

A GENERALISED DICKMAN DISTRIBUTION AND THE NUMBER OF SPECIES IN A NEGATIVE BINOMIAL PROCESS MODEL

YUGUANG IPSEN* *** AND

ROSS A. MALLER * **** *The Australian National University*

SOUDABEH SHEMEHSAVAR,** *University of Tehran*

Abstract

We derive the large-sample distribution of the number of species in a version of Kingman's Poisson–Dirichlet model constructed from an α -stable subordinator but with an underlying negative binomial process instead of a Poisson process. Thus it depends on parameters $\alpha \in (0, 1)$ from the subordinator and $r > 0$ from the negative binomial process. The large-sample distribution of the number of species is derived as sample size $n \rightarrow \infty$. An important component in the derivation is the introduction of a two-parameter version of the Dickman distribution, generalising the existing one-parameter version. Our analysis adds to the range of Poisson–Dirichlet-related distributions available for modeling purposes.

Keywords: Generalised Poisson–Dirichlet laws; negative binomial point process; number of species; Ewens sampling formula; size-biased sampling; generalised Dickman distribution

2010 Mathematics Subject Classification: Primary 60G51; 60G52; 60G55
Secondary 60G57; 62E20; 62P10

1. Introduction

Kingman in [18] suggested a way of constructing random distributions on the unit simplex by ranking the jumps of a subordinator up to a specified (possibly random) time, then normalising them by the value of the subordinator at that time. Taking the subordinator to be a driftless gamma subordinator generates his well known Poisson–Dirichlet distribution $PD(\theta)$, which was later shown to be intimately connected with the famous Ewens sampling formula in genetics [6]. Another of Kingman's distributions, denoted by PD_α , arises when a driftless α -stable subordinator with parameter $\alpha \in (0, 1)$ is used instead of the gamma subordinator. These distributions and the methodologies associated with them have subsequently had an enormous impact in many areas of application, ranging from the excursion theory of stochastic processes to the theory of random partitions, random graphs and networks, probabilistic number theory, machine learning, Bayesian statistics, and a variety of others.

Received 15 December 2019; revision received 25 August 2020.

* Postal address: Research School of Finance, Actuarial Studies and Statistics, Australian National University, Canberra, ACT, 0200, Australia.

** Postal address: School of Mathematics, Statistics & Computer Sciences, University of Tehran. Email address: Shemehsavar@ut.ac.ir

*** Email address: Yuguang.Ipsen@anu.edu.au

**** Email address: Ross.Maller@anu.edu.au

© The Author(s), 2021. Published by Cambridge University Press on behalf of Applied Probability Trust.

Ipsen and Maller [12] generalised Kingman’s PD_α class in another direction, namely, to the two-parameter $PD_\alpha^{(r)}$ class, defined for $\alpha \in (0, 1)$, and $r > 0$. Like PD_α , the class $PD_\alpha^{(r)}$ is based on an α -stable subordinator, but the extra parameter r arises from a connection with the negative binomial point process introduced by Gregoire [9] (whereas PD_α is associated with a Poisson point process). Ipsen, Maller and Shemehsavar [14] developed connections between various Poisson–Dirichlet models by letting $r \rightarrow \infty$ and $\alpha \downarrow 0$ in $PD_\alpha^{(r)}$, while Ipsen, Shemehsavar and Maller [16] fitted $PD_\alpha^{(r)}$ to gene and species sampling data, demonstrating the utility of allowing the extra parameter r in a data analysis.

An aspect of particular interest in practice is the distribution of $K_n(\alpha, r)$, the number of distinct species observed in a sample of size n from the $PD_\alpha^{(r)}$ distribution. Herein we derive the asymptotic distribution (as $n \rightarrow \infty$) of $K_n(\alpha, r)$ for fixed α and r , and discuss how it depends on the parameters α and r .

The relevance of the Dickman function [5] and the corresponding distribution to the theory of the Poisson–Dirichlet distribution has been observed before (see, e.g., Arratia, Barbour and Tavaré [1] (pp. xi, 14, 76), Watterson [28] (Equation (27) and the material above it), and Watterson and Guess [29] (Equation (3.2.4)), but in our development a generalised version of it plays a particularly significant role. More details are in Section 5.

2. Asymptotic distribution of the number of species

In a sample of size n from $PD_\alpha^{(r)}$, the *blocks count vector* is

$$\mathbf{M}_n = \mathbf{M}_n(\alpha, r) = (M_1(\alpha, r), M_2(\alpha, r), \dots, M_n(\alpha, r)),$$

where $M_j(\alpha, r)$ counts the number of allele types or species having j representatives in the sample. The number of distinct species observed in the sample is $K_n(\alpha, r) := M_1(\alpha, r) + \dots + M_n(\alpha, r)$. In the Poisson–Dirichlet model $PD(\theta)$, the probability of any particular realisation of the blocks count vector is given by the Ewens sampling formula in [6].

For the $PD_\alpha^{(r)}$ model, a formula is obtained in [14] for the distribution of the $(n+1)$ -vector $(\mathbf{M}_n(\alpha, r), K_n(\alpha, r))$, where $\mathbf{M}_n(\alpha, r)$ takes values among all n -vectors of nonnegative integers $\mathbf{m} = (m_1, \dots, m_n)$ satisfying $\sum_{j=1}^n jm_j = n$, while $K_n(\alpha, r)$ takes values $k = \sum_{j=1}^n m_j \in \mathbb{N}_n := \{1, 2, \dots, n\}$. Equation (5.11) of [14] gives the formula

$$\mathbb{P}(\mathbf{M}_n(\alpha, r) = \mathbf{m}, \quad K_n(\alpha, r) = k) = n \int_0^\infty \frac{r^{(k)} \lambda^{\alpha k - 1}}{\Psi(\lambda)^{r+k}} \prod_{j=1}^n \frac{1}{m_j!} (F_j(\lambda))^{m_j} d\lambda, \quad (2.1)$$

where $r^{(k)} = r(r+1) \dots (r+k-1)$,

$$\Psi(\lambda) = 1 + \alpha \int_0^1 (1 - e^{-\lambda z}) z^{-\alpha - 1} dz, \quad (2.2)$$

and

$$F_j(\lambda) := \frac{\alpha}{j!} \int_0^\lambda z^{j-\alpha-1} e^{-z} dz, \quad j \in \mathbb{N}_n, \quad \lambda > 0. \quad (2.3)$$

Both $\mathbf{M}_n(\alpha, r)$ and $K_n(\alpha, r)$ depend on the two parameters $\alpha \in (0, 1)$ and $r > 0$. These are kept fixed in the large-sample analysis (as $n \rightarrow \infty$) which follows. Equation (2.1) can be compared with an analogous formula, Equation (4.14) on page 81 of [23], which is based on a Poisson rather than negative binomial construction.

Defining

$$A_{nk} := \left\{ \mathbf{m} = (m_1, \dots, m_n) : m_j \geq 0, \sum_{j=1}^n jm_j = n, \sum_{j=1}^n m_j = k \right\} \quad (2.4)$$

for $k \in \mathbb{N}_n$, $n \in \mathbb{N}$, and summing over the m_j , we can reduce (2.1) to the following formula for the distribution of $K_n(\alpha, r)$ in a sample of size n from $\text{PD}_\alpha^{(r)}$:

$$\mathbb{P}(K_n(\alpha, r) = k) = n \int_0^\infty \frac{r^{(k)} \lambda^{\alpha k - 1}}{\Psi(\lambda)^{r+k}} \sum_{\mathbf{m} \in A_{nk}} \prod_{j=1}^n \frac{1}{m_j!} (F_j(\lambda))^{m_j} d\lambda, \quad k \in \mathbb{N}_n. \quad (2.5)$$

Our aim herein is to derive the limiting large-sample distribution of $K_n(\alpha, r)$ in the $\text{PD}_\alpha^{(r)}$ model, working from (2.5).

To state the result, we need to introduce for each $\lambda > 0$ a subordinator $(Y_t(\lambda))_{t>0}$ having Laplace transform

$$\mathbb{E}e^{-\tau Y_t(\lambda)} = \exp\left(-t \int_0^1 (1 - e^{-\tau y}) \Pi_\lambda(dy)\right), \quad t > 0, \tau > 0, \quad (2.6)$$

with Lévy measure

$$\Pi_\lambda(dy) := \frac{\alpha y^{-\alpha-1} dy}{\Gamma(1-\alpha)} (\mathbf{1}_{\{0 < y < \lambda \leq 1\}} + \mathbf{1}_{\{0 < y < 1 < \lambda\}}). \quad (2.7)$$

Theorem 2.1. *Each $Y_t(\lambda)$, $t > 0$, $\lambda > 0$, has a continuous bounded density which we denote by $f_{Y_t(\lambda)}(y)$, $y > 0$. The asymptotic distribution of $n^{-\alpha} K_n(\alpha, r)$ as $n \rightarrow \infty$ can be written in terms of it as*

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}\left(0 < \frac{K_n(\alpha, r)}{n^\alpha} \leq y\right) &= \frac{1}{\Gamma(r)\Gamma^r(1-\alpha)} \\ &\times \int_{x=0}^y \int_{\lambda>0} x^{r-1} f_{Y_x(\lambda)}(1) \exp\left(-\frac{x(\lambda^{-\alpha} \vee 1)}{\Gamma(1-\alpha)}\right) \lambda^{-\alpha r - 1} d\lambda dx, \quad y > 0. \end{aligned} \quad (2.8)$$

Formula (2.8) looks rather forbidding, but as we will see in Section 6, it can be simplified and written in a form amenable to numerical computation.

The proof of Theorem 2.1 proceeds by a number of steps as set out next.

