Contributed Discussion

Giovanni Rebaudo[*], Augusto Fasano[†], Beatrice Franzolini[‡], and Peter Müller[§]

We congratulate the authors for a very interesting paper, which provides a concrete contribution to the Bayesian nonparametric (BNP) literature. Their work provides an efficient method to evaluate the sensitivity of posterior quantities of interest – computed through variational Bayes approximations – to the prior distribution of the mixing weights in Bayesian discrete mixture models. The authors argue for sensitivity checks to uncover possible non-robustness of the results to prior settings. Importantly, they develop an easy-to-use and efficient method to do it in BNP mixtures. In the following, we illustrate our comments on the most widely used construction discussed by the authors: the Dirichlet process mixture (DPM) model (Lo, 1984), namely

$$X_i \mid \theta_i \overset{\text{ind}}{\sim} \mathcal{P}(\cdot|\theta_i), \quad \theta_i \mid \tilde{P} \overset{\text{iid}}{\sim} \tilde{P}, \quad \tilde{P} \sim \mathrm{DP}(\alpha, \mathcal{P}_{\text{base}}). \tag{1}$$

Following the authors, in such a setting one of the goals is to perform inference about the random number of clusters, G_{cl}, that is the number of occupied mixture components $\mathcal{P}$ in a sample of size N, with a particular focus on its expected value. In this regard, it is worth noticing that robustness is relevant just in terms of the specific quantity of interest or decision of the analysis.

First, we agree with the authors that sensitivity analysis to prior assumptions should be done routinely if the prior specification is driven by mathematical convenience or heuristics, as often is in BNP models. This would highlight the influence of the specific assumptions on the results of the analysis, pointing to which of them should be justified more strongly. After assessing sensitivity, the next fundamental question is: *how can we justify probabilistic assumptions in the challenging infinite/high-dimensional Bayesian settings?* In principle, one possibility could be provided by the subjective Bayesian paradigm. According to it, the prior should reflect *a priori* opinions on quantities of interest. Those are more easily elicited when expressed directly in terms of observable values (see e.g., Fortini and Petrone, 2016). However, this approach can be particularly challenging in the BNP world due to the infinite/high-dimensionality of the parameter space. One way to tackle this issue and elicit prior assumptions consists in working with prior predictive distributions or with the *a priori* expected value of the number of clusters (see e.g., De Blasi et al., 2015). However, posterior inference strongly depends also on other *a priori* assumptions, such as the choice of the mixture kernel and the base measure in DPM. Another way to justify the choice of the prior is in terms of the properties of the summaries of interest (e.g., consistency of G_{cl}) assuming an ideal frequentist truth (Nobile, 1994; Miller and Harrison, 2018; Ascolani et al., 2022). Furthermore, different prior settings can be also specified in a given dataset by tuning

[*]University of Texas at Austin, TX, USA, giovanni.rebaudo@austin.utexas.edu
[†]Collegio Carlo Alberto, Torino, Italy, augusto.fasano@carloalberto.org
[‡]Agency for Science, Technology and Research, Singapore, franzolini@pm.me
[§]University of Texas at Austin, TX, USA, pmueller@math.utexas.edu

the hyperparameters in terms of predictive accuracy, e.g. via cross-validation. All of the above methods – as well as other possible alternatives, including popular empirical-based approaches (Liu, 1996; McAuliffe et al., 2006) – require to specify some subjective assumptions (e.g., homogeneity assumptions, prior distribution, data generating process, loss function). The implications of such assumptions are challenging to assess, pointing toward the need for further research, especially in the BNP mixture framework.

Second, considering the sensitivity of the stick-breaking prior to values of α in (1), it would be interesting to assess how the specification of a prior for the concentration parameter could increase robustness. The use of a prior on α leads to a mixing measure that is itself a mixture in the sense of Antoniak (1974). Ideally, this would allow learning from the data which values of α are most appropriate for the data at hand. Consequently, it would be very interesting to investigate how the results and the sensitivity checks proposed by the authors could be embedded in such a framework.

