\boldsymbol{\Pi}'$ from $q(\boldsymbol{\Pi}'|\boldsymbol{\Pi}^{(t)})$.
2. From $\boldsymbol{\Pi}'$, calculate the proposed joint probabilities $\boldsymbol{p}'$ (for the observed table).
3. From $\boldsymbol{p}'$, calculate $\boldsymbol{\lambda}'$ and the non-zero elements $\vec{\boldsymbol{\lambda}}'$.
4. Set $\boldsymbol{\xi}' = \boldsymbol{\Pi}'_\xi$; where $\boldsymbol{\Pi}'_\xi$ is a pre-specified subset of $\boldsymbol{\Pi}'$ of dimension $\dim(\boldsymbol{\Pi}) - \dim(\vec{\boldsymbol{\lambda}})$.
5. Accept the proposed move with probability $\alpha = \min(1, A)$ with

$$A = \frac{f(\boldsymbol{n}|\boldsymbol{\Pi}')f\left(\vec{\boldsymbol{\lambda}}'\right)f(\boldsymbol{\xi}')q(\boldsymbol{\Pi}^{(t)}|\boldsymbol{\Pi}')}{f(\boldsymbol{n}|\boldsymbol{\Pi}^{(t)})f\left(\vec{\boldsymbol{\lambda}}^{(t)}\right)f(\boldsymbol{\xi})q(\boldsymbol{\Pi}'|\boldsymbol{\Pi}^{(t)})} \times \text{abs}\left(\frac{\mathcal{J}\left(\boldsymbol{\Pi}^{(t)}, \vec{\boldsymbol{\lambda}}^{(t)}, \boldsymbol{\xi}^{(t)}\right)}{\mathcal{J}\left(\boldsymbol{\Pi}', \vec{\boldsymbol{\lambda}}', \boldsymbol{\xi}'\right)}\right) \quad (1)$$

$\boldsymbol{\Pi}_\xi = \boldsymbol{\xi}$, and $\mathcal{J} = \mathcal{J}(\boldsymbol{\Pi}, \vec{\boldsymbol{\lambda}}, \boldsymbol{\xi})$ is the determinant of the jacobian matrix of the transformation $\boldsymbol{\Pi} = g(\vec{\boldsymbol{\lambda}}, \boldsymbol{\xi})$.

Probability Based Independence Sampler (PBIS)

Efficient proposal:

$$q(\boldsymbol{\Pi}'|\boldsymbol{\Pi}^{(t)}) = f_q\left(\boldsymbol{\Pi}'|\boldsymbol{n}^{\mathcal{A}}\right) f\left(\boldsymbol{n}^{\mathcal{A}}|\boldsymbol{\Pi}^{(t)}, \boldsymbol{n}\right),$$

where $\boldsymbol{n}^{\mathcal{A}}$ is an augmented table.
We exploit the conditional conjugate approach of Ntzoufras and Tarantola (2013).
We consider as a "prior" $f_q(\boldsymbol{\Pi})$ a product of Dirichlet distributions obtaining a conjugate "posterior" distribution $f_q(\boldsymbol{\Pi}'|\boldsymbol{n}'^{\mathcal{A}})$.
The acceptance rate in (1) becomes equal to

$$A = \frac{f\left(\boldsymbol{n}^{\mathcal{A}(t)}|\boldsymbol{\Pi}'\right)f\left(\vec{\boldsymbol{\lambda}}'\right)f_q\left(\boldsymbol{\Pi}^{(t)}|\boldsymbol{n}^{\mathcal{A}(t)}\right)}{f\left(\boldsymbol{n}'^{\mathcal{A}}|\boldsymbol{\Pi}^{(t)}\right)f\left(\vec{\boldsymbol{\lambda}}^{(t)}\right)f_q\left(\boldsymbol{\Pi}'|\boldsymbol{n}'^{\mathcal{A}}\right)} \times \text{abs}\left(\frac{\mathcal{J}\left(\boldsymbol{\Pi}^{(t)}, \vec{\boldsymbol{\lambda}}^{(t)}, \boldsymbol{\xi}^{(t)}\right)}{\mathcal{J}\left(\boldsymbol{\Pi}', \vec{\boldsymbol{\lambda}}', \boldsymbol{\xi}'\right)}\right).$$

