Contributed Discussion

Marta Catalano* and Francesco Gaffi†

We would like to congratulate the authors for a fine piece of work that cleverly links the stick-breaking procedure to random trees in the context of defining discrete random probability measures. They observe that, in the classic stick-breaking (Sethuraman, 1994), at every iteration one keeps breaking the *same* piece of the stick, with the length of the leftover parts giving the random weights. Hence, by representing the iterative breaking procedure with a binary tree, whose each node contains the random position of a break, they encompass classic stick-breaking random probability measures in the wider family of *tree stick-breaking* ones. In this framework, stick-breaking random probability measures correspond to *lopsided trees*, whose every left child is a leaf node. In contrast, the authors study the properties of tree stick-breaking random probability measures obtained with *balanced trees*, admitting leaf nodes just in the last level. Namely, these trees represent a stick-breaking procedure that entails splitting *both* the pieces obtained from each of the breaks in the previous iteration. Incorporating covariate dependence in the splitting random variables via a logit-normal scheme, they obtain dependent random probability measures which are employed as mixing measures in mixture models.

Specifically, given a probability measure G_0, the process $\{G_{\mathbf{x}} : \mathbf{x} \in \mathcal{X}\}$ of covariate-dependent tree stick-breaking random probability measures is defined as the common-atoms (MacEachern, 2000; Quintana et al., 2022) random probability measures $G_{\mathbf{x}}(\cdot) = \sum_{\boldsymbol{\varepsilon}\in B(\tau)} W_{\mathbf{x},\boldsymbol{\varepsilon}}\delta_{\theta_{\boldsymbol{\varepsilon}}}(\cdot)$ with $\theta_{\boldsymbol{\varepsilon}} \overset{iid}{\sim} G_0$, and where, for a binary tree τ, the set $B(\tau)$ contains the localization of the leaf nodes in terms of the (unique) splitting path that leads to them (*e.g.*, $\boldsymbol{\varepsilon} = (1, 1, 0)$ identifies the node obtained going right, right, left down the tree starting from the root). The random weights are given by $W_{\mathbf{x},\boldsymbol{\varepsilon}} = \prod_{\ell=1}^{m} V_{\mathbf{x},(\varepsilon_1,\ldots,\varepsilon_\ell)}^{1-\varepsilon_\ell}(1 - V_{\mathbf{x},(\varepsilon_1,\ldots,\varepsilon_\ell)})^{\varepsilon_\ell}$ where $\boldsymbol{\varepsilon} = (\varepsilon_1 \ldots, \varepsilon_m)$, $m \in \mathbb{N}$ is the number of levels of τ and $V_{\mathbf{x},\boldsymbol{\varepsilon}}$ are the splitting random variables, for any node location $\boldsymbol{\varepsilon}$, having generic marginal distribution $F_{\mathbf{x},\boldsymbol{\varepsilon}}$ and satisfying the tail-free conditions (Freedman, 1963; Fabius, 1964), conditionally on $\mathbf{x}$. Theorem 1 states that, for any measurable set A, $Var(G_{\mathbf{x}}(A)) = (G_0(A) - G_0^2(A))a_{\mathbf{x},\mathbf{x}}$ where $a_{\mathbf{x},\mathbf{x}} = \sum_{k=1}^{K} E(W_{\mathbf{x},k}^2)$, with K being the number of leaves in τ and using a slight abuse in the notation for the jumps $W_{\mathbf{x},k}$, which exploits the natural bijection between $B(\tau)$ and $\{1, \ldots, K\}$. In Theorem 2, the authors show that if τ is a balanced tree and $F_{\mathbf{x},\boldsymbol{\varepsilon}}$ does not dependent on the node $\boldsymbol{\varepsilon}$, then $a_{\mathbf{x},\mathbf{x}}$ goes to zero as the total number of leaves diverges, for every $\mathbf{x}$. In other terms,

$$G_{\mathbf{x}}(A) \to G_0(A)$$

almost surely as $K \to \infty$. This implies that the choice of the total number of leaves K has an impact on the variance of the random probability, thus affecting the confidence in the prior mean distribution G_0. We wonder if it could be interesting to preserve a non-vanishing variance, especially in contexts where the choice of m is dictated by other

