A simple proof of Pitman-Yor's Chinese restaurant process from its stick-breaking representation

Caroline Lawless, Julyan Arbel

A simple proof of Pitman–Yor's Chinese restaurant process from its stick-breaking representation

Caroline Lawless and Julyan Arbel

Univ. Grenoble Alpes, Inria, CNRS, LJK, 38000 Grenoble, France

Abstract

For a long time, the Dirichlet process has been the gold standard discrete random measure in Bayesian nonparametrics. The Pitman–Yor process provides a simple and mathematically tractable generalization, allowing for a very flexible control of the clustering behaviour. Two commonly used representations of the Pitman–Yor process are the stick-breaking process and the Chinese restaurant process. The former is a constructive representation of the process which turns out very handy for practical implementation, while the latter describes the partition distribution induced. Obtaining one from the other is usually done indirectly with use of measure theory. In contrast, we provide here an elementary proof of Pitman–Yor's Chinese Restaurant process from its stick-breaking representation.

1 Introduction

The Pitman–Yor process defines a rich and flexible class of random probability measures which was developed by Perman et al. (1992) and further investigated by Pitman (1995), Pitman and Yor (1997). It is a simple generalization of the Dirichlet process (Ferguson, 1973), whose mathematical tractability contributed to its popularity in machine learning theory (Caron et al., 2017), probabilistic models for linguistic applications (Teh, 2006, Wood et al., 2011), excursion theory (Perman et al., 1992, Pitman and Yor, 1997), measure-valued diffusions in population genetics (Petrov, 2009, Feng and Sun, 2010), combinatorics (Vershik et al., 2004, Kerov, 2006) and statistical physics (Derrida, 1981).

Its most prominent role is perhaps in Bayesian nonparametric statistics where it is used as a prior distribution, following the work of Ishwaran and James (2001). Applications in this setting embrace a variety of inferential problems, including species sampling (Favaro et al., 2009, Navarrete et al., 2008, Arbel et al., 2017), survival analysis and graphical models in genetics (Jara et al., 2010, Ni et al., 2018), image segmentation (Sudderth and Jordan, 2009), curve estimation (Canale et al., 2017), exchangeable feature allocations (Battiston et al., 2018) and time-series and econometrics (Caron et al., 2017, Bassetti et al., 2014).

Last but not least, the Pitman–Yor process is also employed in the context of nonparametric mixture modeling, thus generalizing the celebrated Dirichlet process mixture model of Lo (1984). Nonparametric mixture models based on the Pitman–Yor process are characterized by a more flexible parameterization than the Dirichlet process mixture model, thus

allowing for a better control of the clustering behaviour (De Blasi et al., 2015). In addition, see Ishwaran and James (2001), Favaro and Walker (2013), Arbel et al. (2018) for posterior sampling algorithms, Scricciolo et al. (2014), Miller and Harrison (2014) for asymptotic properties, and Scarpa and Dunson (2009), Canale et al. (2017) for spike-and-slab extensions.

The Pitman–Yor process has the following stick-breaking representation: if $\boldsymbol{v}_i \overset{\text{ind}}{\sim} \text{Beta}(1-d, \alpha+id)$ for $i = 1, 2, \ldots$ with $d \in (0, 1)$ and $\alpha > -d$, if $\boldsymbol{\pi}_j = \boldsymbol{v}_j \prod_{i=1}^{j-1}(1 - \boldsymbol{v}_i)$ for $j = 1, 2, \ldots$, and if $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots \overset{\text{iid}}{\sim} H$, then the discrete random probability measure

$$\boldsymbol{P} = \sum_{j=1}^{\infty} \boldsymbol{\pi}_j \delta_{\boldsymbol{\theta}_j} \tag{1}$$

is distributed according to the Pitman–Yor process, $\text{PY}(\alpha, d, H)$, with concentration parameter α, discount parameter d, and base distribution H.

