Bayesian Analysis (0000) **00**, Number 0, pp. 1

Discussion on "Latent Nested Nonparametric Priors" by Camerlenghi et al. (2019)

Li Ma*

I congratulate the authors on this excellent paper that provides deep insights into nested discrete processes, especially in revealing the distributional properties of the random partitions induced by these processes. Moreover, the authors skillfully embed the classical idea from Müller et al. (2004) into nested completely random measures to form a new class of latent nested nonparametric processes (LNNPs), thereby resolving a key difficulty in applying the nested DP (Rodríguez et al., 2008) and its variant (Rodriguez and Dunson, 2014) when tied observations are present across data sets that are otherwise different. I shall discuss two challenges that a practitioner of the LNNP might face and how the LNNP might be further generalized to address them. Because nested discrete processes involve two layers of clustering structures—one at the sample level and the other at the observation level, to avoid confusion, I shall distinguish the two types by referring to them as sample cluster (SC) and observation cluster (OC) respectively.

To see the two challenges, note that by constructing each sampling distribution as a weighted average between a shared measure and a SC-specific idiosyncratic measure, the LNNP assumes that (i) samples that belong to the same SC must share exactly the same sampling distribution without any within-SC sample-to-sample variability; (ii) samples from different SCs must share the same relative weights over the OCs induced by the shared measure. Both of these assumptions can be too restrictive in applications. Next I shall consider each of them in turn.

Incorporating within-SC variation. Sample-to-sample variation is prevalent in applications. Even otherwise similar data sets collected under the same controllable conditions will inevitably display variability to various extents from each other. In the current context, not incorporating such variation within-SC can lead to many small or even singleton SCs when the number of observations in each sample grows. This was previously pointed out by MacEachern (2008) in his illuminating discussion on the nested DP.

It appears that the flexible nested process framework proposed in this paper is capable of incorporating such within-SC variation with some extension. In particular, for each SC, we can introduce a dispersion parameter—which can be scalar, multivariate, or even infinite-dimensional—that characterizes how samples within the SC differ. Specifically, let Ω_d denote the (Borel measurable) space of all possible values of the dispersion parameter. Let the Poisson random measure be defined as $\tilde{N} = \sum_{i\geq 1} \delta_{(J_i,G_i,w_i)}$ on $\mathbb{R}^+ \times \mathrm{P}_{\mathbb{X}} \times \Omega_d$ with a mean intensity function $\nu(\mathrm{d}s, \mathrm{d}p, \mathrm{d}v) = C\rho(s)\mathrm{d}s\, Q(\mathrm{d}p \times \mathrm{d}v)$, where Q is now a probability measure on the product space $\mathrm{P}_{\mathbb{X}} \times \Omega_d$.

Accordingly, define a CRM $\tilde{\mu} = \sum_{i\geq 1} J_i \delta_{(G_i,w_i)}$, and its normalized version

$$\tilde{q} := \frac{\tilde{\mu}}{\tilde{\mu}(\mathrm{P}_{\mathbb{X}} \times \Omega_d)}.$$

*Department of Statistical Science, Duke University, Durham, NC 27708, USA. li.ma@duke.edu

 DOI: 0000

Given $\tilde{q}$, we can generate K sampling distributions $\tilde{p}_1, \tilde{p}_2, \ldots \tilde{p}_K$ hierarchically from their respective SCs with the corresponding within-SC variation as follows. First, we generate the SC centers (e.g., the mean distributions) and the within-SC dispersions

$$(\tilde{p}_{01}, \tilde{v}_1), \ldots, (\tilde{p}_{0K}, \tilde{v}_K) \,|\, \tilde{q} \sim \tilde{q}^K.$$

Given the SC centers and within-SC dispersions, we generate the sampling distributions

$$\tilde{p}_i \,|\, \tilde{p}_{0i}, \tilde{v}_i \overset{\text{ind}}{\sim} F(\tilde{p}_{0i}, \tilde{v}_i) \qquad \text{for } i = 1, 2, \ldots, K,$$

where $F(p, v)$ represents a "location-scale" family in the space of probability distributions with location (or center) p and scale (or dispersion) v.

