An elementary derivation of the Chinese restaurant process from Sethuraman's stick-breaking process

Jeffrey W. Miller*
Harvard University, Department of Biostatistics

January 4, 2018

Abstract

The Chinese restaurant process and the stick-breaking process are the two most commonly used representations of the Dirichlet process. However, the usual proof of the connection between them is indirect, relying on abstract properties of the Dirichlet process that are difficult for nonexperts to verify. This short note provides a direct proof that the stick-breaking process gives rise to the Chinese restaurant process, without using any measure theory.

1 Introduction

Sethuraman (1994) showed that the Dirichlet process has the following stick-breaking representation: if $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots \overset{\text{iid}}{\sim} \text{Beta}(1, \alpha)$, $\boldsymbol{\pi}_k = \boldsymbol{v}_k \prod_{i=1}^{k-1}(1 - \boldsymbol{v}_i)$ for $k = 1, 2, \ldots$, and $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots \overset{\text{iid}}{\sim} H$, then the random discrete measure

$$\boldsymbol{P} = \sum_{k=1}^{\infty} \boldsymbol{\pi}_k \delta_{\boldsymbol{\theta}_k} \tag{1}$$

is distributed according to the Dirichlet process $\text{DP}(\alpha, H)$ with concentration parameter α and base distribution H. This representation has been instrumental in the development of many nonparametric models (MacEachern, 1999, 2000; Hjort, 2000; Ishwaran and Zarepour, 2000; Ishwaran and James, 2001; Griffin and Steel, 2006; Dunson and Park, 2008; Chung and Dunson, 2009; Rodriguez and Dunson, 2011; Broderick et al., 2012), has facilitated the understanding of these models (Favaro et al., 2012; Teh et al., 2007; Thibaux and Jordan, 2007; Paisley et al., 2010), and is central to various inference algorithms (Ishwaran and James, 2001; Blei and Jordan, 2006; Papaspiliopoulos and Roberts, 2008; Walker et al., 2007; Kalli et al., 2011).

*The author gratefully acknowledges support from the National Science Foundation (NSF) grant DMS-1045153 and the National Institutes of Health (NIH) grant 5R01ES017436.

It is well-known that, as shown by Antoniak (1974), the Dirichlet process induces a distribution on partitions as follows: if $\boldsymbol{P} \sim \mathrm{DP}(\alpha, H)$ where H is nonatomic (i.e., $H(\{\theta\}) = 0$ for any θ), $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n | \boldsymbol{P} \overset{\text{iid}}{\sim} \boldsymbol{P}$, and $\boldsymbol{C}$ is the partition of $\{1, \ldots, n\}$ induced by $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, then

$$\mathbb{P}(\boldsymbol{C} = C) = \frac{\alpha^{|C|}\Gamma(\alpha)}{\Gamma(\alpha + n)} \prod_{c \in C} \Gamma(|c|). \tag{2}$$

The sequential sampling process corresponding to this partition distribution is known as the Chinese restaurant process, or Blackwell–MacQueen urn process.

The following key fact is a direct consequence of these two results (Sethuraman's and Antoniak's): if $\boldsymbol{\pi} = (\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots)$ is defined as above, $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n | \boldsymbol{\pi} \overset{\text{iid}}{\sim} \boldsymbol{\pi}$, and $\boldsymbol{C}$ is the partition induced by $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n$, then the distribution of $\boldsymbol{C}$ is given by Equation 2. This can be seen by noting that when H is nonatomic, the distribution of $\boldsymbol{C}$ is the same as when it is induced by $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n | \boldsymbol{P}$.

While this key fact follows directly from the results of Sethuraman and Antoniak, the proofs of their results are rather abstract and are not easy to verify, especially for those without expertise in measure theory. The purpose of this note is to provide a proof of this connection between the Chinese restaurant process and the stick-breaking process using only elementary, non-measure-theoretic arguments. Our proof is completely self-contained and does not rely on any properties of the Dirichlet process or other theoretical results.