The Pitman–Yor process induces the following partition distribution: if $\boldsymbol{P} \sim \text{PY}(\alpha, d, H)$, for some nonatomic probability distribution H, we observe data $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n | \boldsymbol{P} \overset{\text{iid}}{\sim} \boldsymbol{P}$, and $\boldsymbol{C}$ is the partition of the first n integers $\{1, \ldots, n\}$ induced by data, then

$$\mathbb{P}(\boldsymbol{C} = C) = \frac{d^{|C|}}{(\alpha)_{(n)}} \left(\frac{\alpha}{d}\right)_{(|C|)} \prod_{c \in C}(1 - d)_{(|c|-1)}, \tag{2}$$

where the multiplicative factor before the product in (2) is also commonly (and equivalently) written as $(\prod_{i=1}^{|C|-1} \alpha + id)/(\alpha)_{(n-1)}$ in the literature. When the discount parameter d is set to zero, the Pitman–Yor process reduces to the Dirichlet process and the partition distribution (2) boils down to the celebrated Chinese Restaurant process (CRP, see Antoniak, 1974). By abuse of language, we call the partition distribution (2) the Pitman–Yor's CRP. Under the latter partition distribution, the number of parts in a partition C of n elements, $k_n = |C|$, grows to infinity as a power-law of the sample size, n^d (see Pitman, 2003, for details). This Pitman–Yor power-law growth is more in tune with most of empirical data (Clauset et al., 2009) than the logarithmic growth induced by the Dirichlet process CRP, $\alpha \log n$.

The purpose of this note is to provide a simple proof of Pitman–Yor's CRP (2) from its stick-breaking representation (1) (Theorem 2.1). This generalizes the derivation by Miller (2018) who obtained the Dirichlet process CRP (Antoniak, 1974) from Sethuraman's stick-breaking representation (Sethuraman, 1994). In doing so, we also provide the marginal distribution of the allocation variables vector (3) in Proposition 2.2.

2 Partition distribution from stick-breaking

Suppose we make n observations, $z_1, \ldots, z_n$. We denote the set $\{1, \ldots, n\}$ by $[n]$. Our observations induce a partition of $[n]$, denoted $C = \{c_1, \ldots, c_{k_n}\}$ where $c_1, \ldots, c_{k_n}$ are disjoint sets and $\bigcup_{i=1}^{k_n} c_i = [n]$, in such a way that z_i and z_j belong to the same partition if and only if $z_i = z_j$. We denote the number of parts in the partition C by $k_n = |C|$ and we denote the number of elements in partition j by $|c_j|$. We use bold font to represent random variables.

We write $(x)_{(n)} = \prod_{j=0}^{n-1}(x+j)$ to denote the rising factorial.

Theorem 2.1. *Suppose*

$$\boldsymbol{v}_i \overset{\text{ind}}{\sim} \text{Beta}(1-d, \alpha + id) \text{ for } i = 1, 2, \dots,$$

$$\boldsymbol{\pi}_j = \boldsymbol{v}_j \prod_{i=1}^{j-1}(1-\boldsymbol{v}_i) \text{ for } j = 1, 2, \dots$$

Let allocation variables be defined by

$$\boldsymbol{z}_1, \dots, \boldsymbol{z}_n | \boldsymbol{\pi} = \pi \overset{\text{iid}}{\sim} \pi, \ \text{meaning}, \ \mathbb{P}(\boldsymbol{z}_i = j \mid \pi) = \pi_j, \tag{3}$$

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

$$\mathbb{P}(\boldsymbol{C} = C) = \frac{d^{|C|}}{(\alpha)_{(n)}} \left(\frac{\alpha}{d}\right)_{(|C|)} \prod_{c \in C}(1-d)_{(|c|-1)}.$$

The proof of Theorem 2.1 follows the lines of Miller (2018)'s derivation. We need the next two technical results, which we will prove in Section 3. Let C_z denote the partition $[n]$ induced by z for any $z \in \mathbb{N}^n$. Let k_n be the number of parts in the partition. We define $m(z) = \max\{z_1, \dots, z_n\}$, and $g_j(z) = \#\{i : z_i \geq j\}$.

