Exchangeable random partitions and random discrete probability measures: a brief tour guided by the Dirichlet Process

Notes for Oxford Statistics Grad Lecture

Benjamin Bloem-Reddy
benjamin.bloem-reddy@stats.ox.ac.uk

15 February, 2018

These notes are a work in progress; if you find any mistakes, please let me know.

1 Overview and preliminaries

This lecture is intended to be an introductory overview of exchangeable random partitions, random discrete probability measures, and the connections between them—namely, that they are two sides of the same coin, and that in many cases one may move between the two via known characterizations. For the sake of clarity and of time, I will skip most of the technicalities. Further background and details can be found in the references given throughout the notes.

I will focus on two concrete examples: the Chinese Restaurant Process (CRP) and the Dirichlet Process (DP). The CRP is the canonical model for exchangeable random partitions, and the DP is the canonical discrete random probability measure, each satisfying basically every nice property one could desire. This is no coincidence: the DP can be obtained as the limit of the CRP, and the CRP can be obtained as the predictive process of the DP.

Often, a property satisfied by the DP yields a generalization that leads to a larger class of stochastic processes that satisfy some variation of the original property. The DP is the only process that satisfies each of these properties, making it something of a unicorn, and a starting point for further reading. (I've tried to cite liberally for this purpose.)

Preliminaries. Throughout, I will assume a common probability space $(\Omega, \mathcal{A}, \mathbb{P})$. A random measure $\mathbf{W}$ on a measurable space $(S, \mathcal{S})$ is technically defined as a transition kernel from $(\Omega, \mathcal{A})$ to $(S, \mathcal{S})$. That is, a mapping $W : \Omega \times \mathcal{S} \mapsto \mathbb{R}_+$ is a random measure if $\omega \mapsto \mathbf{W}(\omega, A)$ is a random variable for each $A \in \mathcal{S}$ and if $A \mapsto \mathbf{W}(\omega, A)$ is a measure on $(S, \mathcal{S})$ for each $\omega \in \Omega$. If the space $(S, \mathcal{S})$ is "nice" (i.e., standard Borel)—which in practice it always is—the existence of a unique (up to null sets) $\mathbf{W}$ with the properties we want is guaranteed. If you want to dig into this, any graduate-level textbook on measure theoretic probability will cover it in detail; I highly recommend [Çin11]. This is the end of measure-theoretic technicalities; for our purposes, it is enough to think of $\mathbf{W}$ as a measure-valued random variable.

A *random discrete measure* is a random measure that is almost surely discrete, and therefore has the representation

$$\mathbf{W}(\,\cdot\,) = \sum_{j=1}^{\infty} W_j \delta_{X_j^*}(\,\cdot\,) \quad \text{for} \quad X_j^* \in S\ . \tag{1}$$

(X_j) are called the *atom locations* and (W_j) the *atom weights* of $\mathbf{W}$. A random measure with i.i.d. (X_j) that are also independent of the atom weights is called *homogeneous*. All random measures will be assumed to be homogeneous in these notes. A *random discrete probability measure* is a random discrete measure such that $\sum_j W_j = 1$, almost surely. To distinguish between the two, P_j is used for the weights of a random discrete probability measure.

The Dirichlet Process (DP). The DP is a random discrete probability measure first introduced by [Fer73]. It can be (and is) defined in a number of ways. This in itself hints at how special the DP is; as we will see, a number of properties are unique to the DP and can be used to define it. The following is perhaps the most uninformative (though also the most common). A DP on $(S, \mathcal{S})$ is specified by a *base probability distribution* H on S and a parameter $\theta > 0$. The DP so defined is stochastic process whose realizations are discrete probability measures that satisfy the following property almost surely: Given *any* finite partition $\{A_i\}_{i=1}^k$ of S, $P \sim \mathrm{DP}(\theta, H)$ is such that

$$(P(A_1), \dots, P(A_k)) \overset{\mathrm{d}}{=} (D_1, \dots, D_k) \quad \text{where} \quad (D_1, \dots, D_k) \sim \mathrm{Dir}(\theta H(A_1), \dots, \theta H(A_k))\ .$$

