Bayesian Analysis (2013) **8**, Number 3, pp. 539–542

Comment on Article by Scutari

Christine B. Peterson * and Francesco C. Stingo †

We congratulate the author for a well-written article on a problem of clear and increasing scientific relevance. The approach taken here is not only relevant for prior specification and analysis of posterior distributions in graphical models, but has a much wider breadth as explained below.

In his work on Bayesian inference of graphical models, the author adopts a novel focus on prior formulations over binary or trinary indicators of edge inclusion in undirected or directed graphs, and examines relevant properties of these distributions including entropy and variability. The insights of this paper provide a valuable starting point for further exploration of the proposed priors both in terms of theoretical properties and practical application to real or simulated data.

Although we do not see how the proposed parametrization reduces the dimension of the parameter space, we consider the approach taken in this work a novel point of view both in terms of defining prior distributions over a set of dependent binary and trinary random variables and in terms of analyzing the posterior distributions of those variables.

1 Generalization beyond graphical models

Since multivariate Bernoulli and trinomial distributions can be applied to any setting where the parameters of interest are binary or trinary, the theoretical properties explored in this paper are not limited to the framework of graph inference and can be considered in more general settings. In the context of Bayesian variable selection, the latent indicators of variable inclusion are modeled as multivariate binary random variables. While binary variables are more common in statistical modeling, a number of recent papers on modeling gene and protein expression rely on trinary variables to capture states of underexpression, normal expression, and overexpression (Parmigiani et al. 2002; Telesca et al. 2012; Xu et al. 2012).

2 Choice of prior parameters

Treatment of the Bernoulli or trinomial prior parameters warrants further discussion. Although the author expresses a preference for priors favoring sparsity, he does not directly address the issue of multiple testing. Scott and Berger (2010) argue that fixing the prior probability of inclusion in the context of variable selection fails to account for multiplicity, and that it is preferable to place a prior distribution on the inclusion probabilities. Because the edge selection problem resembles the variable selection problem in

*Department of Statistics, Rice University, Houston, TX, cbpeterson@gmail.com
†Department of Biostatistics, University of Texas MD Anderson Cancer Center, Houston, TX, fra.stingo@gmail.com

 DOI:10.1214/13-BA842

many respects, literature on prior formulations in this context are relevant, particularly when dealing with the multivariate Bernoulli distribution. For example, several recent papers have explored the use of Ising priors to model dependency among the Bernoulli indicators of inclusion in the setting of Bayesian variable selection (Li and Zhang 2010; Stingo and Vannucci 2011). While this class of prior only describes first and second order moments, for many practical problems this level of detail is sufficient.

3 Impact of entropy on model selection

In many applications, we would like to identify a single "best" model. One option is the maximum a posteriori (MAP) estimate, which is the graph with highest posterior probability. For undirected graphs, another common choice is the median model, which consists of all edges with marginal posterior probability greater than 0.5 (Barbieri and Berger 2004). It would be interesting to explore how the entropy of the posterior distribution of the binary and trinary indicators relates to the choice of the MAP and median model. In particular, it would be nice to determine if lower posterior entropy results in sharper selection under these approaches.

4 Applicability

We look forward to seeing applications of the proposed approach to real problems and data. Our understanding is that, in order to assess features of the posterior distributions such as entropy and variability, a large enough number of samples needs to be drawn from the posterior distribution. The proposed approach will then suffer from the same restrictions in terms of scalability as most Bayesian approaches for large networks. For example, these computational restrictions make it hard to use this approach to obtain the optimal tuning parameter value τ^* of Section 4 when the analyzed network becomes large. Moreover, we would like to understand how to interpret the variability measures introduced in equation (16). While these measures can provide a ranking of posterior distributions, it's very hard to give a practical interpretation of the distance between two distributions.

References

Barbieri, M. and Berger, J. (2004). "Optimal predictive model selection." *The Annals of Statistics*, 32(3): 870–897. 540

Li, F. and Zhang, N. (2010). "Bayesian variable selection in structured high-dimensional covariate spaces with applications in genomics." *Journal of the American Statistical Association*, 105(491): 1202–1214. 540

Parmigiani, G., Garrett, E., Anbazhagan, R., and Gabrielson, E. (2002). "A statistical framework for expression based molecular classification in cancer." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4): 717–736. 539

Scott, J. and Berger, J. (2010). "Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem." *Annals of Statistics*, 38(5): 2587–2619. 539

Stingo, F. and Vannucci, M. (2011). "Variable selection for discriminant analysis with Markov random field priors for the analysis of microarray data." *Bioinformatics*, 27(4): 495–501. 540

Telesca, D., Müller, P., Kornblau, S., Suchard, M., and Ji, Y. (2012). "Modeling protein expression and protein signaling pathways." *Journal of the American Statistical Association*, 107(500): 1372–1384. 539

Xu, Y., Zhang, J., Yuan, Y., Mitra, R., Müller, P., and Ji, Y. (2012). "A Bayesian graphical model for integrative analysis of TCGA data." *2012 IEEE International Workshop on Genomic Signal Processing and Statistics (GENSIPS)*, 135–138. 539