Proposition 2.2. *For any $z \in \mathbb{N}^n$, the marginal distribution of the allocation variables vector $\boldsymbol{z} = (\boldsymbol{z}_1, \dots, \boldsymbol{z}_n)$ is given by*

$$\mathbb{P}(\boldsymbol{z} = z) = \frac{1}{(\alpha)_{(n)}} \prod_{c \in C_z} \frac{\Gamma(|c|+1-d)}{\Gamma(1-d)} \prod_{j=1}^{m(z)} \frac{\alpha + (j-1)d}{g_j(z) + \alpha + (j-1)d}.$$

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

$$\sum_{z \in \mathbb{N}^n} \mathbb{1}(C_z = C) \prod_{j=1}^{m(z)} \frac{\alpha + (j-1)d}{g_j(z) + \alpha + (j-1)d} = \frac{d^{|C|}}{\prod_{c \in C}(|c|-d)} \left(\frac{\alpha}{d}\right)_{(|C|)}.$$

Proof of Theorem 2.1.

$$\mathbb{P}(\boldsymbol{C} = C) = \sum_{z \in \mathbb{N}^n} \mathbb{P}(\boldsymbol{C} = C | \boldsymbol{z} = z)\mathbb{P}(\boldsymbol{z} = z)$$

$$\stackrel{(a)}{=} \sum_{z \in \mathbb{N}^n} \mathbb{1}(C_z = C)\frac{1}{(\alpha)_{(n)}}\prod_{c \in C_z}\frac{\Gamma(|c| + 1 - d)}{\Gamma(1 - d)}\prod_{j=1}^{m(z)}\frac{\alpha + (j - 1)d}{g_j(z) + \alpha + (j - 1)d}$$

$$= \frac{1}{(\alpha)_{(n)}}\prod_{c \in C}\frac{\Gamma(|c| + 1 - d)}{\Gamma(1 - d)}\sum_{z \in \mathbb{N}^n}\mathbb{1}(C_z = C)\prod_{j=1}^{m(z)}\frac{\alpha + (j - 1)d}{g_j(z) + \alpha + (j - 1)d}$$

$$\stackrel{(b)}{=} \frac{1}{(\alpha)_{(n)}}\prod_{c \in C}\frac{\Gamma(|c| + 1 - d)}{\Gamma(1 - d)}\frac{d^{|C|}}{\prod_{c \in C}(|c| - d)}\left(\frac{\alpha}{d}\right)_{(|C|)}$$

$$\stackrel{(c)}{=} \frac{1}{(\alpha)_{(n)}}\prod_{c \in C}(1 - d)_{(|c|-1)}\prod_{c \in C}(|c| - d)\frac{d^{|C|}}{\prod_{c \in C}(|c| - d)}\left(\frac{\alpha}{d}\right)_{(|C|)}$$

$$= \frac{d^{|C|}}{(\alpha)_{(n)}}\left(\frac{\alpha}{d}\right)_{(|C|)}\prod_{c \in C}(1 - d)_{(|c|-1)},$$

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

3 Proofs of the technical results

3.1 Additional lemmas

We require the following additional lemmas.

Lemma 3.1. *For $a + c > 0$, and $b + d > 0$, if $\boldsymbol{y} \sim \text{Beta}(a, b)$, then $\mathbb{E}[\boldsymbol{y}^c(1 - \boldsymbol{y})^d] = \frac{B(a+c,b+d)}{B(a,b)}$ where B denotes the beta function.*

Proof.

$$\mathbb{E}[\boldsymbol{y}^c(1 - \boldsymbol{y})^d] = \int_0^1 y^c(1 - y)^d\frac{1}{B(a, b)}y^{a-1}(1 - y)^{b-1}\mathrm{d}y$$

$$= \frac{1}{B(a, b)}\int_0^1 y^{a+c-1}(1 - y)^{b+d-1}\mathrm{d}y$$

$$= \frac{B(a + c, b + d)}{B(a, b)}.$$

$\qquad\square$

Let S_{k_n} denote the set of $k_n!$ permutations of $[k_n]$. The following lemma is key for proving Lemma 2.3.

4

Lemma 3.2. *For any $n_1, \ldots, n_{k_n} \in \mathbb{N}$,*

$$\sum_{\sigma \in S_{k_n}} \prod_{i=1}^{k_n} \frac{1}{a_i(\sigma) - (k_n - i + 1)d} = \frac{1}{\prod_{i=1}^{k_n}(n_i - d)}$$