The finite-dimensional distributions of a DP on S are indexed by all of the finite partitions of S; they are coherent such that they satisfy the requirements of Kolmogorov's extension theorem, which guarantees the existence of the DP. Ferguson showed that the DP is conjugate with itself (a property that makes inference easy for the Hierarchical DP [Teh+06]) and has a simple predictive distribution. That is, suppose $P \sim \mathrm{DP}(\theta, H)$ and n samples are drawn $X_1, \dots, X_n \sim P$. Then the posterior of P is

$$\mathbb{P}[P \in \,\cdot\, | X_1, \dots, X_n] = \mathrm{DP}\left(\theta + n, \frac{\theta H + \sum_{i=1}^n \delta_{X_i}}{\theta + n}\right) \tag{2}$$

and the predictive distribution is

$$\mathbb{P}[X_{n+1} \in \,\cdot\, | X_1, \dots, X_n] = \frac{\theta}{\theta + n} H(\,\cdot\,) + \frac{1}{\theta + n} \sum_{i=1}^n \delta_{X_i}(\,\cdot\,)\ . \tag{3}$$

These basic properties will be used to develop a number of interesting characterizations of the DP, which are *highlighted in blue.* *

2 Exchangeable random partitions of $\mathbb{N}$

We begin by constructing random partitions from a sequence. This is more intuitive than working directly with partitions; properties of partitions can be translated into properties of sequences, and vice versa. Although it is possible to define distributions directly on random partitions [Pit06], in practice they are most often generated by a sequence. The exposition here owes much to Sections 3 and 4 of [Pit96b].

Species sampling and clustering. Suppose that a random sample $\mathbf{X} = X_1, X_2, \dots$ is drawn from a continuous spectrum of colors, or species. Assume that X_i takes values in some measurable space $(S, \mathcal{S})$, and that we keep track of the unique species X_j^*. We may do so by noting the *arrival time* of the jth new color, for each j,

$$T_j := \inf\{n : X_n \notin \{X_1, \dots, X_{n-1}\}\}\ , \tag{4}$$

and setting $X_j^* = X_{T_j}$. A sequence $\mathbf{X}_n = X_1, \ldots, X_n$ will contain a random number of unique species, denoted K_n. Let the number of occurrences of the jth species to appear in $\mathbf{X}_n$ be denoted by $n_{j,n}$:

$$n_{j,n} := \sum_{i=1}^{n} \mathbb{1}\{X_i = X_j^*\} \,. \tag{5}$$

$\mathbf{X}_n$ induces a random partition of $[n] := \{1, \ldots, n\}$ by clustering observations of the same species:

$$\Pi(\mathbf{X}_n) := \Pi_n = \{B_1, \ldots, B_{K_n}\} \quad \text{where} \quad i \in B_j \iff X_i = X_j^* \,. \tag{6}$$

Note that if the values of $\mathbf{X}_n$ are all distinct (i.e., every n is an arrival time), then the partition consists of n *singleton* blocks: $\Pi(\mathbf{X}_n) = \{\{1\}, \ldots, \{n\}\}$. We'll ignore this case as uninteresting, but note that all technical results must take the pure singletons case (and the single block case) into account.[1]

A *prediction rule* tells us how to generate a new sample X_{n+1}, given $\mathbf{X}_n$. For instance, assume that $X_1 \sim \nu(\,\bullet\,)$, where ν is a non-atomic distribution on S, and proceed as follows:

$$\mathbb{P}_\theta[X_{n+1} \in \bullet \,|\mathbf{X}_n] = \frac{\theta}{n+\theta}\nu(\,\bullet\,) + \sum_{j=1}^{K_n} \frac{n_{j,n}}{n+\theta}\delta_{X_j^*}(\,\bullet\,) \quad \text{with} \quad \theta > 0 \,. \tag{7}$$

This is the prediction rule of the *Chinese Restaurant Process* (CRP). The scheme and the name are attributed to Pitman and Dubins [Pit06], who thought of the process as customers arriving at a Chinese restaurant with a potentially infinite number of tables, and blocks in the partition corresponding to tables at which customers are sitting. With probability $\frac{\theta}{n+\theta}$ we sample a new species, otherwise we sample an existing species with probability proportional to its number of previous occurrences.

