Bayesian Estimation of Graphical Log-Linear Marginal Models

Stima Bayesiana di Modelli Grafici Log-Lineari Marginali

Claudia Tarantola, Ioannis Ntzoufras and Monia Lupparelli

Abstract Bayesian methods for graphical log-linear marginal models have not been developed as much as traditional frequentist approaches. The likelihood function cannot be analytically expressed in terms of the marginal log-linear interactions, but only in terms of cell counts or probabilities. No conjugate analysis is feasible, and MCMC methods are needed. We present a fully automatic and efficient MCMC strategy for quantitative learning, based on the DAG representation of the model. While the prior is expressed in terms of the marginal log-linear interactions, the proposal is on the probability parameter space. In order to obtain an efficient algorithm, we use as proposal values draws from a Gibbs sampling on the probability parameters.

Abstract *I metodi bayesiani per l'analisi di modelli grafici log-lineari marginali non sono stati sviluppati allo stesso modo di quelli frequentisti. La funzione di verosimiglianza non può essere espressa analiticamente attraverso i parametri log-lineari marginali, ma solamente in termini di frequenze o probabilità di cella. Non è possibile effettuare analisi coniugata, rendendo necessario l'utilizzo di metodi MCMC. Presentiamo una strategia MCMC per l'apprendimento quantitivo, completamente automatica ed efficiente, basata sulla rappresentazione del modello in termini di DAG. Mentre la prior è espressa in termini dei parametri marginali log-lineari, la proposal è sullo spazio delle probabilità. Al fine di ottenere un algoritmo efficiente, usiamo come proposal i valori ottenuti applicando un campionamento di Gibbs sullo spazio delle probabilità.*

Key words: DAG Representation, Marginal Log-Linear Parameterisation, Markov Chain Monte Carlo, Probability Based Sampler

Claudia Tarantola
University of Pavia, Italy, e-mail: claudia.tarantola@unipv.it

Ioannis Ntzoufras
Athens University of Economics and Business, Greece, e-mail: ntzoufras@aueb.gr

Monia Lupparelli
University of Bologna, Italy, e-mail: monia.lupparelli@unibo.it

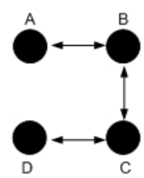
1 Introduction

Statistical models defined by imposing restrictions on marginal distributions of contingency tables have received considerable attention in economics and social sciences; for a thorough review see [2]. In particular standard log-linear models have been extended by [1] to allow the analysis of marginal distributions in contingency tables. This wider class of models is known as the class of marginal log-linear models. In these models, the log-linear interactions are estimated using the frequencies of appropriate marginal contingency tables, and are expressed in terms of marginal log-odds ratios. Following [1], the parameter vector λ of the marginal log-linear interactions can be obtained as

$$\lambda = C \log\left(M \mathrm{vec}(p)\right), \tag{1}$$

where $\mathrm{vec}(p)$ is the vector of joint probabilities, C is a contrast matrix and M specifies from which marginal each element of λ is calculated. A standard log-linear model is obtained from (1) setting M equal to the identity matrix and C to the inverse of the design matrix. Marginal log-linear models have been used by [4] to provide a parameterisation for discrete graphical models of marginal independence. A graphical log-linear marginal model is defined by zero constraints on specific log-linear interactions. It can be represented by a bi-directed graph like the one in Figure 1, where a missing edge indicates that the corresponding variables are marginally independent; for the related notation and terminology see [3] and [4] .

Fig. 1 A bi-directed graph



Despite the increasing interest in the literature for graphical log-linear marginal models, Bayesian analysis has not been developed as much as traditional methods. The main reasons are the following. Graphical log-linear marginal models belong to curved exponential families that are difficult to handle from a Bayesian perspective. Posterior distributions cannot be directly obtained, and MCMC methods are needed. The likelihood cannot be analytically expressed as a function of the marginal log-linear interactions, but only in terms of cell counts or probabilities. Hence, an iterative procedure should be implemented to calculate the cell probabilities, and consequently the model likelihood. Another important point is that, in order to have a well-defined model of marginal independence, we need to construct an algorithm which generates parameter values that lead to a joint probability distribution with compatible marginals.

A possibility is to follow the approach presented in [5], where a Gibbs sampler based on a probability parameterisation of the model is presented. Even if using this approach one can obtain as a by-product the distribution of the log-linear interactions, if the focus is on marginal log-odds a prior should be directly specified for these parameters. Additionally, and more importantly, if any prior information exists for log-odds then we need to work directly using the log-linear parameterisation. For instance, symmetry constraints, vanishing high-order associations or further prior information about the joint and marginal distributions can be easily specified by setting linear constraints on marginal log-linear interactions, instead of non-linear multiplicative constraints on the probability space. For the previous reasons, in [6] a novel MCMC strategy (the probability based sampler) is introduced. In the probability based sampler the prior is expressed in terms of the marginal log-linear interactions, while the proposal is defined on the probability parameter space. Efficient proposal values are obtained via the conditional conjugate approach of [5]. The corresponding proposal density on the marginal log-linear interactions is obtained by implementing standard theory about functions of random variables. For more details on the methodology and the obtained results, see the extended version of this work ([6]).