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

Proof. Consider the process of sampling without replacement k_n times from an urn containing k_n balls. The balls have sizes $n_1 - d, \ldots, n_{k_n} - d$, and the probability of drawing ball i is proportional to its size $n_i - d$. Thus for any permutation $\sigma \in S_{k_n}$ we have that

$$p(\sigma_1) = \frac{n_{\sigma_1} - d}{n - td} = \frac{n_{\sigma_1} - d}{a_1(\sigma) - td},$$

$$p(\sigma_2 | \sigma_1) = \frac{n_{\sigma_2} - d}{n - n_{\sigma_1} - (k_n - 1)d} = \frac{n_{\sigma_2} - d}{a_2(\sigma) - (k_n - 1)d},$$

$$p(\sigma_i | \sigma_1, \ldots, \sigma_{i-1}) = \frac{n_{\sigma_i} - d}{n - n_{\sigma_1} - \cdots - n_{\sigma_{i-1}} - (k_n - i + 1)d} = \frac{n_{\sigma_i} - d}{a_i(\sigma) - (k_n - i + 1)d}.$$

Therefore,

$$p(\sigma) = p(\sigma_1)p(\sigma_2|\sigma_1) \cdots p(\sigma_{k_n}|\sigma_1, \ldots, \sigma_{k_n-1}) = \prod_{i=1}^{k_n} \frac{n_{\sigma_i} - d}{a_i(\sigma) - (k_n - i + 1)d}. \tag{4}$$

This way, we construct a distribution on S_{k_n}. We know that $\sum_{\sigma \in S_{k_n}} p(\sigma) = 1$. Applying this to Equation (4) and dividing both sides by $(n_{\sigma_1} - d) \cdots (n_{\sigma_{k_n}} - d) = (n_1 - d) \cdots (n_{k_n} - d)$ gives the result. $\square$

Lemma 3.3. *Let $b_i \in \mathbb{N}$ for $i \in \{1, \ldots, k_n\}$ and let $b_0 = 0$. We define $\bar{b}_i = b_0 + b_1 + \cdots + b_i$. Then*

$$\prod_{i=1}^{k_n} \sum_{b_i \in \mathbb{N}} \frac{(\frac{\alpha}{d} + \bar{b}_{i-1})_{(b_i)}}{(\frac{a_i + \alpha}{d} + \bar{b}_{i-1})_{(b_i)}} = \frac{(\frac{\alpha}{d})_{(k_n)}}{\prod_{i=1}^{k_n}(\frac{a_i}{d} - (k_n + 1 - i))}.$$

Proof. Let A_j denote the intermediate sum $A_j = \prod_{i=j}^{k_n} \sum_{b_i \in \mathbb{N}} \frac{(\frac{\alpha}{d} + \bar{b}_{i-1})_{(b_i)}}{(\frac{a_i + \alpha}{d} + \bar{b}_{i-1})_{(b_i)}}$. We show by induction decreasing from $j = k_n$ to $j = 0$ that

$$A_j = \frac{(\frac{\alpha}{d} + \bar{b}_{j-1})_{(k_n - j + 1)}}{\prod_{i=j}^{k_n}(\frac{a_i}{d} - (k_n + 1 - i))}. \tag{5}$$

When $j = k_n$ we have

$$A_{k_n} = \sum_{b_{k_n} \in \mathbb{N}} \frac{(\frac{\alpha}{d} + \bar{b}_{k_n-1})_{(b_{k_n})}}{(\frac{a_{k_n} + \alpha}{d} + \bar{b}_{k_n-1})_{(b_{k_n})}} = \sum_{b_{k_n} \in \mathbb{N}} \mathbb{E}[X^{b_{k_n}}]$$

5

where $X \sim \text{Beta}(\frac{\alpha}{d} + \bar{b}_{k_n-1}, \frac{a_{k_n}}{d})$. We have that

$$\sum_{b_{k_n} \in \mathbb{N}} \mathbb{E}[X^{b_{k_n}}] = \mathbb{E}\Big[\sum_{b_{k_n} \in \mathbb{N}} X^{b_{k_n}}\Big] = \mathbb{E}\left[\frac{X}{1-X}\right] = \frac{\alpha + d\bar{b}_{k_n-1}}{a_{k_n} - d},$$

due to Lemma 3.1, which proves the initialization for (5).