2.1. Outline of the proof of Theorem 2.1

Deriving (2.8) from (2.5) requires an extensive analysis whose basic ingredients are as follows:

- The formula (2.5) appears far from transparent but in fact possesses a lot of structure. (It's not even clear that (2.5) defines a proper probability mass function, i.e., sums to 1 over $k \in \mathbb{N}_n$. That this is so is demonstrated in [14].) A first step is to notice a Poissonian component in (2.5) and observe that the term summed over $\mathbf{m} \in A_{nk}$ can be interpreted as giving rise to a joint Poisson probability (see (3.1) below). Conditioning on one of the Poisson components then produces a multinomial probability ((3.2) below) and a marginal Poisson probability.

- The unpromising-looking first factor under the integral sign in (2.5) combines with the marginal Poisson to produce a negative binomial probability, modulo a correction factor ((3.7) below).
- A useful trick, used also in [27], is to write the multinomial probability as a probability involving a sum of independent and identically distributed (i.i.d.) random variables $(X_{in}(\lambda))_{i \geq 1}$ ((3.9) below), resulting in the representation (3.26) for the distribution of $K_n(\alpha, r)$ as a product of this probability, the negative binomial probability, and the correction factor, all functions of $\lambda > 0$, and integrated over $\lambda > 0$.
- The next step is to rescale $K_n(\alpha, r)$ and find the limits of the probabilities under the integral sign in (3.26). None of the representations from (2.5)–(3.26) give much clue as to an appropriate scaling for $K_n(\alpha, r)$, but we recall that for the corresponding quantity (the number of species observed) in Kingman's PD_α model, the correct scaling is by n^α , and this results in a limiting Mittag-Leffler distribution for that quantity; see [23, Theorem 3.8, p. 68]. So we are tempted to try this scaling for $K_n(\alpha, r)$ here. Less obvious however is that to make this work, it's necessary to change variable from λ to λn in the integral in (3.26), thereby giving the integral in (3.27).
- Following these manipulations, we need to find the limits as $n \rightarrow \infty$ of the functions in the integrand in (3.27). This is done in Proposition 3.2. The limit of the correction factor is easily obtained in the form of an exponential, and the negative binomial probability can be handled using Stirling's formula (the limit in fact is a gamma distribution). For the probability involving the i.i.d. random variables $(X_{in}(\lambda))$ we apply a classical limit theorem for sums of a triangular array (the distribution of the X_{in} depends on n) and then require a local version of this for i.i.d. discrete lattice random variables, which we derive directly.
- Having found the limits of the functions in the integrand in (3.27), we formally take the limit as $n \rightarrow \infty$ under the integral sign, obtaining the right-hand side of (2.8). But rather than try to justify this interchange of limit and integral, the approach we adopt is to show that the limiting function is in fact a probability density function, i.e., integrates to 1. This suffices to give convergence in distribution in (2.8). Our route to showing it goes by way of developing a kind of generalised Dickman distribution which is of interest in itself. This is done in Section 5 and is used to complete the proof of Theorem 2.1.
- Finally we derive expressions for the moments of the distribution in (2.8), and some computation-friendly formulae for it, and give some concluding comments, in Section 6. Some tables of moments and plots of distributions are in an appendix.

3. Ingredients for the proof of Theorem 2.1

To initiate the programme outlined in the previous section, write the right-hand side of (2.5) as

$$n \int_0^\infty \frac{r^{(k)} \lambda^{\alpha k - 1}}{\Psi(\lambda)^{r+k}} \sum_{\mathbf{m} \in A_{nk}} P(N_j(\lambda) = m_j, j \in \mathbb{N}_n) \exp\left(\sum_{j=1}^n F_j(\lambda)\right) d\lambda, \quad (3.1)$$

where, for each $\lambda > 0$, the $N_j(\lambda)$ are independent Poisson($F_j(\lambda)$) random variables, $j \in \mathbb{N}_n$. In A_{nk} we add over integers $m_j \geq 0$ such that $\sum_{j=1}^n jm_j = n$ and $\sum_{j=1}^n m_j = k$. For brevity, write $e_n(\lambda) := \exp(\sum_{j=1}^n F_j(\lambda))$. Then from (3.1) we have

$$\begin{aligned} \mathbb{P}(K_n(\alpha, r) = k) &= n \int_0^\infty \frac{r^{(k)} \lambda^{\alpha k - 1}}{\Psi(\lambda)^{r+k}} \mathbb{P}\left(\sum_{j=1}^n jN_j(\lambda) = n, \sum_{j=1}^n N_j(\lambda) = k\right) e_n(\lambda) d\lambda \\ &= n \int_0^\infty \frac{r^{(k)} \lambda^{\alpha k - 1}}{\Psi(\lambda)^{r+k}} \mathbb{P}\left(\sum_{j=1}^n jM_{nkj}(\lambda) = n\right) \mathbb{P}\left(\sum_{j=1}^n N_j(\lambda) = k\right) e_n(\lambda) d\lambda. \end{aligned} \quad (3.2)$$

Here, for each $\lambda > 0$, $n \in \mathbb{N}$, $k \in \mathbb{N}_n$, $\mathbf{M}_{nk}(\lambda) := (M_{nkj}(\lambda))_{j \in \mathbb{N}_n}$ is a multinomial (k , $\mathbf{p}_n(\lambda)$) vector with

$$\mathbf{p}_n(\lambda) = (p_{nj}(\lambda))_{j \in \mathbb{N}_n} = \left(\frac{F_j(\lambda)}{\sum_{\ell=1}^n F_\ell(\lambda)} \right)_{j \in \mathbb{N}_n}, \quad (3.3)$$

which results from conditioning on the sum $\sum_{j=1}^n N_j(\lambda)$ of Poisson random variables in (3.2). Specifically, the distribution of $\mathbf{M}_{nk}(\lambda)$ is given by

$$\mathbb{P}(\mathbf{M}_{nk}(\lambda) = (m_1, \dots, m_n)) = \frac{k!}{m_1! \dots m_n!} p_{n,1}(\lambda)^{m_1} \dots p_{n,n}(\lambda)^{m_n},$$

for $m_j \geq 0$, $j \in \mathbb{N}_n$, with $\sum_{j=1}^n m_j = k$. We can rewrite (3.2) as

$$\mathbb{P}(K_n(\alpha, r) = k) = n \int_0^\infty \frac{r^{(k)} \lambda^{\alpha k - 1}}{\Psi(\lambda)^{r+k}} \mathbb{P}\left(\sum_{j=1}^n jM_{nkj}(\lambda) = n\right) \mathbb{P}\left(\text{Pois}\left(\sum_{j=1}^n F_j(\lambda)\right) = k\right) e_n(\lambda) d\lambda. \quad (3.4)$$

In this expression, by (2.3),

$$\sum_{j=1}^n F_j(\lambda) = \alpha \int_0^\lambda \sum_{j=1}^n \frac{z^j}{j!} z^{-\alpha-1} e^{-z} dz = \lambda^{-\alpha} (\Psi_n(\lambda) - 1),$$

where we define

$$\Psi_n(\lambda) := 1 + \lambda^\alpha \sum_{j=1}^n F_j(\lambda) = 1 + \alpha \lambda^\alpha \int_0^\lambda \sum_{j=1}^n \frac{z^j}{j!} z^{-\alpha-1} e^{-z} dz. \quad (3.5)$$

Also define the ratio

$$\ell_n(\lambda) := \frac{\Psi_n(\lambda) - 1}{\Psi(\lambda) - 1} = \frac{\int_0^\lambda \sum_{j=1}^n (z^j/j!) z^{-\alpha-1} e^{-z} dz}{\int_0^\lambda z^{-\alpha-1} (1 - e^{-z}) dz} \leq 1, \quad \lambda > 0. \quad (3.6)$$

In (3.4), recall $e_n(\lambda) := \exp\left(\sum_{j=1}^n F_j(\lambda)\right)$, and consider the component

$$\begin{aligned} & \frac{r^{(k)} \lambda^{\alpha k-1}}{\Psi(\lambda)^{r+k}} \mathbb{P}\left(\text{Pois}\left(\sum_{j=1}^n F_j(\lambda)\right) = k\right) e_n(\lambda) \\ &= \frac{\Gamma(r+k)}{\Gamma(r)} \frac{\lambda^{-1}}{\Psi(\lambda)^{r+k}} \frac{\left(\lambda^\alpha \sum_{j=1}^n F_j(\lambda)\right)^k}{k!} \exp\left(-\sum_{j=1}^n F_j(\lambda)\right) \times e_n(\lambda) \\ &= \frac{\Gamma(r+k)}{\Gamma(r) k!} \frac{\lambda^{-1}}{\Psi(\lambda)^{r+k}} (\Psi_n(\lambda) - 1)^k \\ &= \frac{\Gamma(r+k)}{\Gamma(r) k!} \frac{\lambda^{-1}}{\Psi(\lambda)^r} \left(\frac{\Psi(\lambda) - 1}{\Psi(\lambda)}\right)^k \left(\frac{\Psi_n(\lambda) - 1}{\Psi(\lambda) - 1}\right)^k \\ &= \ell_n(\lambda)^k \lambda^{-1} \mathbb{P}\left(\text{Negbin}\left(r, \frac{1}{\Psi(\lambda)}\right) = k\right). \end{aligned}$$

Here $\text{Negbin}(r, 1/\Psi(\lambda))$ is a negative binomial random variable with parameter $r > 0$ and success probability $1/\Psi(\lambda)$. So (3.4) can be written as

$$\mathbb{P}(K_n(\alpha, r) = k) = n \int_0^\infty \ell_n(\lambda)^k \mathbb{P}\left(\sum_{j=1}^n jM_{nkj}(\lambda) = n\right) \mathbb{P}\left(\text{Negbin}\left(r, \frac{1}{\Psi(\lambda)}\right) = k\right) \frac{d\lambda}{\lambda}. \tag{3.7}$$

The multinomial vector $\mathbf{M}_{nk}(\lambda) = (M_{nkj}(\lambda))_{j \in \mathbb{N}_n}$ has moment generating function

$$\mathbb{E}\left(\prod_{j=1}^n \exp(\theta_j M_{nkj}(\lambda))\right) = \left(\sum_{j=1}^n e^{\theta_j} p_{nj}(\lambda)\right)^k,$$

where $\theta_j > 0, j \in \mathbb{N}_n$. Set $\theta_j = \theta \times j, \theta > 0$, in this to get

$$\mathbb{E} \exp\left(\theta \sum_{j=1}^n jM_{nkj}(\lambda)\right) = \left(\sum_{j=1}^n e^{\theta \times j} p_{nj}(\lambda)\right)^k =: \mathbb{E} \exp\left(\theta \sum_{i=1}^k X_{in}(\lambda)\right),$$

where $(X_{in}(\lambda))_{1 \leq i \leq k}$ are i.i.d. with

$$\mathbb{P}(X_{1n}(\lambda) = j) = p_{nj}(\lambda), \quad j \in \mathbb{N}_n. \tag{3.8}$$

So we see that

$$\mathbb{P}\left(\sum_{j=1}^n jM_{nkj}(\lambda) = n\right) = \mathbb{P}\left(\sum_{i=1}^k X_{in}(\lambda) = n\right). \tag{3.9}$$

We need information on the limiting behaviour of the $X_{in}(\lambda)$.

