YUGUANG F. IPSEN AND ROSS A. MALLER

NEGATIVE BINOMIAL CONSTRUCTION OF RANDOM DISCRETE DISTRIBUTIONS ON THE INFINITE SIMPLEX

The Poisson-Kingman distributions, $\mathrm{PK}(\rho)$, on the infinite simplex, can be constructed from a Poisson point process having intensity density ρ or by taking the ranked jumps up till a specified time of a subordinator with Lévy density ρ, as proportions of the subordinator. As a natural extension, we replace the Poisson point process with a negative binomial point process having parameter $r > 0$ and Lévy density ρ, thereby defining a new class $\mathrm{PK}^{(r)}(\rho)$ of distributions on the infinite simplex. The new class contains the two-parameter generalisation $\mathrm{PD}(\alpha, \theta)$ of [13] when $\theta > 0$. It also contains a class of distributions derived from the trimmed stable subordinator. We derive properties of the new distributions, with particular reference to the two most well-known PK distributions: the Poisson-Dirichlet distribution $\mathrm{PK}(\rho_\theta)$ generated by a Gamma process with Lévy density $\rho_\theta(x) = \theta e^{-x}/x$, $x > 0$, $\theta > 0$, and the random discrete distribution, $\mathrm{PD}(\alpha, 0)$, derived from an α-stable subordinator.

1. INTRODUCTION

The random discrete distributions on the infinite simplex constructed from Poisson point processes possess both elegant theoretical properties and wide applicability to many real-world problems. The construction is as follows. Let $\mathbb{X} = \sum_{i=1}^{\infty} \delta_{\Delta_{(i)}}$, where δ_x denotes a point mass at $x \in \mathbb{R}$, be a Poisson point process on $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ with ordered points $\Delta_{(1)} \geq \Delta_{(2)} \geq \cdots$ and intensity measure $\Pi(\cdot)$, satisfying

$$(1.1) \qquad \Pi\{(0, \infty)\} = \infty, \ \Pi\{(x, \infty)\} < \infty \text{ for each } x > 0, \text{ and } \int_0^1 x \, \Pi(\mathrm{d}x) < \infty.$$

Denote the sum of points in $\mathbb{X}$ by $T(\mathbb{X}) = \sum_i \Delta_{(i)}$. Then condition (1.1) ensures that $\mathbb{P}(0 < T(\mathbb{X}) < \infty) = 1$. Assume that Π admits a density ρ, so that $\Pi(\mathrm{d}x) = \rho(x)\mathrm{d}x$ for any $x > 0$. Define $\mathbf{V}$ to be the ordered jumps normalised by their sum:

$$(1.2) \qquad \mathbf{V} = (V_1, V_2, \ldots) = \left(\frac{\Delta_{(1)}}{T(\mathbb{X})}, \frac{\Delta_{(2)}}{T(\mathbb{X})}, \cdots \right).$$

Then $\mathbf{V}$ follows a Poisson-Kingman distribution with density ρ, denoted by $\mathrm{PK}(\rho)$, following the terminology and notation in [12]. The two most well-known Poisson-Kingman distributions are the Poisson-Dirichlet distribution, $\mathrm{PK}(\rho_\theta)$, where $\rho_\theta(x) = \theta e^{-x}/x$, $x > 0$, $\theta > 0$, introduced in [9], and the random discrete distribution derived from an α-stable subordinator, $\mathrm{PK}(\rho_\alpha)$, with $\rho_\alpha(x) = C\alpha x^{-\alpha-1}$, $x > 0$, for some $C > 0$ and $0 < \alpha < 1$. [1] This distribution is also known as $\mathrm{PD}(\alpha, 0)$ in [13], in which a two-parameter generalisation $\mathrm{PD}(\alpha, \theta)$ is constructed. These distributions have applications ranging from the modelling of gene frequencies in population genetics through to models

---

[1]We will always use the Greek letters θ and α, and only these, to parametrise the ρ_θ and ρ_α as just defined, so there should be no confusion between them.

for prior distributions in Bayesian nonparametric statistics as well as, recently, in the machine learning community; for example, [2, 3, 7, 10].

The construction of $\mathrm{PK}(\rho)$ from a Poisson point process is equivalent to that from a subordinator. Let $X = (X_t)_{t \geq 0}$ be a driftless subordinator with Lévy density ρ. Write the jump process of X as $(\Delta X_t := X_t - X_{t-})_{t>0}$, and its jumps up till time $t > 0$ arranged in decreasing order as $\Delta X_t^{(1)} \geq \Delta X_t^{(2)} \geq \cdots$. Then by the definition in (1.2), the vector

$$(1.3) \qquad \left(\frac{\Delta X_t^{(1)}}{X_t}, \frac{\Delta X_t^{(2)}}{X_t}, \cdots \right) \sim \mathrm{PK}(t\rho),$$

where "$\sim$" stands for "has the distribution of".

As a natural extension of (1.3), one can consider the subordinator X after omitting a fixed number, $r \in \mathbb{N}$, of its largest jumps, and forming the analogous normalised vector of ratios. The corresponding r-trimmed subordinator is $^{(r)}X_t := X_t - \sum_{i=1}^r \Delta X_t^{(i)}$. When X is an α-stable subordinator, we denote the generalised class of distributions by $\mathrm{PD}_\alpha^{(r)}$, so that

$$(1.4) \qquad \left(\frac{\Delta X_1^{(r+1)}}{^{(r)}X_1}, \frac{\Delta X_1^{(r+2)}}{^{(r)}X_1}, \cdots \right) \sim \mathrm{PD}_\alpha^{(r)}.$$

There is a close connection between the trimmed stable subordinator and the negative binomial point process of [5] which we develop in Section 5. The "trimming" concept, of removing the r largest points, is very natural in this context, but having defined the new distributions, there is no need to keep r as an integer, and any value $r > 0$ can be allowed (as we do herein).

Returning to the general situation, this suggests replacing the Poisson construction with a negative binomial construction. Alternatively to the trimming rationale mentioned in the previous paragraph, this generalisation is in the same spirit as replacing a Poisson by a negative binomial distribution in the statistical analysis of discrete data, to cater for overdispersion in the data. In any event, this procedure produces a new class of random discrete distributions, parametrised by an extra parameter $r > 0$, which includes the important special case $\mathrm{PD}(\alpha, \theta)$ for $\theta > 0$, hence, in particular, the classes $\mathrm{PD}(\alpha, 0)$ and $\mathrm{PD}(0, \theta)$, as well as $\mathrm{PD}_\alpha^{(r)}$. We go on to study stick-breaking representations and other properties of the new class.

The paper is organised as follows. Section 2 sets up the negative binomial construction and defines the new Poisson-Kingman class, denoted by $\mathrm{PK}^{(r)}(\rho)$. We note the connection of this class to subordinated Lévy processes as well as to the $\mathrm{PD}(\alpha, \theta)$ distribution when $\theta > 0$. From previous Poisson formulae, we then derive the joint density of the size-biased random permutation in Section 3 and note its interesting consequences. In Section 4, we give the stick-breaking representation for important cases of the new class (see Theorem 4.1) and make explicit how it differs from the original α-stable case. Section 5 defines a special case $\mathrm{PD}_\alpha^{(r)}$ of the new family that arises naturally from trimmed stable subordinators as discussed above. The related stick-breaking property is investigated in Theorem 5.1 and a characterisation as a shifted sequence from $\mathrm{PD}(\alpha, 0)$ is obtained in Lemma 5.1. Finally, the paper ends with a brief discussion of potential applications of the new class in Section 6.