The Dirichlet Process. Observe that (7) may be differently obtained as follows. Sample $P \sim \mathrm{DP}(\theta, \nu)$ and $X_1, \ldots, X_n \sim P$. The predictive distribution (3), re-written to account for multiple observations of the same atom, is precisely (7). *Hence, the CRP is the unique predictive (or urn) process associated with the DP.* Properties of the CRP are hence also sampling properties of the DP. *

Another property unique to the DP can be described by considering the probability of observing a new species in an exchangeable species sampling model. *The CRP (and thus the DP) is the only exchangeable species sampling model in which the probability of observing a new species depends only on n (and not K_n or $(n_{j,n})$) [De +15].*

Exchangeability. It is not hard to show that $\mathbf{X}_n$ sampled from (7) is exchangeable for every n; hence, $\mathbf{X}$ is *infinitely exchangeable*. The distinction may seem unnecessary, but there are plenty of statistical models that are finitely exchangeable that are not infinitely exchangeable; the difference typically amounts to a lack of *coherence*, also known as *consistency under marginalization* (e.g., [BCL11; Bet+16]), which has implications for valid inference across datasets of difference sizes. I will assume going forward that all exchangeable distributions are infinitely exchangeable.

A random partition Π_n is said to be exchangeable if its distribution is invariant under permutations of $[n]$; $\Pi := (\Pi_n)_{n\geq 1}$ is an exchangeable partition of $\mathbb{N}$ if Π_n is exchangeable for every n. Clearly, $\Pi(\mathbf{X})$ is exchangeable if $\mathbf{X}$ is exchangeable.

EPPFs and Gibbs-type partitions. Consider the distribution of the partition Π_n induced by $\mathbf{X}_n$ sampled from

[1] The fact that its probability is so easy to compute can be useful for proving technical results. For example, see [GP06].

(7):

$$\mathbb{P}_\theta[\Pi_n] = \mathbb{P}_\theta\left[\cap_{j=1}^{K_n}(X_i = X_j^*)\ \forall i \in B_j, K_n = k\right] = \frac{\theta^{k-1}\prod_{j=1}^{k}(n_{j,n}-1)!}{\prod_{i=1}^{n-1}(\theta+i)} \tag{8}$$
$$= \mathbb{P}_\theta[n_{j,1},\dots,n_{j,k}, K_n = k]$$
$$:= p(n_{j,1},\dots,n_{j,k})\ .$$

p is called the *exchangeable partition probability function* (EPPF) for obvious reasons: It is the probability of the partition Π_n, symmetric in its arguments—the block sizes of Π_n—and thus invariant under permutations of the blocks and of the elements of $[n]$.

Note that the EPPF is defined on partitions and does not require that Π_n is induced by a sequence. (In fact, an EPPF can be used to generate a species sampling sequence $\mathbf{X}$; see Section 3.2 of [Pit96b].)

An EPPF (and the corresponding model of exchangeable partitions) is said to be *Gibbs-type* if the EPPF has the form

$$\mathbb{P}_\alpha[\Pi_n = \{n_{1,n},\dots,n_{K_n,n}\}] = V_{n,K_n}\prod_{j=1}^{K_n}\frac{\Gamma(n_{j,n}-\alpha)}{\Gamma(1-\alpha)}\ , \tag{9}$$

for some $-\infty \leq \alpha < 1$ and a sequence of coefficients $V_{n,k}$, $1 \leq k \leq n$ that control the distribution of the number of blocks. A bit of algebra shows that (8) is of this form, with $\alpha = 0$. *Gnedin and Pitman [GP06] showed that* ***all*** *Gibbs-type partitions with $\alpha = 0$ are mixtures over θ of the CRP.* The $\alpha \in (0,1)$ regime has fascinating properties, with connections to stable processes [HJL07; HJL18] and to another interesting class of models, the Poisson–Kingman partitions [Pit03]. *

Further characterizations of the CRP. The CRP and its two-parameter generalization have a number of other unique properties described in terms of random deletion and regeneration of partitions and compositions [GHP09; Gne10].

Kingman's paintbox. The construction of an exchangeable random partition from an exchangeable sequence is useful. Can every exchangeable partition be constructed from an exchangeable sequence? Kingman [Kin78a; Kin78b] showed that the answer is "yes", via the following *paintbox* construction of Π (this description closely follows [GHP09]). Suppose we have a sequence of random variables $\mathbf{P}^{\downarrow} = (P_1^{\downarrow}, P_2^{\downarrow}, \dots)$, called *ranked frequencies*, such that

$$1 \geq P_1^{\downarrow} \geq P_2^{\downarrow} \geq \cdots \geq 0 \quad \text{and} \quad P_* := 1 - \sum_{j=1}^{\infty} P_j^{\downarrow} \geq 0\ . \tag{10}$$

