Journal of the Korean Statistical Society (2020) 49:389–394
https://doi.org/10.1007/s42952-019-00008-w

Online ISSN 2005-2863
Print ISSN 1226-3192

RESEARCH ARTICLE

A new proof of the stick-breaking representation of Dirichlet processes

Jaeyong Lee[1] · **Steven N. MacEachern**[2]

Received: 8 December 2018 / Accepted: 8 August 2019 / Published online: 1 January 2020

Abstract
The stick-breaking representation is one of the fundamental properties of the Dirichlet process. It represents the random probability measure as a discrete random sum whose weights and atoms are formed by independent and identically distributed sequences of beta variates and draws from the normalized base measure of the Dirichlet process parameter. It is used extensively in posterior simulation for statistical models with Dirichlet processes. The original proof of Sethuraman (Stat Sin 4:639–650, 1994) relies on an indirect distributional equation and does not encourage an intuitive understanding of the property. In this paper, we give a new proof of the stick-breaking representation of the Dirichlet process that provides an intuitive understanding of the theorem. The proof is based on the posterior distribution and self-similarity properties of the Dirichlet process.

Keywords Dirichlet process · Stick-breaking representation · Sethuraman's representation

1 Introduction

Sethuraman (1994) proved a fundamental property of the Dirichlet process (Ferguson 1973, DP)—namely that it can be constructed from two sequences of independent and identically distributed (i.i.d.) variates. The result is known as the stick-breaking representation (construction) theorem or Sethuraman's representation theorem of the Dirichlet process.

Let $\mathcal{X}$ be a complete and separable metric space with Borel σ-field $\mathcal{B}$ and

$$\alpha = M \cdot F_0 \tag{1}$$

✉ Jaeyong Lee
leejyc@gmail.com

1 Department of Statistics, Seoul National University, Seoul, South Korea

2 Department of Statistics, The Ohio State University, Columbus, USA

a finite nonnull measure on $\mathcal{X}$, where $M > 0$ and F_0 is a probability measure on $\mathcal{X}$. The distribution of a random probability measure F on $\mathcal{X}$ is $DP(\alpha)$, if for any measurable partition $B_1, B_2, \ldots, B_k$ of $\mathcal{X}$,

$$(F(B_1), F(B_2), \ldots, F(B_k)) \sim Dirichlet(\alpha(B_1), \alpha(B_2), \ldots, \alpha(B_k)).$$

The stick-breaking representation is given as follows.

Theorem 1.1 (Sethuraman's representation) *Suppose* $\theta_1, \theta_2, \ldots \overset{i.i.d.}{\sim} F_0$, $u_1, u_2, \ldots \overset{i.i.d.}{\sim} Beta(1, M)$, *and* θ_i*'s and* u_i*'s are independent of each other. Then,*

$$F = \sum_{j=1}^{\infty} \left(u_j \prod_{l<j} (1 - u_l) \right) \delta_{\theta_j} \sim DP(\alpha), \tag{2}$$

where δ_a *is a degenerate probability measure at* a.

The significance of the theorem is the following. First, the theorem relates the distribution of random probability to two sequences of i.i.d. random variables, beta random variables and those from the base measure of the Dirichlet process. The theorem's main purpose is to relate a complicated infinite dimensional object, the random probability measure, to more familiar objects, i.i.d. sequences of random variables. Second, the theorem is constructive, i.e., it shows how to construct the Dirichlet process from elementary random variables. Third, the theorem gives another proof that the DP is discrete with probability one. Fourth, the arbitrary nature of $\mathcal{X}$ expands the definition of the DP to allow a broad range of θ_i that need not lie in a finite dimensional Euclidean space.

This simple constructive representation becomes the later basis for posterior computation (Ishwaran and James 2001; Neal 2000; Blei and Jordan 2006) and for extension of the model to dependent random probability measures (MacEachern 2000; Rodriguez et al. 2008; Chung and Dunson 2011).

The original proof given in Sethuraman (1994) relies on an inductive property of the Dirichlet distribution. Although the statement of the theorem is simple, the proof does not allow intuitive understanding. Broderick et al. (2013) showed another route for the proof of the sticking-breaking representation, which is based on the Chinese restaurant process and recursive application of de Finetti's theorem to binary Pólya urn sequences. Recently, Miller (2018) gave another proof of the Chinese restaurant process from the stick-breaking representation.

In this paper, we give a new proof of the stick-breaking representation theorem which is more intuitive and easier to understand. The proof is based on two basic properties of the DP, the posterior distribution for i.i.d. samples and the self-similarity property.