While popular random measures such as the DP and the Pitman-Yor process can serve as this "location-scale" family, they are limited in that their dispersion parameter is either scalar or otherwise low-dimensional, and thus they cannot characterize flexible within-SC variation. A particularly flexible "location-scale" family is the Pólya tree (PT), which has an infinite-dimensional dispersion parameter with $\Omega_d = [0, \infty)^\infty$ as pointed out in Berger and Guglielmi (2001). Following this route, Christensen and Ma (In press) demonstrated a special case of the above nested model with $\rho(s) = s^{-1}e^{-s}$ and Q the product of an "adaptive PT" and a "stochastically increasing shrinkage" hyperprior on the dispersion parameter, both introduced in Ma (2017).

Allowing different relative weights on shared OCs. In many applications, the samples belonging to different SCs share OCs, and this is indeed one of the key motivations for the authors to introduce the LNNP as nested processes alone do not allow this feature. By introducing a shared component and building each sampling distribution as a weighted average of a shared and an idiosyncratic component in the style of Müller et al. (2004), the LNNP assumes that the idiosyncratic components do not share any OCs whereas the shared component must endow all the shared OCs with exactly the same relative weights among them. These constraints can be overly restrictive in practice. For example, Soriano and Ma (2019) considered the application of flow cytometry, where the observations are blood cells and each OC corresponds to a cell subtype (e.g., T-cells, B-cells, etc.). Different subtypes of patients (the SCs) will share some of the same cell subtypes, or one should hope so as they are all humans! In other words, the actual SCs might differ only in the weights of some OCs, not their identities.

The strategy that Soriano and Ma (2019) employed to address this issue is to let all the samples share a common set of OCs, but introduce shared and idiosyncratic components only in generating the weights of these OCs. This way, while the shared component still corresponds to the OCs with the same relative weights across SCs as in the LNNP, the idiosyncratic components now allow SCs to have distinct weights on common OCs. (Note that this strategy still allows SCs to have unique OCs in that an SC without a certain OC will just have a very small weight on that OC.) As Surya Tokdar pointed out through personal communications, this strategy is essentially a (limiting) version of a Müller et al. (2004) style mixture of a shared DP and idiosyncratic DPs generated from a hierarchical DP (Teh et al., 2006). Unlike the LNNP, this "LHDP" does not allow inference on the partition at the sample level. It is of interest to investigate how to allow different weights on shared OCs across SCs in the LNNP in such a way that maintains its ability to carry out the partitioning or clustering on the samples.

References

Berger, J. O. and Guglielmi, A. (2001). "Bayesian and Conditional Frequentist Testing of a Parametric Model versus Nonparametric Alternatives." *Journal of the American Statistical Association*, 96(453): 174–184.
URL http://www.jstor.org/stable/2670357

Christensen, J. and Ma, L. (In press). "A Bayesian hierarchical model for related densities using Polya trees." *Journal of the Royal Statistical Society. Series B.*, arXiv:1710.01702.

Ma, L. (2017). "Adaptive Shrinkage in Pólya Tree Type Models." *Bayesian Anal.*, 12(3): 779–805.
URL https://doi.org/10.1214/16-BA1021

MacEachern, S. N. (2008). "Discussion of "The nested Dirichlet process" by A.E. Gelfand, D.B. Dunson and A. Rodriguez." *Journal of the American Statistical Association*, 103: 1149–1151.

Müller, P., Quintana, F., and Rosner, G. (2004). "A method for combining inference across related nonparametric Bayesian models." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(3): 735–749.

Rodriguez, A. and Dunson, D. B. (2014). "Functional clustering in nested designs: Modeling variability in reproductive epidemiology studies." *Ann. Appl. Stat.*, 8(3): 1416–1442.
URL https://doi.org/10.1214/14-AOAS751

Rodríguez, A., Dunson, D. B., and Gelfand, A. E. (2008). "The Nested Dirichlet Process." *Journal of the American Statistical Association*, 103(483): 1131–1154.
URL http://dx.doi.org/10.1198/016214508000000553

Soriano, J. and Ma, L. (2019). "Mixture Modeling on Related Samples by ψ -Stick Breaking and Kernel Perturbation." *Bayesian Anal.*, 14(1): 161–180.
URL https://doi.org/10.1214/18-BA1106

Teh, Y. W., Jordan, M. I., Beal, M. J., and Blei, D. M. (2006). "Hierarchical Dirichlet processes." *Journal of the American Statistical Association*, 101(476).