Thus, intervals $[l_k, r_k)$ of length $P_k^{\downarrow}$, with $l_k = \sum_{j=1}^{k-1} P_j^{\downarrow}$, $r_k = l_k + P_k^{\downarrow}$, partition the unit interval $[0,1)$, with a final interval $[1-P_*, 1)$ if $P_* > 0$. Consider all points in each interval as having the same unique color, identified by l_k: $c(u) = l_k$ for all $u \in [l_k, r_k)$; if $P_* > 0$, let $c(u) = u$ for all $u \in [1-P_*, 1)$ (these generate singleton blocks, also called *dust*). This coloring of points in the unit interval is called Kingman's paintbox representation of $\mathbf{P}^{\downarrow}$. It can be used to generate Π by sampling an infinite sequence of uniform random variables $U_i \sim \text{Uniform}[0,1]$, assigning color $c(U_i)$, and generating the partition Π_n by

$$\Pi(\mathbf{X}_n) := \Pi_n = \{B_1,\dots,B_{K_n}\} \quad \text{where} \quad i \in B_j \iff c(U_i) = l_j\ .$$

It's clear that this will generate an exchangeable partition of $\mathbb{N}$; it's less clear that *any* exchangeable partition can be generated this way. Kingman showed that this is the case. Intuition via de Finetti?

Theorem 2.1 (Kingman [Kin78a; Kin78b]; restated from [GHP09])**.** *Each exchangeable partition Π of $\mathbb{N}$ generates a sequence of ranked frequencies $\mathbf{P}^{\downarrow}$ such that the conditional distribution of Π given these frequencies* *is that of the color partition of $\mathbb{N}$ derived from $\mathbf{P}^{\downarrow}$ by Kingmans paintbox construction. The EPPF associated with Π determines the distribution of $\mathbf{P}^{\downarrow}$, and vice versa.*

Kingman's representation shows that the study of exchangeable random partitions is equivalent to the study of random discrete probability measures, which we take up in the next section.

Representations of exchangeable structures. The paintbox representation lets us express all dependence between elements of Π via $\mathbf{P}^{\downarrow}$; given $\mathbf{P}^{\downarrow}$, the variables $c(U_i)$ are conditionally independent. This type of construction of an exchangeable structure—in terms of i.i.d. uniform random variables passed through a random function—is extremely useful for statistical inference and probabilistic analysis, and appears in various forms for data like arrays and stationary processes. See [Ald09; OR15].

Uses of random partition models and further reading. In practice, models for exchangeable random partitions are most often used as priors for Bayesian nonparametric latent clustering models; Jessie Wu will talk next week about these applications (and others). Random partition models also have substantial roots in the population genetics literature, where Ewens' sampling formula, equivalent to (8), arises from the study the frequencies of alleles in a sample from a population. See [Cra16] for a recent review. Prediction rules like (7) have been used in biological applications to estimate quantities like the expected number of new species in the next m samples, and related quantities. See, for example, [FNT16]. Generalizations of the CRP prediction rule are its two-parameter extension [Pit96b] and more complicated constructions [IJ03].

3 Random discrete probability measures

The ranked frequencies $\mathbf{P}^{\downarrow} = (P_1^{\downarrow}, P_2^{\downarrow}, \dots)$ are typically difficult to work with analytically. Instead, we define the frequencies as the limiting proportions of blocks in Π labeled in order of appearance, $P_j = \lim_{n\to\infty} \frac{n_{j,n}}{n}$. (These limits exist almost surely by Theorem 2.1.) How might we generate $\mathbf{P} = (P_1, P_2, \dots)$? Consider sampling from Kingman's paintbox induced by some $\mathbf{P}^{\downarrow}$, and assume for simplicity that $P_* = 0$. The probability that the length of the interval containing U_1 is P_1 is just the interval's length, P_1. This is known as a *size-biased sample* from $\mathbf{P}^{\downarrow}$. P_2 is the length of the interval corresponding to the next U_i that is not in the same interval as U_1, which is a size-biased sample of $\mathbf{P}^{\downarrow} \setminus P_1$. Proceeding in this fashion, it is clear that the probability of a *random size-biased permutation* of $\mathbf{P}^{\downarrow}$ is

$$\mathbb{P}[P_1, P_2, \dots | \mathbf{P}^{\downarrow}] = P_1 \frac{P_2}{1-P_1} \frac{P_3}{1-P_2-P_1} \cdots \frac{P_k}{1-\sum_{j=1}^{k-1} P_j} \cdots \ . \tag{11}$$

It is usually more convenient to specify distributions over random discrete probability measures in terms of their size-biased representations. Size-biasing essentially "smooths" the hard ordering restrictions on the ranked frequencies: P_j will larger than P_{j+1} *on average*, but it relaxes the strict ordering requirement.