Proposition 3.1. For each $\lambda > 0$ and $h > 0$, we have the following:

(a)

$$\lim_{n \rightarrow \infty} n^\alpha \mathbb{P}(X_{1n}(\lambda n) \geq hn) = \begin{cases} 0, & \lambda < h < 1 \text{ or } h > 1, \\ \frac{h^{-\alpha} - \lambda^{-\alpha}}{\Gamma(1 - \alpha)}, & h < \lambda < 1, \\ \frac{h^{-\alpha} - 1}{\Gamma(1 - \alpha)}, & h < 1 < \lambda; \end{cases} \quad (3.10)$$

(b)

$$\lim_{n \rightarrow \infty} n^{\alpha-1} \mathbb{E}(X_{1n}(\lambda n) \mathbf{1}_{\{X_{1n}(\lambda n) < hn\}}) = \frac{\alpha}{\Gamma(2 - \alpha)} (\lambda^{1-\alpha} \wedge (h^{1-\alpha} \wedge 1)); \quad (3.11)$$

(c)

$$\lim_{n \rightarrow \infty} n^{\alpha-2} \mathbb{E}(X_{1n}^2(\lambda n) \mathbf{1}_{\{X_{1n}(\lambda n) < hn\}}) = \frac{\alpha}{(2 - \alpha)\Gamma(1 - \alpha)} (\lambda^{1-\alpha} \wedge (h^{1-\alpha} \wedge 1)). \quad (3.12)$$

Proof of Proposition 3.1: Throughout, keep $\lambda > 0$ and $h > 0$. For Part (a) we can compute, according to (3.8), for $0 < h \leq 1$,

$$n^\alpha \mathbb{P}(X_{1n}(\lambda n) \geq hn) = n^\alpha \sum_{j=\lfloor hn \rfloor}^n p_{nj}(\lambda n) = \frac{n^\alpha \sum_{j=\lfloor hn \rfloor}^n F_j(\lambda n)}{\sum_{j=1}^n F_j(\lambda n)} \quad (3.13)$$

(the left-hand side of (3.13) is 0 for $h > 1$). In the denominator of (3.13), by (2.3),

$$\sum_{j=1}^n F_j(\lambda n) = \alpha \int_0^{\lambda n} \sum_{j=1}^n \frac{z^j}{j!} z^{-\alpha-1} e^{-z} dz \uparrow \alpha \int_0^\infty z^{-\alpha-1} (1 - e^{-z}) dz = \Gamma(1 - \alpha), \quad (3.14)$$

as $n \rightarrow \infty$. For the numerator of (3.13), consider

$$\begin{aligned} n^\alpha \sum_{j=\lfloor hn \rfloor}^n F_j(\lambda n) &= \alpha n^\alpha \int_0^{\lambda n} \sum_{j=\lfloor hn \rfloor}^n \frac{z^j}{j!} z^{-\alpha-1} e^{-z} dz \\ &= \alpha \int_0^\lambda \sum_{j=\lfloor hn \rfloor}^n \frac{(nz)^j}{j!} e^{-nz} z^{-\alpha-1} dz \\ &= \alpha \int_0^\lambda \mathbb{P}(\lfloor hn \rfloor \leq \text{Poiss}(nz) \leq n) z^{-\alpha-1} dz. \end{aligned} \quad (3.15)$$

When $0 < \delta < h$, by Chebyshev’s inequality,

$$\begin{aligned} &\alpha \int_0^\delta \mathbb{P}(\text{Poiss}(nz) > hn) z^{-\alpha-1} dz \\ &= \alpha \int_0^\delta \mathbb{P}(\text{Poiss}(nz) - nz > n(h - z)) z^{-\alpha-1} dz \\ &\leq \alpha \int_0^\delta \frac{(nz) z^{-\alpha-1} dz}{(n(h - z))^2} \leq \frac{\delta^{1-\alpha}}{n(h - \delta)^2} \rightarrow 0, \text{ as } n \rightarrow \infty. \end{aligned}$$

For each $z > 0$, $\text{Pois}(nz)/nz \xrightarrow{P} 1$ as $n \rightarrow \infty$ by the weak law of large numbers. Thus by dominated convergence

$$\alpha \int_{\delta}^{\lambda} \mathbb{P}\left(\frac{\lfloor hn \rfloor}{nz} \leq \frac{\text{Pois}(nz)}{nz} \leq \frac{n}{nz}\right) z^{-\alpha-1} dz \rightarrow \alpha \int_{\delta}^{\lambda} \mathbf{1}_{\{h < z < 1\}} z^{-\alpha-1} dz.$$

Letting $\delta \downarrow 0$ we obtain the right-hand side of (3.10).

For Part (b) we calculate, from (3.3) and (3.8), that

$$n^{\alpha-1} \mathbb{E}(X_{1n}(\lambda n) \mathbf{1}_{\{X_{1n}(\lambda n) < hn\}}) = n^{\alpha-1} \sum_{j=1}^{\lfloor hn \rfloor} j p_{nj}(\lambda n) = \frac{n^{\alpha-1} \sum_{j=1}^{\lfloor hn \rfloor} j F_j(\lambda n)}{\sum_{j=1}^n F_j(\lambda n)}. \tag{3.16}$$

The denominator in (3.16) tends to $\Gamma(1 - \alpha)$, by (3.14). In the numerator, by (2.3),

$$\begin{aligned} n^{\alpha-1} \sum_{j=1}^{\lfloor hn \rfloor} j F_j(\lambda n) &= \alpha n^{\alpha-1} \int_0^{\lambda n} \sum_{j=1}^{\lfloor hn \rfloor} \frac{z^{j-1}}{(j-1)!} z^{-\alpha} e^{-z} dz \\ &= \alpha \int_0^{\lambda} \sum_{j=1}^{\lfloor hn \rfloor} \frac{(nz)^{j-1}}{(j-1)!} z^{-\alpha} e^{-nz} dz \\ &= \alpha \int_0^{\lambda} \mathbb{P}(\text{Pois}(nz) \leq \lfloor hn \rfloor - 1) z^{-\alpha} dz \\ &= \alpha \int_0^{\lambda} \mathbb{P}\left(\frac{\text{Pois}(nz)}{nz} \leq \frac{\lfloor hn \rfloor - 1}{nz}\right) z^{-\alpha} dz. \end{aligned} \tag{3.17}$$

In this, consider values of $h \leq 1$ and $h > 1$ separately. For $h > 1$ the sum on the left-hand side of (3.17) should be replaced by the sum over $1 \leq j \leq n$ and h in (3.17) by 1. Then, again using the fact that $\text{Pois}(nz)/nz \xrightarrow{P} 1$ for each $z > 0$, along with dominated convergence, we get

$$\lim_{n \rightarrow \infty} n^{\alpha-1} \sum_{j=1}^{\lfloor n \rfloor} j F_j(\lambda n) = \alpha \int_0^{\lambda} \mathbf{1}_{\{0 < z < 1\}} z^{-\alpha} dz = \alpha \int_0^{\lambda \wedge 1} z^{-\alpha} dz. \tag{3.18}$$

For $0 \leq h \leq 1$, similarly,

$$\lim_{n \rightarrow \infty} n^{\alpha-1} \sum_{j=1}^{\lfloor hn \rfloor} j F_j(\lambda n) = \alpha \int_0^{\lambda} \mathbf{1}_{\{0 < z < h\}} z^{-\alpha} dz = \alpha \int_0^{\lambda \wedge h} z^{-\alpha} dz. \tag{3.19}$$

Dividing by $\Gamma(1 - \alpha)$ (from (3.14)) gives the right-hand side of (3.11).

