Contributed Discussion

Xenia Miscouridou* and Francesca Panero†

We congratulate the authors for the interesting and novel work on the evaluation of the sensitivity of a set of stick-breaking priors via mean-field variational Bayes. The paper focuses on Dirichlet process mixtures (DPM), a popular prior distribution in many applications. The widespread use of DPMs makes it vital to be able to understand their properties and the implications of their use. This method has the potential to become part of the toolkit of statisticians who would like to pursue applications under the Bayesian nonparametric (BNP) framework.

Our discussion focuses on possible extensions of the current work to non stick-breaking priors. We motivate why these random probability measures deserve to be considered for a similar sensitivity analysis and suggest a possible way to adapt the framework of Giordano et al.'s using some recently developed finite approximations of completely random measures.

The authors provide a computational tool to quickly and automatically assess the sensitivity to prior specification of variational Bayes (VB) approximations in the particular case of some stick-breaking priors. They focus on the canonical Dirichlet process mixture model, heavily used in topic modelling and clustering, and suggest that the methods apply directly to any discrete BNP model that admits a truncated stick-breaking construction with independent and identically distributed (iid) proportions $(\nu_k)_k$. The Dirichlet process (DP) is arguably the most widely used discrete random probability measure admitting a stick-breaking representation. It belongs to a wider family of species sampling models known as Gibbs-type priors (see, for example, De Blasi et al. (2013)), which are characterised by a particular form of the exchangeable partition probability function. Other models within this family are the Pitman-Yor (PY) process, the normalised σ-stable process, the normalised generalised gamma process (NGGP) and the uniform process (Wallach et al. (2010)).

Gibbs-type priors do not necessarily admit a stick-breaking representation. As already mentioned in the Invited Discussion by J. Griffin and M. Kalli, an exciting result would be to obtain an extension of the sensitivity analysis provided by Giordano et al.'s to the broader family of Gibbs-type models. Expanding on this, we will highlight in the following paragraphs the notable clustering properties of some random probability measures other than the Dirichlet process, and suggest a possible way to address the sensitivity analysis despite the lack of stick-breaking representations.

The generalised gamma process (GGP) (Hougaard, 1986; Brix, 1999), also known

*Department of Mathematics, Imperial College London, 180 Queen's Gate, South Kensington, London SW7 2AZ, x.miscouridou@imperial.ac.uk

†Department of Statistics, London School of Economics and Political Science, 69 Aldwych, London WC2B 4RR, f.panero@lse.ac.uk

as (exponentially) tilted stable process, has mean measure

$$\rho(dw) = \frac{1}{\Gamma(1-\sigma)} w^{-1-\sigma} e^{-\tau w} dw, \tag{1}$$

where $\sigma \in (0,1)$ and $\tau \geq 0$, or $\sigma \leq 0$ and $\tau > 0$.

Clustering models based on the DP or PY priors can only describe clusters whose size grows linearly with the sample size n. Di Benedetto et al. (2021) propose a class of random partition models based on the GGP which is able to generate partitions whose cluster sizes grow sublinearly with n, a property known as microclustering. In particular, their model offers a power-law growth of cluster sizes with exponent in $(0, 1)$. While Di Benedetto et al. (2021) employed an MCMC approach for inference, a variational approach has computational and practical advantages. For a variational framework one needs to consider approximations to these distributions. Lee et al. (2016, 2017) proceed in this direction by introducing and using finite dimensional approximations of the GGP (and other infinite measures). Precisely, they use the BFRY (Devroye and James, 2014) distributions to approximate the infinite measures for power-law mixture models and graphs with power-law degree distribution within a mean-field variational inference framework. For some context on BFRY distributions,[1] recall that a BFRY$(\tau), \tau \in (0,1)$ random variable X is characterised by a density function $f_\tau(x) = \tau(1-e^{-x})/(\Gamma(1-\tau)x^{1+\tau})$, $x > 0$. It is infinitely divisible and can conveniently sampled as a ratio of gamma and uniform random variables. Heading a step further, Lee et al. (2022) generalise the BFRY priors giving more generic series representations and iid approximations for both the GGP and stable beta process. This suggests the following question: can we adapt the proposed sensitivity toolbox to the case of series representations and iid approximations proposed in Lee et al. (2017, 2022) to cover these interesting applications? In this way one would obtain a sensitivity analysis for microclustering or other applications of the GGP which have a power-law behaviour.