3.1 Stick-breaking constructions

How might we specify a distribution on $\mathbf{P}$ and sample from it? Assume for simplicity that $P_* = 0$ so that $\sum_j P_j = 1$. To start, we'd like the simplest possible construction: Experience indicates that for the purposes of analysis and inference, more independence is better. Sampling P_j i.i.d. won't do the trick; an infinite sum of i.i.d. positive variables diverges almost surely. We might think instead of composing i.i.d. $[0,1]$-valued random variables in a way that ensures that $\sum_j P_j = 1$.

A simple way to do so is as follows. Sample $V_1, V_2, \ldots$ i.i.d. from some distribution F on $[0, 1]$, and set $P_1 = V_1$ and

$$P_k := V_k \prod_{j=1}^{k-1} (1 - V_j) \; . \tag{12}$$

To understand this construction, consider $P_2 = V_2(1 - V_1)$. $1 - V_1$ is the size of the portion of the unit interval that remains after V_1 is "broken off". P_2 is just the V_2-sized break from that remainder, and so on for $k > 2$. Clearly, this will satisfy $\sum_{j=1}^{k} P_j \leq 1$ for all k, and the sum gets arbitrarily close to 1 as $k \to \infty$.

Remarkably, McCloskey [McC65] (see also [Pit96a]) showed that $\mathbf{P}$ constructed this way corresponds to the limiting frequencies of an exchangeable random partition Π in order of appearance if and only if F is the $\text{Beta}(1, \theta)$ distribution, in which case the distribution of Π is given by the EPPF of the CRP (8).

Another remarkable result due to Pitman [Pit96a] is that if we relax the requirement that the V_j be i.i.d. to simply requiring them to be independent, then only $V_j \sim \text{Beta}(1 - \alpha, \theta + j\alpha)$, for $\alpha \in [0, 1)$, $\theta > -\alpha$ will produce a sequence $\mathbf{P}$ corresponding to the limiting frequencies of an exchangeable random partition in order of appearance. This choice of stick-breaking construction is known in the Bayesian nonparametrics literature as the Pitman–Yor process [IJ01]. For more on its properties, see [Pit96a; PY97].

Of course, one may define the weights of a random discrete probability measure with any arbitrary sequence $V_j \sim F_j$, and with V_j not necessarily independent, as long as $V_j \in [0, 1)$. This will not always correspond to a size-biased representation of the limiting frequencies of an exchangeable random partition, but it will define a valid discrete probability measure with which to sample an exchangeable random partition via Kingman's paintbox. An example of this type of construction with independent $V_j \sim \text{Beta}(a_j, b_j)$ is the so-called Beta–Stacy process [WM97].

The Dirichlet Process. Sethuraman [Set94] constructively defined the DP as precisely the $\text{Beta}(1, \theta)$ stick-breaking process. Its uniqueness as *the only i.i.d. stick-breaking construction corresponding to an exchangeable random partition* is a consequence of properties that uniquely characterize the Gamma distribution, which itself is connected to the property of *neutrality* (more on this in the next section).

3.2 Normalized completely random measures

A different construction of random discrete probability measures that uses independence is by normalizing a *completely random measure* (CRM). I won't get into the details of CRMs here, but it is enough to say that a random measure (not necessarily a probability measure) $\mathbf{W}$ on $(S, \mathcal{S})$ is completely random if, for any two disjoint sets $A, A' \subset S$, $\mathbf{W}(A) \perp\!\!\!\perp \mathbf{W}(A')$. This is equivalent to a Poisson point process on $S \times \mathbb{R}_+$. Kingman's monograph [Kin93] is a beautiful introduction to all things Poisson process, including CRMs. A Poisson process is typically defined by its mean measure μ (sometimes called the parameter measure, or the Lévy measure): $\mathbb{E}[\mathbf{W}(A)] = \mu(A)$ for $A \subseteq S \times \mathbb{R}_+$. Let $T := \sum_{j=1}^{\infty} W_j$ be the (random) total mass of $\mathbf{W}$. A sufficient condition for T to be almost surely finite is

$$\int_{\mathbb{R}_+} \int_S \min\{1, w\} \mu(dw, dx) < \infty \; . \tag{13}$$

Define $P_j = W_j / T$. It is clear that the probability weights P_j depend on each other only through the total mass. Normalized CRMs have been used to great success in a range of applications due to the fact that posterior sampling often has analytic updates that make use of the fact that the P_j's are only dependent through T [Jam05; JLP09; FT13; LFT15].