For Part (c), keep $0 < h < 1$ at first. Then by (3.3) and (3.8),

$$\begin{aligned} n^{\alpha-2} \mathbb{E}(X_{1n}^2(\lambda n) \mathbf{1}_{\{X_{1n}(\lambda n) < hn\}}) &= n^{\alpha-2} \sum_{j=1}^{\lfloor hn \rfloor} j^2 p_{nj}(\lambda n) \\ &= \frac{n^{\alpha-2} \sum_{j=1}^{\lfloor hn \rfloor} j^2 F_j(\lambda n)}{\sum_{j=1}^n F_j(\lambda n)} \sim \frac{n^{\alpha-2} \sum_{j=1}^{\lfloor hn \rfloor} j^2 F_j(\lambda n)}{\Gamma(1 - \alpha)}, \text{ as } n \rightarrow \infty \end{aligned} \tag{3.20}$$

(using (3.14)). In the numerator of the last expression,

$$\begin{aligned}
 n^{\alpha-2} \sum_{j=1}^{\lfloor hn \rfloor} j^2 F_j(\lambda n) &= \alpha n^{\alpha-2} \int_0^{\lambda n} \sum_{j=1}^{\lfloor hn \rfloor} \frac{j^2 z^j}{j!} z^{-\alpha-1} e^{-z} dz \\
 &= \alpha n^{-2} \int_0^\lambda \sum_{j=1}^{\lfloor hn \rfloor} \frac{j^2 (nz)^j}{j!} z^{-\alpha-1} e^{-nz} dz \\
 &= \alpha n^{-2} \int_0^\lambda \sum_{j=1}^{\lfloor hn \rfloor} \frac{j(nz)^j}{(j-1)!} z^{-\alpha-1} e^{-nz} dz \quad (\text{keep } n > 2) \\
 &= \alpha n^{-2} \int_0^\lambda \left(\sum_{j=2}^{\lfloor hn \rfloor} \frac{j(nz)^j}{(j-1)!} + nz \right) z^{-\alpha-1} e^{-nz} dz \\
 &= \alpha n^{-2} \int_0^\lambda \left(\sum_{j=2}^{\lfloor hn \rfloor} \frac{(nz)^{j-2}}{(j-2)!} (nz)^2 + \sum_{j=2}^{\lfloor hn \rfloor} \frac{(nz)^j}{(j-1)!} + nz \right) z^{-\alpha-1} e^{-nz} dz.
 \end{aligned}$$

Here again we see Poisson distributions and can write the last expression as

$$\begin{aligned}
 &\alpha \int_0^\lambda \mathbb{P}(\text{Poiss}(nz) \leq \lfloor hn \rfloor - 2) z^{1-\alpha} dz \\
 &\quad + \alpha n^{-2} \int_0^\lambda \sum_{j=1}^{\lfloor hn \rfloor} \frac{(nz)^{j-1}}{(j-1)!} (nz) z^{-\alpha-1} e^{-nz} dz \\
 &= \alpha \int_0^\lambda \mathbf{1}_{\{z < h\}} z^{1-\alpha} dz + o(1) + \alpha n^{-1} \int_0^\lambda \mathbb{P}(\text{Poiss}(nz) \leq \lfloor hn \rfloor - 1) z^{-\alpha} dz.
 \end{aligned}$$

As $n \rightarrow \infty$, this tends to

$$\alpha \int_0^\lambda \mathbf{1}_{\{z < h\}} z^{1-\alpha} dz = \frac{\alpha}{2-\alpha} (\lambda^{2-\alpha} \wedge h^{2-\alpha}). \tag{3.21}$$

Dividing this by $\Gamma(1-\alpha)$ (from (3.14)) gives the right-hand side of (3.12) when $0 < h < 1$. For $h > 1$, similar analysis gives a limit of $\alpha/(2-\alpha)(\lambda^{2-\alpha} \wedge 1)$. Combining these gives the right-hand side of (3.12) and completes the proof of Proposition 3.1. \square

We next list some properties of the subordinator $(Y_t(\lambda))_{t>0}$ whose Laplace transform is in (2.6) with the canonical measure in (2.7). From that measure we can calculate, for $z > 0$,

$$\begin{aligned}
 \liminf_{z \downarrow 0} \frac{1}{z^{2-\alpha}} \int_{-z}^z y^2 \Pi_\lambda(dy) &= \frac{\alpha}{1-\alpha} \liminf_{z \downarrow 0} \frac{1}{z^{2-\alpha}} \int_0^{1 \vee z} y^{1-\alpha} dy \\
 &= \frac{\alpha}{(1-\alpha)(2-\alpha)} > 0,
 \end{aligned} \tag{3.22}$$

so by a result of Orey (see [26, p. 190]), each $Y_t(\lambda)$ has a C^∞ density all of whose derivatives tend to 0 at ∞ . Write

$$\mathbb{E} e^{i\theta Y_t(\lambda)} = \exp \left(t \int_0^1 (e^{i\theta y} - 1) \frac{k_\lambda(y)}{y} dy \right), \quad \theta \in \mathbb{R}, \tag{3.23}$$

where

$$k_\lambda(y) := \frac{\alpha y^{-\alpha}}{\Gamma(1-\alpha)} (\mathbf{1}_{\{0 < y < \lambda \leq 1\}} + \mathbf{1}_{\{0 < y < 1 < \lambda\}}) \tag{3.24}$$

is left-continuous and decreasing on $(0, \infty)$. So by [26, Theorem 28.4, p. 191], $(Y_t(\lambda))_{t>0}$ is a self-decomposable Lévy process for each $\lambda > 0$, and by a similar argument as in [26, Lemma 28.5, p. 191], we can deduce that

$$|\mathbb{E}e^{i\theta Y_t(\lambda)}| = o(|\theta|^{-\beta}), \quad \text{uniformly in } \lambda > 0, \text{ as } |\theta| \rightarrow \infty, \text{ for all } \beta > 0. \tag{3.25}$$

Now proceeding with the main proof, from (3.7) and (3.9) we get

$$\mathbb{P}(K_n(\alpha, r) = k) = n \int_0^\infty \ell_n(\lambda)^k \mathbb{P}\left(\sum_{i=1}^k X_{in}(\lambda) = n\right) \mathbb{P}\left(\text{Negbin}\left(r, \frac{1}{\Psi(\lambda)}\right) = k\right) \frac{d\lambda}{\lambda}. \tag{3.26}$$

Change variable from λ to λn to write this as

$$\begin{aligned} \mathbb{P}(K_n = k) &= n \int_0^\infty \ell_n(\lambda n)^k \mathbb{P}\left(\sum_{i=1}^k X_{in}(\lambda n) = n\right) \mathbb{P}\left(\text{Negbin}\left(r, \frac{1}{\Psi(\lambda n)}\right) = k\right) \frac{d\lambda}{\lambda} \\ &=: f_n(k). \end{aligned} \tag{3.27}$$

Then for $0 < a < b$,

$$\begin{aligned} \mathbb{P}(an^\alpha < K_n(\alpha, r) \leq bn^\alpha) &= \sum_{an^\alpha < x \leq bn^\alpha} \mathbb{P}(K_n = \lfloor x \rfloor) \\ &= \int_{\lceil an^\alpha \rceil}^{\lfloor bn^\alpha \rfloor} f_n(\lfloor x \rfloor) dx = n^\alpha \int_{a_n}^{b_n} f_n(\lfloor xn^\alpha \rfloor) dx, \end{aligned}$$

where $a_n := \lceil an^\alpha \rceil / n^\alpha$ and $b_n := \lfloor bn^\alpha \rfloor / n^\alpha$. Thus we can write

$$\begin{aligned} \mathbb{P}(an^\alpha < K_n(\alpha, r) \leq bn^\alpha) &= n^{\alpha+1} \int_{a_n}^{b_n} dx \int_{\lambda>0} (\ell_n(\lambda n))^{\lfloor xn^\alpha \rfloor} \\ &\quad \times \mathbb{P}\left(\sum_{i=1}^{\lfloor xn^\alpha \rfloor} X_{in}(\lambda n) = n\right) \mathbb{P}\left(\text{Negbin}\left(r, \frac{1}{\Psi(\lambda n)}\right) = \lfloor xn^\alpha \rfloor\right) \frac{d\lambda}{\lambda}. \end{aligned} \tag{3.28}$$

We have $a_n \rightarrow a, b_n \rightarrow b$, and we need to find the limits of the ℓ_n term and the probabilities in (3.28).

Proposition 3.2. *Holding $x > 0$ and $\lambda > 0$ fixed, we have*

(a)

$$\lim_{n \rightarrow \infty} n \mathbb{P}\left(\sum_{i=1}^{\lfloor xn^\alpha \rfloor} X_{in}(\lambda n) = n\right) = f_{Y_x(\lambda)}(1); \tag{3.29}$$

(b)

$$\lim_{n \rightarrow \infty} n^\alpha \mathbb{P}\left(\text{Negbin}\left(r, \frac{1}{\Psi(\lambda n)}\right) = \lfloor xn^\alpha \rfloor\right) = \frac{x^{r-1} \lambda^{-\alpha r}}{\Gamma(r) \Gamma^r(1-\alpha)} \exp\left(-\frac{x \lambda^{-\alpha}}{\Gamma(1-\alpha)}\right); \tag{3.30}$$

(c)

$$\lim_{n \rightarrow \infty} (\ell_n(\lambda n))^{\lfloor xn^\alpha \rfloor} = \exp \left(- \frac{x(1 - \lambda^{-\alpha}) \mathbf{1}_{\{\lambda > 1\}}}{\Gamma(1 - \alpha)} \right). \tag{3.31}$$

Consequently,

(d)

$$\begin{aligned} \lim_{n \rightarrow \infty} n^\alpha (\ell_n(\lambda n))^{\lfloor xn^\alpha \rfloor} \mathbb{P} \left(\text{Negbin} \left(r, \frac{1}{\Psi(\lambda n)} \right) = \lfloor xn^\alpha \rfloor \right) \\ = \frac{x^{r-1} \lambda^{-\alpha r}}{\Gamma(r) \Gamma^r(1 - \alpha)} \exp \left(- \frac{x(\lambda^{-\alpha} \vee 1)}{\Gamma(1 - \alpha)} \right). \end{aligned} \tag{3.32}$$

Proof of Proposition 3.2: We keep $x > 0$ and $\lambda > 0$ fixed throughout this proof.

(a) We begin by finding the limiting distribution of the sum

$$n^{-1} \sum_{i=1}^{\lfloor xn^\alpha \rfloor} X_{in}(\lambda n)$$

as $n \rightarrow \infty$, using a classical limit theorem for sums of a triangular array. Thus, we verify Conditions (i), (ii), and (iii) of [17, Corollary 15.16, p. 297]. It suffices to set $x = 1$ for this. Those conditions can be read from (3.10)–(3.12) of Proposition 3.1 as follows.

