Bayesian Analysis (2013) **8**, Number 3, pp. 533–538

Comment on Article by Scutari

Adrian Dobra *

This is an interesting and thought-provoking paper which focuses on defining new prior distributions on graphical structures. Priors on graphs represent a key component of any Bayesian approach for graphical models, hence the identification of new prior distributions for graphs is a very important topic. The author proceeds by modeling the possible edges of a graph through appropriate joint probability distributions. This idea receives a good treatment in this writing, but it is certainly not as novel as the author might seem to suggest by not mentioning many other papers who used various priors on graphs which are different from the uniform prior given in equation (2) page 2 of the paper. In fact, the Bayesian literature dedicated to graphical models has a longstanding track of using priors that encourage sparsity in order to increase interpretability and avoid the risk of overfitting. Some of these priors are constructed precisely by treating edges as random variables. In the context of DAGs, a typical prior specification starts with the traditional Bayesian variable selection prior for regressions which is defined by assuming a constant probability of inclusion β of each variable x_i, $i \in V = \{1, 2, \ldots, p\}$, in the regression model. This leads to a prior $\Pr(k) \propto (\beta/(1-\beta))^k$ associated with a regression with k predictors. Independent priors for regressions in the compositional specification of a DAG D,

$$x_i = \sum_{j \in pa(i)} \gamma_{ij} x_j + \epsilon_i, \text{ for each } i \in V,$$

where $pa(i)$ are the parents of vertex i in D, lead to the following prior for D (see, for example, Dobra et al. (2004)):

$$\Pr(D) \propto (\beta/(1-\beta))^{\sum_{i=1}^{p} \#pa(i)}.$$

The DAG D becomes sparser as $\binom{p}{2}\beta$ gets smaller. In the context of Gaussian graphical models, a usual choice is the uniform prior $\Pr(G) \propto 1$. Alternative priors on $\mathcal{G}_p$, the set of graphs with p vertices, have been developed by assuming a constant probability of inclusion $\beta \in (0,1)$ of each edge. This leads to a prior for a graph $G \in \mathcal{G}_p$ (Dobra et al. 2004; Jones et al. 2005)

$$\Pr(G) \quad \propto \quad (\beta/(1-\beta))^{size(G)}, \qquad (1)$$

where $size(G)$ is the number of edges in G. Sparse graphs receive high prior probabilities when $\binom{p}{2}\beta$ is small. By assuming $\beta \sim \mathsf{Beta}(a, b)$, Carvalho and Scott (2009) integrate out β in (1) to obtain the following prior on $\mathcal{G}_p$:

$$\Pr(G) \quad \propto \quad B\left(a + size(G), b + \binom{p}{2} - size(G)\right)/B(a,b), \qquad (2)$$

*Department of Statistics, Department of Biobehavioral Nursing and Health Systems, and Center for Statistics and the Social Sciences, University of Washington adobra@uw.edu

 DOI:10.1214/13-BA841