The Dirichlet Process. The DP can be constructed by normalizing the Gamma process. Before making this precise, consider the finite-dimensional Dirichlet distribution. Let $(D_1, \dots, D_k)$ denote a random vector with distribution Dir$(\alpha_1, \dots, \alpha_k)$. A Dir$(\alpha_1, \dots, \alpha_k)$-distributed random vector can be constructed via the identity *

$$(D_1, \dots, D_k) \stackrel{\mathrm{d}}{=} \left(\frac{G_1}{\sum_{j=1}^k G_j}, \dots, \frac{G_k}{\sum_{j=1}^k G_j} \right) \quad \text{with} \quad G_j \sim_{\text{ind}} \Gamma(\alpha_j, 1) \ .$$

The Gamma distribution's unique properties extend to the stochastic process level, and impart the DP with some fascinating properties.

Denote by G_a a Gamma$(a, 1)$ random variable. Suppose $G_a \perp\!\!\!\perp G_b$. Then

$$\frac{G_a}{G_a + G_b} \perp\!\!\!\perp (G_a + G_b) \quad \text{and} \quad G_a + G_b \stackrel{\mathrm{d}}{=} G_{a+b} \ . \tag{14}$$

Lukacs [Luk55] showed that this property is unique to the Gamma distribution.

This also holds at the stochastic process level, and yields a unique characterization of the DP. In particular, the DP can be constructed by normalizing the jumps of the Gamma process, whose mean measure $\mu(dw, ds) = \rho(dw)\nu(ds)$ has jump component

$$\rho(dw) = \alpha w^{-1} e^{-\beta w} \ .$$

The Gamma process has the property of *infinite activity*, meaning that there are an infinite number of jumps on any bounded set. However, most of the weights are very small (this can be made precise); as a result, $T < \infty$ almost surely.

The DP is constructed by normalizing the jumps, $P_j = W_j/T$, which are independent of the total mass due to (14); *the DP is the only normalized CRM with this property.*

Neutrality. A random probability vector $\mathbf{P}_k = (P_1, \dots, P_k)$, with $P_j \in [0, 1]$ and $\sum_{j=1}^k P_j \leq 1$, is *neutral to the right* (NTR) if the sequence of relative increments

$$\mathbf{R}_r := \left(P_1, \frac{P_2}{1 - P_1}, \dots, \frac{P_k}{1 - \sum_{j=1}^{k-1} P_j} \right) \tag{15}$$

is a vector of independent random variables.

A random probability vector is *neutral to the left* (NTL) if the sequence of relative increments

$$\mathbf{R}_\ell := \left(P_1, \frac{P_2}{P_1 + P_2}, \dots, \frac{P_k}{\sum_{j=1}^k P_j} \right) \tag{16}$$

is a vector of independent random variables.

A NTL vector can be obtained from an NTR vector, and vice versa, by reversing the order of the elements of the vector. It is important to note that neutrality relies on an ordering of the vector. A NTR (NTL) vector with randomly permuted entries may not be NTR (NTL). Both of these concepts can be extended to stochastic processes defined on $\mathbb{R}$ [Dok74], which imposes a natural ordering on the relative increments. NTR processes have been used as priors for Bayesian nonparametric survival analysis [Hjo90] and other applications [WM97], and extended to processes defined on arbitrary spaces $\mathbb{R} \times \mathcal{X}$ [Jam06]. NTL processes are less common, but have appeared in more recent work related to Gibbs-type partitions [GS07] and preferential attachment graphs [BO17].

It is straightforward to show via the Gamma process construction that the DP defined on $\mathbb{R}$ is both NTR and NTL. *This property in fact characterizes the DP as the only stochastic process that is both NTR and NTL [Dok74].* Because of this, the DP is neutral with respect to any order imposed on the space on which it is defined. *

For example, note that the relative increments correspond to the probabilities involved in generating a random size-biased permutation (11) of a probability weight sequence. A natural question is whether a random discrete probability measure is NTR under random size-biased ordering. A bit of algebra also shows that the stick-breaks in (12) are $V_k = \frac{P_k}{1-\sum_{j=1}^{k-1} P_j}$, implying that for the DP,

$$\mathbf{R}_r = (V_1, \dots, V_k) \quad \text{with} \quad V_j \sim_{\text{ind}} \text{Beta}(1, \theta) \ .$$

Extending this idea, one can show that a random discrete probability measure is NTR under random size-biased ordering if and only if it has a stick-breaking construction with independent stick-breaks.