First, recalling the definition of Π_λ in (2.7), we note that (3.10) implies

$$\lim_{n \rightarrow \infty} n^\alpha \mathbb{P}(n^{-1} X_{1n}(\lambda n) \in dy) = \Pi_\lambda(dy), \quad y > 0, \tag{3.33}$$

which is Condition (i) of [17, Corollary 15.16].

For Condition (ii), we can deduce from (3.11) and using the definition of Π_λ in (2.7) that

$$\lim_{n \rightarrow \infty} n^\alpha \mathbb{E} \left(\frac{X_{1n}(\lambda n)}{n} \mathbf{1}_{\{X_{1n}(\lambda n) < hn\}} \right) = \int_{0 < y \leq h} y \Pi_\lambda(dy) = b - \int_{h < y \leq 1} y \Pi_\lambda(dy), \tag{3.34}$$

where $b = \int_{0 < y \leq 1} y \Pi_\lambda(dy)$. The right-hand side of (3.34) is in the form required by Condition (ii) of [17].

For Condition (iii) of [17, Corollary 15.16] we require

$$\lim_{n \rightarrow \infty} n^\alpha \mathbb{E} \left(\frac{X_{1n}^2(\lambda n)}{n^2} \mathbf{1}_{\{X_{1n}(\lambda n) < hn\}} \right) = a + \int_{0 < y \leq h} y^2 \Pi_\lambda(dy) \tag{3.35}$$

for a finite constant a and all $h > 0$. That this holds, in fact with $a = 0$, can be deduced from (3.12).

With these three conditions satisfied, [17, Corollary 15.16] then gives that the normed sum

$$n^{-1} \sum_{i=1}^{\lfloor n^\alpha \rfloor} X_{in}(\lambda n)$$

converges in distribution to the infinitely divisible distribution $id(a, b, \Pi_\lambda)$, in Kallenberg’s notation. This distribution has characteristic exponent (see [17, Corollary 15.8, p. 291]) given by

$$\begin{aligned} i\theta b - \frac{1}{2}\theta^2 a + \int_{\mathbb{R}\setminus\{0\}} (e^{i\theta y} - 1 - i\theta y \mathbf{1}_{\{0 < y < 1\}}) \Pi_\lambda(dy) \\ = i\theta \left(b - \int_{0 < y < 1} y \Pi_\lambda(dy) \right) + \int_{\mathbb{R}\setminus\{0\}} (e^{i\theta y} - 1) \Pi_\lambda(dy) \end{aligned} \tag{3.36}$$

(since $a = 0$). Here $b = \int_{0 < y < 1} y \Pi_\lambda(dy)$, so the first term on the right-hand side of (3.36) equals 0. The limit distribution thus has characteristic exponent $\int_{\mathbb{R}\setminus\{0\}} (e^{i\theta y} - 1) \Pi_\lambda(dy)$, and hence is the distribution of $Y_1(\lambda)$ having Laplace transform (2.6) for $t = 1$.

Continuing with the proof of Part (a), we have that the characteristic function of the normed sum $n^{-1} \sum_{i=1}^{\lfloor n^\alpha \rfloor} X_{in}(\lambda n)$ also converges, so we can write

$$\lim_{n \rightarrow \infty} \phi_{\lambda n}^{\lfloor n^\alpha \rfloor}(\theta/n) = \lim_{n \rightarrow \infty} \mathbb{E} \exp \left(\frac{i\theta}{n} \sum_{i=1}^{\lfloor n^\alpha \rfloor} X_{in}(\lambda n) \right) = \mathbb{E} e^{i\theta Y_1(\lambda)}. \tag{3.37}$$

For (3.29) we need a local version of this convergence, given as Lemma 3.1.

Lemma 3.1. *For each $x > 0$ and $\lambda > 0$, (3.29) holds.*

We defer the proof of Lemma 3.1 to Appendix A. Assuming it, we have completed the proof of Part (a) of Proposition 3.2.

(b) For (3.30) write

$$\mathbb{P} \left(\text{Negbin} \left(r, \frac{1}{\Psi(\lambda n)} \right) = \lfloor x n^\alpha \rfloor \right) = \frac{\Gamma(r + \lfloor x n^\alpha \rfloor)}{\Gamma(r) \lfloor x n^\alpha \rfloor!} \frac{1}{\Psi(\lambda n)^r} \left(1 - \frac{1}{\Psi(\lambda n)} \right)^{\lfloor x n^\alpha \rfloor} \tag{3.38}$$

and use Stirling’s approximation to get

$$\frac{\Gamma(r + \lfloor x n^\alpha \rfloor)}{\lfloor x n^\alpha \rfloor!} \sim (x n^\alpha)^{r-1} \quad \text{as } n \rightarrow \infty.$$

Also, recalling (2.2),

$$\begin{aligned} n^{-\alpha} \Psi(\lambda n) &= n^{-\alpha} \left(1 + \alpha(\lambda n)^\alpha \int_0^{\lambda n} z^{-\alpha-1} (1 - e^{-z}) dz \right) \\ &\rightarrow \alpha \lambda^\alpha \int_0^\infty z^{-\alpha-1} (1 - e^{-z}) dz = \lambda^\alpha \Gamma(1 - \alpha), \quad \text{as } n \rightarrow \infty. \end{aligned} \tag{3.39}$$

So from (3.38)

$$\begin{aligned} &\mathbb{P} \left(\text{Negbin} \left(r, \frac{1}{\Psi(\lambda n)} \right) = \lfloor x n^\alpha \rfloor \right) \\ &\sim \frac{(x n^\alpha)^{r-1}}{\Gamma(r)} \frac{1}{(n^\alpha \lambda^\alpha \Gamma(1 - \alpha))^r} \left(1 - \frac{1}{\Psi(\lambda n)} \right)^{\lfloor x n^\alpha \rfloor} \\ &\sim \frac{n^{-\alpha} x^{r-1} \lambda^{-\alpha r}}{\Gamma(r) \Gamma^r(1 - \alpha)} \exp \left(-\frac{x \lambda^{-\alpha}}{\Gamma(1 - \alpha)} \right), \end{aligned}$$

Here we used (3.39) and

$$\begin{aligned} \lim_{n \rightarrow \infty} \left(1 - \frac{1}{\Psi(\lambda n)}\right)^{\lfloor xn^\alpha \rfloor} &= \lim_{n \rightarrow \infty} \left(1 - \frac{x}{(xn^\alpha)n^{-\alpha}\Psi(\lambda n)}\right)^{\lfloor xn^\alpha \rfloor} \\ &= \exp\left(-\frac{x\lambda^{-\alpha}}{\Gamma(1-\alpha)}\right). \end{aligned}$$

Thus we have proved (3.30).

(c) From (3.6) we can write

$$\begin{aligned} \ell_n(\lambda n) &= \frac{\int_0^{\lambda n} \sum_{j=1}^n (z^j/j!)z^{-\alpha-1}e^{-z}dz}{\int_0^{\lambda n} z^{-\alpha-1}(1-e^{-z})dz} \\ &= 1 - \frac{\int_0^{\lambda n} \sum_{j=n+1}^\infty (z^j/j!)z^{-\alpha-1}e^{-z}dz}{\int_0^{\lambda n} z^{-\alpha-1}(1-e^{-z})dz}. \end{aligned} \tag{3.40}$$

The numerator here is

$$n^{-\alpha} \int_0^\lambda \sum_{j=n+1}^\infty \frac{(nz)^j}{j!} e^{-nz} z^{-\alpha-1} dz = n^{-\alpha} \int_0^\lambda \mathbb{P}(\text{Poiss}(nz) > n) z^{-\alpha-1} dz. \tag{3.41}$$

When $0 < \lambda < 1$, the right-hand side is, by Chebyshev’s inequality,

$$\begin{aligned} &n^{-\alpha} \int_0^\lambda \mathbb{P}(\text{Poiss}(nz) - nz > n(1-z))z^{-\alpha-1} dz \\ &\leq n^{-\alpha} \int_0^1 \frac{(nz)z^{-\alpha-1}}{(n(1-z))^2} dz = o(n^{-\alpha}), \text{ as } n \rightarrow \infty, \end{aligned}$$

so the right-hand side of (3.41) is asymptotic to

$$n^{-\alpha} \mathbf{1}_{\{\lambda > 1\}} \int_1^\lambda \mathbb{P}\left(\frac{\text{Poiss}(nz)}{nz} > \frac{1}{z}\right) z^{-\alpha-1} dz \sim n^{-\alpha} \mathbf{1}_{\{\lambda > 1\}} \frac{1 - \lambda^{-\alpha}}{\alpha}.$$

Since the denominator in (3.40) tends to $\Gamma(1-\alpha)/\alpha$, we have

$$(\ell_n(\lambda n))^{\lfloor xn^\alpha \rfloor} = \left(1 - \frac{x(1-\lambda^{-\alpha})\mathbf{1}_{\{\lambda > 1\}}}{xn^\alpha \Gamma(1-\alpha)(1+o(1))}\right)^{\lfloor xn^\alpha \rfloor},$$

and the right-hand side here tends to the right-hand side of (3.31).

Finally, to prove (3.32) simply multiply (3.30) and (3.31) together. □

Now to continue with the proof of Theorem 2.1, return to (3.28) and, formally, take the limits in (3.29) and (3.32) through the integral in (3.28) to get the expression on the right-hand side of (2.8).

Justifying this interchange directly by dominated convergence seems difficult, so we take an indirect approach. The idea is to show that the right-hand side of (2.8) is in fact a probability density function, i.e., integrates to 1. This will complete the proof of the theorem by the following argument.