2. The Negative Binomial Construction

Throughout, we assume a background probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is given and all random variables are measurable mappings from $(\Omega, \mathcal{A}, \mathbb{P})$ to some appropriate space. We follow the exposition in [14, Ch.3] for the point process setup. Denote the space of integer-valued Radon measures on $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ endowed with its Borel σ-algebra by

$(\mathbb{M}, \mathcal{M})$, and let $\mathcal{F}^+$ be the set of nonnegative measurable functions on $(0, \infty)$. We use the abbreviation PPP(Π) (or PPP(ρ)) throughout to denote a Poisson point process with intensity measure Π (or density ρ).

Given a measure Π on $\mathcal{B}(\mathbb{R}_+)$, locally finite at infinity, and a parameter $r > 0$, a negative binomial point process (NBPP) on $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ generated by Π and r, as defined in [5], is a measurable mapping $\mathbb{B}^{(r)}$ from $(\Omega, \mathcal{A}, \mathbb{P})$ to $(\mathbb{M}, \mathcal{M})$ characterised by its Laplace functional

$$(2.1) \qquad \Phi(f) = \mathbb{E}\Big(e^{-\mathbb{B}^{(r)}(f)}\Big) = \Big(1 + \int_{\mathbb{R}_+} (1 - e^{-f(x)})\Pi(\mathrm{d}x)\Big)^{-r}, \ f \in \mathcal{F}^+.$$

Denote the distribution of such a $\mathbb{B}^{(r)}$ by $\mathcal{BN}(r, \Pi)$ or $\mathcal{BN}(r, \rho)$ if Π admits a density ρ. From (2.1), $\mathbb{B}^{(r)}$ with distribution $\mathcal{BN}(r, \Pi)$ can be regarded as a Poisson point process with randomised intensity measure $\Gamma_r\Pi$, where Γ_r is an independent Gamma(r,1) random variable.

The "negative binomial" terminology arises as follows. Let $\mathbb{B}^{(r)}$ be a point process on $\mathbb{R}_+$ distributed as $\mathcal{BN}(r, \Pi)$ for some $r \in \mathbb{N}$, and let $B_1, \ldots, B_n$ be a sequence of pairwise disjoint bounded Borel sets on $\mathbb{R}_+$. Then the numbers of points of $\mathcal{BN}(r, \Pi)$ in $B_1, \ldots, B_n$ follow an n-variate negative binomial distribution NB$(r, q_0, q_1, \ldots, q_n)$, with

$$q_0 = \frac{1}{1 + \Pi\{\bigcup_{i=1}^n B_i\}} \quad \text{and} \quad q_i = \frac{\Pi(B_i)}{1 + \Pi\{\bigcup_{i=1}^n B_i\}} \ \text{ for } i = 1, \ldots, n.$$

Write $\mathbb{B}^{(r)}$ in the form

$$\mathbb{B}^{(r)} = \sum_{i \in \mathbb{N}} \delta_{J_{(i)}},$$

where $J_{(1)} \geq J_{(2)} \geq \cdots$ are the ordered points in $\mathbb{B}^{(r)}$, and denote by $T(\mathbb{B}^{(r)}) := \sum_{i \in \mathbb{N}} J_{(i)}$ the sum of the points in $\mathbb{B}^{(r)}$. Since Π satisfies (1.1), $\mathbb{P}(0 < T(\mathbb{B}^{(r)}) < \infty) = 1$ for all $r > 0$. We now define from $\mathcal{BN}(r, \Pi)$ a random discrete distribution on the infinite simplex using a similar procedure as in (1.2).

Definition 2.1 (PK$^{(r)}(\rho)$). *For each $r > 0$ let $\mathbb{B}^{(r)}$ be distributed as $\mathcal{BN}(r, \Pi)$ and suppose Π admits a density ρ. The vector*

$$(2.2) \qquad \mathbf{W} := (W_1, W_2, \ldots,) = \Big(\frac{J_{(1)}}{T(\mathbb{B}^{(r)})}, \frac{J_{(2)}}{T(\mathbb{B}^{(r)})}, \cdots \Big),$$

is said to follow a Poisson-Kingman distribution generated by $\mathcal{BN}(r, \rho)$, which we denote as PK$^{(r)}(\rho)$.

Just as $\mathcal{BN}(r, \rho)$ can be characterised as a PPP($\Gamma_r\rho$), so an equivalent construction of the PK$^{(r)}(\rho)$ sequence can be made from Gamma subordinated Lévy processes. Let $X = (X_t)_{t>0}$ be a driftless subordinator with Lévy density ρ and let $(\sigma_r)_{r>0}$ be an independent gamma subordinator, i.e., a subordinator having Lévy density $\Pi_\sigma(\mathrm{d}z) = e^{-z}z^{-1}\mathrm{d}z$, $z > 0$. Denote the ranked jumps of X up till time $t > 0$ by $\Delta X_t^{(1)} > \Delta X_t^{(2)} > \cdots$. From (1.3) we immediately obtain

$$(2.3) \qquad \Big(\frac{\Delta X_{\sigma_r}^{(1)}}{X_{\sigma_r}}, \frac{\Delta X_{\sigma_r}^{(2)}}{X_{\sigma_r}}, \frac{\Delta X_{\sigma_r}^{(3)}}{X_{\sigma_r}}, \ldots \Big) \sim \mathrm{PK}^{(r)}(\rho).$$

One may generalise this further by considering PPP($\xi\rho$) with any positive random variate ξ replacing Γ_r. See for example [8] for a generalisation of the Indian Buffet process using scaled subordinators. Our emphasis on the negative binomial class is due to its natural derivation from an r-trimmed stable subordinator (see Section 5), but it also turns out that PK$^{(r)}(\rho)$ defined in (2.2) includes the well-known two parameter Poisson-Dirichlet distributions PD(α, θ), when $\theta > 0$.

Class $\mathrm{PD}(\alpha, \theta)$ is defined in [13] through a stick-breaking representation of size-biased permutations of vectors having those distributions. Thus, $\mathbf{V} = (V_i)_{i \geq 1}$ is said to have law $\mathrm{PD}(\alpha, \theta)$ with size-biased permutation $\tilde{\mathbf{V}} = (\tilde{V}_i)_{i \geq 1}$ if for each $n \in \mathbb{N}$

$$\tilde{V}_n = U_n \prod_{i=1}^{n-1} (1 - U_i),$$

where the U_i are independent and distributed as $\mathrm{Beta}(\theta + i\alpha, 1 - \alpha)$. When $\theta > 0$, $\mathrm{PD}(\alpha, \theta)$ has a subordinator representation in terms of the generalised Gamma subordinator. Take $0 < \alpha < 1$ and denote the Lévy density of the generalised Gamma subordinator as

$$\rho_G(x) = \frac{\alpha}{\Gamma(1 - \alpha)} x^{-\alpha - 1} e^{-x}, \quad x > 0.$$

Let $(X_t)_{t \geq 0}$ be a subordinator with Lévy density ρ_G and let $(\sigma_r)_{r > 0}$ be an independent gamma subordinator. Then, by [13, Prop.21],

$$(2.4) \qquad \mathbf{V} = \Big(\frac{\Delta X_T^{(1)}}{X_T}, \frac{\Delta X_T^{(2)}}{X_T}, \dots \Big) \sim \mathrm{PD}(\alpha, \theta) \quad \text{if} \quad T = \sigma_{\theta/\alpha}.$$

Comparing (2.3) with (2.4), we see that for each $\theta > 0$, $\mathrm{PK}^{(\theta/\alpha)}(\rho_G)$ has the same law as $\mathrm{PD}(\alpha, \theta)$. In particular, X_T is independent of $\mathbf{V}$ in (2.4).