References

[Ald09] David J. Aldous. "More Uses of Exchangeability: Representations of Complex Random Structures". In: (2009). eprint: 0909.4339. URL: http://arxiv.org/abs/0909.4339.

[BCL11] Peter J. Bickel, Aiyou Chen, and Elizaveta Levina. "The method of moments and degree distributions for network models". In: *Ann. Statist.* 39.5 (2011), pp. 2280–2301.

[Bet+16] Brenda Betancourt et al. "Flexible Models for Microclustering with Application to Entity Resolution". In: *Advances in Neural Information Processing Systems 29*. Ed. by D. D. Lee et al. Curran Associates, Inc., 2016, pp. 1417–1425.

[BO17] Benjamin Bloem-Reddy and Peter Orbanz. "Preferential Attachment and Vertex Arrival Times". In: (Oct. 2017). eprint: 1710.02159. URL: https://arxiv.org/abs/1710.02159.

[Çin11] Erhan Çinlar. *Probability and Stochastics*. Graduate Texts in Mathematics. Springer-Verlag New York, 2011.

[Cra16] Harry Crane. "The Ubiquitous Ewens Sampling Formula (with discussion and a rejoinder by the author)". In: *Statist. Sci.* 31.1 (Feb. 2016), pp. 1–19.

[De +15] P. De Blasi et al. "Are Gibbs-Type Priors the Most Natural Generalization of the Dirichlet Process?" In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37.2 (Feb. 2015), pp. 212–229.

[Dok74] Kjell Doksum. "Tailfree and Neutral Random Probabilities and Their Posterior Distributions". In: *Ann. Probab.* 2.2 (Apr. 1974), pp. 183–201.

[Fer73] T. S. Ferguson. "A Bayesian Analysis of some Nonparametric Problems". In: *Ann. Statist.* 1.2 (1973).

[FNT16] Stefano Favaro, Bernardo Nipoti, and Yee Whye Teh. "Rediscovery of Good–Turing estimators via Bayesian nonparametrics". In: *Biometrics* 72.1 (2016), pp. 136–145.

[FT13] Stefano Favaro and Yee Whye Teh. "MCMC for Normalized Random Measure Mixture Models". In: *Statistical Science* 28.3 (Aug. 2013), pp. 335–359.

[GHP09] Alexander Gnedin, Chris Haulk, and Jim Pitman. "Characterizations of exchangeable partitions and random discrete distributions by deletion properties". In: (Sept. 2009). eprint: 0909.3642. URL: https://arxiv.org/abs/0909.3642.

[Gne10] Alexander V. Gnedin. "Regeneration in random combinatorial structures". In: *Probab. Surveys* 7 (2010), pp. 105–156.

[GP06] Alexander Gnedin and Jim Pitman. "Exchangeable Gibbs partitions and Stirling triangles". In: *Journal of Mathematical Sciences* 138.3 (2006), pp. 5674–5685. ISSN: 1573-8795.

[GS07] Robert C. Griffiths and Dario Spanò. "Record Indices and Age-Ordered Frequencies in Exchangeable Gibbs Partitions". In: *Electron. J. Probab.* 12 (2007), pp. 1101–1130.

[HJL07] Man-Wai Ho, Lancelot F. James, and John W. Lau. "Gibbs Partitions (EPPF's) Derived From a Stable Subordinator are Fox H and Meijer G Transforms". In: (Aug. 2007). eprint: 0708.0619. URL: https://arxiv.org/abs/0708.0619.

[HJL18] Man-Wai Ho, Lancelot F. James, and John W. Lau. "Gibbs Partitions, Riemann-Liouville Fractional Operators, Mittag-Leffler Functions, and Fragmentations Derived From Stable Subordinators". In: (Feb. 2018). eprint: 1802.05352. URL: https://arxiv.org/abs/1802.05352.

[Hjo90] N. L. Hjort. "Nonparametric Bayes Estimators Based on Beta Processes in Models for Life History Data". In: *Ann. Statist.* 18 (1990), pp. 1259–1294.

