

Sociological Methodology

<http://smx.sagepub.com/>

Comment: Graphical Marginal Log-Linear Models--A Bayesian Perspective

Claudia Tarantola, Ioannis Ntzoufras and Monia Lupparelli

Sociological Methodology 2013 43: 109

DOI: 10.1177/0081175013493395

The online version of this article can be found at:

<http://smx.sagepub.com/content/43/1/109>

Published by:



<http://www.sagepublications.com>

On behalf of:



AMERICAN SOCIOLOGICAL ASSOCIATION

American Sociological Association

Additional services and information for *Sociological Methodology* can be found at:

Email Alerts: <http://smx.sagepub.com/cgi/alerts>

Subscriptions: <http://smx.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

>> [Version of Record](#) - Sep 5, 2013

[What is This?](#)

mobility, estimation of voting transitions, and voting behavior. For details, see <http://www.stat.unipg.it/forcina/English.html>.

COMMENT: GRAPHICAL MARGINAL LOG-LINEAR MODELS—A BAYESIAN PERSPECTIVE

*Claudia Tarantola**,
Ioannis Ntzoufras[†], and
Monia Lupparelli[‡]

*University of Pavia, Italy

[†]Athens University of Economics and

[‡]University of Bologna, Italy

Corresponding Author: Claudia Tarantola, claudia.tarantola@unipv.it

DOI: 10.1177/0081175013493395

In the last few years statistical models defined by imposing restrictions on marginal distributions of contingency tables have received considerable attention in the social sciences and economics; see Németh and Rudas (this volume, p. 70–100) and references herein. In particular standard log-linear models have been extended to allow the analysis of marginal distributions in contingency tables. This new class of models, named marginal log-linear (MLL) models, has also been used to provide parameterizations for discrete graphical models.

Németh and Rudas present a nice overview on how discrete graphical models can be useful instruments for social scientists. Particular emphasis is given to the MLL parameterization (denoted by λ) for discrete graphical models. They adopt a frequentist approach and show, through several examples, how this parameterization can be applied for the analysis of DAG and chain graph models; complementary detailed material for the implementation in R is available on the authors' web page at <http://nemethr.web.elte.hu/discrete-graphical-models/>.

Their paper highlights how the choice of suitable parameterizations is an important issue. The vector of joint probabilities is perhaps the

most intuitive parameter for multinomial distributions, although there are several works, such as Roverato, Lupparelli, and La Rocca (forthcoming), investigating other parameterizations. The increasing attention for novel parameterizations is driven by the fact that the following five properties are desirable, especially in graphical modeling contexts:

1. Defined by a smooth mapping from the simplex of joint probabilities
2. Straightforward and useful interpretation
3. Linear constrained parameters defining independences between variables
4. Variation independence
5. Upward compatibility

The MLL parameterization proposed by the authors for DAGs satisfies properties 1 through 4 and partially property 5, which is fulfilled by the multivariate logistic parameterization representing a special case in the class of MLL models. Properties 3 and 4 are especially appealing in a Bayesian context. Nevertheless, the MLL parameterizations defined for several classes of graphical models do not always satisfy variation independence; see Lupparelli, Marchetti, and Bergsma (2009), Rudas, Bergsma, and Németh (2010), and Evans and Richardson (forthcoming).

Even if MLL parameters have a straightforward interpretation in terms of associations in marginal distributions, developing a Bayesian analysis for them is still an open problem. The lack of variation independence may cause problems and inconsistencies for the implementation of MCMC methods. Furthermore, conjugate analysis is not feasible. For these reasons, Ntzoufras and Tarantola (2012, forthcoming) present a conjugate and conditional conjugate Bayesian analysis based on probability parameters. Working directly with probabilities leads to simple Monte Carlo schemes through which the MLL parameters are always computable.

For comparative purposes, since the authors do not consider the Bayesian approach, we illustrate the methodology developed by Ntzoufras and Tarantola (2012, forthcoming) for the data set analyzed by the authors in Example 5 through two DAG models, which we name in the discussion that follows M1 and M2, respectively. The results are coherent with the analysis presented by the authors.

