Bayesian Analysis (2021) **16**, Number 1, pp. 301–370

Centered Partition Processes: Informative Priors for Clustering (with Discussion)

Sally Paganin[*], Amy H. Herring[†], Andrew F. Olshan[‡], David B. Dunson[§], and The National Birth Defects Prevention Study

Abstract. There is a very rich literature proposing Bayesian approaches for clustering starting with a prior probability distribution on partitions. Most approaches assume exchangeability, leading to simple representations in terms of Exchangeable Partition Probability Functions (EPPF). Gibbs-type priors encompass a broad class of such cases, including Dirichlet and Pitman-Yor processes. Even though there have been some proposals to relax the exchangeability assumption, allowing covariate-dependence and partial exchangeability, limited consideration has been given on how to include concrete prior knowledge on the partition. For example, we are motivated by an epidemiological application, in which we wish to cluster birth defects into groups and we have prior knowledge of an initial clustering provided by experts. As a general approach for including such prior knowledge, we propose a Centered Partition (CP) process that modifies the EPPF to favor partitions close to an initial one. Some properties of the CP prior are described, a general algorithm for posterior computation is developed, and we illustrate the methodology through simulation examples and an application to the motivating epidemiology study of birth defects.

Keywords: Bayesian clustering, Bayesian nonparametrics, centered process, Dirichlet Process, exchangeable probability partition function, mixture model, product partition model.

Contributed Discussion

Tommaso Rigon[*], Emanuele Aliverti[†], Massimiliano Russo[‡], and Bruno Scarpa[§]

We congratulate the authors on an interesting paper, which provides a concrete contribution in Bayesian nonparametric methods. The proposed centered partition (CP) process $p(\boldsymbol{c} \mid \boldsymbol{c}_0)$ is an exponential contamination of a baseline process $p_0(\boldsymbol{c})$ towards a fixed partition $\boldsymbol{c}_0$. The authors suggest a Gibbs-type specification for the baseline distribution $p_0(\boldsymbol{c})$, since this class displays a nice balance between flexibility and complexity (Lijoi et al., 2007). The CP informs the clustering process exploiting existing prior knowledge about the partition.

The CP process is defined as $p(\boldsymbol{c} \mid \boldsymbol{c}_0) \propto p_0(\boldsymbol{c}) \exp\{-\psi d(\boldsymbol{c}, \boldsymbol{c}_0)\}$, with $\psi > 0$ being a penalization parameter, and $d(\boldsymbol{c}, \boldsymbol{c}_0)$ being a metric between partitions, such as the Variation of Information (VI). The CP process can be also interpreted as a *generalized Bayes posterior*, in the sense of Bissiri et al. (2016). Within such a framework, the baseline distribution $p_0(\boldsymbol{c})$ represents the prior belief about an unknown partition, whereas $\boldsymbol{c}_0$ is regarded as a data point. Moreover, in the generalized Bayes terminology the distance $d(\boldsymbol{c}, \boldsymbol{c}_0)$ is the *loss function*, meaning that the parameter $\psi > 0$ balances the importance of the observations relative to the prior. This perspective leads to an alternative interpretation of CP processes, where $p(\boldsymbol{c} \mid \boldsymbol{c}_0)$ can be regarded as the posterior belief about the partition conditionally on the observation $\boldsymbol{c}_0$.

Such a generalized Bayes interpretation leads to interesting modeling extensions. In many practical contexts, it might be difficult to select a single $\boldsymbol{c}_0$ encapsulating our prior knowledge about the partition. Instead, it might be easier to identify several plausible partitions that well describe the phenomenon under consideration. For example, in the application considered by the authors, different investigators could provide equally plausible mechanistic groups of the birth defects $\boldsymbol{c}_{0,1}, \ldots, \boldsymbol{c}_{0,S}$. Following Bissiri et al. (2016), it is natural to include all these representative partitions in an additive manner, namely

$$p(\boldsymbol{c} \mid \boldsymbol{c}_{0,1}, \ldots, \boldsymbol{c}_{0,S}) \propto p_0(\boldsymbol{c}) \exp\left\{-\psi \sum_{s=1}^{S} d(\boldsymbol{c}, \boldsymbol{c}_{0,s})\right\}. \tag{1}$$

The above conditional distribution can be regarded as the posterior distribution of $\boldsymbol{c}$ given the observations $\boldsymbol{c}_{0,1}, \ldots, \boldsymbol{c}_{0,S}$. As $\psi \to 0$ the distribution $p(\boldsymbol{c} \mid \boldsymbol{c}_{0,1}, \ldots, \boldsymbol{c}_{0,S})$ converges to the baseline law $p_0(\boldsymbol{c})$. However, when $\psi \to \infty$ then $p(\boldsymbol{c} \mid \boldsymbol{c}_{0,1}, \ldots, \boldsymbol{c}_{0,S})$ converges to a discrete distribution function placing mass over the set of partitions

[*]Department of Economics, Management & Statistics, University of Milano-Bicocca, Milano, Italy, tommaso.rigon@unimib.it

[†]Department of Economics, University Ca' Foscari, Venezia, Italy, emanuele.aliverti@unive.it

[‡]Harvard–MIT Center for Regulatory Science, Harvard Medical School and Department of Data Science Dana-Farber Cancer Institute, Boston, USA, m_russo@hms.harvard.edu

[§]Department of Statistical Sciences and Department of Mathematics "Tullio Levi-Civita", University of Padova, Padova, Italy, bruno.scarpa@unipd.it