We now consider the case of an arbitrary j, greater than 0 and less than k_n. By the induction hypothesis, we have that Equation (5) holds for $j + 1$, that is

$$A_{j+1} = \frac{(\frac{\alpha}{d} + \bar{b}_j)_{(k_n-j)}}{\prod_{i=j+1}^{k_n}(\frac{a_i}{d} - (k_n + 1 - i))}.$$

Therefore,

$$\begin{aligned}
A_j &= \sum_{b_j \in \mathbb{N}} \frac{(\frac{\alpha}{d} + \bar{b}_{j-1})_{(b_j)}}{(\frac{a_j+\alpha}{d} + \bar{b}_{j-1})_{(b_j)}} \prod_{i=j+1}^{k_n} \sum_{b_i \in \mathbb{N}} \frac{(\frac{\alpha}{d} + \bar{b}_{i-1})_{(b_i)}}{(\frac{a_i+\alpha}{d} + \bar{b}_{i-1})_{(b_i)}} \\
&= \sum_{b_j \in \mathbb{N}} \frac{(\frac{\alpha}{d} + \bar{b}_{j-1})_{(b_j)}}{(\frac{a_j}{d} + \frac{\alpha}{d} + \bar{b}_{j-1})_{(b_j)}} \frac{(\frac{\alpha}{d} + \bar{b}_j)_{(k_n-j)}}{\prod_{i=j+1}^{k_n}(\frac{a_i}{d} - (k_n + 1 - i))}
\end{aligned}$$

Rearranging the rising factorials in the numerator, we can write

$$\begin{aligned}
\left(\frac{\alpha}{d} + \bar{b}_{j-1}\right)_{(b_j)}\left(\frac{\alpha}{d} + \bar{b}_j\right)_{(k_n-j)} &= \left(\frac{\alpha}{d} + \bar{b}_{j-1}\right)_{(b_j)}\left(\frac{\alpha}{d} + \bar{b}_{j-1} + b_j\right)_{(k_n-j)} \\
&= \left(\frac{\alpha}{d} + \bar{b}_{j-1}\right)_{(b_j+k_n-j)} \\
&= \left(\frac{\alpha}{d} + \bar{b}_{j-1}\right)_{(k_n-j)}\left(\frac{\alpha}{d} + \bar{b}_{j-1} + k_n - j\right)_{(b_j)}
\end{aligned}$$

and thus factorize the terms independent of b_j in order to obtain

$$A_j = \frac{(\frac{\alpha}{d} + \bar{b}_{j-1})_{(k_n-j)}}{\prod_{i=j+1}^{k_n}(\frac{a_i}{d} - (k_n + 1 - i))} \sum_{b_j \in \mathbb{N}} \frac{(\frac{\alpha}{d} + \bar{b}_{j-1} + k_n - j)_{(b_j)}}{(\frac{a_j}{d} + \frac{\alpha}{d} + \bar{b}_{j-1})_{(b_j)}}.$$

The sum above can be rewritten, using $X \sim \text{Beta}(\frac{\alpha}{d} + \bar{b}_{j-1} + (k_n - j), \frac{a_j}{d} - (k_n - j))$, as

$$\sum_{b_j \in \mathbb{N}} \mathbb{E}[X^{b_j}] = \mathbb{E}\left[\frac{X}{1-X}\right] = \frac{\frac{\alpha}{d} + \bar{b}_{j-1} + (k_n - j)}{\frac{a_j}{d} - (k_n + 1 - j)}.$$

Putting this all together,

$$A_j = \frac{(\frac{\alpha}{d} + \bar{b}_{j-1} + (k_n - j))}{\frac{a_j}{d} - (k_n + 1 - j)} \frac{(\frac{\alpha}{d} + \bar{b}_{j-1})_{(k_n-j)}}{\prod_{i=j+1}^{k_n}(\frac{a_i}{d} - (k_n + 1 - i))}$$

$$= \frac{(\frac{\alpha}{d} + \bar{b}_{j-1})_{(k_n - j + 1)}}{\prod_{i=j}^{k_n}(\frac{a_i}{d} - (k_n + 1 - i))}$$

which proves the desired result for j. By induction, this result is true for all $j \in \{1, \ldots, k_n\}$. Letting $j = 1$ gives the result stated in the lemma, since $\bar{b}_0 = b_0 = 0$. $\qquad\square$