2 Probability Based independence Sampler

Following the notation of [5], we can divide the class of graphical log-linear marginal models in two major categories: homogeneous and non-homogeneous models. Both type of models are shown to be compatible, in terms of independencies, with a certain DAG representation (augmented DAG). Nevertheless, while homogeneous models can be generated via a DAG with the same vertex set, for non-homogeneous ones it is necessary to include some additional latent variables. The advantage of the augmented DAG representation is that the joint probability over the augmented variable space (including both observed and latent variables) can be written using the standard DAG factorisation. We parameterise the augmented DAG via a set Π of marginal and conditional probability parameters on which, following [5], we implement a conjugate analysis based on products of Dirichlet distributions. Once a suitable prior is assigned on the marginal log-linear interaction parameters, a Metropolis-Hastings algorithm can be used to obtain a sample from the posterior distribution. For $t = 1, \dots, T$, we repeat the following steps

1. propose Π' from $q(\Pi'|\Pi^{(t)})$, where $\Pi^{(t)}$ is the value of Π at t iteration;
2. from Π' calculate via marginalisation the proposed joint probabilities p' for the observed table;
3. from p', calculate λ' using (1) and then obtain the corresponding non-zero elements $\overrightarrow{\lambda}'$;
4. set $\xi' = \Pi'_\xi$; where Π'_ξ is a pre-specified subset of Π' of dimension $\dim(\Pi) - \dim(\overrightarrow{\lambda})$;
5. accept the proposed move with probability $\alpha = \min(1, A)$

$$A = \frac{f(n|\Pi')f\big(\overrightarrow{\lambda}'\big)q(\Pi^{(t)}|\Pi')}{f(n|\Pi^{(t)})f\big(\overrightarrow{\lambda}^{(t)}\big)q(\Pi'|\Pi^{(t)})} \times \text{abs}\left(\frac{\mathscr{J}\left(\Pi^{(t)}, \overrightarrow{\lambda}^{(t)}, \xi^{(t)}\right)}{\mathscr{J}\left(\Pi', \overrightarrow{\lambda}', \xi'\right)}\right), \qquad (2)$$

where abs$(\cdot)$ stands for the absolute value, and $\mathscr{J} = \mathscr{J}(\Pi, \overrightarrow{\lambda}, \xi)$ is the determinant of the jacobian matrix of the transformation $\Pi = g(\overrightarrow{\lambda}, \xi)$. The construction of the Jacobian matrix is facilitated by the augmented DAG representation of the model. Note that the ratio $f(\xi')/f(\xi^{(t)})$ cancels out from the acceptance rate since we set $f(\xi_\iota) = I_{\{0<\xi_\iota<1\}}$.

6. If the move is accepted, we set $\Pi^{(t+1)} = \Pi'$, $\xi^{(t+1)} = \xi'$, and $\overrightarrow{\lambda}^{(t+1)} = \overrightarrow{\lambda}'$ otherwise we set $\Pi^{(t+1)} = \Pi^{(t)}$ and $\overrightarrow{\lambda}^{(t+1)} = \overrightarrow{\lambda}^{(t)}$.

In order to obtain a high acceptance rate it is crucial the choice of the proposal density $q(\Pi'|\Pi^{(t)})$. As discussed in [6], an efficient proposal is $q(\Pi'|\Pi^{(t)}) = f_q\big(\Pi'|n^{\mathscr{A}}\big)f\big(n^{\mathscr{A}}|\Pi^{(t)}, n\big)$, where $n^{\mathscr{A}}$ is an augmented table. Exploiting the conditional conjugate approach of [5] we consider as a “prior” $f_q(\Pi)$ a product of Dirichlet distributions obtaining a conjugate “posterior” distribution $f_q(\Pi'|n'^{\mathscr{A}})$. The acceptance rate in (2) becomes equal to

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

In the following, we will refer to this approach as the *probability-based independence sampler* (PBIS). Although PBIS simplifies the MCMC scheme, the parameter space is still considerably extended by considering the augmented frequency table $n^{\mathscr{A}}$. This algorithm can be further simplified by using as proposal a random permutation of the MCMC output obtained applying the Gibbs sampling of [5]. The acceptance rate becomes

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

This sampler is named the *prior-adjustment* algorithm (PAA) due to its characteristic to correct for the differences between the prior distributions used under the two parameterisations.

