Bayesian Analysis (2013) **8**, Number 3, pp. 543–548

Comment on Article by Scutari

Hao Wang *

Scutari's paper studies properties of the distribution of graphs $p(\mathcal{G})$. This is an interesting angle because it differs from many works that focus on distributions over parameter spaces for a given graph $p(\Theta \mid \mathcal{G})$. The paper's investigation of $p(\mathcal{G})$ centers around its implied covariance matrix $\mathbf{\Sigma} = \text{Cov}(\mathcal{G})$. The major theoretical results, as I see it, concern eigenvalues of $\mathbf{\Sigma}$ as well as variance and covariance elements of $\mathbf{\Sigma}$ in the *maximum entropy* case for DAGs (i.e., a uniform distribution over all DAGs). While these results are certainly very worth noting by their own intellectual merits, what practical difference they might make is unclear to me. Eigenvalues of $\mathbf{\Sigma}$ might be hard to interpret in terms of their intuitive connections with underlying graph structures. The maximum entropy case is somehow limited as it is rarely the case for posterior graph distributions and is also often less preferred than sparser cases for prior graph distributions. More discussions on the implications of these theoretical results on real data analysis will be very helpful.

The more general point raised by the paper is more interesting to me. It calls attention to deeper investigation on statistical properties of distributions of graphs $p(\mathcal{G})$. In the literature of my own research topic of Gaussian graphical models (Dempster 1972), existing studies usually only focus on a point estimation of $\mathcal{G}$ from $p(\mathcal{G})$ – the mean or the mode of $p(\mathcal{G})$ is often used to represent prior belief or to summarize posterior information. The paper's framework extends this sort of simple summary to the covariance matrix $\mathbf{\Sigma}$ of $p(\mathcal{G})$. It is then tempting to ask what will be gained from these extra efforts. Specific questions include how to construct a prior $p(\mathcal{G})$ with a consideration beyond the implied mean or mode graphs, and how to put $\mathbf{\Sigma}$ into a perspective that better illustrates graph structures than a point estimate alone.

I attempt to explore these questions in this discussion from a more applied point of view than the paper. In addition, I have some doubts about the paper's argument of using variability measures in choosing learning algorithms or hyperparameters. The context of my discussion is Gaussian graphical models under a fully Bayesian treatment (Jones et al. 2005). Generalizations of the following points might be made to other undirected graphs or even DAGs too.

1 Distribution $p(\mathcal{G})$ and its covariance matrix

Similar to the paper, I use the edge set to represent a graph $\mathcal{G}$. Let e_{ij} be the binary edge inclusion indictor variable, that is, $e_{ij} = 1$ if there is an edge between nodes i and j in $\mathcal{G}$, and $e_{ij} = 0$ otherwise. Then the set of $k = p(p-1)/2$ binary variables $\mathcal{E} = \{e_{ij}\}_{1 \leqslant i < j \leqslant p}$ can be used in place of $\mathcal{G}$. The distribution of graphs is $p(\mathcal{E})$ and the implied $k \times k$ covariance matrix is $\mathbf{\Sigma} = \text{Cov}(\mathcal{E})$.

*Department of Statistics, University of South Carolina, Columbia, SC, U.S.A. haowang@sc.edu

 DOI:10.1214/13-BA840