Remark 2.1. It is possible to enlarge the current class $\mathrm{PK}^{(r)}(\cdot)$ to include $\mathrm{PD}(\alpha, \theta)$ when $0 > \theta > -\alpha$ by using a 2-variate mixing measure $\eta_2(\cdot, \cdot)$ on $(0, \infty) \times (0, \infty)$ such that

$$\mathrm{PK}(\rho, \eta_2(\cdot, \cdot)) = \int_{s > 0} \int_{t > 0} \mathrm{PK}(s\rho \,|\, t) \, \eta_2(\mathrm{d}s, \mathrm{d}t).$$

Then the mixed class $\mathrm{PK}(\rho, \eta(\cdot))$ defined in Definition 3 of [12] is included in the enlarged class with mixing measure $\eta_2(\mathrm{d}s, \mathrm{d}t) = \delta_1(ds) \, \eta(dt)$ trivially. So we have a generalisation of $\mathrm{PD}(\alpha, \theta)$ for the entire parameter range.

In the following sections, many interesting properties of $\mathrm{PK}^{(r)}(\rho)$ are derived through the corresponding Poisson formulae in [11] and [13].

3. Joint Density of the Size-biased Permutation

The *size-biased permutation* of a sequence $\mathcal{J} := (J_{(i)})_{i \in \mathbb{N}}$, denoted by $(\tilde{J}_i)_{i \in \mathbb{N}}$, is defined as follows. Conditional on $\mathcal{J}$, $\tilde{J}_1$ takes value $J_{(i)}$ with probability $J_{(i)}/T(\mathbb{B}^{(r)})$; for $n \geq 1$, conditional on $\mathcal{J}$ and $\{\tilde{J}_1, \dots, \tilde{J}_n\}$, $\tilde{J}_{n+1}$ takes value $J_{(j)} \in \mathcal{J} \setminus \{\tilde{J}_1, \dots, \tilde{J}_n\}$ with probability $J_{(j)}/\big(T(\mathbb{B}^{(r)}) - \sum_{i=1}^{n} \tilde{J}_i\big)$.

Let the sum of the points remaining after each size-biased pick from $\mathbb{B}^{(r)}$ be the sequence $(^{(r)}T_n)$, $n \geq 0$; thus, $^{(r)}T := {}^{(r)}T_0 := T(\mathbb{B}^{(r)})$, and, for each $n \in \mathbb{N}$,

$$(3.1) \qquad {}^{(r)}T_n := {}^{(r)}T_{n-1} - \tilde{J}_n.$$

For each $r > 0$, denote the density of the random variable $T(\mathbb{B}^{(r)})$ by

$$(3.2) \qquad g_r(t) := \mathbb{P}\big(T(\mathbb{B}^{(r)}) \in \mathrm{d}t\big)/\mathrm{d}t = \mathbb{P}\big({}^{(r)}T \in \mathrm{d}t\big)/\mathrm{d}t, \ t > 0.$$

By (2.1), g_r exists and satisfies, for each $\lambda > 0$,

$$\int_0^\infty e^{-\lambda x} g_r(x) \mathrm{d}x = \Big(1 + \int_0^\infty (1 - e^{-\lambda x}) \Pi(\mathrm{d}x) \Big)^{-r}.$$

Furthermore, we always have

$$(3.3) \qquad g_r(t) = \int_{v > 0} f^v(t) \mathbb{P}(\Gamma_r \in \mathrm{d}v)$$

where f^v is the density of the sum of Poisson points $T(\mathbb{X})$ with Lévy density $v\rho$.

Since $\mathrm{PK}^{(r)}(\rho) = \mathrm{PK}(\Gamma_r \rho)$, we can derive the joint densities of the sequence of remaining sums $(^{(r)}T_i, i \geq 0)$ by applying the Poisson formulae in [11]. (This is also derived from first principles through Palm characterisation in [6].) Write the ascending factorial of base r and order $n \in \mathbb{N}$ as $r^{[n]} = r(r+1)(r+2) \cdots (r+n-1)$ with $r^{[0]} = 1$.

Theorem 3.1. *Let $r > 0$ and $\Theta(x) := x\rho(x)$, $x > 0$.*

(i) *The joint density of $\left(^{(r)}T, {}^{(r)}T_1, {}^{(r)}T_2, \ldots, {}^{(r)}T_n\right)$ with respect to Lebesgue measure is, for $t_0 > t_1 > \cdots > t_n > 0$ and $n \in \mathbb{N}$,*

$$(3.4) \qquad f(t_0, t_1, \ldots, t_n) = r^{[n]} g_{r+n}(t_n) \prod_{i=0}^{n-1} \frac{\Theta(t_i - t_{i+1})}{t_i}.$$

(ii) *$\{^{(r)}T, {}^{(r)}T_1, {}^{(r)}T_2, \ldots\}$ is a non-homogeneous Markov chain with transition density*

$$(3.5) \qquad \mathbb{P}\left(^{(r)}T_{n+1} \in t_1 \,\big|\, {}^{(r)}T_n = t\right) = (r+n) \frac{\Theta(t - t_1)}{t} \frac{g_{r+n+1}(t_1)}{g_{r+n}(t)} \, dt_1,$$

for each $n = 0, 1, 2, \ldots$, $t > 0$ and $0 < t_1 < t$.

Proof of Theorem 3.1. Recall in (3.3) that f^v is the density of the sum of Poisson points with Lévy density $v\rho(x)$. By [11, Theorem 2.1], the joint density of the corresponding remaining sum $(T_i, i = 0, \ldots, n)$ is

$$f^v(t_n) \prod_{i=1}^{n} \frac{v^n \Theta(t_{i-1} - t_i)}{t_i}.$$

Randomising v by a Gamma$(r, 1)$ distribution, we see that $\left(^{(r)}T_i, i = 0, \ldots, n\right)$ has joint density

$$\int_{v>0} f^v(t_n) v^n \mathbb{P}(\Gamma_r \in dv) \prod_{i=1}^{n} \frac{\Theta(t_{i-1} - t_i)}{t_i}.$$

Here we note by (3.3) that

$$\int_{v>0} f^v(t_n) v^n \mathbb{P}(\Gamma_r \in dv) = \frac{\Gamma(r+n)}{\Gamma(r)} \int_{v>0} f^v(t_n) \mathbb{P}(\Gamma_{r+n} \in dv) = r^{[n]} g_{r+n}(t_n).$$

This proves (3.4). Part (ii) follows immediately from Part (i) as

$$\mathbb{P}\left(^{(r)}T_{n+1} \in dt_{n+1} \,\big|\, {}^{(r)}T = t_0, {}^{(r)}T_1 = t_1, {}^{(r)}T_2 = t_2, \ldots, {}^{(r)}T_n = t_n\right)$$
$$= (r+n) \frac{\Theta(t_n - t_{n+1})}{t_n} \frac{g_{r+n+1}(t_{n+1})}{g_{r+n}(t_n)} dt_{n+1},$$

which does not depend on $t_0, t_1, \ldots, t_{n-1}$. Thus (3.5) is established. $\qquad \square$

We next write down some interesting consequences of Theorem 3.1. Suppose $\mathbf{W} = (W_i)_{i \in \mathbb{N}}$ is defined as in (2.2) with size-biased permutation of (W_i) as $(\tilde{W}_i)$. Let the distribution of $\mathbf{W}$ conditional on $\{T(\mathbb{B}^{(r)}) = t\}$ be $\mathrm{PK}^{(r)}(\rho \,|\, t)$.