Prior Adjustment Algorithm (PAA)

We simplify PBIS as follows:

Step 1: Run the Gibbs sampler of Ntzoufras and Tarantola (2013) to obtain a sample from the joint probability distribution of the observed variables .

Step 2: Use the sample of step 1 (or sub-sample of it) as a proposal in the general Metropolis-Hastings algorithm).

The acceptance rate becomes

$$A = \frac{f(\vec{\boldsymbol{\lambda}}')f_q(\boldsymbol{\Pi}^{(t)})}{f(\vec{\boldsymbol{\lambda}}^{(t)})f_q(\boldsymbol{\Pi}')} \times \text{abs}\left(\frac{\mathcal{J}\left(\boldsymbol{\Pi}^{(t)}, \vec{\boldsymbol{\lambda}}^{(t)}, \boldsymbol{\xi}^{(t)}\right)}{\mathcal{J}\left(\boldsymbol{\Pi}', \vec{\boldsymbol{\lambda}}', \boldsymbol{\xi}'\right)}\right).$$

8. Simulation Study

Model: Same as in Graph of Section 2.

We present results for

- A specific single dataset &
- 100 Simulated datasets.

Simulation Plan: True Values

Marginal	Active interactions
AC	$\lambda_\emptyset^{AC} = -1.40, \lambda_A^{AC}(2) = -0.15, \lambda_C^{AC}(2) = 0.10$
AD	$\lambda_B^{AD}(2) = 0.12,$
BD	$\lambda_D^{BD}(2) = -0.09,$
ACD	$\lambda_{CD}^{ACD}(2,2) = 0.20,$
ABD	$\lambda_{AB}^{ABD}(2,2) = -0.15,$
ABCD	$\lambda_{BC}^{ABCD}(2,2) = -0.30, \lambda_{ABC}^{ABCD}(2,2,2) = 0.15,$
	$\lambda_{BCD}^{ABCD}(2,2,2) = -0.10, \lambda_{ABCD}^{ABCD}(2,2,2) = 0.07.$

Zero interactions: $\lambda_{AC}^{AC} = \lambda_{AD}^{AD}(2,2) = \lambda_{BD}^{BD}(2,2) = \lambda_{ACD}^{ACD}(2,2,2) = \lambda_{ABD}^{ABD}(2,2,2) = 0$

