Bayesian Analysis (2013) **8**, Number 3, pp. 549–552

Rejoinder

Marco Scutari *

I would like to thank Hao Wang, Adrian Dobra, Christine Peterson and Francesco Stingo for the insightful comments and critiques they contributed to the discussion. The material contained in the paper originated in large part as the theoretical core of my Ph.D. thesis (Scutari 2011), and served the purpose of improving my understanding of the workings of prior and posterior distributions on graph structures as much as that of exploring novel applications. As a result, and as the discussants have observed, the paper provides a useful starting point for further developments while not focusing on specific applications such as prior specification or the analysis of real-world data.

The discussants' remarks highlight key strengths and limitations in the material, and suggest several useful directions for future research. In the following, I will concentrate on four topics that were touched on in all discussions: the development of new priors, sampling random graphs from non-uniform distributions, the applications and interpretation of the covariance matrix and the entropy of $\mathrm{P}(\mathcal{G}(\mathcal{E}))$ and $\mathrm{P}(\mathcal{G}(\mathcal{E}) \mid \mathcal{D})$, and the role of structure learning in graphical modelling.

1 Developing new prior distributions

In the paper much attention is devoted to the uniform prior and the *maximum entropy* case. As remarked by Dobra, other choices are available in the literature that are more flexible and tailored to real-world data. Additional examples inspired by genetics and systems biology are presented, for instance, in Imoto et al. (2003), Werhli and Husmeier (2007) and Mukherjee and Speed (2008). The reason for investigating the uniform prior is two-fold. First of all, it is a limit case in terms of entropy and therefore it is useful as a term of comparison along with *maximum entropy* distributions. Furthermore, the uniform prior is a *de facto* standard for $\mathrm{P}(\mathcal{G}(\mathcal{E}))$ in computer science and artificial intelligence literature on Bayesian networks, to the point that sometimes its use is not even mentioned explicitly but is implied by the fact that imaginary sample size is the only hyperparameter.

Developing new priors using the second order moments of $\mathrm{P}(\mathcal{G}(\mathcal{E}))$ (i.e. arc and edge correlations) in addition to first order moments (i.e. arc and edge probabilities) presents significant challenges due to the number of parameters involved. As the discussants pointed out, achieving sparsity and addressing the need for multiplicity adjustment while keeping hyperparameter specification simple is a difficult task. In my thesis, I addressed a related problem, the regularisation of the covariance matrix of $\mathrm{P}(\mathcal{G}(\mathcal{E}) \mid \mathcal{D})$ with the shrinkage estimators from Ledoit and Wolf (2003) and Schäfer and Strimmer (2005). Such estimators have a Bayesian interpretation and can be used to achieve sparsity by shrinking $\mathrm{diag}(\Sigma)$ and (in turn) edge and arc probabilities towards zero

*Genetics Institute, University College London, United Kingdom, m.scutari@ucl.ac.uk

 DOI:10.1214/13-BA845