The rest of the paper is organized as follows. In Sect. 2, we give a new proof for the stick-breaking representation theorem. The lemmas used in the proof are given in Sect. 3.

2 A new proof of the theorem

In this section, we give an another proof of Sethuraman's representation theorem that is more intuitive. In this proof, we use two properties of the Dirichlet process: the posterior distribution for i.i.d. samples and the self-similarity property. The posterior distribution of DP with an i.i.d. sample is given in Ferguson (1973, Theorem 1). If $F \sim DP(\alpha)$ and $X|F \sim F$, then the posterior distribution of F given $X = x$ is again DP and

$$F|X = x \sim DP(\alpha + \delta_x). \tag{3}$$

By the self-similarity property of the DP, the posterior of F given $X = x$ can be represented as

$$F \stackrel{d}{=} u\delta_x + (1-u)G, \tag{4}$$

where $u \sim Beta(1, M)$, $G \sim DP(\alpha)$, and u and G are independent. Ghosh and Ramamoorthi (2003, P. 93.) state (4) for the finite dimensional Dirichlet distribution, but the same proof goes through to prove (4) for the general Dirichlet process. We use (3) and (4) as basic building blocks of the proof.

Consider a sequence of random variables $\theta_1, \theta_2, \ldots$ whose distribution is described below. Let $F \sim DP(\alpha)$ where $\alpha = M \cdot F_0$, $M > 0$ and F_0 is a probability measure on $\mathcal{X}$. Given F, sample θ_1 from F. Given θ_1, by (4) the distribution of F can be represented as

$$F \stackrel{d}{=} u_1\delta_{\theta_1} + (1-u_1)G_1,$$

where $u_1 \sim Beta(1, M)$, $G_1 \sim DP(\alpha)$ and u_1 and G_1 are independent. Given θ_1 and F, sample θ_2 from G_1. By applying (4) again, given θ_1 and θ_2, $G_1 \stackrel{d}{=} u_2\delta_{\theta_2} + (1-u)G_2$, where $u_2 \sim Beta(1, M)$, $G_2 \sim DP(\alpha)$ and u_2 and G_2 are independent. Combining the results of F and G_1, we get that given θ_1 and θ_2,

$$\begin{aligned} F &\stackrel{d}{=} u_1\delta_{\theta_1} + (1-u_1)\big(u_2\delta_{\theta_2} + (1-u_2)G_2\big) \\ &\stackrel{d}{=} u_1\delta_{\theta_1} + u_2(1-u_1)\delta_{\theta_2} + (1-u_1)(1-u_2)G_2, \end{aligned}$$

where $u_1, u_2 \stackrel{i.i.d.}{\sim} Beta(1, M)$, $G_2 \sim DP(\alpha)$ and (u_1, u_2) is independent of G_2. Given θ_1, θ_2 and F, sample θ_3 from G_2, and so on. Repeating the same argument, we obtain that given $\theta_1, \theta_2, \ldots, \theta_k$,

$$F \stackrel{d}{=} u_1\delta_{\theta_1} + u_2(1-u_1)\delta_{\theta_2} + \cdots + \left(u_k \prod_{l<k}(1-u_l)\right)\delta_{\theta_k} + \left(\prod_{l\le k}(1-u_l)\right)G_k, \tag{5}$$

where $u_1, \ldots, u_k \stackrel{i.i.d.}{\sim} Beta(1, M)$, $G_k \sim DP(\alpha)$ and u_i's and G_k are independent.

To complete the proof, we use the following two lemmas whose proofs are given in the next section.

Lemma 2.1 *The sequence $\theta_1, \theta_2, \ldots$ is marginally an i.i.d. sequence from F_0.*

Lemma 2.2 *Given $\theta_1, \theta_2, \ldots$, the conditional distribution of F is given by*

$$F \overset{d}{=} \sum_{j=1}^{\infty} \left[u_j \prod_{l<j} (1-u_l) \right] \delta_{\theta_j}, \tag{6}$$

where $u_1, u_2, \ldots \overset{i.i.d.}{\sim} Beta(1, M)$.

Lemma 2.2 implies the θ_i's and $u_i's$ are independent of each other, and together with Lemma 2.1 it implies the conclusion of the theorem. □

Remark 2.3 In the proof, we use the form of the posterior with DP prior and the self-similarity property of DP, and thus implicitly assume the existence of DP. Sethuraman (1994) proved the stick-breaking representation without assuming the existence of DP and thus can be considered as an alternative proof of the existence theorem. Our proof, however,is not an alternative proof of the existence of DP.