Figure 1: *ESS per second of CPU time for the single simulated dataset*

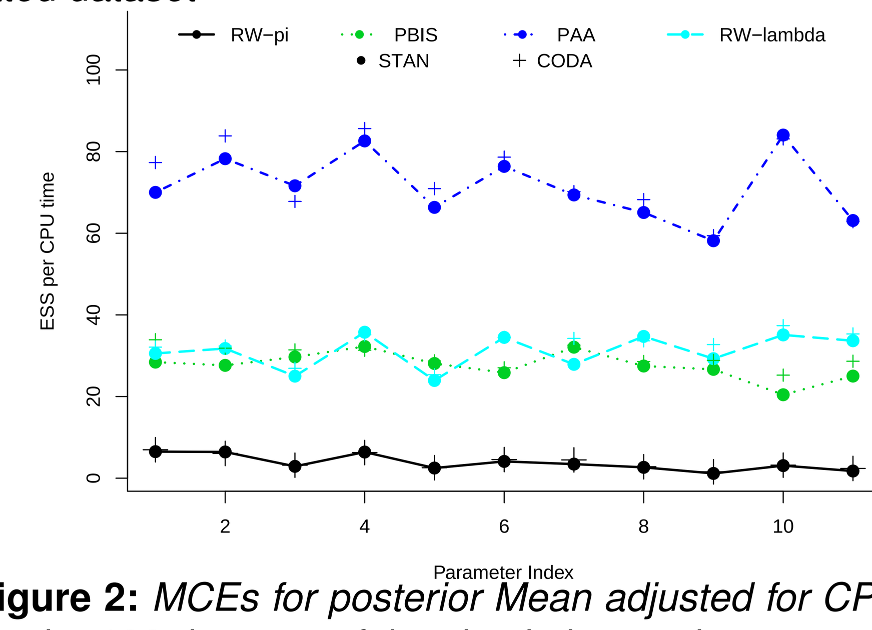
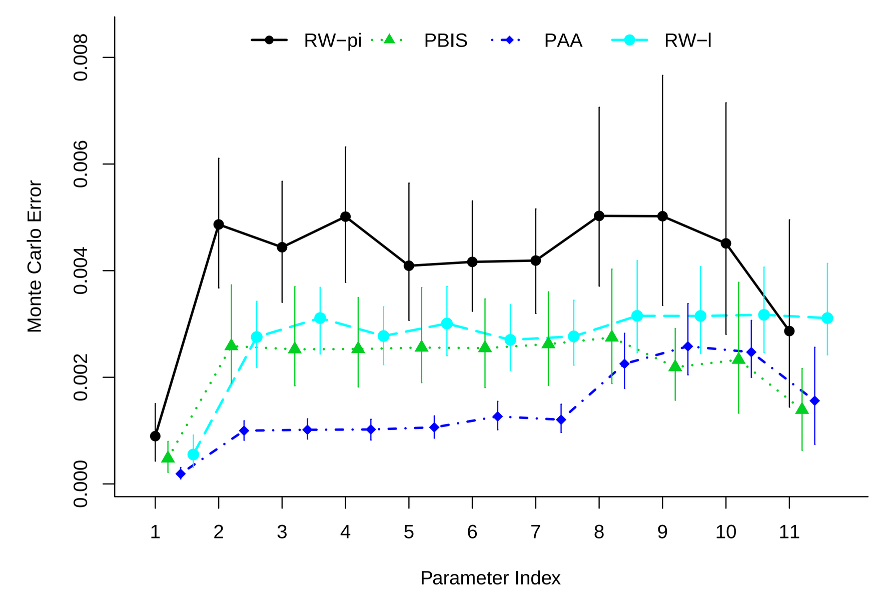


Figure 2: *MCEs for posterior Mean adjusted for CPU time for the 100 datasets of the simulation study*



9. Conclusions

- We proposed two MCMC methods for estimating graphical log-linear marginal models.
- PBIS is exact but more demanding.
- PAA is approximate but efficient (faster) with similar resutls to other methods.

Funding Details

Partially funded by the RC/AUEB (Research Fund for publications of the AUEB Faculty members) and by the Dept. of Economics and Management of University of Pavia.

References

Bergsma, W. P. and Rudas, T. (2002). Marginal log-linear models for categorical data. *Annals of Statistics*, **30**, 140-159.

Dellaportas, P. and Forster, J.J.(1999). Markov Chain Monte Carlo model determination for hierarchical and graphical log-linear models. *Biometrika*, **86**, 615-633.

Lupparelli, M., Marchetti, G. M., Bergsma, W. P. (2009). Parameterization and fitting of bi-directed graph models to categorical data. *Scandinavian journal of Statistics* **36**, 559-576.

Ntzoufras, I. and Tarantola, C. (2013). Conjugate and Conditional Conjugate Bayesian Analysis of Discrete Graphical Models of Marginal Independence. *Computational Statistics & Data Analysis*, **66**, 161-177.

Cox, D. R. and Wermuth, N. (1993). Linear dependencies represented by chain graphs (with discussion). *Statistical Science*, **8**, 204-218, 247-277.