Take $a = 0$ and $b = y > 0$ in (3.27), and write, for $y > 0$,

$$\mathbb{P}\left(\frac{K_n(\alpha, r)}{n^\alpha} \leq y\right) = \int_0^y \mathbb{P}\left(\frac{K_n(\alpha, r)}{n^\alpha} \in dx\right) = \int_{x=0}^y \int_{\lambda>0} f_n(x, \lambda) dx,$$

where $f_n(x, \lambda)$ is the integrand in (2.8). Then by Fatou’s lemma

$$\begin{aligned} \liminf_{n \rightarrow \infty} \mathbb{P}\left(\frac{K_n(\alpha, r)}{n^\alpha} \leq y\right) &= \liminf_{n \rightarrow \infty} \int_{x=0}^y \int_{\lambda>0} f_n(x, \lambda) dx d\lambda \\ &\geq \int_{x=0}^y \int_{\lambda>0} \liminf_{n \rightarrow \infty} f_n(x, \lambda) dx d\lambda = \int_{x=0}^y \int_{\lambda>0} \lim_{n \rightarrow \infty} f_n(x, \lambda) dx d\lambda, \end{aligned}$$

where the limit $\lim_{n \rightarrow \infty} f_n(x, \lambda)$ exists as the product of the limits (3.29) and (3.32) in Proposition 3.2. Similarly

$$\begin{aligned} \liminf_{n \rightarrow \infty} \mathbb{P}\left(\frac{K_n(\alpha, r)}{n^\alpha} > y\right) &= \liminf_{n \rightarrow \infty} \int_{x=y}^\infty \int_{\lambda>0} f_n(x, \lambda) dx d\lambda \\ &\geq \int_{x=y}^\infty \int_{\lambda>0} \liminf_{n \rightarrow \infty} f_n(x, \lambda) dx d\lambda = \int_{x=y}^\infty \int_{\lambda>0} \lim_{n \rightarrow \infty} f_n(x, \lambda) dx d\lambda. \end{aligned}$$

Suppose we know

$$\int_{x=0}^\infty \int_{\lambda>0} \lim_{n \rightarrow \infty} f_n(x, \lambda) dx d\lambda = 1. \tag{3.42}$$

Then

$$\begin{aligned} \limsup_{n \rightarrow \infty} \mathbb{P}\left(\frac{K_n(\alpha, r)}{n^\alpha} \leq y\right) &= 1 - \liminf_{n \rightarrow \infty} \mathbb{P}\left(\frac{K_n(\alpha, r)}{n^\alpha} > y\right) \\ &= 1 - \liminf_{n \rightarrow \infty} \int_{x=y}^\infty \int_{\lambda>0} f_n(x, \lambda) dx d\lambda \\ &\leq 1 - \int_{x=y}^\infty \int_{\lambda>0} \lim_{n \rightarrow \infty} f_n(x, \lambda) dx d\lambda \\ &= \int_{x=0}^y \int_{\lambda>0} \lim_{n \rightarrow \infty} f_n(x, \lambda) dx d\lambda, \end{aligned}$$

and from these we deduce

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{K_n(\alpha, r)}{n^\alpha} \leq y\right) = \int_0^y \int_{\lambda>0} \lim_{n \rightarrow \infty} f_n(x, \lambda) dx d\lambda.$$

Since $\lim_{n \rightarrow \infty} f_n(x, \lambda)$ is the integrand in (2.8), this completes the proof of (2.8) subject to proving (3.42). This is done in the next two sections.

4. The negative binomial point process and its sum

We need some concepts concerning a negative binomial point process which are set out in [14]. We refer to that paper for further background and details. The dependence of the various

quantities on the parameters α and r is not always made explicit in that paper. Here we need to make it explicit for clarity.

Let $\mathbb{B}^{(r)}$ be a negative binomial point process with measure $\alpha x^{-\alpha-1} dx \mathbf{1}_{\{0 < x \leq 1\}}$, for $r > 0$, $\alpha \in (0, 1)$, in the sense of Gregoire [9], and define $^{(\alpha,r)}T$ to be the sum of the points in $\mathbb{B}^{(r)}$. The variable $^{(\alpha,r)}T$ has a density

$$g_{\alpha,r}(t) = \mathbb{P}({}^{(\alpha,r)}T \in dt)/dt, \quad t > 0, \tag{4.1}$$

and a Laplace transform

$$\mathbb{E}(e^{-\tau \times {}^{(\alpha,r)}T}) = \int_0^\infty e^{-\tau t} g_{\alpha,r}(t) dt = \left(1 + \alpha \int_0^1 (1 - e^{-\tau x}) x^{-\alpha-1} dx \right)^{-r}, \tag{4.2}$$

for $\tau > 0, r \in \mathbb{N}$. This implies the convolution formula

$${}^{(\alpha,r)}T * {}^{(\alpha,s)}T \stackrel{D}{=} {}^{(\alpha,r+s)}T, \quad r, s, > 0 \tag{4.3}$$

(independent copies on the left-hand side). Let $G_{\alpha,r}(t) = \int_0^t g_{\alpha,r}(y) dy, t \geq 0$, be the cumulative distribution function of $^{(\alpha,r)}T$. The next lemma connects these ideas with the result of Theorem 2.1.

Lemma 4.1. *The integral on the right-hand side of (2.8), taken with $y = \infty$, equals*

$$1 + \frac{1}{\alpha r} g_{\alpha,r}(1) - G_{\alpha,r}(1). \tag{4.4}$$

Proof of Lemma 4.1: Split the λ -integral on the right-hand side of (2.8) into two components: one component over $\lambda \in (0, 1]$ and the other over $\lambda > 1$. For the integral over $\lambda \in (1, \infty)$ we compute (with $y = \infty$ in (2.8))

$$\begin{aligned} & \frac{1}{\Gamma(r)\Gamma^r(1-\alpha)} \int_{x=0}^\infty \int_{\lambda=1}^\infty x^{r-1} f_{Y_x(\lambda)}(1) \exp\left(-\frac{x}{\Gamma(1-\alpha)}\right) \lambda^{-\alpha r-1} d\lambda dx \\ &= \frac{1}{\alpha r \Gamma(r)\Gamma^r(1-\alpha)} \int_{x=0}^\infty x^{r-1} f_{Y_x(\lambda)}(1) \exp\left(-\frac{x}{\Gamma(1-\alpha)}\right) dx. \end{aligned} \tag{4.5}$$

By (2.6) and (2.7), each $Y_x(\lambda)$ has characteristic function

$$\mathbb{E}e^{i\theta Y_x(\lambda)} = \exp\left(\frac{x}{\Gamma(1-\alpha)} \int_0^{\lambda \wedge 1} (e^{i\theta y} - 1) \alpha y^{-\alpha-1} dy\right), \quad \theta \in \mathbb{R}. \tag{4.6}$$

Taking $\lambda > 1$ in (4.6), the right-hand side of (4.5) is, by Fourier inversion,

$$\begin{aligned} & \frac{1}{\alpha r \Gamma(r)\Gamma^r(1-\alpha)} \int_{x=0}^\infty x^{r-1} \exp\left(-\frac{x}{\Gamma(1-\alpha)}\right) \\ & \times \frac{1}{2\pi} \int_{\theta=-\infty}^\infty e^{-i\theta} \exp\left(\frac{x}{\Gamma(1-\alpha)} \int_0^1 (e^{i\theta y} - 1) \alpha y^{-\alpha-1} dy\right) dx d\theta \\ &= \frac{1}{2\pi \alpha r} \int_{\theta=-\infty}^\infty \frac{e^{-i\theta} d\theta}{\left(1 - \int_0^1 (e^{i\theta y} - 1) \alpha y^{-\alpha-1} dy\right)^r} \\ &= \frac{1}{\alpha r} g_{\alpha,r}(1). \end{aligned} \tag{4.7}$$

The last line follows from setting $\tau = -i\theta$ in (4.2) and observing that we have the Fourier inverse of $g_{\alpha,r}$ at 1. For the integral over $\lambda \in (0, 1]$ we compute

$$\begin{aligned} & \frac{1}{2\pi\Gamma(r)\Gamma^r(1-\alpha)} \int_{x=0}^\infty \int_{\lambda=0}^1 \int_{\theta=-\infty}^\infty e^{-i\theta} x^{r-1} \exp\left(-\frac{x\lambda^{-\alpha}}{\Gamma(1-\alpha)}\right) \\ & \times \exp\left(\frac{x}{\Gamma(1-\alpha)} \int_0^\lambda (e^{i\theta y} - 1)\alpha y^{-\alpha-1} dy\right) \lambda^{-\alpha r-1} d\lambda dx d\theta \\ & = \int_{\lambda=0}^1 \frac{1}{2\pi} \int_{\theta=-\infty}^\infty \frac{e^{-i\theta} \lambda^{-\alpha r-1} d\lambda d\theta}{\left(\lambda^{-\alpha} - \lambda^{-\alpha} \int_0^1 (e^{i\theta y} - 1)\alpha y^{-\alpha-1} dy\right)^r} \\ & = \int_{\lambda=0}^1 \frac{1}{2\pi} \int_{\theta=-\infty}^\infty \frac{e^{-i\theta/\lambda} \lambda^{-2} d\lambda d\theta}{\left(1 - \int_0^1 (e^{i\theta y} - 1)\alpha y^{-\alpha-1} dy\right)^r} \\ & = \int_{\lambda=0}^1 \lambda^{-2} g_{\alpha,r}(1/\lambda) d\lambda \\ & = \int_{\lambda=1}^\infty g_{\alpha,r}(\lambda) d\lambda = 1 - \int_{\lambda=0}^1 g_{\alpha,r}(\lambda) d\lambda = 1 - G_{\alpha,r}(1). \end{aligned} \tag{4.8}$$

Adding (4.7) and (4.8) gives (4.4). □

Thus to achieve our aim of showing that the right-hand side of (2.8) defines a proper distribution, we have to show that the expression in (4.4) equals 1. We do this by developing a connection with the theory of Dickman functions.