Remark 2.4 Since the proofs of the existence of DP and the posterior of DP in Ferguson (1973) can be carried out under the assumption of complete and separable $\mathcal{X}$ and these facts are used in the proof, we assume $\mathcal{X}$ is a complete and separable metric space.

Remark 2.5 A referee pointed out that a similar proof can be done using the indirect distributional equation used in the proof of Sethuraman (1994),

$$F \overset{d}{=} u\delta_\theta + (1-u)G, \tag{7}$$

where $F, G \sim DP(\alpha)$, $u \sim Beta(1, M)$ and $\theta \sim F_0$. By repeated applications of equation (7), one gets

$$F \overset{d}{=} \sum_{i=1}^{k} \left(u_i \prod_{j<i} (1-u_j) \right) \delta_{\theta_i} + \prod_{j=1}^{k} (1-u_j) G.$$

Similarly to the proof of Lemma 2.2, one can show (2). This proof is the same as the original proof of Sethuraman (1994) except that (7) is used instead of the distributional relation of the finite dimensional Dirichlet distribution.

3 Proofs of lemmas

In this section, we give the proofs of the lemmas used in the previous section.

Proof of lemma 2.1 First, marginally $\theta_1 \sim F_0$. We prove the lemma by induction. Suppose $\theta_1, \ldots, \theta_k$, are i.i.d. from F_0. We will show that θ_{k+1} given $\theta_1, \ldots, \theta_k$ follows F_0. For a measurable $A \subset \mathcal{X}$,

$$\mathbb{P}(\theta_{k+1} \in A | \theta_1, \ldots, \theta_k) = \mathbb{E}\big[\mathbb{P}(\theta_{k+1} \in A | \theta_1, \ldots, \theta_k, F) | \theta_1, \ldots, \theta_k\big]$$

$$= \mathbb{E}\big[G_k(A)|\theta_1, \dots, \theta_k\big]$$
$$= F_0(A).$$

The equalities follow from the definition of θ_k and equation (5). This completes the proof. □

Proof of lemma 2.2 Let $u_1, u_2, \dots \overset{i.i.d.}{\sim} Beta(1, M)$, $\theta_1, \theta_2 \dots \overset{i.i.d.}{\sim} F_0$, $G_k \overset{i.i.d.}{\sim} DP(\alpha)$, and all these random elements are independent. Define

$$F_k^* = u_1\delta_{\theta_1} + u_2(1-u_1)\delta_{\theta_2} + \cdots + \left[u_k \prod_{l<k}(1-u_l)\right]\delta_{\theta_k} + \left[\prod_{l\le k}(1-u_l)\right]G_k.$$

Then, given $\theta_1, \dots, \theta_k$ F_k^* has the same distribution as F for all $k \geq 1$, by equation (5). Let $F^* = \sum_{j=1}^{\infty}\big[u_j \prod_{l<j}(1-u_l)\big]\delta_{\theta_j}$. We will show that the distribution of F given θ_j's is the same as that of F^* given θ_j's. For this, it suffices to show that for all disjoint measurable sets $B_1, B_2, \dots, B_m \subset \mathcal{X}$, the conditional distribution of $(F(B_1), \dots, F(B_m))$ given θ_j's is the same as that of $(F^*(B_1), \dots, F^*(B_m))$. Since the distributions of $(F(B_1), \dots, F(B_m))$ and $(F^*(B_1), \dots, F^*(B_m))$ are supported on a bounded set, equality of moments of all orders implies equality of the distributions. Thus, we need to show that for all $n_1, \dots, n_m \geq 0$,

$$\begin{aligned}&\mathbb{E}\big[F(B_1)^{n_1} \times \cdots \times F(B_m)^{n_m}\big|\theta_1, \theta_2, \dots\big]\\ &\quad= \mathbb{E}\big[F^*(B_1)^{n_1} \times \cdots \times F^*(B_m)^{n_m}\big|\theta_1, \theta_2, \dots\big],\ a.s.\end{aligned} \tag{8}$$