Corollary 3.1. *For each $r > 0$ and $t > 0$, the following statements hold.*

(i) *The joint density of $(\tilde{J}_1, {}^{(r)}T)$ is*

$$(3.6) \qquad \mathbb{P}\left(\tilde{J}_1 \in dv, {}^{(r)}T \in dt\right) = r\frac{v}{t} \rho(v) \, g_{r+1}(t-v) \, dv \, dt, \ t > 0, \ 0 < v < t.$$

(ii) *g_r satisfies the following integral recursion equation:*

$$(3.7) \qquad g_r(t) = r \int_0^t \rho(v) g_{r+1}(t-v) \frac{v}{t} dv, \quad t > 0.$$

(iii) *The first size-biased pick from* $\mathrm{PK}^{(r)}(\rho \,|\, t)$ *has density*

$$(3.8) \qquad \tilde{f}_r(w \,|\, t) = \mathbb{P}\big(\tilde{W}_1 \in \mathrm{d}w \,\big|\, {}^{(r)}T = t\,\big)/\mathrm{d}w = r\, tw\, \rho(tw)\, \frac{g_{r+1}(t\bar{w})}{g_r(t)},$$

where $\bar{w} = 1 - w$ *and* $w \in (0,1)$.

(iv) *Let* $G_{i+1} = \tilde{J}_{i+1}/{}^{(r)}T_i$. *For each* $i = 0, 1, \ldots$, $w \in (0,1)$,

$$(3.9) \quad \mathbb{P}\big(G_{i+1} \in \mathrm{d}w \,\big|\, {}^{(r)}T_0, \ldots, {}^{(r)}T_i = t\,\big) = \mathbb{P}\big(G_{i+1} \in \mathrm{d}w \,\big|\, {}^{(r)}T_i = t\,\big) = \tilde{f}_{r+i}(w \,|\, t)\mathrm{d}w.$$

Proof of Corollary 3.1. (i) The lefthand side of (3.6) equals $\mathbb{P}\big({}^{(r)}T \in \mathrm{d}t, {}^{(r)}T_1 \in \mathrm{d}(t-v)\big)$. Apply the density formulae in (3.4) with $f(t, t-v)$ to get the righthand side of (3.6).

(ii) Integrate (3.6) with respect to $\mathrm{d}v$ to obtain (3.7).

(iii) Noting that $\mathbb{P}({}^{(r)}T \in \mathrm{d}t) = g_r(t)\mathrm{d}t$, (3.8) follows from (3.6) by a change of variable.

(iv) It can be read from (3.5) that

$$\mathbb{P}\big(G_{i+1} \in \mathrm{d}w \,\big|\, {}^{(r)}T_i = t\,\big) = (r+i)\Theta(tw)\frac{g_{r+i+1}(t\bar{w})}{g_{r+i}(t)}\mathrm{d}w\,.$$

Comparing this with (3.8), we obtain (3.9). $\qquad\qquad\square$

Comparing (3.8) and (3.9), we see that, conditional on $\{{}^{(r)}T_i = t\}$, G_{i+1} has the same density as the first size-biased pick from $\mathrm{PK}^{(r+i)}(\rho \,|\, t)$. This means that the $(i+1)$st size-biased pick from the remaining point process $\mathbb{B}_i^{(r)} := \mathbb{B}^{(r)} - \sum_{j=1}^{i} \delta_{\tilde{J}_j}$ has the same distribution as the first size-biased pick from an independent point process $\mathbb{B}^{(r+i)}$ after conditioning on their sums. This gives a characterisation of the sequence obtained by removing the first k size-biased jumps and then renormalising it as

$$\tilde{\mathbf{W}}_k := \left(\frac{\tilde{J}_{k+1}}{{}^{(r)}T_k}, \frac{\tilde{J}_{k+2}}{{}^{(r)}T_k}, \frac{\tilde{J}_{k+3}}{{}^{(r)}T_k}, \ldots \right),$$

as stated in the next corollary.

Corollary 3.2. *For each* $t > 0$, $r > 0$ *and* $k \in \mathbb{N}$, *we have, for* $w_i \in (0,1)$, $i \in \mathbb{N}$,

$$(3.10) \quad \mathbb{P}\left(\frac{\tilde{J}_{k+i}}{{}^{(r)}T_k} \in \mathrm{d}w_i, i \geq 1 \,\bigg|\, {}^{(r)}T_k = t\right) = \mathbb{P}\left(\frac{\tilde{J}_i}{{}^{(r+k)}T_0} \in \mathrm{d}w_i, i \geq 1 \,\bigg|\, {}^{(r+k)}T_0 = t\right).$$

Proof of Corollary 3.2. We compute the finite dimensional distribution for two terms. The general case is similar. Fix $t > 0$, $r > 0$ and $k \in \mathbb{N}$ and recall (3.9). Then the lefhand side of (3.10) is

$$\mathbb{P}\left(\frac{\tilde{J}_{k+1}}{{}^{(r)}T_k} \in \mathrm{d}w_1, \frac{\tilde{J}_{k+2}}{{}^{(r)}T_k} \in \mathrm{d}w_2 \,\bigg|\, {}^{(r)}T_k = t\right)$$

$$= \mathbb{P}\left(\frac{\tilde{J}_{k+2}}{{}^{(r)}T_k} \in \mathrm{d}w_2 \,\bigg|\, {}^{(r)}T_k = t, \tilde{J}_{k+1} = tw_1\right)\mathbb{P}\left(\frac{\tilde{J}_{k+1}}{{}^{(r)}T_k} \in \mathrm{d}w_1 \,\bigg|\, {}^{(r)}T_k = t\right)$$

$$= \mathbb{P}\left(\frac{\tilde{J}_{k+2}}{{}^{(r)}T_{k+1}} \in \frac{\mathrm{d}w_2}{\bar{w}_1} \,\bigg|\, {}^{(r)}T_{k+1} = t\bar{w}_1, {}^{(r)}T_k = t\right)\tilde{f}_{r+k}(w_1 \,|\, t)\,\mathrm{d}w_1$$

$$= \tilde{f}_{r+k+1}\left(\frac{w_2}{\bar{w}_1} \,\bigg|\, t\bar{w}_1\right)\tilde{f}_{r+k}(w_1 \,|\, t)\,\mathrm{d}w_1\,\mathrm{d}w_2\,.$$