3.2 Proof of Proposition 2.2 and Lemma 2.3

Proof of Proposition 2.2. For simplicity, we fix the allocation variable vector to a value z and denote $m(z)$ by m and $g_j(z)$ by g_j. We have

$$\mathbb{P}(\boldsymbol{z} = z | \pi_1, \ldots, \pi_m) = \prod_{i=1}^{n} \pi_{z_i} = \prod_{j=1}^{m} \pi_j^{e_j}$$

where $e_j = \#\{i : z_i = j\}$. Thus,

$$\mathbb{P}(\boldsymbol{z} = z | v_1, \ldots, v_m) = \prod_{j=1}^{m} \left(v_j \prod_{i=1}^{j-1}(1 - v_i) \right)^{e_j} = \prod_{j=1}^{m} v_j^{e_j}(1 - v_j)^{f_j}$$

where $f_j = \#\{i : z_i > j\}$. Therefore,

$$\mathbb{P}(\boldsymbol{z} = z) = \int \mathbb{P}(\boldsymbol{z} = z | v_1, \ldots, v_m) p(v_1, \ldots, v_m) \mathrm{d}v_1 \cdots \mathrm{d}v_m$$

$$= \int \left(\prod_{j=1}^{m} v_j^{e_j}(1 - v_j)^{f_j} \right) p_1(v_1) \cdots p_m(v_m) \mathrm{d}v_1 \cdots \mathrm{d}v_m$$

$$= \prod_{j=1}^{m} \int v_j^{e_j}(1 - v_j)^{f_j} p_j(v_j) \mathrm{d}v_j$$

$$\overset{(a)}{=} \prod_{j=1}^{m} \frac{B(e_j + 1 - d, f_j + \alpha + jd)}{B(1 - d, \alpha + jd)}$$

$$= \prod_{j=1}^{m} \frac{\Gamma(e_j + 1 - d)\Gamma(f_j + \alpha + jd)\Gamma(\alpha + (j-1)d) + 1}{\Gamma(e_j + f_j + \alpha + (j-1)d) + 1)\Gamma(1 - d)\Gamma(\alpha + jd)}$$

$$\overset{(b)}{=} \prod_{j=1}^{m} \frac{\Gamma(e_j + 1 - d)}{\Gamma(1 - d)} \prod_{j=1}^{m} \frac{\Gamma(g_{j+1} + \alpha + jd)}{\Gamma(g_j + \alpha + (j-1)d + 1)} \prod_{j=1}^{m} \frac{\Gamma(\alpha + (j-1)d + 1)}{\Gamma(\alpha + jd)}$$

$$\overset{(c)}{=} \prod_{j=1}^{m} \frac{\Gamma(e_j + 1 - d)}{\Gamma(1 - d)} \prod_{j=1}^{m} \frac{\alpha + (j-1)d}{g_j + \alpha + (j-1)d} \frac{\Gamma(g_{m+1} + \alpha + md)\Gamma(\alpha)}{\Gamma(g_1 + \alpha)\Gamma(\alpha + md)}$$

$$\overset{(d)}{=} \frac{\Gamma(\alpha)}{\Gamma(n + \alpha)} \prod_{c \in C_z} \frac{\Gamma(|c| + 1 - d)}{\Gamma(1 - d)} \prod_{j=1}^{m} \frac{\alpha + (j-1)d}{g_j + \alpha + (j-1)d}$$

where step (a) follows from Lemma 3.1, step (b) since $f_j = g_{j+1}$ and $g_j = e_j + f_j$, step (c) since $\Gamma(x+1) = x\Gamma(x)$, and step (d) since $g_1 = n$ and $g_{m+1} = 0$. $\qquad\square$

Proof of Lemma 2.3. As before, we denote the parts of C by $c_1, \ldots, c_{k_n}$, and we let $k_n = |C|$. We denote the distinct values taken on by $z_1, \ldots, z_n$ by $j_1 < \cdots < j_{k_n}$. We define $j_0 = b_0 = 0$, $b_i = j_i - j_{i-1}$, and $\bar{b}_i = b_0 + \cdots + b_i$ for $i \in \{1, \ldots, k_n\}$. We use the notation $a_i(\sigma) = n_{\sigma_i} + \cdots + n_{\sigma_{k_n}}$, where σ is the permutation of $[k_n]$ such that $c_{\sigma_i} = \{\ell : z_\ell = j_i\}$. Then for any $z \in \mathbb{N}^n$ such that $C_z = C$,