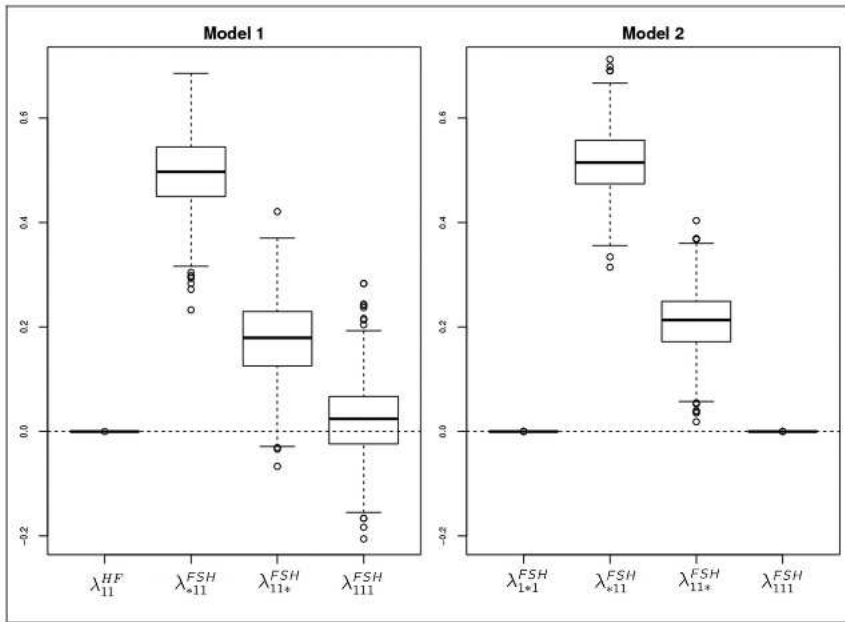


Figure 1. Posterior box plots for the MLL parameters of M1 and M2, respectively. For M1, the parameters are obtained using the sequence of margins HF FSH; for M2, the parameters are obtained using the single margin HFS.

M1 is Markov equivalent to a covariance graph model (Cox and Wermuth 1996) also known as a bidirected graph model (Drton and Richardson 2008). Instead, M2 is Markov equivalent to an undirected graph model. For the analysis of M2 we then adopt the standard log-linear parameterization that is commonly used for these models and is itself a special case in the MLL class.

We examined the saturated model and the two previous models using the Dirichlet priors described by Ntzoufras and Tarantola (2012). For M1, the Bayes factors (BFs) are quite sensitive on the choice of the prior parameters with the unit expected cell prior (Dirichlet with $\alpha_i = 1$) and the Jeffreys prior ($\alpha_i = 1/2$) supporting the saturated model (BFs of M1 versus the saturated equal to 0.38 and 0.82, respectively) while for the third prior (Perks prior with $\alpha_i = 1/8$), M1 is marginally better than the saturated with BF = 3.09. M2 clearly fits the data well with BFs. 12 when compared with the saturated and BFs. 32 when compared with M1 under all prior setups, and the corresponding posterior weight higher than 0.9.

We can easily estimate the posterior distributions of any MLL parameterization using the simple Monte Carlo scheme proposed by Ntzoufras and Tarantola (2012). Under this approach, we generate samples from the Dirichlet posterior distributions of probability parameters, calculate the corresponding joint probabilities, and finally transform them to λ s by simple appropriate transformations. Because the results under all three priors are similar, we present the analysis only for the Perks prior. Figure 1 shows the posterior distributions of the adopted MLL parameterization for M1 and M2. For space reasons, the box plots do not include first-order interactions.

For M1, we observe that the three-way interaction is not significant (since zero is placed in the center of the posterior distribution) while the two-way interaction FS is of marginal significance (the 99 percent posterior interval includes zero). Finally, for M2, posterior results are equivalent to the MLEs. The posterior distribution of the interaction SH seems to be far away from zero, while for FS seems to be of marginal significance (a 99 percent posterior interval 0.04–0.35).

