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

Comment on Article by Paganin et al.*

Isadora Antoniano-Villalobos [†] Cristiano Villa [‡] and Sara Wade[§],

We congratulate the authors on a well written article containing innovative ideas and a compelling application. The authors propose to incorporate subjective prior information on the clustering structure by defining a centred partition process. The proposed family of processes combines well known priors for partitions (specifically exchangeable partition probability functions, EPPFs) with a measure of discrepancy from an initial partition $\boldsymbol{c}_0$, summarizing prior belief.

While it may be difficult to subjectively elicit $\boldsymbol{c}_0$ for large sample sizes, one limitation of the proposed methodology is that the prior calibration strategy is computationally expensive and therefore limited to small sample sizes. It also becomes too expensive to include hyperpriors on key parameters of the EPPF, such as the concentration parameter α of the Dirichlet process, which controls the number of clusters.

We discuss a simple, alternative idea to include the prior clustering information. Specifically, we include the initial partition $\boldsymbol{c}_0$ as a covariate, through dependent Dirichlet processes (e.g. MacEachern, 2000; Dunson and Park, 2008; Griffin and Steel, 2006; Rodriguez and Dunson, 2011) or more generally, dependent normalized random measures (e.g. Griffin and Leisen, 2017; Chen et al., 2013; Lijoi et al., 2014; Griffin et al., 2013). In this case, we view $\boldsymbol{c}_0$ as a categorical covariate, with each element $c_{0,i}$ indicating the cluster allocation of the ith data point in the initial partition. For example, we focus on the dependent normalized weights model proposed in Antoniano-Villalobos et al. (2014) and define:

$$\mathrm{p}(c_i = j|\boldsymbol{c}_0, \boldsymbol{w}, \boldsymbol{p}) = \frac{w_j \mathrm{Cat}(c_{0,i}|\boldsymbol{p}_j)}{\sum_{j'=1}^{\infty} w_{j'} \mathrm{Cat}(c_{0,i}|\boldsymbol{p}_{j'})} = \frac{w_j p_{j,c_{0,i}}}{\sum_{j'=1}^{\infty} w_{j'} p_{j',c_{0,i}}},$$

where $\boldsymbol{w} = (w_1, w_2, \ldots)$ and $\boldsymbol{p} = (\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots)$ are the parameters defining the dependent weights. Specifically, we can assume $\boldsymbol{w}$ follow a stick-breaking construction with mass parameter α, and the $\boldsymbol{p}_j = (p_{j,1}, \ldots, p_{j,k_0})$ are iid with $\boldsymbol{p}_j \sim \mathrm{Dir}(\beta/k_0, \ldots, \beta/k_0)$.

In this construction, we can study the following limiting cases. On one hand, if $\beta \to \infty$, then

$$\boldsymbol{p}_j \to (1/k_0, \ldots, 1/k_0)$$

with probability one. Thus, the prior on the partition $\boldsymbol{c}$ converges to the EPPF induced by the DP with mass parameter α. On the other hand, if $\beta \to 0$, then base measure on $\boldsymbol{p}_j$

*Paganin, S., Herring, A. H., Olshan, A. F., and Dunson, D. B. (2021). "Centered Partition Processes: Informative Priors for Clustering."

†Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University of Venice, Italy isadora.antoniano@unive.it

‡School of Mathematics, Statistics and Physics, Newcastle University, UK Cristiano.Villa@newcastle.ac.uk

§School of Mathematics, University of Edinburgh, UK sara.wade@ed.ac.uk

 DOI: 0000

converges to a uniform discrete distribution over the vertices of the simplex. Moreover, when $\alpha \to 0$ and $\beta \to 0$, $\boldsymbol{c} = \boldsymbol{c}_0$ with probability one.

While not as intuitive as the centred partition process, which includes a parameter ψ to control the tradeoff between the EPPF and subjective information, it is possible to fix the value of α to reflect prior belief in the number of clusters and β to reflect the strength of belief in $\boldsymbol{c}_0$. Moreover, an advantage of this alternative approach is the ability to place hyperpriors on α and β, thus avoiding the calibration issues and making the method more robust to misspecification. This also helps to scale to larger sample sizes.

We highlight an additional advantage of this alternative approach is the ability to simulate from the prior by using a reasonable computable truncation for the mixture weights. This can be exploited to investigate prior sensitivity, callibration or elicitation. Furthermore, prediction for new observations could incorporate prior information regarding clustering, without requiring a recalibration of the prior and subsequent recalculation of the posterior.

References

Antoniano-Villalobos, I., Wade, S., and Walker, S. (2014). "A Bayesian Nonparametric Regression Model With Normalized Weights: A Study of Hippocampal Atrophy in Alzheimer's Disease." *Journal of the American Statistical Association*, 109(506): 477–490. 1

Chen, C., Rao, V., Buntine, W., and Teh, Y. W. (2013). "Dependent normalized random measures." In *International Conference on Machine Learning*, 969–977. PMLR. 1

Dunson, D. and Park, J. (2008). "Kernel Stick-Breaking Processes." *Biometrika*, 95: 307–323. 1

Griffin, J. and Leisen, F. (2017). "Compound Random Measures and their use in Bayesian Nonparametrics." *JRSS B*, 79(2): 525–545. 1

Griffin, J. and Steel, M. (2006). "Order-Based Dependent Dirichlet Processes." *Journal of the American Statistical Association*, 10: 179–194. 1

Griffin, J. E., Kolossiatis, M., and Steel, M. F. (2013). "Comparing distributions by using dependent normalized random-measure mixtures." *Journal of the Royal Statistical Society: SERIES B: Statistical Methodology*, 499–529. 1

Lijoi, A., Nipoti, B., and Prünster, I. (2014). "Bayesian inference with dependent normalized completely random measures." *Bernoulli*, 20(3): 1260–1291. 1

MacEachern, S. (2000). "Dependent Dirichlet Processes." Technical report, Department of Statistics, Ohio State University. 1

Rodriguez, A. and Dunson, D. (2011). "Nonparametric Bayesian Models through Probit Stick-Breaking Processes." *Bayesian Analysis*, 6: 145–178. 1