5. Generalised Dickman functions

The generalised Dickman function, when normalised, occurs naturally as the density of the infinitely divisible random variable Y_a having Laplace transform

$$\mathbb{E}(e^{-\tau Y_a}) = \exp\left(-a \int_0^1 (1 - e^{-\tau y}) dy/y\right), \quad \tau > 0, a > 0. \tag{5.1}$$

(There should be no confusion with the $Y_t(\lambda)$ defined in (2.6).) The descriptor ‘generalised’ was added by [22]; it signifies the inclusion of the parameter $a > 0$ in (5.1), where $a = 1$ defines the Dickman distribution as usually understood. The process $(Y_t)_{t>0}$ is described as the *Dickman subordinator* in [3].

Properties of the Dickman function (also known as the Dickman–de Bruijn function; see [20] for a review) and of its associated distribution have been teased out over the years since its original formulation in [5] in a number-theoretic context. The papers [21] and [22] provide convenient summaries, for our purposes, of these properties, and of the generalised version. In particular, they give the formula for the density $f_{Y_a}(y)$ of Y_a as $f_{Y_a}(y) = e^{-\alpha y} \rho_a(y) / \Gamma(a)$, $y > 0$, where $\gamma = 0.577\dots$ is Euler’s constant,

$$\rho_a(y) = y^{a-1} \quad \text{for } 0 < y \leq 1, \tag{5.2}$$

and $\rho_a(y)$ satisfies a certain differential-delay equation for $y > 1$. There is also the representation

$$Y_a D = U^{1/a}(Y_a + 1) \quad (5.3)$$

(see [22, Proposition 5.1, p. 14]) where the Uniform $[0, 1]$ random variable U and the Y_a on the right-hand side are independent. Notably, with $F_{Y_a}(y) := \mathbb{P}(Y_a \leq y)$, we see from (5.2) that

$$\frac{1}{a} f_{Y_a}(1) = F_{Y_a}(1), \quad (5.4)$$

which can be compared with (4.4).

Suppose we were to let $r \rightarrow \infty$ and $\alpha = \alpha_r \downarrow 0$ in such a way that $\alpha_r r \rightarrow a \in (0, \infty)$, and it were the case that

$$\lim_{r \rightarrow \infty} \frac{1}{\alpha_r r} g_{\alpha_r, r}(x) = \frac{1}{a} f_{Y_a}(x) \quad \text{for } x > 0. \quad (5.5)$$

That this is plausible is indicated by inspection of (4.2) and (5.1). Then from (5.4) and (5.5) we would have

$$\lim_{r \rightarrow \infty} \frac{1}{\alpha_r r} g_{\alpha_r, r}(1) = \frac{1}{a} f_{Y_a}(1) = F_{Y_a}(1) = \lim_{r \rightarrow \infty} G_{\alpha_r, r}(1), \quad (5.6)$$

suggesting perhaps that the identity

$$\frac{1}{\alpha r} g_{\alpha, r}(1) = G_{\alpha, r}(1) \quad (5.7)$$

is true for each $\alpha \in (0, 1)$ and $r > 0$, not just in the limit. This would prove that the expression in (4.4) equals 1.

These heuristics could possibly be made to give a rigorous proof of (5.7), but we do not go down that route; rather, we deal directly with $g_{\alpha, r}(x)$ and show that (5.7) holds by generalising the Dickman relationship, in the next subsection.

5.1. The function $g_{\alpha, r}$ as an (α, r) -generalised Dickman distribution

We proceed by giving an analogue of (5.3) for the negative binomial sums $^{(\alpha, r)}T$, then show that this implies (5.7).

Proposition 5.1.

(i) For each $\alpha \in (0, 1)$ and $r > 0$ we have

$$^{(\alpha, r)}T \stackrel{D}{=} U^{\frac{1}{\alpha r}} \left(^{(\alpha, r+1)}T + 1 \right), \quad (5.8)$$

where the Uniform $[0, 1]$ random variable U and the $^{(\alpha, r+1)}T$ on the right-hand side are independent.

(ii) Consequently, (5.7) is true for each $\alpha \in (0, 1)$ and $r > 0$.

Proof of Proposition 5.1: First we prove Part (i). Throughout this proof we write $a = \alpha r$ for the combination αr , which will occur frequently. With this in mind, we begin by noting that

$$\begin{aligned} & \frac{d}{d u} \left(1 + \alpha \int_0^1 (1 - e^{-\tau u^{1/a} y}) y^{-\alpha-1} dy \right)^{-r} \\ &= -\tau u^{1/a-1} \int_0^1 e^{-\tau u^{1/a} y} y^{-\alpha} dy \times \frac{1}{\left(1 + \alpha \int_0^1 (1 - e^{-\tau u^{1/a} y}) y^{-\alpha-1} dy \right)^{r+1}}, \end{aligned}$$

for $0 < u < 1$. So, integrating by parts,

$$\int_{u=0}^1 \frac{\tau u^{1/a}}{u} \int_{y=0}^1 e^{-\tau u^{1/a} y} y^{-\alpha} dy \times \frac{u du}{\left(1 + \alpha \int_0^1 (1 - e^{-\tau u^{1/a} y}) y^{-\alpha-1} dy\right)^{r+1}}$$

$$= - \frac{1}{\left(1 + \alpha \int_0^1 (1 - e^{-\tau y}) y^{-\alpha-1} dy\right)^r} + \int_{u=0}^1 \frac{du}{\left(1 + \alpha \int_0^1 (1 - e^{-\tau u^{1/a} y}) y^{-\alpha-1} dy\right)^r}.$$

(5.9)

The first term on the right-hand side of (5.9) is $-\mathbb{E}(e^{-\tau \times (\alpha,r)T})$ (recall (4.2)). Rearranging (5.9) gives

$$\mathbb{E}(e^{-\tau \times (\alpha,r)T}) = \int_0^1 \frac{1}{\left(1 + \alpha \int_0^1 (1 - e^{-\tau u^{1/a} y}) y^{-\alpha-1} dy\right)^r}$$

$$\times \left(1 - \frac{\tau u^{1/a}}{1 + \alpha \int_0^1 (1 - e^{-\tau u^{1/a} y}) y^{-\alpha-1} dy} \int_0^1 e^{-\tau u^{1/a} y} y^{-\alpha} dy\right) du.$$

(5.10)

The second term on the right-hand side, in parentheses, equals

$$\frac{1 + \alpha \int_0^1 (1 - e^{-\tau u^{1/a} y}) y^{-\alpha-1} dy - \tau u^{1/a} \int_0^1 e^{-\tau u^{1/a} y} y^{-\alpha} dy}{1 + \alpha \int_0^1 (1 - e^{-\tau u^{1/a} y}) y^{-\alpha-1} dy},$$

and after an integration by parts in the numerator this equals

$$\frac{e^{-\tau u^{1/a}}}{1 + \alpha \int_0^1 (1 - e^{-\tau u^{1/a} y}) y^{-\alpha-1} dy} = e^{-\tau u^{1/a}} \times \mathbb{E}(e^{-\tau u^{1/a} \times (\alpha,1)T}).$$

Substituting back in (5.10) gives

$$\mathbb{E}(e^{-\tau \times (\alpha,r)T}) = \int_0^1 e^{-\tau u^{1/a}} \times \mathbb{E}(e^{-\tau u^{1/a} \times (\alpha,1)T}) \times \mathbb{E}(e^{-\tau u^{1/a} \times (\alpha,r)T}) du$$

$$= \int_0^1 \mathbb{E}(e^{-\tau u^{1/a} (\alpha,r+1)T+1}) du,$$

(5.11)

where the last equality follows from (4.3). So, recalling that $a \equiv \alpha r$, we arrive at (5.8).

We now prove Part (ii). From (5.8) we can write, for $t > 0$,

$$G_{\alpha,r}(t) = \mathbb{P}^{(\alpha,r)T} \leq t) = \mathbb{P}(U^{\frac{1}{\alpha r}} (\alpha,r+1)T + 1) \leq t)$$

$$= \int_0^1 \mathbb{P}^{(\alpha,r+1)T} \leq tu^{-1/a} - 1) du$$

$$= \int_0^1 G_{\alpha,r+1}(tu^{-1/a} - 1) du$$

$$= \alpha t^a \int_{t-1}^\infty G_{\alpha,r+1}(v)(1+v)^{-1-a} dv.$$

(5.12)

Now since $G_{\alpha,r+1}(v) = 0$ for $v < 0$, we have for $0 < t \leq 1$

$$\int_{t-1}^{\infty} G_{\alpha,r+1}(v)(1+v)^{-1-a} dv = \int_0^{\infty} G_{\alpha,r+1}(v)(1+v)^{-1-a} dv =: c_{\alpha,r+1}.$$

It follows that $G_{\alpha,r}(t) = at^a c_{\alpha,r+1}$ for $0 < t \leq 1$; hence

$$g_{\alpha,r}(t) = G'_{\alpha,r}(t) = a^2 c_{\alpha,r+1} t^{a-1} = at^{-1} G_{\alpha}(t), \tag{5.13}$$

for $0 < t \leq 1$. So, recalling that $a \equiv \alpha r$, we have proved (5.7). □

With (5.7), we have completed the proof of Theorem 2.1. □

Remark 5.1. In (5.8), letting $r \rightarrow \infty$ and $\alpha \downarrow 0$ so that $r\alpha \rightarrow a \in (0, \infty)$, we recover (5.3), and taking the same limit in (4.2) we recover (5.1). In this sense the negative binomial sums $^{(\alpha,r)}T$ can be thought of as a two-parameter generalisation of the Dickman random variable Y_a . (This is distinct from a two-parameter generalisation of the Dickman function due to [11], and from the two-parameter generalisation of the Poisson–Dirichlet distribution in [24]. Another generalisation having application in polymer modelling occurs in [3] as a multivariate Dickman–Gaussian combination.)