$$
\prod_{j=1}^{m(z)} \frac{\alpha + (j-1)d}{g_j(z) + \alpha + (j-1)d} = \prod_{j=1}^{m(z)} \frac{\frac{\alpha}{d} + j - 1}{\frac{g_j(z)+\alpha}{d} + j - 1}
$$
$$
= \prod_{i=1}^{k_n} \prod_{j=\bar{b}_{i-1}+1}^{\bar{b}_i} \frac{\frac{\alpha}{d} + j - 1}{\frac{g_j(z)+\alpha}{d} + j - 1} = \prod_{i=1}^{k_n} \frac{(\frac{\alpha}{d} + \bar{b}_{i-1})_{(b_i)}}{(\frac{\alpha+a_i(\sigma)}{d} + \bar{b}_{i-1})_{(b_i)}},
$$

because $g_j(z) = a_i(\sigma)$ for $\bar{b}_{i-1} < j \leq \bar{b}_i$. It follows from the definition of $b = (b_1, ..., b_{k_n})$ and σ that there is a one-to-one correspondence between $\{z \in \mathbb{N}^n : C_z = C\}$ and $\{(\sigma, b) : \sigma \in S_{k_n}, b \in \mathbb{N}^{k_n}\}$. Therefore,

$$
\sum_{z \in \mathbb{N}^n} \mathbb{1}(C_z = C) \prod_{j=1}^{m(z)} \frac{\alpha + (j-1)d}{g_j(z) + \alpha + (j-1)d} = \sum_{\sigma \in S_{k_n}} \sum_{b \in \mathbb{N}^{k_n}} \prod_{i=1}^{k_n} \frac{(\frac{\alpha}{d} + \bar{b}_{i-1})_{(b_i)}}{(\frac{a_i(\sigma)+\alpha}{d} + \bar{b}_{i-1})_{(b_i)}}
$$
$$
= \sum_{\sigma \in S_{k_n}} \prod_{i=1}^{k_n} \sum_{b_i \in \mathbb{N}} \frac{(\frac{\alpha}{d} + \bar{b}_{i-1})_{(b_i)}}{(\frac{a_i(\sigma)+\alpha}{d} + \bar{b}_{i-1})_{(b_i)}}
$$
$$
\overset{(a)}{=} \sum_{\sigma \in S_{k_n}} \prod_{i=1}^{k_n} \frac{(\frac{\alpha}{d})_{(k_n)}}{\frac{a_i(\sigma)}{d} - (k_n - i + 1)}
$$
$$
= d^{k_n} \left(\frac{\alpha}{d}\right)_{(k_n)} \sum_{\sigma \in S_{k_n}} \prod_{i=1}^{k_n} \frac{1}{a_i(\sigma) - (k_n - i + 1)d}
$$
$$
\overset{(b)}{=} \frac{d^{k_n}}{\prod_{c \in C}(|c| - d)} \left(\frac{\alpha}{d}\right)_{(k_n)},
$$

where step (a) follows from Lemma 3.3 and step (b) follows from Lemma 3.2. $\qquad\square$

Acknowledgement

References

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

Arbel, J., De Blasi, P., and Prünster, I. (2018). Stochastic approximations to the Pitman–Yor process. *Bayesian Analysis, in press.*

Arbel, J., Favaro, S., Nipoti, B., and Teh, Y. W. (2017). Bayesian nonparametric inference for discovery probabilities: credible intervals and large sample asymptotics. *Statistica Sinica*, 27:839–858.

Bassetti, F., Casarin, R., and Leisen, F. (2014). Beta-product dependent Pitman–Yor processes for Bayesian inference. *Journal of Econometrics*, 180(1):49 – 72.

Battiston, M., Favaro, S., Roy, D. M., and Teh, Y. W. (2018). A characterization of product-form exchangeable feature probability functions. *The Annals of Applied Probability*, 28(3):1423–1448.

Canale, A., Lijoi, A., Nipoti, B., and Prünster, I. (2017). On the Pitman–Yor process with spike and slab base measure. *Biometrika*, 104(3):681–697.

Caron, F., Neiswanger, W., Wood, F., Doucet, A., and Davy, M. (2017). Generalized Pólya Urn for Time-Varying Pitman-Yor Processes. *Journal of Machine Learning Research*, 18(27):1–32.

Clauset, A., Shalizi, C. R., and Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM review*, 51(4):661–703.