By the martingale convergence theorem, as $k \longrightarrow \infty$,

$$\begin{aligned}&\mathbb{E}\big[F(B_1)^{n_1} \times \cdots \times F(B_m)^{n_m}|\theta_1, \dots, \theta_k\big]\\ &\quad\longrightarrow \mathbb{E}\big[F(B_1)^{n_1} \times \cdots \times F(B_m)^{n_m}|\theta_1, \theta_2 \dots\big],\ \text{a.s.,}\end{aligned}$$

where the right hand side (RHS) is the same as the left hand side (LHS) of (8). On the other hand, since $F_k^*(B_j), F^*(B_j) \leq 1$,

$$\begin{aligned}&\big|\mathbb{E}\big[F_k^*(B_1)^{n_1} \times \cdots \times F_k^*(B_m)^{n_m}\big|\theta_1, \theta_2, \dots\big]\\ &\quad-\mathbb{E}\big[F^*(B_1)^{n_1} \times \cdots \times F^*(B_m)^{n_m}\big|\theta_1, \theta_2, \dots\big]\big|\\ &\le \mathbb{E}\Big[\,\big|F_k^*(B_1)^{n_1} \times \cdots \times F_k^*(B_m)^{n_m} - F^*(B_1)^{n_1} \times \cdots \times F^*(B_m)^{n_m}\big|\ |\theta_1, \theta_2, \dots\Big]\\ &\le \sum_{i=1}^{m} n_i \mathbb{E}\big[|F_k^*(B_i) - F^*(B_i)|\,\big|\theta_1, \theta_2, \dots\big].\end{aligned}$$

The last inequality holds because $|\prod_{i=1}^k a_i - \prod_{i=1}^k b_i| \leq \sum_{i=1}^k |a_i - b_i|$ if $0 \leq |a_i|, |b_i| \leq 1$, for $i = 1, 2, \dots, k$. For any measurable $B \subset \mathcal{X}$,

$$\mathbb{E}\left[|F_k^*(B) - F^*(B)|\,\big|\theta_1, \theta_2, \dots\right]$$

$$= \mathbb{E}\left(\left|\sum_{j=1}^{k}\left[u_j\prod_{l<j}(1-u_l)\right]\delta_{\theta_j}(B) + \left[\prod_{l\le k}(1-u_l)\right]G_k(B)\right.\right.$$
$$\left.\left.- \sum_{j=1}^{\infty}\left[u_j\prod_{l<j}(1-u_l)\right]\delta_{\theta_j}(B)\right| \Big|\theta_1,\theta_2,\ldots\right)$$
$$= \mathbb{E}\left(\left|\left[\prod_{l\le k}(1-u_l)\right]G_k(B) - \sum_{j=k+1}^{\infty}\left[u_j\prod_{l<j}(1-u_l)\right]\delta_{\theta_j}(B)\right| \Big|\theta_1,\theta_2,\ldots\right)$$
$$\le 2\mathbb{E}\left(\prod_{l\le k}(1-u_l)\right) = 2\left(\frac{M}{M+1}\right)^k \longrightarrow 0.$$

Thus, $\lim_{k\to\infty}\sum_{i=1}^{m} n_i\mathbb{E}\big[|F_k^*(B_i) - F^*(B_i)| \,\big|\theta_1,\theta_2,\ldots\big] = 0$ *a.s.* This in turn implies

$$\lim_{k\to\infty}\mathbb{E}\big[F_k^*(B_1)^{n_1}\times\cdots\times F_k^*(B_m)^{n_m}\big] = \mathbb{E}\big[F^*(B_1)^{n_1}\times\cdots\times F^*(B_m)^{n_m}\big]\ a.s.$$

Thus, we have shown (8). This completes the proof. □

Acknowledgements Jaeyong Lee was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korea government(MSIT) (no. 2018R1A2A3074973), and MacEachern was supported in part by the National Science Foundation of the United States.

References

Blei, D. M., & Jordan, M. I. (2006). Variational inference for Dirichlet process mixtures. *Bayesian Analysis*, *1*(1), 121–143.
Broderick, T., Jordan, M. I., & Pitman, J. (2013). Cluster and feature modeling from combinatorial stochastic processes. *Statistical Science*, *28*(3), 289–312.
Chung, Y., & Dunson, D. B. (2011). The local Dirichlet process. *Annals of the Institute of Statistical Mathematics*, *63*(1), 59–80.
Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, *1*(2), 209–230.
Ghosh, J., & Ramamoorthi, R. (2003). *Bayesian nonparametrics*. New York: Springer.
Ishwaran, H., & James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, *96*(453), 161–173.
MacEachern, S. N. (2000). Dependent Dirichlet processes. Unpublished manuscript, Department of Statistics, The Ohio State University, pp 1–40.
Miller, J. W. (2018). An elementary derivation of the Chinese restaurant process from Sethuraman's stick-breaking process. arXiv preprint arXiv:1801.00513.
Neal, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, *9*(2), 249–265.
Rodriguez, A., Dunson, D. B., & Gelfand, A. E. (2008). The nested Dirichlet process. *Journal of the American Statistical Association*, *103*(483), 1131–1154.
Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, *4*, 639–650.