Another useful extension of the idea and techniques developed by the authors is to provide a toolkit that assesses sensitivity to the choice of the kernel or of the base measure of the DP. A common choice of kernel $\mathcal{P}$ and base measure $\mathcal{P}_{\text{base}}$ in (1) are the Gaussian kernel and the conjugate normal-inverse-Wishart base measure, respectively. Such assumptions are typically motivated by mathematical convenience and the choice of the hyperparameters of the base measure is mainly carried out following heuristics. However, posterior inference on the number of clusters strongly depends on such assumptions as shown empirically and theoretically (see e.g., Petralia et al., 2012; Cai et al., 2021; Chandra et al., 2020).

Finally, as anticipated by the authors, it would be very interesting to exploit the general results developed in the work to obtain an easy-to-use tool to check the sensitivity also for mixture models under different prior distributions for the mixing measure in the popular class of Gibbs-type priors (Gnedin and Pitman, 2006; De Blasi et al., 2015) as well as for other approximations arising from different divergences or distances.

To conclude we believe the work by Giordano, Liu, Jordan, and Broderick can stimulate further computational research in the Bayesian community and be applied in many practical situations. We commend them one more time for a remarkable paper.

References

Antoniak, C. E. (1974). "Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems." *Annals of Statistics*, 2: 1152–1174. MR0365969. doi: https://doi.org/10.1214/aos/1176342871. 60

Ascolani, F., Lijoi, A., Rebaudo, G., and Zanella, G. (2022). "Clustering consistency with Dirichlet process mixtures." *Biometrika*, in press. doi: https://doi.org/10.1093/biomet/asac051. 59

Cai, D., Campbell, T., and Broderick, T. (2021). "Finite mixture models do not reliably learn the number of components." In *International Conference on Machine Learning*, PMLR, volume 139, 1158–1169. 60

Chandra, N. K., Canale, A., and Dunson, D. B. (2020). "Escaping the curse of dimensionality in Bayesian model based clustering." *Preprint at arXiv:2006.02700*. doi: https://doi.org/10.48550/arXiv.2006.02700. 60

De Blasi, P., Favaro, S., Lijoi, A., Mena, R. H., Prünster, I., and Ruggiero, M. (2015). "Are Gibbs-type priors the most natural generalization of the Dirichlet process?" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37: 212–229. doi: https://doi.org/10.1109/TPAMI.2013.217. 59, 60

Fortini, S. and Petrone, S. (2016). *Predictive distribution (de Finetti's view)*, 1–9. Wiley StatsRef: Statistics Reference Online. doi: https://doi.org/10.1002/9781118445112.stat07831. 59

Gnedin, A. V. and Pitman, J. (2006). "Exchangeable Gibbs partitions and Stirling triangles." *Journal of Mathematical Sciences* , 138: 5674–5685. MR2160320. doi: https://doi.org/10.1007/s10958-006-0335-z. 60

Liu, J. S. (1996). "Nonparametric hierarchical Bayes via sequential imputations." *Annals of Statistics*, 24: 911–930. MR1401830. doi: https://doi.org/10.1214/aos/1032526949. 60

Lo, A. Y. (1984). "On a class of Bayesian nonparametric estimates: I. density estimates." *Annals of Statistics*, 12: 351–357. MR0733519. doi: https://doi.org/10.1214/aos/1176346412. 59

McAuliffe, J. D., Blei, D. M., and Jordan, M. I. (2006). "Nonparametric empirical Bayes for the Dirichlet process mixture model." *Statistics and Computing*, 16: 5–14. MR2224185. doi: https://doi.org/10.1007/s11222-006-5196-2. 60

Miller, J. W. and Harrison, M. T. (2018). "Mixture models with a prior on the number of components." *Journal of the American Statistical Association*, 113: 340–356. MR3803469. doi: https://doi.org/10.1080/01621459.2016.1255636. 59

Nobile, A. (1994). "Bayesian Analysis of Finite Mixture Distributions." Ph.D. thesis, Carnegie Mellon University. MR2692049. 59

Petralia, F., Rao, V., and Dunson, D. B. (2012). "Repulsive mixtures." In *Advances in Neural Information Processing Systems*, volume 25, 1889–1897. 60