3 Simulation study

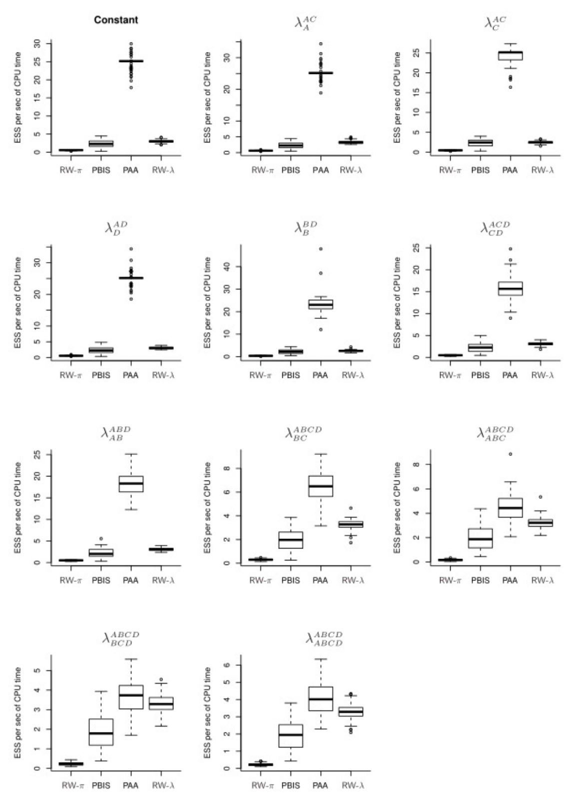
We evaluate the performance of the algorithms presented in Section 2 via a simulation study. We generated 100 samples from the marginal association model represented by the bi-directed graph of Figure 1, and true log-linear interactions given in Table 1. In addition to the algorithms described in Section 2, for comparative purposes, we consider random walks on marginal log-linear interactions λ and on logits of probability parameters π (RW-λ and RW-π respectively). We compare the examined methods in terms of Effective Sample Size (ESS) and Monte Carlo Error (MCE).

Table 1 True Effect Values Used for the Simulation Study

Marginal	Active interactions	Zero interactions
AC	$\lambda_{\emptyset}^{AC} = -1.40, \lambda_{A}^{AC}(2) = -0.15, \lambda_{C}^{AC}(2) = 0.10,$	$\lambda_{AC}^{AC} = 0$
AD	$\lambda_{B}^{AD}(2) = 0.12,$	$\lambda_{BD}^{BD}(2,2) = 0$
BD	$\lambda_{D}^{BD}(2) = -0.09,$	$\lambda_{AD}^{AD}(2,2) = 0$
ACD	$\lambda_{CD}^{ACD}(2,2) = 0.20,$	$\lambda_{ACD}^{ACD}(2,2,2) = 0$
ABD	$\lambda_{AB}^{ABD}(2,2) = -0.15,$	$\lambda_{ABD}^{ABD}(2,2,2) = 0$
ABCD	$\lambda_{BC}^{ABCD}(2,2) = -0.30, \lambda_{ABC}^{ABCD}(2,2,2) = 0.15,$	
	$\lambda_{BCD}^{ABCD}(2,2,2) = -0.10, \lambda_{ABCD}^{ABCD}(2,2,2) = 0.07.$	

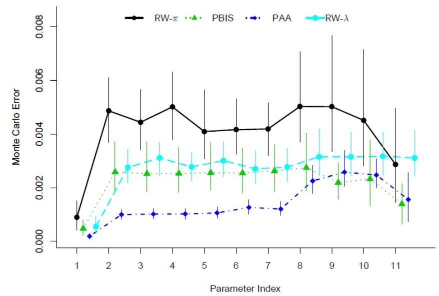
In Figure 2 we report the distribution of the ESS per second of CPU time. PAA is clearly the most efficient among the four methods under consideration.

Fig. 2 ESS per second of CPU time



In Figure 3, for all methods and all marginal log-linear interactions, we represent the time adjusted MCEs for the posterior means. In the graph we represent the 95% error bars of the average time adjusted MCEs for the posterior means. We notice that PAA performs better than all competing methods, since the corresponding MCEs are lower for almost all interactions.

Fig. 3 MCEs for posterior Mean adjusted for CPU time for the simulation study



For a more detailed analysis and a real data application, see Section 5 of [6].

4 Concluding remarks

In this paper we have presented a novel Bayesian methodology for quantitative learning for graphical log-linear marginal models. The main advantages of this approach are the following. It allows us to incorporate in the model prior information on the marginal log-linear interactions. It leads to an efficient and fully automatic setup, and no time consuming and troublesome tuning of MCMC parameters is needed. The authors are planning to extend the method to accommodate fully automatic selection, comparison and model averaging.

References

1. Bergsma, W. P. and Rudas, T. (2002). Marginal log-linear models for categorical data. *Annals of Statistics*, **30**, 140-159.
2. Bergsma, W. O., Croon, M. A. and Hagenaars, J. A. (2009). *Marginal Models. For Dependent, Clustered, and Longitudinal Categorical Data*, Springer
3. Drton, M. and Richardson, T. S. (2008). Binary models for marginal independence. *Journal of the Royal Statistical Society*, Ser. B , **70**, 287-309.
4. Lupparelli, M., Marchetti, G. M., Bersgma, W. P. (2009). Parameterization and fitting of bi-directed graph models to categorical data. *Scandinavian journal of Statistics* **36**, 559-576.
5. 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.
6. Ntzoufras, I. Tarantola, C. and Lupparelli, M. (2018). Probability Based Independence Sampler for Bayesian Quantitative Learning in Graphical Log-Linear Marginal Models, *DEM Working Paper Series*.