Recently, there was a lot of attention on graph modelling with power-law behaviour as these can model real-world graphs with node heterogeneity. Sparse random graphs with power-law degree distributions were originally introduced by Caron and Fox (2017) who used a GGP process prior on the network node parameters. A series of papers (Herlau et al. (2016); Miscouridou et al. (2018); Todeschini et al. (2020); Naik et al. (2021, 2022)) followed, expanding the properties and types of graphs. In all of these works, the parameter σ in Eq (1) is crucial as it tunes the sparsity of the graph and the degree heterogeneity (power-law), therefore a desirable result would be to come up with a similar computational toolbox to evaluate the sensitivity to the GGP process prior focusing on σ for graph modelling.

References

Bertoin, J., Fujita, T., Roynette, B., and Yor, M. (2006). "On a particular class of self-decomposable random variables: the durations of Bessel excursions straddling

[1]The name BFRY was coined in Devroye and James (2014) after the work of Bertoin et al. (2006), who used this random variable in the study of excursion duration in Bessel processes.

independent exponential times." *Probability and Mathematical Statistics*, (26): 315–366. MR2325310. 349

Brix, A. (1999). "Generalized gamma measures and shot-noise Cox processes." *Advances in Applied Probability*, 31(4): 929–953. MR1747450. doi: https://doi.org/10.1239/aap/1029955251. 348

Caron, F. and Fox, E. (2017). "Sparse Graphs using Exchangeable Random Measures." *Journal of the Royal Statistical Society B*, 79: 1295–1366. Part 5. MR3731666. doi: https://doi.org/10.1111/rssb.12233. 349

De Blasi, P., Favaro, S., Lijoi, A., Mena, R. H., Prünster, I., and Ruggiero, M. (2013). "Are Gibbs-type priors the most natural generalization of the Dirichlet process?" *IEEE transactions on pattern analysis and machine intelligence*, 37(2): 212–229. 348

Devroye, L. and James, L. F. (2014). "On simulation and properties of the stable law." *Statistical Methods and Applications*,, 3(23): 307–343. MR3233961. doi: https://doi.org/10.1007/s10260-014-0260-0. 349

Di Benedetto, G., Caron, F., and Teh, Y. W. (2021). "Nonexchangeable random partition models for microclustering." *Annals of Statistics*, 49(4): 1931–1957. MR4319236. doi: https://doi.org/10.1214/20-aos2003. 349

Herlau, T., Schmidt, M. N., and Mørup, M. (2016). "Completely random measures for modelling lock-structured sparse networks." *Advances in Neural Information Processing Systems*, 29. 349

Hougaard, P. (1986). "Survival models for heterogeneous populations derived from stable distributions." *Biometrika*, 73(2): 387–396. MR0855898. doi: https://doi.org/10.1093/biomet/73.2.387. 348

Lee, J., Heaukulani, C., Ghahramani, Z., James, L. F., and Choi, S. (2017). "Bayesian inference on random simple graphs with power law degree distributions." In *Proceedings of the 34th International Conference on Machine Learning*. 349

Lee, J., James, L. F., and Choi, S. (2016). "Finite-dimensional BFRY priors and variational Bayesian inference for power-law models." In *Advances in Neural Information Processing Systems 29 (NIPS 2016)*. 349

Lee, J., Miscouridou, X., and Caron, F. (2022). "A unified construction of series representations and iid approximations of completely random measures." *Bernoulli*. 349

Miscouridou, X., Caron, F., and Teh, Y. W. (2018). "Modelling sparsity, heterogeneity, reciprocity and community structure in temporal interaction data." *Advances in Neural Information Processing Systems*, 31. 349

Naik, C., Caron, F., and Rousseau, J. (2021). "Sparse networks with core-periphery structure." *Electronic Journal of Statistics*, 15(1): 1814–1868. MR4255305. doi: https://doi.org/10.1214/21-ejs1819. 349

Naik, C., Caron, F., Rousseau, J., Teh, Y. W., and Palla, K. (2022). "Bayesian Nonparametrics for Sparse Dynamic Networks." *European Conference on Machine Learning and Data Mining (ECML PKDD)*. 349

Todeschini, A., Miscouridou, X., and Caron, F. (2020). "Exchangeable random measures for sparse and modular graphs with overlapping communities." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(2): 487–520. MR4084173. doi: https://doi.org/10.1111/rssb.12363. 349

Wallach, H. M., Jensen, S. T., Dicker, L., and Heller, K. A. (2010). "An Alternative Prior Process for Nonparametric Bayesian Clustering." In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. 348