References

- Cox, David R. and Nanny Wermuth. 1996. *Multivariate Dependencies*. London, England: Chapman and Hall.
- Drton, Mathias and Thomas S. Richardson. 2008. "Binary Models of Marginal Independence." *Journal of the Royal Statistical Society, Series B*, 70(2):287–309.
- Evans, Robin J. and Thomas S. Richardson. Forthcoming. "Marginal Log-linear Parameters for Graphical Markov Models." *Journal of the Royal Statistical Society, Series B*.
- Lupparelli, Monia, Giovanni M. Marchetti, and Wicher P. Bergsma. 2009. "Parameterization and Fitting of Discrete Bi-directed Graph Models." *Scandinavian Journal of Statistics* 36:559–76.
- Ntzoufras, Ioannis and Claudia Tarantola. 2012. "Bayesian Analysis of Marginal Log-linear Graphical Models for Three-way Contingency Tables." Technical report. Department of Political Economy and Quantitative Methods. University of Pavia.
- Ntzoufras, Ioannis and Claudia Tarantola. Forthcoming. "Conjugate and Conditional Conjugate Bayesian Analysis of Discrete Graphical Models of Marginal Independence." *Computational Statistics and Data Analysis*, DOI: 10.1016/j.csda.2013.04.005.
- Roverato, Alberto, Monia Lupparelli, and Luca La Rocca. Forthcoming. "Log-mean Linear Models for Binary Data." *Biometrika*, DOI: 10.1093/biomet/ass080.
- Rudas, Tamás, Wicher Bergsma, and Renáta Németh. 2010. "Marginal Log-linear Parameterization of Conditional Independence Models." *Biometrika* 97(4): 1006–12.

Author Biographies

Claudia Tarantola is assistant professor of statistics at the Department of Economics and Management of University of Pavia. She graduated in "Political Economy" from L. Bocconi University, Milan and did her PhD in "Methodological Statistics" at the University of Trento. She spent part of her PhD program at the Department Mathematics of University of Bristol, she was a post-doc at department of Statistics of Athens University of Economics and Business. Her main researches interest are Categorical data analysis, Graphical models, Bayesian models for the analysis of financial data, Marginal Models, Markov Chain Monte Carlo methods, and she has various publication on these topics. She has taken part in various research programs, both National and International. She is a member of the Italian Statistical Society. She served as a referee for various scientific journals, including *Biometrika*, *Journal of the Royal Statistical Society series B*, and *Journal of Statistical Planning and Inference* among others. Further details can be found on her website http://economia.unipv.it/pagp/pagine_personali/ctaranto/tarantola.htm.

Monia Lupparelli is assistant professor of statistics at the Department of Statistical Sciences of University of Bologna since 2008. She graduated in Economics from University of Perugia, and in 2005 she got her PhD in "Applied Statistics" at University of Florence. Her main research interests are Graphical Models, Categorical Data Analysis, Latent Variables models, and she has various publications on these topics. Since 2005, she has taken part in various research programs. She is a member of the Italian Statistical Society. She served as a referee for various international journals. Further details can be found on her website <http://www2.stat.unibo.it/lupparelli/>.

Ioannis Ntzoufras is associate professor of statistics at Athens University of Economics and Business (Greece). He holds a Bachelor degree in Statistics and Insurance Science from the University of Piraeus (Greece) and a Masters degree in Statistics with Applications in Medicine from the University of Southampton (UK); both of them with distinction. He obtained his PhD in Statistics from Athens University of Economics and Business (Greece). He has published more than 34 articles in refereed journals such as the *Statistical Science*, *Statistics and Computing*, *Psychometrika*, and *Annals of Applied Statistics*. His research interests include the areas of Bayesian and computational statistics, variable selection methods, statistical modeling, analysis of psychometric and sports data among others. His book on "Bayesian modelling using WinBUGS" is a best seller in its category and it received honorary mention on the subject of Mathematics in PROSE awards for books published in 2009. He served as a referee in more than 20 scientific journal including *Biometrika*, *The Annals of Statistics*, and *Statistics and Computing* among others. More details can be found in <http://stat-athens.aueb.gr/~jbn/publications.htm>.