Here $\bar{w}_1 := 1 - w_1$. We can also compute the corresponding finite dimensional distribution from the righthand side of (3.10) as

$$
\mathbb{P}\Big(\frac{\tilde{J}_1}{(r+k)T} \in \mathrm{d}w_1,\ \frac{\tilde{J}_2}{(r+k)T} \in \mathrm{d}w_2 \,\Big|\, {}^{(r+k)}T = t\Big)
$$

$$
= \ \mathbb{P}\Big(\frac{\tilde{J}_2}{(r+k)T} \in \mathrm{d}w_2 \,\Big|\, {}^{(r+k)}T = t, \tilde{J}_1 = tw_1\Big) \mathbb{P}\Big(\frac{\tilde{J}_1}{(r+k)T} \in \mathrm{d}w_1 \,\Big|\, {}^{(r+k)}T = t\Big)
$$

$$
= \ \mathbb{P}\Big(\frac{\tilde{J}_2}{(r+k)T_1} \in \frac{\mathrm{d}w_2}{\bar{w}_1} \,\Big|\, {}^{(r+k)}T_1 = t\bar{w}_1,\ {}^{(r+k)}T = t\Big) \tilde{f}_{r+k}(w_1 \,|\, t)\, \mathrm{d}w_1
$$

$$
= \ \tilde{f}_{r+k+1}\Big(\frac{w_2}{\bar{w}_1} \,\Big|\, t\bar{w}_1\Big) \tilde{f}_{r+k}(w_1 \,|\, t)\, \mathrm{d}w_1\, \mathrm{d}w_2.
$$

Comparing these proves (3.10) for two terms, and analogously for n terms, hence for an infinite number of terms, which may be rearranged into decreasing order. This gives (3.10) as stated. $\hfill\square$

4. Stick-Breaking Representations

Let $\mathbf{V}$ be defined as in (1.2) with size-biased permutation $\tilde{\mathbf{V}} = (\tilde{V}_1, \tilde{V}_2, \ldots)$. Remarkable stick-breaking properties exist for $\tilde{\mathbf{V}}$ when $\mathbf{V}$ is distributed as $\mathrm{PK}(\rho_\theta)$ or $\mathrm{PK}(\rho_\alpha)$. Recall that for each $x > 0$,

$$
\rho_\theta(x) = \theta e^{-x}/x,\ \theta > 0,\ \text{and}\ \rho_\alpha(x) = C\alpha x^{-\alpha-1},\ 0 < \alpha < 1,\ C > 0.
$$

By [11, Thm.1.2], the nth term in $\tilde{\mathbf{V}}$ can be written in the product form

$$
\tilde{V}_n = (1 - U_n) \prod_{i=1}^{n-1} U_i, \quad n \in \mathbb{N}
$$

(with $\prod_{i=1}^{0} \equiv 1$). When $\mathbf{V}$ follows a $\mathrm{PK}(\rho_\theta)$ distribution, the (U_i) are i.i.d. $\mathrm{Beta}(\theta, 1)$ variables. When $\mathbf{V}$ is distributed as $\mathrm{PK}(\rho_\alpha)$, the (U_i) are independent $\mathrm{Beta}(i\alpha, 1 - \alpha)$ variables.

To derive the corresponding stick-breaking representations for vectors having distributions in $\mathrm{PK}^{(r)}(\rho_\theta)$ and $\mathrm{PK}^{(r)}(\rho_\alpha)$, recall the sequence of sums remaining after successive size-biased picks, $({}^{(r)}T_i)$, $i \in \mathbb{N}$, defined in (3.1), and denote the successive *residual fractions* by

$$
{}^{(r)}U_i := \frac{{}^{(r)}T_i}{{}^{(r)}T_{i-1}},\ i \in \mathbb{N}.
$$

Then the nth term in $\tilde{\mathbf{W}}$ can be represented as

$$
(4.1) \qquad\qquad \tilde{W}_n = (1 - {}^{(r)}U_n) \prod_{i=1}^{n-1} {}^{(r)}U_i,\ n \in \mathbb{N}.
$$

Theorem 4.1 (Stick-Breaking). *Fix $r > 0$.*

(i) Suppose $\mathbf{W}$ is distributed as $\mathrm{PK}^{(r)}(\rho_\theta)$. Then the stick-breaking factors in (4.1) are distributed as

$$
(4.2) \qquad\qquad \big({}^{(r)}U_i\big)_{i \in \mathbb{N}} \overset{\mathrm{D}}{=} \big(B_i(\Gamma_r\theta, 1)\big)_{i \in \mathbb{N}},
$$

where, for each $v > 0$, $\big(B_i(v\theta, 1)\big)_{i \in \mathbb{N}}$ are i.i.d. $\mathrm{Beta}(v\theta, 1)$ variables, and Γ_r is a $\mathrm{Gamma}(r, 1)$ random variable, independent of the (B_i).

(ii) Suppose $\mathbf{W}$ is distributed as $\mathrm{PK}^{(r)}(\rho_\alpha)$. Then the corresponding $({}^{(r)}U_i)$ in (4.1) are distributed as $\mathrm{Beta}(i\alpha, 1 - \alpha)$, $i \geq 1$, not depending on r.

Proof of Theorem 4.1. We make use of the fact that $\mathrm{PK}^{(r)}(\rho)$ has the same law as $\mathrm{PK}(\Gamma_r\rho)$. Recall from [11, Theorem 2.1] that if $\mathbf{V}$ has law $\mathrm{PK}(v\rho_\theta)$, then the corresponding residual fractions (U_i) are independent $\mathrm{Beta}(v\theta, 1)$ variables. Randomising v by an independent $\mathrm{Gamma}(r, 1)$ random variable, we obtain (4.2).

Next suppose $\mathbf{V}$ has law $\mathrm{PK}(v\rho_\alpha)$. Then the corresponding residual fractions (U_i) are independent $\mathrm{Beta}(i\alpha, 1 - \alpha)$ variables, whose distribution does not depend on v. Thus, randomising v does not change the distribution. $\square$

By Theorem 4.1, $\mathrm{PK}^{(r)}(\rho_\alpha)$ comprises the same laws as $\mathrm{PK}(\rho_\alpha)$ for all values of $r > 0$. Thus, as a characteristic of $\mathrm{PK}(\rho_\alpha)$,

$$^{(r)}T_i \quad \text{is independent of} \quad \left(^{(r)}U_1, \ldots, {}^{(r)}U_i\right), \quad i \in \mathbb{N}.$$

However the sequence $\left(^{(r)}T, {}^{(r)}T_1, \ldots\right)$ of remaining sums has different dynamics, as will be elucidated next. Let $\rho = \rho_\alpha$ for the rest of this section. Recall the definition of g_r in (3.2).