De Blasi, P., Favaro, S., Lijoi, A., Mena, R. H., Prünster, I., and Ruggiero, M. (2015). Are Gibbs-type priors the most natural generalization of the Dirichlet process? *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 37(2):212–229.

Derrida, B. (1981). Random-energy model: An exactly solvable model of disordered systems. *Physical Review B*, 24(5):2613.

Favaro, S., Lijoi, A., Mena, R., and Prünster, I. (2009). Bayesian non-parametric inference for species variety with a two-parameter Poisson–Dirichlet process prior. *J. R. Stat. Soc. Ser. B*, 71:993–1008.

Favaro, S. and Walker, S. G. (2013). Slice sampling σ-stable Poisson-Kingman mixture models. *Journal of Computational and Graphical Statistics*, 22(4):830–847.

Feng, S. and Sun, W. (2010). Some diffusion processes associated with two parameter Poisson–Dirichlet distribution and Dirichlet process. *Probability theory and related fields*, 148(3-4):501–525.

Ferguson, T. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2):209–230.

Ishwaran, H. and James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *J. Amer. Statist. Assoc.*, 96:161–173.

Jara, A., Lesaffre, E., De Iorio, M., and Quintana, F. (2010). Bayesian semiparametric inference for multivariate doubly-interval-censored data. *Ann. Appl. Stat.*, 4(4):2126–2149.

Kerov, S. V. (2006). Coherent random allocations, and the Ewens-Pitman formula. *Journal of Mathematical sciences*, 138(3):5699–5710.

Lo, A. (1984). On a class of Bayesian nonparametric estimates: I. Density estimates. *The Annals of Statistics*, 12(1):351–357.

Miller, J. W. (2018). An elementary derivation of the Chinese restaurant process from Sethuraman's stick-breaking process. *arXiv preprint arXiv:1801.00513*.

Miller, J. W. and Harrison, M. T. (2014). Inconsistency of Pitman-Yor process mixtures for the number of components. *The Journal of Machine Learning Research*, 15(1):3333–3370.

Navarrete, C., Quintana, F. A., and Mueller, P. (2008). Some issues in nonparametric Bayesian modeling using species sampling models. *Statistical Modelling*, 8(1):3–21.

Ni, Y., Müller, P., Zhu, Y., and Ji, Y. (2018). Heterogeneous reciprocal graphical models. *Biometrics*, 74(2):606–615.

Perman, M., Pitman, J., and Yor, M. (1992). Size-biased sampling of Poisson point processes and excursions. *Probability Theory and Related Fields*, 92(1):21–39.

Petrov, L. (2009). Two-parameter family of diffusion processes in the Kingman simplex. *Functional Analysis and Its Applications*, 43:279–296.

Pitman, J. (1995). Exchangeable and partially exchangeable random partitions. *Probability Theory and Related Fields*, 102(2):145–158.

Pitman, J. (2003). Poisson-Kingman partitions. *Lecture Notes-Monograph Series*, pages 1–34.

Pitman, J. and Yor, M. (1997). The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *The Annals of Probability*, 25(2):855–900.

Scarpa, B. and Dunson, D. B. (2009). Bayesian hierarchical functional data analysis via contaminated informative priors. *Biometrics*, 65(3):772–780.

Scricciolo, C. et al. (2014). Adaptive Bayesian Density Estimation in L^p-metrics with Pitman-Yor or Normalized Inverse-Gaussian Process Kernel Mixtures. *Bayesian Analysis*, 9(2):475–520.

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

Sudderth, E. B. and Jordan, M. I. (2009). Shared segmentation of natural scenes using dependent Pitman-Yor processes. In *Advances in Neural Information Processing Systems 21*, pages 1585–1592. Curran Associates, Inc.

Teh, Y. W. (2006). A hierarchical Bayesian language model based on Pitman-Yor processes. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, pages 985–992. Association for Computational Linguistics.

Vershik, A., Yor, M., and Tsilevich, N. (2004). On the Markov–Krein identity and quasi-invariance of the gamma process. *Journal of Mathematical Sciences*, 121(3):2303–2310.

Wood, F., Gasthaus, J., Archambeau, C., James, L., and Teh, Y. W. (2011). The sequence memoizer. *Communications of the ACM*, 54(2):91–98.