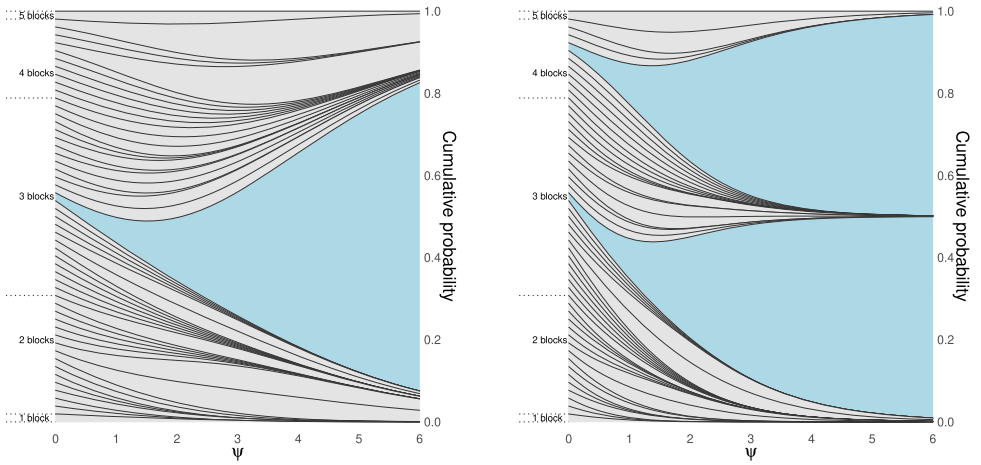

Figure 1: Prior probabilities of the 52 partitions of $N = 5$ elements for the CP process with $p_0(\boldsymbol{c}) \propto 1$. Left panel corresponds to the CP process centered on a single partition $\boldsymbol{c}_0 = \{1, 2\}\{3, 4\}\{5\}$. Right panel refers to a CP process centered on two partitions: $\boldsymbol{c}_{0,1} = \{1, 2\}\{3, 4\}\{5\}$ and $\boldsymbol{c}_{0,2} = \{1\}\{2\}\{3, 4\}\{5\}$. The cumulative probabilities across different values of the penalization parameter ψ are joined to form the curves, so that the probability of a given partition corresponds to the area between the curves. Blue areas correspond to the centering partitions $\boldsymbol{c}_0$ (left plot), and $\boldsymbol{c}_{0,1}$, $\boldsymbol{c}_{0,2}$ (right plot).

$\hat{\boldsymbol{c}}_1, \ldots, \hat{\boldsymbol{c}}_M$, corresponding to the minimizers of

$$\min_{\boldsymbol{c}} \sum_{s=1}^{S} d(\boldsymbol{c}, \boldsymbol{c}_{0,s}),$$

where M represents the number of solutions of the above minimization problem. Broadly speaking, each $\hat{\boldsymbol{c}}_m$, for $m = 1, \ldots, M$, is an "average" partition summarizing the information contained in the observations $\boldsymbol{c}_{0,1}, \ldots, \boldsymbol{c}_{0,S}$. Hence, the distribution $p(\boldsymbol{c} \mid \boldsymbol{c}_{0,1}, \ldots, \boldsymbol{c}_{0,S})$ can be arguably regarded as a CP process with multiple centers $\hat{\boldsymbol{c}}_1, \ldots, \hat{\boldsymbol{c}}_M$. Such a generalization of the CP is fairly straightforward and it might have useful practical implications, especially if there is uncertainty about the fixed partition $\boldsymbol{c}_0$. In addition, the Gibbs sampling devised by Paganin et al. (2021) can be easily modified to account for this extension.

In Figure 1 we reproduce Figure 2 of Paganin et al. (2021) and we illustrate the effect of our multi-centers extension. We compare the model of Paganin et al. (2021) when $p_0(\boldsymbol{c}) \propto 1$ and $\boldsymbol{c}_0 = \{1, 2\}\{3, 4\}\{5\}$, with the extension in (1) when $p_0(\boldsymbol{c}) \propto 1$, $S = 2$, and $\boldsymbol{c}_{0,1} = \{1, 2\}\{3, 4\}\{5\}$, $\boldsymbol{c}_{0,2} = \{1\}\{2\}\{3, 4\}\{5\}$. Larger values of ψ increase the prior probability assigned to $\boldsymbol{c}_0$ in the left panel, and to each of the centers $\hat{\boldsymbol{c}}_1, \ldots, \hat{\boldsymbol{c}}_M$ in the right panel. These centers represent the partitions that are more similar in terms of VI to $\boldsymbol{c}_{0,1}$ and $\boldsymbol{c}_{0,2}$. In this specific scenario, the centers $\hat{\boldsymbol{c}}_1, \ldots, \hat{\boldsymbol{c}}_M$ actually coincide with the data points $\boldsymbol{c}_{0,1}, \boldsymbol{c}_{0,2}$ and $S = M = 2$, but this is not always the case.

References

Bissiri, P., Holmes, C., and Walker, S. (2016). "A general framework for updating belief distributions." *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 78(5): 1103–1130. MR3557191. doi: https://doi.org/10.1111/rssb.12158. 348

Lijoi, A., Mena, R., and Prünster, I. (2007). "Controlling the reinforcement in Bayesian non-parametric mixture models." *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 69(4): 715–740. MR2370077. doi: https://doi.org/10.1111/j.1467-9868.2007.00609.x. 348

Paganin, S., Herring, A. H., Olshan, A. F., Dunson, D. B., et al. (2021). "Centered partition processes: Informative priors for clustering." *Bayesian Analysis*. 349