*Luiss University, Italy, mcatalano@luiss.it
†University of Maryland, College Park, USA, fgaffi@umd.edu

specific needs. In our intuition, the vanishing variance effect is caused by the fact that, as m grows, one keeps splitting all the stick pieces without retaining any, so eventually there is no weight large enough to counterbalance the base measure component in the sequential sampling from the random probability measure. This could possibly be prevented by using distributions of $V_{\mathbf{x},\boldsymbol{\varepsilon}} = V_{\mathbf{x},(\varepsilon_1,\ldots,\varepsilon_\ell)}$ that, still not strictly depending on $\boldsymbol{\varepsilon}$, depend on the leaf level m, and shrinking their value as m increases. This will ensure that at least some weights are sufficiently large for any fixed m. For example, one could consider $V_{\mathbf{x},\boldsymbol{\varepsilon}} = \tilde{V}_{\mathbf{x}}/m$, where the distribution of $\tilde{V}_{\mathbf{x}}$ does not depend on $\boldsymbol{\varepsilon}$ or m. By denoting $c = \mathbb{E}[\tilde{V}_{\mathbf{x}}]$ and $d = \mathbb{E}[\tilde{V}_{\mathbf{x}}^2]$, and reasoning as in the proof of Theorem 2, for any $\boldsymbol{\varepsilon} \in B(\tau)$, $\mathbb{E}[W^2_{\mathbf{x},\varepsilon}]$ is equal to

$$\begin{aligned}\prod_{l=1}^{m} \mathbb{E}[V^2_{\mathbf{x},\varepsilon_1\cdots\varepsilon_{l-1}}]^{1-\varepsilon_l}\mathbb{E}[(1-V_{\mathbf{x},\varepsilon_1\cdots\varepsilon_{l-1}})^2]^{\varepsilon_l} &= \prod_{l=1}^{m}\left(\frac{\mathbb{E}[\tilde{V}_{\mathbf{x}}^2]}{m^2}\right)^{1-\varepsilon_l}\left(\frac{\mathbb{E}[(m-\tilde{V}_{\mathbf{x}})^2]}{m^2}\right)^{\varepsilon_l}\\ &= \frac{1}{m^{2m}} d^{m-\sum_{l=1}^{m}\varepsilon_l}(m^2+d-2mc)^{\sum_{l=1}^{m}\varepsilon_l}.\end{aligned}$$

Since for any $k \in \{0,1,\ldots,m\}$, the tree τ has $\binom{m}{k}$ leaves $(\varepsilon_1,\ldots,\varepsilon_m)$ such that $\sum_{l=1}^{m}\varepsilon_l = k$,

$$a_{\mathbf{x},\mathbf{x}} = \frac{1}{m^{2m}}\sum_{k=0}^{m}\binom{m}{k} d^{m-k}(m^2+d-2mc)^k = \left(\frac{d+m^2+d-2mc}{m^2}\right)^m.$$

To compute the limit as $m\to\infty$, we observe that $a_{\mathbf{x},\mathbf{x}} = (1-f(m))^m$, for $f(m) = (2cm-2d)/m^2$. Thus,

$$\lim_{m\to+\infty} a_{\mathbf{x},\mathbf{x}} = e^{-2c} = e^{-2\mathbb{E}[\tilde{V}_{\mathbf{x}}]}.$$

Recalling that in a balanced tree $m = \log_2(K)$, this determines a sequence of random probabilities $G_{\mathbf{x},K}$ that could have the benefit of ensuring a non-deterministic limit as K diverges, with a natural prior elicitation of $\mathbb{E}[\tilde{V}_{\mathbf{x}}]$ in terms of the confidence on G_0.

The authors nicely pinpoint two peculiarities of the classic stick-breaking random probability measures as directly linked to the lopsided structure of their corresponding tree, and show how such peculiarities disappear in the case of balanced tree stick-breaking random probability measures. If, on one hand, adopting the m-dependent splitting variable distributions suggested above would induce non-vanishing cross-covariance across covariates, as in the lopsided case, we wonder what would be the effect of this approach on posterior uncertainty quantification. Does the balanced structure still ensure an improved posterior inference as empirically showed by the authors?

In conclusion, we are deeply grateful to the authors for the work showcased in this paper, since it naturally spurs important discussions on which properties of dependent discrete random probability measure are desirable for improving inference, while paving the route for further investigations on the connections between such objects and random trees.

References

Fabius, J. (1964). "Asymptotic Behavior of Bayes' Estimates." *The Annals of Mathematical Statistics*, 35(2): 846–856. MR0162325. doi: https://doi.org/10.1214/aoms/1177703584. 1201

Freedman, D. A. (1963). "On the Asymptotic Behavior of Bayes' Estimates in the Discrete Case." *The Annals of Mathematical Statistics*, 34(4): 1386–1403. MR0158483. doi: https://doi.org/10.1214/aoms/1177703871. 1201

MacEachern, S. N. (2000). "Dependent Dirichlet processes." *Technical Report*, The Ohio State University. 1201

Quintana, F. A., Müller, P., Jara, A., and MacEachern, S. N. (2022). "The Dependent Dirichlet Process and Related Models." *Statistical Science*, 37(1). MR4371095. doi: https://doi.org/10.1214/20-sts819. 1201

Sethuraman, J. (1994). "A constructive definition of Dirichlet priors." *Statistica Sinica*, 4(2): 639–650. MR1309433. 1201