Proposition 4.1. *Fix $r > 0$ and $\rho = \rho_\alpha$. Then*

$$(4.3) \quad \mathbb{P}\left(^{(r)}T_n \in \mathrm{d}t\right) = L_n t^{-n\alpha} g_{r+n}(t)\mathrm{d}t, \text{ where } L_n = r^{[n]}(C\Gamma(1 - \alpha))^n \frac{\Gamma(n\alpha + 1)}{\Gamma(n + 1)},$$

and thus

$$(4.4) \quad \frac{\mathbb{P}\left(^{(r)}T_n \in \mathrm{d}t\right)}{\mathbb{P}\left(^{(r+n)}T \in \mathrm{d}t\right)} = L_n t^{-n\alpha}.$$

Proof of Proposition 4.1. Fix $r > 0$. For each $n \in \mathbb{N}$, we first derive the joint density of $\left(^{(r)}T_n, {}^{(r)}U_1, {}^{(r)}U_2, \ldots, {}^{(r)}U_n\right)$ by change of variables using the joint density of $\left(^{(r)}T_0, {}^{(r)}T_1, {}^{(r)}T_2, \ldots, {}^{(r)}T_n\right)$ in (3.4) with $\rho = \rho_\alpha$. For simplicity, we only consider the case $n = 2$. An analogous derivation holds for $n > 2$. For $t_2 > 0$, $0 < u_i < 1$, $1 \leq i \leq 2$, the joint density of $\left(^{(r)}T_2, {}^{(r)}U_1, {}^{(r)}U_2\right)$ is

$$h(t_2, u_1, u_2) = f\left(\frac{t_2}{u_1 u_2}, \frac{t_2}{u_2}, t_2\right) t_2^2 u_1^{-2} u_2^{-3};$$

where f is defined in (3.4) and $t_2^2 u_1^{-2} u_2^{-3}$ is the Jacobian from the change of variables. Expanding the expression in (3.4) with $\Theta(x) = C\alpha x^{-\alpha}$, we get $h(t_2, u_1, u_2)$ equal to

$$r^{[2]} g_{r+2}(t_2) u_1^{-1} u_2^{-1} \Theta\left(\frac{t_2}{u_1 u_2}\bar{u}_1\right)\Theta\left(\frac{t_2}{u_2}\bar{u}_2\right)$$

$$= r^{[2]} g_{r+2}(t_2)(C\alpha)^2 t_2^{-2\alpha}\left(u_2^{2\alpha-1}\bar{u}_2^{-\alpha}\right)\left(u_1^{\alpha-1}\bar{u}_1^{-\alpha}\right)$$

$$(4.5) \quad = \frac{r^{[2]}}{K_2} g_{r+2}(t_2) t_2^{-2\alpha} \cdot \left[\frac{\Gamma(1 + \alpha)}{\Gamma(2\alpha)\Gamma(1 - \alpha)} u_2^{2\alpha-1}\bar{u}_2^{-\alpha}\right]\left[\frac{\Gamma(1)}{\Gamma(\alpha)\Gamma(1 - \alpha)} u_1^{\alpha-1}\bar{u}_1^{-\alpha}\right]$$

where

$$K_2 = \frac{\prod_{i=0}^{1}\Gamma(1 + i\alpha)}{(C\alpha)^2\Gamma^2(1 - \alpha)\prod_{i=1}^{2}\Gamma(i\alpha)}.$$

Integrate (4.5) with respect to u_1, u_2 to get (4.3). Equality (4.4) is immediate from (4.3). $\square$

The next corollary gives a moment formula for $^{(r+n)}T$, $n \in \mathbb{N}$.

Corollary 4.1. *Fix $r > 0$ and keep $\rho = \rho_\alpha$. For each $n \in \mathbb{N}$,*

$$(4.6) \quad L_n^{-1} = \mathbb{E}\left(^{(r+n)}T^{-n\alpha}\right) = \frac{\Gamma(n + 1)}{\Gamma(n\alpha + 1)}\frac{1}{(C\Gamma(1 - \alpha))^n r^{[n]}} = \frac{1}{r^{[n]}}\mathbb{E}\left(T^{-n\alpha}\right).$$

Proof of Corollary 4.1. For each $n \in \mathbb{N}$, integrate (4.3) to get

$$1 = \int_0^\infty L_n \, g_{r+n}(t_n) \, t_n^{-n\alpha} \, \mathrm{d}t_n = L_n \, \mathbb{E}\big((r+n)T^{-n\alpha}\big),$$

with $L_n = r^{[n]} \, (C\Gamma(1-\alpha))^n \, \Gamma(n\alpha+1)/\Gamma(n+1)$. This gives the first and second equalities in (4.6). From [11, Eq. (2.n)],

$$\mathbb{E}(T^{-n\alpha}) = \frac{\Gamma(n+1)}{\Gamma(n\alpha+1)(C\Gamma(1-\alpha))^n}.$$

Comparing this with the value for L_n^{-1}, we get the last equality in (4.6). $\square$

5. $\mathrm{PD}_\alpha^{(r)}$ Arising from a Trimmed Stable Process

Recall the distribution $\mathrm{PD}_\alpha^{(r)}$ derived from the trimmed α-stable subordinator in (1.4). Let $\Delta_1 > \Delta_2 > \ldots$ be the ordered jumps of an α-stable subordinator $(S_t, 0 < t < 1)$. Then $\sum_i \delta_{\Delta_i}$ forms a Poisson process with intensity measure $\Lambda(\mathrm{d}x) := \rho_\alpha(x)\mathrm{d}x$. For $r \in \mathbb{N}$, denoting the r-trimmed process up till time 1 by $^{(r)}S_1 = S_1 - \sum_{i=1}^r \Delta_i$, we have

$$(5.1) \qquad (V_n^{(r)}, n \geq 1) := \left(\frac{\Delta_{r+1}}{^{(r)}S_1}, \frac{\Delta_{r+2}}{^{(r)}S_1}, \frac{\Delta_{r+3}}{^{(r)}S_1}, \ldots \right) \sim \mathrm{PD}_\alpha^{(r)}.$$

In this section we derive some analogous properties for $\mathrm{PD}_\alpha^{(r)}$. First note that $\mathrm{PD}_\alpha^{(r)}$ has the same law as $\mathrm{PK}^{(r)}(\rho_\alpha^*)$, for $\rho_\alpha^*(x) := \alpha x^{-\alpha-1}\mathbf{1}_{0<x<1}$. To see this, we know from [13, Lemma 24], that, conditionally on Δ_r, the point process

$$(5.2) \qquad \mathbb{B}^T := \sum_{i \in \mathbb{N}} \delta_{J_{(i)}}, \quad \text{where } J_{(i)} = \Delta_{r+i}/\Delta_r,$$

is a Poisson process with intensity measure $(\Delta_r)^{-\alpha}\Lambda(\mathrm{d}x)\mathbf{1}_{\{0<x<1\}}$. Since $(\Delta_r)^{-\alpha} \overset{\mathrm{D}}{=} \Gamma_r/C$, $\mathbb{B}^T$ is a negative binomial point process, $\mathcal{BN}(r, \rho_\alpha^*)$. Thus the vector

$$\left(\frac{J_{(1)}}{^{(r)}S_1/\Delta_r}, \frac{J_{(2)}}{^{(r)}S_1/\Delta_r}, \frac{J_{(3)}}{^{(r)}S_1/\Delta_r}, \ldots \right),$$

which is equal to the lefthand side of (5.1), has distribution $\mathrm{PK}^{(r)}(\rho_\alpha^*)$.

Due to the restriction to $(0,1)$ in ρ_α^*, $\mathrm{PD}_\alpha^{(r)}$ is distinct from $\mathrm{PD}(\alpha, 0)$ (unlike for $\mathrm{PK}^{(r)}(\rho_\alpha)$). The next theorem gives a characterisation of the stick breaking sequences as in (4.1).