In previous work, Broderick et al. (2013) used De Finetti's theorem to provide an elegant inductive derivation of the stick-breaking process from the Chinese restaurant process. Also, Paisley (2010) showed by elementary calculations that if the base distribution H is a discrete distribution on $\{1, \ldots, K\}$, and $\boldsymbol{P}$ is defined by the stick-breaking process as in Equation 1, then $(\boldsymbol{P}(1), \ldots, \boldsymbol{P}(K)) \sim \text{Dirichlet}(\alpha H(1), \ldots, \alpha H(K))$; thus, despite the similar sounding title of the article by Paisley (2010), the result shown there is altogether different from what we show here.

2 Main result

We use $[n]$ to denote the set $\{1, \ldots, n\}$, and $\mathbb{N}$ to denote $\{1, 2, 3, \ldots\}$. As is standard, we represent a partition of $[n]$ as a set $C = \{c_1, \ldots, c_t\}$ of nonempty disjoint sets $c_1, \ldots, c_t$ such that $\bigcup_{i=1}^t c_i = [n]$. Thus, $t = |C|$ is the number of parts in the partition, and $|c|$ is the number of elements in a given part $c \in C$. We say that C is the partition of $[n]$ induced by $z_1, \ldots, z_n$ if it has the property that for any $i, j \in [n]$, i and j belong to the same part $c \in C$ if and only if $z_i = z_j$. We use bold font to denote random variables.

Theorem 2.1. *Suppose*

$$\begin{aligned}
&\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots \overset{\text{iid}}{\sim} \text{Beta}(1, \alpha) \\
&\boldsymbol{\pi}_k = \boldsymbol{v}_k \prod_{i=1}^{k-1} (1 - \boldsymbol{v}_i) \ \textit{for } k = 1, 2, \ldots, \\
&\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n | \boldsymbol{\pi} = \pi \overset{\text{iid}}{\sim} \pi, \ \textit{that is, } \mathbb{P}(\boldsymbol{z}_i = k \mid \pi) = \pi_k,
\end{aligned}$$

and $\boldsymbol{C}$ is the partition of $[n]$ induced by $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n$. Then

$$\mathbb{P}(\boldsymbol{C} = C) = \frac{\alpha^{|C|}\Gamma(\alpha)}{\Gamma(n+\alpha)} \prod_{c \in C} \Gamma(|c|).$$

Our proof of the theorem relies on the following lemmas. Let us abbreviate $z = (z_1, \ldots, z_n)$. Given $z \in \mathbb{N}^n$, let C_z denote the partition $[n]$ induced by z.

Lemma 2.2. *For any $z \in \mathbb{N}^n$,*

$$\mathbb{P}(\boldsymbol{z} = z) = \frac{\Gamma(\alpha)}{\Gamma(n+\alpha)} \Big(\prod_{c \in C_z} \Gamma(|c|+1) \Big) \Big(\prod_{k=1}^{m} \frac{\alpha}{g_k + \alpha} \Big)$$

where $m = \max\{z_1, \ldots, z_n\}$ and $g_k = \#\{i : z_i \geq k\}$.

The proofs of the lemmas will be given in Section 3. We use $\mathbb{1}(\cdot)$ to denote the indicator function, that is, $\mathbb{1}(E) = 1$ if E is true, and $\mathbb{1}(E) = 0$ otherwise.

Lemma 2.3. *For any partition C of $[n]$,*

$$\frac{\alpha^{|C|}}{\prod_{c \in C} |c|} = \sum_{z \in \mathbb{N}^n} \mathbb{1}(C_z = C) \prod_{k=1}^{m(z)} \frac{\alpha}{g_k(z) + \alpha}$$

where $m(z) = \max\{z_1, \ldots, z_n\}$ and $g_k(z) = \#\{i : z_i \geq k\}$.

Proof of Theorem 2.1.

