Supplementary Material for Bayesian Inference in Nonparanormal Graphical Models

Jami Jackson Mulgrave and Subhashis Ghosal

Proof of Consistency Theorems

Let $p_{f,\mu,\sigma}$ stand for the density of X so that $f(X) \sim \mathrm{N}(\mu, \sigma)$, i.e.,

$$p_{f,\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp[-\frac{1}{2\sigma^2}(f(x) - \mu)^2] f'(x), \quad 0 < x < 1.$$

Then $X_1, \dots, X_n$ are i.i.d. samples from p_{f_0,μ_0,σ_0}. It follows that $X_1^*, \dots, X_{n^*}^*$ are i.i.d. samples from the density

$$p^*_{f_0,\mu_0,\sigma_0}(x) = \frac{p_{f_0,\mu_0,\sigma_0}(x)}{\int_\delta^{1-\delta} p_{f_0,\mu_0,\sigma_0}(u)du} \mathbb{1}\{x \in [\delta, 1-\delta]\}.$$

Note that n^* is random, but n^*/n converges a.s. to a positive constant, and hence for posterior consistency, it suffices to treat n^* as a deterministic subsequence of n. Indeed, to simplify notations, we shall drop the asterics to write n in place of n^* and $X_1, \dots, X_n$ in place of $X_1^*, \dots, X_{n^*}^*$.

In view of Schwartz's theorem on posterior consistency (see Section 6.4 of Ghosal and van der Vaart (2017) for details), it will follow that the posterior distribution is consistent with respect to the weak topology on densities provided that the prior distribution has the Kullback-Leibler property at the true density: for all $\epsilon > 0$,

$$\Pi\{(f, \mu, \sigma) : K(p^*_{f_0,\mu_0,\sigma_0}, p^*_{f,\mu,\sigma}) < \epsilon\} > 0, \tag{1}$$

where $K(p, q) = \int p \log(p/q)$ is the Kullback-Leibler divergence. Let $I = [\delta, 1-\delta]$, $P_{f,\mu,\sigma}$ stand for the probability measure corresponding to $p_{f,\mu,\sigma}$ and P_{f_0,μ_0,σ_0} stand for that corresponding to p_{f_0,μ_0,σ_0}. Then we can write $K(p^*_{f_0,\mu_0,\sigma_0}, p^*_{f,\mu,\sigma})$ as

$$\int_I p^*_{f_0,\mu_0,\sigma_0} \log \frac{p_{f_0,\mu_0,\sigma_0}}{p_{f,\mu,\sigma}} + \log \frac{P_{f,\mu,\sigma}(I)}{P_{f_0,\mu_0,\sigma_0}(I)}. \tag{2}$$

Note that, expressing $f_0(X) \sim \mathrm{N}(\mu, \sigma^2)$, we have that $P_{f_0,\mu_0,\sigma_0}(I) = \Phi((f_0(1-\delta) - \mu_0)/\sigma_0) - \Phi((f_0(\delta) - \mu_0)/\sigma_0)$. Similarly, $P_{f,\mu,\sigma}(I) = \Phi((f(1-\delta) - \mu)/\sigma) - \Phi((f(\delta) - \mu)/\sigma)$. Therefore if $|\mu - \mu_0|$, $|\sigma - \sigma_0|$ and $\sup\{|f(x) - f_0(x)| : x \in I\}$ are sufficiently small, then the second term in (2) can be made arbitrarily small. We can bound $\int_I p^*_{f_0,\mu_0,\sigma_0} \log(p_{f_0,\mu_0,\sigma_0}/p_{f,\mu,\sigma})$ by

$$|\log \sigma - \log \sigma_0| + \sup_{x \in I} \Big| \frac{(f(x) - \mu)^2}{2\sigma^2} - \frac{(f_0(x) - \mu_0)^2}{2\sigma_0^2} \Big| + \sup_{x \in I} |\log f'(x) - \log f_0'(x)|. \tag{3}$$

Clearly, (3), and hence also (2), can be made arbitrarily small by making $|\mu - \mu_0|$, $|\sigma - \sigma_0|$, $\sup\{|f(x) - f_0(x)| : x \in I\}$ and $\sup\{|f'(x) - f_0'(x)| : x \in I\}$ all smaller than ϵ for a sufficiently small ϵ. Under the assumed condition on the prior distribution on μ, σ and f, the preceding event has positive probability.

Note that weak consistency implies posterior consistency of the corresponding cumulative distribution function F^* with respect to the uniform distance in view of Pólya's theorem, i.e.

$$\Pi(\sup\{|F^*(x) - F_0^*(x)| : x \in I\} > \epsilon | X_1^*, \ldots, X_{n^*}^*) \to 0 \text{ a.s. for any } \epsilon > 0.$$

Since the posterior distribution of (π^-, π^+) is consistent at (π_0^-, π_0^+), from (4.1), it follows that the posterior distribution of F is consistent at F_0 with respect to the uniform pseudo-distance on I, i.e.

$$\Pi(\sup\{|F(x) - F_0(x)| : x \in I\} > \epsilon | X_1^*, \ldots, X_{n^*}^* \; n_-^*, n_+^*) \to 0 \text{ a.s. for any } \epsilon > 0.$$

In view of the relations (4.3) and (4.4), it then follows that the posterior distributions of (μ, σ) is consistent at (μ_0, σ_0). Now using the fact that Φ^{-1} is uniformly continuous on compact subsets of $(0, 1)$, it follows from (4.2) that

$$\Pi(\sup\{|f(x) - f_0(x)| : x \in I\} > \epsilon | X_1^*, \ldots, X_{n^*}^* \; n_-^*, n_+^*) \to 0 \text{ a.s. for any } \epsilon > 0.$$

This completes the proof of theorem.

To prove the corollary, we observe that as the true transformation f_0 is strictly increasing and continuously differentiable, it is uniformly approximable on compact subintervals of $(0, 1)$ by a linear combination of B-splines with strictly monotone increasing coefficients, by applying Lemma 1(b) of Shen and Ghosal (2015) to the derivative function, where the derivative is also uniformly approximable on compact intervals. Now since the truncated normal distribution has positive density on a small neighborhood of a strictly increasing coefficient vector, the condition of prior positivity is fulfilled. Thus posterior consistency holds under the B-spline series prior with respect to the uniform pseudo-distance on any compact subset of $(0, 1)$.

References

Ghosal, S. and van der Vaart, A. (2017). *Fundamentals of Nonparametric Bayesian Inference*. Cambridge Series in Statistical and Probabilistic Mathematics (44). Cambridge University Press, Cambridge.

Shen, W. and Ghosal, S. (2015). Adaptive Bayesian procedures using random series priors: adaptive Bayesian procedures. *Scandinavian Journal of Statistics*, 42(4):1194–1213.