Theorem 5.1. *The joint distribution of $^{(r)}T_n$ and $^{(r)}U_1, \, ^{(r)}U_2, \ldots, \, ^{(r)}U_n$ for $\mathrm{PD}_\alpha^{(r)}$ can be written as*

$$(5.3) \qquad \big(^{(r)}T_n, \, ^{(r)}U_1, \, ^{(r)}U_2, \ldots, \, ^{(r)}U_n\big) \overset{\mathrm{D}}{=} \big(Y_{d(U_1,\ldots,U_n)}, U_1, U_2, \ldots, U_n\big),$$

where the (U_i) are independent $Beta(i\alpha, 1 - \alpha)$ rvs,

$$d(u_1, \ldots, u_n) := \min_{1 \leq i \leq n} \prod_{j=i}^n u_j/\bar{u}_i,$$

with $\bar{u}_i := 1 - u_i$, and for each $c > 0$, $Y_c := (^{(r+n)}T)^{-n\alpha}\mathbf{1}_{\{^{(r+n)}T<c\}}$.

Proof of Theorem 5.1. Let $r > 0$ and $\rho = \rho_\alpha^*$. Similar to (4.5), we can derive for $t_n > 0$, $0 < u_i < 1$, $1 \leq i \leq n$, and $n \in \mathbb{N}$, the joint density of $\left({}^{(r)}T_0, {}^{(r)}T_1, {}^{(r)}T_2, \ldots, {}^{(r)}T_n \right)$ as

$$h(t_n, u_1, \ldots, u_n) = \frac{r^{[n]}}{K_n} g_{r+n}^*(t_n) t_n^{-n\alpha}$$

$$\times \prod_{i=1}^{n} \frac{\Gamma(i\alpha + 1 - \alpha)}{\Gamma(i\alpha)\Gamma(1-\alpha)} u_i^{i\alpha - 1} \bar{u}_i^{-\alpha} \mathbf{1}_{\{t_n < \prod_{j=i}^{n} u_j/\bar{u}_i\}}$$

$$(5.4) \qquad = \frac{r^{[n]}}{K_n} t_n^{-n\alpha} g_{r+n}^*(t_n) \mathbf{1}_{\{t_n < d(u_1, \ldots, u_n)\}} \times \prod_{i=1}^{n} \beta_{i\alpha, 1-\alpha}(u_i).$$

Here g_r^* is the corresponding density of the sum of points in $\mathcal{BN}(r, \rho_\alpha^*)$, $\beta_{a,b}$ is the density of a Beta(a,b) distribution, and

$$K_n = \frac{\prod_{i=0}^{n-1} \Gamma(1 + i\alpha)}{\alpha^n \Gamma^n(1-\alpha) \prod_{i=1}^{n} \Gamma(i\alpha)} = \frac{\Gamma(n+1)}{\Gamma^n(1-\alpha)\Gamma(n\alpha + 1)}.$$

The indicator function in (5.4) reflects the restriction of x to the interval $(0,1)$ in ρ_α^*. $\quad \square$

Remark 5.1. (i) By integrating (5.4), we get the identity

$$K_n = r^{[n]} \int_{u_1 = 0}^{1} \cdots \int_{u_n = 0}^{1} \int_{t_n = 0}^{d(u_1, \ldots, u_n)} t_n^{-n\alpha} g_{r+n}^*(t_n) \mathrm{d}t_n$$

$$\times \prod_{i=1}^{n} \beta_{i\alpha, 1-\alpha}(u_i) \mathrm{d}u_1 \cdots \mathrm{d}u_n.$$

(ii) For a stick-breaking representation, as in (4.1), the size-biased permutation of $(V_n^{(r)})$, denoted by $(\tilde{V}_n^{(r)})$, can be written as

$$(5.5) \qquad \tilde{V}_n^{(r)} = (1 - {}^{(r)}U_n) \prod_{i=1}^{n-1} {}^{(r)}U_i.$$

The joint distribution of $({}^{(r)}U_i)_{1 \leq i \leq n}$ can be computed from (5.3), in which we note that $U_1, U_2, \ldots, U_n$ are individually independent, but dependence overall is introduced via the connection with the Y term. In this respect the result is different from the $\mathrm{PD}(\alpha, 0)$ situation, as we would expect, but the distribution of $\tilde{V}_n^{(r)}$ as given by (5.5) is sufficiently explicit to enable computations or simulations.

(iii) Although motivated by the idea of trimming an integer number r of large jumps, our formulae once derived are valid for $r > 0$, and available for modelling purposes in this generality.

(iv) We may set $r = 0$ in (5.1) to have the distribution of $\mathrm{PD}_\alpha^{(r)}$ reduce to that of $\mathrm{PD}(\alpha, 0)$. But we cannot take $r = 0$ in (5.2) with the idea that the size-biased distribution associated with $\mathrm{PD}_\alpha^{(r)}$ might then reduce to the one associated with $\mathrm{PD}(\alpha, 0)$. Note that $\mathbb{B}^T$ is not defined for $r = 0$ (its points Δ_{r+i}/Δ_i are not defined for $r = 0$). Setting $r = 0$ in (5.4), which results from an analysis of $\mathbb{B}^T$, is not permissible.

By restricting r to be an integer, we can further construct a vector $(V_n^{(r)})$ from independent beta random variables and characterise the law of the sequences in $\mathrm{PD}_\alpha^{(r)}$ as a shifted version of $\mathrm{PD}(\alpha, 0)$ which can in turn be characterised by a change of measure formula.

Lemma 5.1. (i) *Let $(R_i, i \geq 1)$ be a sequence of independent $Beta((r+i)\alpha, 1)$ variables. Define*

$$V_n^{(r)} := \frac{\prod_{i=1}^{n-1} R_i}{1 + R_1 + R_1 R_2 + \cdots} = Y_n \prod_{i=1}^{n-1}(1 - Y_i),$$

with stick-breaking factors $Y_n = (1 + \Sigma_n)^{-1}$ *and* $\Sigma_n = R_n + R_n R_{n+1} + \cdots$. *Then* $(V_n^{(r)}, n \geq 1)$ *has law* $\mathrm{PD}_\alpha^{(r)}$.

(ii) *Let* $(V_i^{(r)})$ *be distributed as* $\mathrm{PD}_\alpha^{(r)}$ *and* (V_i) *be distributed as* $\mathrm{PD}(\alpha, 0)$. *Then for any nonnegative measurable function* f, *we have*

$$\mathbb{E}\{f(V_1^{(r)}, V_2^{(r)}, \ldots)\} = \mathbb{E}\Big\{ \frac{\mathfrak{E}_1^r}{r!} f(V_1, V_2, \ldots) \Big\},$$

where $\mathfrak{E}_1 = \lim_{n \to \infty} n V_n^\alpha / V_1^\alpha$, *and the limit holds almost surely and in pth mean for all* $p \geq 1$.

Proof of Lemma 5.1. (i) Recall that $(\Delta_1, \Delta_2, \ldots)$ comprise the points of a Poisson process with intensity measure $\Lambda(\mathrm{d}x) = \rho_\alpha(x)\mathrm{d}x$. Write $R_n = \Delta_{r+n+1}/\Delta_{r+n}$. Then by [13, Prop.8], the sequence of successive ratios $(R_1, R_2, \ldots)$ is of independent $\mathrm{Beta}((r+n)\alpha, 1)$ variables.