[IJ01] H. Ishwaran and L. F. James. "Gibbs Sampling Methods for Stick-Breaking Priors". In: *Journal of the American Statistical Association* 96.453 (2001), pp. 161–173.

[IJ03] Hemant Ishwaran and Lancelot F. James. "GENERALIZED WEIGHTED CHINESE RESTAURANT PROCESSES FOR SPECIES SAMPLING MIXTURE MODELS". In: *Statistica Sinica* 13.4 (2003), pp. 1211–1235. ISSN: 10170405, 19968507. URL: http://www.jstor.org/stable/24307169.

[Jam05] Lancelot F. James. "Bayesian Poisson process partition calculus with an application to Bayesian Lévy moving averages". In: *Ann. Statist.* 33.4 (Aug. 2005), pp. 1771–1799.

[Jam06] Lancelot F. James. "Poisson calculus for spatial neutral to the right processes". In: *Ann. Statist.* 34.1 (Feb. 2006), pp. 416–440.

[JLP09] Lancelot F. James, Antonio Lijoi, and Igor Prünster. "Posterior Analysis for Normalized Random Measures with Independent Increments". In: *Scandinavian Journal of Statistics* 36.1 (2009), pp. 76–97.

[Kin78a] J. F. C. Kingman. "Random Partitions in Population Genetics". In: *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 361.1704 (1978), pp. 1–20.

[Kin78b] J. F. C. Kingman. "The Representation of Partition Structures". In: *Journal of the London Mathematical Society* s2-18.2 (1978), pp. 374–380.

[Kin93] J. F. C. Kingman. *Poisson Processes*. Vol. 3. Oxford Studies in Probability. New York: Oxford University Press, 1993.

[LFT15] María Lomelí, Stefano Favaro, and Yee Whye Teh. "A hybrid sampler for Poisson-Kingman mixture models". In: *Advances in Neural Information Processing Systems 28*. Ed. by C. Cortes et al. Curran Associates, Inc., 2015, pp. 2161–2169.

[Luk55] Eugene Lukacs. "A Characterization of the Gamma Distribution". In: *Ann. Math. Statist.* 26.2 (June 1955), pp. 319–324.

[McC65] J. W. McCloskey. "A Model for the Distribution of Individuals by Species in an Environment". PhD thesis. 1965.

[OR15] Peter Orbanz and Daniel M. Roy. "Bayesian Models of Graphs, Arrays and Other Exchangeable Random Structures". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37.2 (Feb. 2015), pp. 437–461.

[Pit03] Jim Pitman. "Poisson-Kingman partitions". In: *Statistics and science: a Festschrift for Terry Speed*. Ed. by Darlene R. Goldstein. Vol. Volume 40. Lecture Notes–Monograph Series. Beachwood, OH: Institute of Mathematical Statistics, 2003, pp. 1–34.

[Pit06] Jim Pitman. *Combinatorial Stochastic Processes*. Vol. 1875. Ecole d'Eté de Probabilités de Saint-Flour XXXII. Springer-Verlag Berlin Heidelberg, 2006.

[Pit96a] Jim Pitman. “Random discrete distributions invariant under size-biased permutation”. In: *Adv. in Appl. Probab.* 28.2 (1996), pp. 525–539.

[Pit96b] Jim Pitman. “Some developments of the Blackwell-MacQueen urn scheme”. In: *Statistics, probability and game theory*. Ed. by T. S. Ferguson, L. S. Shapley, and J. B. MacQueen. Vol. Volume 30. Lecture Notes–Monograph Series. Hayward, CA: Institute of Mathematical Statistics, 1996, pp. 245–267.

[PY97] Jim Pitman and Marc Yor. “The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator”. In: *Ann. Probab.* 25.2 (Apr. 1997), pp. 855–900.

[Set94] J. Sethuraman. “A constructive definition of Dirichlet priors”. In: *Statist. Sinica* 4 (1994), pp. 639–650.

[Teh+06] Yee Whye Teh et al. “Hierarchical Dirichlet Processes”. In: *Journal of the American Statistical Association* 101.476 (2006), pp. 1566–1581.

[WM97] Stephen G Walker and Pietro Muliere. “Beta-Stacy processes and a generalization of the Pólya-urn scheme”. In: *Ann. Statist.* 25.4 (Aug. 1997), pp. 1762–1780.