$$\begin{aligned}
\mathbb{P}(\boldsymbol{C} = C) &= \sum_{z \in \mathbb{N}^n} \mathbb{P}(\boldsymbol{C} = C \mid \boldsymbol{z} = z) \mathbb{P}(\boldsymbol{z} = z) \\
&\overset{\text{(a)}}{=} \sum_{z \in \mathbb{N}^n} \mathbb{1}(C_z = C) \frac{\Gamma(\alpha)}{\Gamma(n+\alpha)} \Big(\prod_{c \in C_z} \Gamma(|c|+1) \Big) \Big(\prod_{k=1}^{m(z)} \frac{\alpha}{g_k(z) + \alpha} \Big) \\
&= \frac{\Gamma(\alpha)}{\Gamma(n+\alpha)} \Big(\prod_{c \in C} \Gamma(|c|+1) \Big) \sum_{z \in \mathbb{N}^n} \mathbb{1}(C_z = C) \Big(\prod_{k=1}^{m(z)} \frac{\alpha}{g_k(z) + \alpha} \Big) \\
&\overset{\text{(b)}}{=} \frac{\Gamma(\alpha)}{\Gamma(n+\alpha)} \Big(\prod_{c \in C} \Gamma(|c|+1) \Big) \frac{\alpha^{|C|}}{\prod_{c \in C} |c|} \\
&\overset{\text{(c)}}{=} \frac{\Gamma(\alpha)}{\Gamma(n+\alpha)} \Big(\prod_{c \in C} \Gamma(|c|) \Big) \alpha^{|C|}
\end{aligned}$$

where (a) is by Lemma 2.2, (b) is by Lemma 2.3, and (c) is since $\Gamma(|c|+1) = |c|\Gamma(|c|)$. $\square$

3 Proofs of lemmas

Proof of Lemma 2.2. Letting $e_k = \#\{i : z_i = k\}$, we have

$$\mathbb{P}(\boldsymbol{z} = z \mid \pi_1, \ldots, \pi_m) = \prod_{i=1}^{n} \pi_{z_i} = \prod_{k=1}^{m} \pi_k^{e_k}$$

and thus

$$\mathbb{P}(\boldsymbol{z} = z \mid v_1, \ldots, v_m) = \prod_{k=1}^{m} \Big(v_k \textstyle\prod_{i=1}^{k-1}(1 - v_i)\Big)^{e_k} = \displaystyle\prod_{k=1}^{m} v_k^{e_k}(1 - v_k)^{f_k}$$

where $f_k = \#\{i : z_i > k\}$. Therefore,

$$\begin{aligned}
\mathbb{P}(\boldsymbol{z} = z) &= \int \mathbb{P}(\boldsymbol{z} = z \mid v_1, \ldots, v_m) p(v_1, \ldots, v_m) dv_1 \cdots dv_m \\
&= \int \Big(\prod_{k=1}^{m} v_k^{e_k}(1 - v_k)^{f_k}\Big) p(v_1) \cdots p(v_m) dv_1 \cdots dv_m \\
&= \prod_{k=1}^{m} \int v_k^{e_k}(1 - v_k)^{f_k} p(v_k) dv_k \\
&\overset{\text{(a)}}{=} \prod_{k=1}^{m} \alpha B(e_k + 1,\ f_k + \alpha) \\
&= \prod_{k=1}^{m} \frac{\alpha \Gamma(e_k + 1)\Gamma(f_k + \alpha)}{\Gamma(e_k + f_k + \alpha + 1)} \\
&\overset{\text{(b)}}{=} \prod_{k=1}^{m} \frac{\alpha \Gamma(e_k + 1)\Gamma(g_{k+1} + \alpha)}{\Gamma(g_k + \alpha + 1)} \\
&\overset{\text{(c)}}{=} \Big(\prod_{k=1}^{m} \Gamma(e_k + 1)\Big)\Big(\prod_{k=1}^{m} \frac{\alpha}{g_k + \alpha}\Big)\Big(\prod_{k=1}^{m} \frac{\Gamma(g_{k+1} + \alpha)}{\Gamma(g_k + \alpha)}\Big) \\
&= \Big(\prod_{c \in C_z} \Gamma(|c| + 1)\Big)\Big(\prod_{k=1}^{m} \frac{\alpha}{g_k + \alpha}\Big) \frac{\Gamma(\alpha)}{\Gamma(n + \alpha)}
\end{aligned}$$

where step (a) holds since

$$\int x^r (1 - x)^s \operatorname{Beta}(x|1, \alpha) dx = \frac{B(r + 1,\ s + \alpha)}{B(1, \alpha)} = \alpha B(r + 1,\ s + \alpha),$$

step (b) since $f_k = g_{k+1}$ and $g_k = e_k + f_k$, and step (c) since $\Gamma(x + 1) = x\Gamma(x)$. □

Let S_t denote the set of $t!$ permutations of $[t]$.