Since the sequence

$$\Big(\frac{\Delta_{r+1}}{\sum_{i \geq r+1} \Delta_i}, \frac{\Delta_{r+2}}{\sum_{i \geq r+1} \Delta_i}, \cdots \Big) \sim \mathrm{PD}_\alpha^{(r)},$$

the nth term can be expressed as

$$\frac{\Delta_{r+n}/\Delta_{r+1}}{\Delta_{r+1}/\Delta_{r+1} + \Delta_{r+2}/\Delta_{r+1} + \cdots} = \frac{\prod_{i=1}^{n-1} R_i}{1 + R_1 + R_1 R_2 + \ldots}$$

$$= \frac{\Delta_{r+n}}{\sum_{i \geq n} \Delta_{r+i}} \cdot \frac{\sum_{i \geq n} \Delta_{r+i}}{\sum_{i \geq n-1} \Delta_{r+i}} \cdots \frac{\sum_{i \geq 2} \Delta_{r+i}}{\sum_{i \geq 1} \Delta_{r+i}}$$

$$= Y_n \prod_{i=1}^{n-1} (1 - Y_i)$$

where $Y_n = \Delta_{r+n} / \sum_{i \geq n} \Delta_{r+i} = (1 + \Sigma_n)^{-1}$.

(ii) First consider the homogeneous Poisson point process $\sum_i \delta_{\Gamma_i}$, where $\Gamma_n := \sum_{i=1}^n E_i$ with $(E_i, i \geq 1)$ independent unit exponential random variables. Then for any nonnegative measurable function f, the shifted sequence $(\Gamma_{r+i}, i \geq 1)$ can be characterised by a change of measure from the original process as

$$\mathbb{E}\{f(\Gamma_{r+1}, \Gamma_{r+2}, \ldots)\} = \mathbb{E}\Big\{ \frac{\Gamma_1^r}{r!} f(\Gamma_1, \Gamma_2, \ldots) \Big\}.$$

Since for each $i \geq 1$, $\Delta_i \overset{\mathrm{D}}{=} \overline{\Lambda}^{\leftarrow}(\Gamma_i)$, where $\overline{\Lambda}^{\leftarrow}(x) = C^{1/\alpha} x^{-1/\alpha}$, then for each nonnegative measurable function f, we again have

$$\mathbb{E}\{f(\Delta_{r+1}, \Delta_{r+2}, \ldots)\} = \mathbb{E}\Big\{ \frac{\mathfrak{E}_1^r}{r!} f(\Delta_1, \Delta_2, \ldots) \Big\}$$

where $\mathfrak{E}_1 = \Lambda(\Delta_1, \infty)$. Thus, normalising the jumps, we still get

$$\mathbb{E}\{f(V_1^{(r)}, V_2^{(r)}, \ldots)\} = \mathbb{E}\Big\{ \frac{\mathfrak{E}_1^r}{r!} f(V_1, V_2, \ldots) \Big\}$$

where $\mathfrak{E}_1 := \Lambda(\Delta_1, \infty) = C \Delta_1^{-\alpha}$. To write $C \Delta_1^\alpha$ as a function of $(V_1, V_2, \ldots)$, we note that by [13, Prop. 10], $\lim_{n \to \infty} n V_n^\alpha = C S_1$ almost surely and in pth mean for $p \geq 1$. Thus, we can write

$$\mathfrak{E}_1 = C \Delta_1^{-\alpha} = \frac{C}{V_1^\alpha S_1^\alpha} = \frac{\lim_{n \to \infty} n V_n^\alpha}{V_1^\alpha},$$

concluding the proof of Lemma 5.1. □

6. Discussion: Applications

We mention some possible applications of our results. A common tool in linguistics studies is the "Zipf plot": a plot of log frequencies of words, against their log-ranks. [4] show such plots for some word counts in the Penn Wall St. journal. In their Figure 4, half a dozen or so of the most frequent words appear as outliers, while the rest conform closely to a $\text{PD}(\alpha, 0)$ fit. This suggests that a $\text{PD}_\alpha^{(r)}$ distribution might provide a better fit.

A similar situation occurs in [15], who shows "capital distribution curves" (a log plot of normalized stock capitalizations ranked in descending order, against their log-ranks) for over 20 countries listed on the NASDAQ stock exchange. The curves appear to be very well fitted by a $\text{PD}(\alpha, 0)$ distribution over much of their range, but with a small number of the largest stocks as outliers – as we might expect from this kind of data.

Known difficulties arise in fitting the general 2-parameter $\text{PD}(\alpha, \theta)$ distribution to data; the maximum likelihood estimator of θ is inconsistent ([1, Lemma 5.7]). Introducing the extra parameter r in $\text{PK}^{(r)}(\rho)$ may help to improve estimation of θ, as well as allowing extra flexibility in data description.

In general, we expect that our generalised $\text{PK}^{(r)}(\cdot)$ distribution could be used to extend analyses which are implicitly based on underlying Poisson point processes, to negative binomial point processes, and thereby reveal interesting features of data. More research along these lines would certainly prove profitable.

Acknowledgements

References

1. M.A. Carlton, *Applications of the two-parameter poisson-dirichlet distribution*, Unpublished Ph.D. thesis, Department of Statistics, University of California, Los Angeles, 1999.
2. P. Donnelly, *Partition structures, Polya urns, the Ewens sampling formula, and the ages of alleles*, Theor Popul Biol. **30** (1986), 271–288.
3. W.J. Ewens, *The sampling theory of selectively neutral alleles*, Theoret. Popul. Biol. **3** (1972), 87–112.
4. S. Goldwater, T.L. Griffiths, and M. Johnson, *Producing power-law distributions and damping word frequencies with two-stage language models,*, J. Mach. Learn. Res. **12** (2011), 2335–2382.
5. Gérard Gregoire, *Negative binomial distributions for point processes*, Stochastic Process. Appl. **16** (1984), no. 2, 179–188.
6. Yuguang F. Ipsen and Ross A. Maller, *Generalised Poisson-Dirichlet distributions and the negative binomial point process*, arXiv:1611.09980 (2017).
7. H. Ishwaran and L. F. James, *Gibbs sampling methods for stick-breaking priors*, J. Amer. Statist. Ass. **96** (2001), 161–173.
8. L. F. James, P. Orbanz, and Y. W. Teh, *Scaled subordinators and generalizations of the Indian buffet process*, arXiv:1510.073 (2015).
9. J. F. C. Kingman, *Random discrete distributions*, J. R. Stat. Soc. Series B Stat. Methodol. **37** (1975), no. 1, 1–22.
10. K.L. Lim, W. Buntine, C. Chen, and L. Du, *Nonparametric Bayesian topic modelling with the hierarchical Pitman-Yor processes*, International J. of Approximate Reasoning **78** (2016), 172–191.
11. Mihael Perman, Jim Pitman, and Marc Yor, *Size-biased sampling of Poisson point processes and excursions*, Probab. Theory Related Fields **92** (1992), no. 1, 21–39.
12. Jim Pitman, *Poisson-Kingman partitions*, Lect. Notes Monogr. Ser. **40** (2003), 1–34.
13. Jim Pitman and Marc Yor, *The two-parameter Poisson–Dirichlet distribution derived from a stable subordinator*, Annals of Probability **25** (1997), no. 2, 855–900.
14. Sidney I. Resnick, *Extreme values, regular variation, and point processes*, Springer-Verlag, 1987.
15. Sergey Sosnovskiy, *On financial applications of the two-parameter Poisson-Dirichlet distribution*, arXiv:1501.01954v3 (2015).

Research School of Finance, Actuarial Studies & Statistics, Australian National University, Canberra, Australia

E-mail address: Yuguang.Ipsen@anu.edu.au

E-mail address: Ross.Maller@anu.edu.au