Lemma 3.1. *For any $n_1, \ldots, n_t \in \mathbb{N}$,*

$$\sum_{\sigma \in S_t} \frac{1}{a_1(\sigma) \cdots a_t(\sigma)} = \frac{1}{n_1 \cdots n_t}$$

where $a_i(\sigma) = n_{\sigma_i} + n_{\sigma_{i+1}} + \cdots + n_{\sigma_t}$.

Proof. Consider an urn containing t balls of various sizes—specifically, suppose the balls are labeled $1, \ldots, t$ and have sizes $n_1, \ldots, n_t$. Consider the process of sampling without replacement t times from the urn, supposing that the probability of drawing any given ball is proportional to its size. This defines a distribution on permutations $\sigma \in S_t$ such that, letting $n = \sum_{i=1}^{t} n_i$,

$$\begin{aligned}
p(\sigma_1) &= \frac{n_{\sigma_1}}{n} = \frac{n_{\sigma_1}}{a_1(\sigma)}, \\
p(\sigma_2|\sigma_1) &= \frac{n_{\sigma_2}}{n - n_{\sigma_1}} = \frac{n_{\sigma_2}}{a_2(\sigma)}, \\
p(\sigma_3|\sigma_1, \sigma_2) &= \frac{n_{\sigma_3}}{n - n_{\sigma_1} - n_{\sigma_2}} = \frac{n_{\sigma_3}}{a_3(\sigma)},
\end{aligned}$$

and so on. Therefore, since $n_{\sigma_1} \cdots n_{\sigma_t} = n_1 \cdots n_t$,

$$p(\sigma) = p(\sigma_1)p(\sigma_2|\sigma_1) \cdots p(\sigma_t|\sigma_1, \ldots, \sigma_{t-1}) = \frac{n_1 \cdots n_t}{a_1(\sigma) \cdots a_t(\sigma)}. \tag{3}$$

Since $p(\sigma)$ is a distribution on S_t by construction, we have $\sum_{\sigma \in S_t} p(\sigma) = 1$; applying this to Equation 3 and dividing both sides by $n_1 \cdots n_t$ gives the result. $\square$

Proof of Lemma 2.3. Let $t = |C|$, and suppose $c_1, \ldots, c_t$ are the parts of C. For $\sigma \in S_t$, define $a_i(\sigma) = |c_{\sigma_i}| + \cdots + |c_{\sigma_t}|$. For any $z \in \mathbb{N}^n$ such that $C_z = C$, if $k_1 < \cdots < k_t$ are the distinct values taken on by $z_1, \ldots, z_n$, then

$$\begin{aligned}
\prod_{k=1}^{m(z)} \frac{\alpha}{g_k(z) + \alpha} &= \Big(\frac{\alpha}{g_{k_1}(z) + \alpha}\Big)^{k_1} \Big(\frac{\alpha}{g_{k_2}(z) + \alpha}\Big)^{k_2 - k_1} \cdots \Big(\frac{\alpha}{g_{k_t}(z) + \alpha}\Big)^{k_t - k_{t-1}} \\
&= \Big(\frac{\alpha}{a_1(\sigma) + \alpha}\Big)^{d_1} \Big(\frac{\alpha}{a_2(\sigma) + \alpha}\Big)^{d_2} \cdots \Big(\frac{\alpha}{a_t(\sigma) + \alpha}\Big)^{d_t}
\end{aligned}$$

where $d_i = k_i - k_{i-1}$, with $k_0 = 0$, and σ is the permutation of $[t]$ such that $c_{\sigma_i} = \{j : z_j = k_i\}$. Note that the definition of $d = (d_1, \ldots, d_t)$ and σ sets up a one-to-one correspondence (that is, a bijection) between $\{z \in \mathbb{N}^n : C_z = C\}$ and $\{(\sigma, d) : \sigma \in S_t,\ d \in \mathbb{N}^t\}$. Therefore,

$$\begin{aligned}
\sum_{z \in \mathbb{N}^n} \mathbb{1}(C_z = C) \prod_{k=1}^{m(z)} \frac{\alpha}{g_k(z) + \alpha} &= \sum_{\sigma \in S_t} \sum_{d \in \mathbb{N}^t} \prod_{i=1}^{t} \Big(\frac{\alpha}{a_i(\sigma) + \alpha}\Big)^{d_i} \\
&= \sum_{\sigma \in S_t} \prod_{i=1}^{t} \sum_{d_i \in \mathbb{N}} \Big(\frac{\alpha}{a_i(\sigma) + \alpha}\Big)^{d_i} \\
&\overset{\text{(a)}}{=} \sum_{\sigma \in S_t} \prod_{i=1}^{t} \frac{\alpha}{a_i(\sigma)} \\
&\overset{\text{(b)}}{=} \frac{\alpha^t}{\prod_{i=1}^{t} |c_i|} = \frac{\alpha^t}{\prod_{c \in C} |c|}
\end{aligned}$$

where step (a) follows from the geometric series, $\sum_{k=1}^{\infty} x^k = 1/(1-x) - 1$ for $x \in [0, 1)$, and step (b) is by Lemma 3.1. $\square$

References

C. E. Antoniak. Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *The Annals of Statistics*, 2(6):1152–1174, 1974.

D. M. Blei and M. I. Jordan. Variational inference for Dirichlet process mixtures. *Bayesian Analysis*, 1(1):121–143, 2006.

T. Broderick, M. I. Jordan, and J. Pitman. Beta processes, stick-breaking and power laws. *Bayesian Analysis*, 7(2):439–476, 2012.

T. Broderick, M. I. Jordan, and J. Pitman. Cluster and feature modeling from combinatorial stochastic processes. *Statistical Science*, 28(3):289–312, 2013.

Y. Chung and D. B. Dunson. Nonparametric Bayes conditional distribution modeling with variable selection. *Journal of the American Statistical Association*, 104(488), 2009.

D. B. Dunson and J.-H. Park. Kernel stick-breaking processes. *Biometrika*, 95(2):307–323, 2008.

S. Favaro, A. Lijoi, and I. Pruenster. On the stick-breaking representation of normalized inverse Gaussian priors. *Biometrika*, 99(3):663–674, 2012.

J. E. Griffin and M. J. Steel. Order-based dependent Dirichlet processes. *Journal of the American Statistical Association*, 101(473):179–194, 2006.

N. L. Hjort. Bayesian analysis for a generalised Dirichlet process prior. *Technical Report, University of Oslo*, 2000.

H. Ishwaran and L. F. James. Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, 96(453), 2001.

H. Ishwaran and M. Zarepour. Markov chain Monte Carlo in approximate Dirichlet and beta two-parameter process hierarchical models. *Biometrika*, 87(2):371–390, 2000.

M. Kalli, J. E. Griffin, and S. G. Walker. Slice sampling mixture models. *Statistics and Computing*, 21(1):93–105, 2011.

S. N. MacEachern. Dependent nonparametric processes. In *ASA Proceedings of the Section on Bayesian Statistical Science*, pages 50–55, 1999.

S. N. MacEachern. Dependent Dirichlet processes. *Unpublished manuscript, Department of Statistics, The Ohio State University*, 2000.

J. Paisley. A simple proof of the stick-breaking construction of the Dirichlet Process. Technical report, Princeton University, Department of Computer Science, August 2010.

J. W. Paisley, A. K. Zaas, C. W. Woods, G. S. Ginsburg, and L. Carin. A stick-breaking construction of the beta process. In *Proceedings of the 27th International Conference on Machine Learning*, pages 847–854, 2010.

O. Papaspiliopoulos and G. O. Roberts. Retrospective Markov chain Monte Carlo methods for Dirichlet process hierarchical models. *Biometrika*, 95(1):169–186, 2008.

A. Rodriguez and D. B. Dunson. Nonparametric Bayesian models through probit stick-breaking processes. *Bayesian Analysis*, 6(1), 2011.

J. Sethuraman. A constructive definition of Dirichlet priors. *Statistica Sinica*, 4:639–650, 1994.

Y. W. Teh, D. Görür, and Z. Ghahramani. Stick-breaking construction for the Indian buffet process. In *International Conference on Artificial Intelligence and Statistics*, pages 556–563, 2007.

R. Thibaux and M. I. Jordan. Hierarchical beta processes and the Indian buffet process. In *International Conference on Artificial Intelligence and Statistics*, pages 564–571, 2007.

S. G. Walker, A. Lijoi, and I. Prünster. On rates of convergence for posterior distributions in infinite-dimensional models. *The Annals of Statistics*, 35(2):738–746, 2007.