Differentiating (5.12) we obtain, for $t > 1$, the density in the form

$$tg_{\alpha,r}(t) = \alpha r(G_{\alpha,r}(t) - G_{\alpha,r+1}(t-1)), \tag{5.14}$$

and differentiating this further gives

$$tg'_{\alpha,r}(t) + (1 - \alpha r)g_{\alpha,r+1}(t) + \alpha rg_{\alpha,r+1}(t-1) = 0, \quad t > 1. \tag{5.15}$$

This can be compared with the corresponding differential-delay equation for the Dickman function (see [22, Equation (5.10)]). See Section 6 of [22] for further interesting discussion.

6. Properties of the limiting distribution

Let $K = K(\alpha, r)$ be the limiting random variable of $n^{-\alpha}K_n(\alpha, r)$ as $n \rightarrow \infty$, whose distribution is given by the right-hand side of (2.8), and let $f_{Y_t(\lambda)}(y)$, $y > 0$, be the density of the subordinator $(Y_t(\lambda))_{t>0}$ whose Laplace transform is in (2.6). In this section we derive some properties of this distribution, first getting it in a form more amenable to numerical computation, then, in Subsection 6.2, deriving formulae for the moments of the distribution.

6.1. The limiting distribution as a gamma mixture and subordinated truncated stable process

As before, consider the λ -integral on the right-hand side of (2.8) in two components: one over $\lambda \in (0, 1]$ and the other over $\lambda > 1$. From (2.6), when $0 < \lambda \leq 1$, $x > 0$, $\theta \in \mathbb{R}$,

$$\begin{aligned} \mathbb{E}(e^{i\theta Y_x(\lambda)}) &= \exp\left(x \int_0^\lambda (e^{i\theta y} - 1)\alpha y^{-\alpha-1} dy / \Gamma(1 - \alpha)\right) \\ &= \exp\left(x\lambda^{-\alpha} \int_0^1 (e^{i\theta\lambda y} - 1)\alpha y^{-\alpha-1} dy / \Gamma(1 - \alpha)\right) \\ &= \mathbb{E}(e^{i\theta\lambda Y_{x\lambda^{-\alpha}}(1)}), \end{aligned} \tag{6.1}$$

so by the inversion formula for absolutely continuous distributions [26, p. 9],

$$\begin{aligned}
 f_{Y_x(\lambda)}(1) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\theta} \mathbb{E}(e^{i\theta Y_x(\lambda)}) d\theta \\
 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\theta} \mathbb{E}(e^{i\theta \lambda Y_{x\lambda^{-\alpha}}(1)}) d\theta \\
 &= \frac{1}{2\pi \lambda} \int_{-\infty}^{\infty} e^{-i\theta/\lambda} \mathbb{E}(e^{i\theta Y_{x\lambda^{-\alpha}}(1)}) d\theta \\
 &= \lambda^{-1} f_{Y_{x\lambda^{-\alpha}}(1)}(\lambda^{-1}).
 \end{aligned}
 \tag{6.2}$$

Substituting in (2.8) and changing variable from $x/\Gamma(1-\alpha)$ to x , we obtain

$$\frac{1}{\Gamma(r)} \int_{x=0}^{y/\Gamma(1-\alpha)} x^{r-1} \int_{\lambda=0}^1 f_{\tilde{Y}_{x\lambda^{-\alpha}}}(\lambda^{-1}) e^{-x\lambda^{-\alpha}} \lambda^{-\alpha r-2} d\lambda dx,
 \tag{6.3}$$

where now $(\tilde{Y}_t)_{t>0} = (Y_{t\Gamma(1-\alpha)}(1))_{t>0}$ is a subordinator with Laplace transform

$$\mathbb{E}(e^{-\tau \tilde{Y}_t}) = \exp\left(-t \int_0^1 (1 - e^{-\tau y}) \alpha y^{-\alpha-1} dy\right), \quad \tau > 0, t \geq 0,
 \tag{6.4}$$

not depending on λ . To avoid carrying the factor $\Gamma(1-\alpha)$ along, we replace $y/\Gamma(1-\alpha)$ by y in (6.3), so that we are now dealing with the cumulative distribution function of $K(\alpha, r)\Gamma(1-\alpha)$. After a change of variable from $x\lambda^{-\alpha}$ to x , the expression in (6.3), with this replacement, equals

$$\begin{aligned}
 &\frac{1}{\Gamma(r)} \int_{\lambda=0}^1 \int_{x=0}^{y\lambda^{-\alpha}} x^{r-1} f_{\tilde{Y}_x}(\lambda^{-1}) e^{-x\lambda^{-2}} d\lambda dx \\
 &= \frac{1}{\Gamma(r)} \left[\int_{x=0}^y \int_{\lambda=0}^1 + \int_{x=y}^{\infty} \int_{\lambda=0}^{(y/x)^{1/\alpha}} \right] x^{r-1} f_{\tilde{Y}_x}(\lambda^{-1}) e^{-x\lambda^{-2}} d\lambda dx,
 \end{aligned}
 \tag{6.5}$$

after an interchange of integration order. Let Γ_r be a Gamma(r) random variable with density

$$f_{\Gamma_r}(x) = \frac{x^{r-1} e^{-x}}{\Gamma(r)}, \quad x > 0.$$

Changing variable from λ^{-1} to λ , we can write the right-hand side of (6.5) as

$$\begin{aligned}
 &\left[\int_{x=0}^y \int_{\lambda=1}^{\infty} + \int_{x=y}^{\infty} \int_{\lambda=(x/y)^{1/\alpha}}^{\infty} \right] f_{\tilde{Y}_x}(\lambda) f_{\Gamma_r}(x) d\lambda dx \\
 &= \int_{x=0}^y (1 - F_{\tilde{Y}_x}(1)) f_{\Gamma_r}(x) dx + \int_{x=y}^{\infty} (1 - F_{\tilde{Y}_x}((x/y)^{1/\alpha})) f_{\Gamma_r}(x) dx,
 \end{aligned}
 \tag{6.6}$$

where $F_{\tilde{Y}_x}$ is the cumulative distribution function of \tilde{Y}_x . This is for the component over $0 < \lambda \leq 1$.

For the integral over $\lambda \in (1, \infty)$ in (2.8) (with y replaced by $y/\Gamma(1-\alpha)$), we compute

$$\begin{aligned} & \frac{1}{\Gamma(r)} \int_{x=0}^y \int_{\lambda=1}^{\infty} x^{r-1} f_{\tilde{Y}_x}(1) e^{-x\lambda^{-\alpha r-1}} d\lambda dx \\ &= \frac{1}{\alpha r \Gamma(r)} \int_{x=0}^y x^{r-1} f_{\tilde{Y}_x}(1) e^{-x} dx = \frac{1}{\alpha r} \int_{x=0}^y f_{\tilde{Y}_x}(1) f_{\Gamma_r}(x) dx. \end{aligned} \tag{6.7}$$

Adding (6.6) and (6.7) gives an expression for the cumulative distribution function of $K(\alpha, r)\Gamma(1-\alpha)$ as a kind of gamma mixture:

$$\begin{aligned} \mathbb{P}(K(\alpha, r)\Gamma(1-\alpha) \leq y) &= \int_{x=0}^y \left(1 - F_{\tilde{Y}_x}(1) + \frac{1}{\alpha r} f_{\tilde{Y}_x}(1) \right) f_{\Gamma_r}(x) dx \\ &+ \int_{x=y}^{\infty} \left(1 - F_{\tilde{Y}_x}\left(\frac{x}{y}\right)^{1/\alpha} \right) f_{\Gamma_r}(x) dx. \end{aligned} \tag{6.8}$$

This can also be written in terms of the subordinated variable \tilde{Y}_{Γ_r} , where now $(\tilde{Y}_t)_{t>0}$ and Γ_r are taken to be independent. The standardised process $(\tilde{Y}_t)_{t>0}$ has the Laplace transform in (6.4), which is that of a truncated stable process. Notice that the effects of the parameters α and r are well separated in (6.8), which is helpful for numerical computation purposes. Equation (6.8) and the subordinated representation \tilde{Y}_{Γ_r} are also useful for simulating versions of the distribution.

Some further reductions are helpful for numerical computations. The Gil-Pelaez [8] inversion formula gives for the cumulative distribution function of the subordinator (\tilde{Y}_t) the expression

$$\begin{aligned} F_{\tilde{Y}_t}(x) &= \frac{1}{2} - \frac{1}{\pi} \int_0^{\infty} \mathcal{I} \left[e^{-i\theta x} \mathbb{E} \left(e^{-i\theta \tilde{Y}_t} \right) \right] \frac{d\theta}{\theta} \\ &= \frac{1}{2} + \frac{1}{\pi} \int_0^{\infty} \sin \left(\theta x - t\theta^\alpha \int_0^\theta \alpha z^{-\alpha-1} \sin z dz \right) e^{-t\theta^\alpha \int_0^\theta (1-\cos z)\alpha z^{-\alpha-1} dz} \frac{d\theta}{\theta}. \end{aligned} \tag{6.9}$$

6.2. Moments of the limiting distribution

An advantage of the limiting random variable $K(\alpha, r)$ in (6.8) is that it has finite moments of all orders, whereas the Mittag-Leffler distribution of order α (the analogous limiting distribution for PD_α) has finite moments only of order less than α .

Proposition 6.1.

(a) For $q > 0$

$$\mathbb{E}(K^q(\alpha, r)) = \frac{\Gamma(r+q)\Gamma^q(1-\alpha)}{\Gamma(r)} \mathbb{E} \left(((\alpha, r+q)T)^{-\alpha q} \right), \tag{6.10}$$

where

(b)

$$\mathbb{E} \left(((\alpha, r+q)T)^{-\alpha q} \right) = \frac{1}{\Gamma(\alpha q)} \int_0^{\infty} \frac{\lambda^{\alpha q-1} d\lambda}{(1 + \alpha \int_0^1 (1 - e^{-\lambda x}) x^{-\alpha-1} dx)^{r+q}}. \tag{6.11}$$

Proof of Proposition 6.1: (a) This is a modification of the proof of Lemma 4.1. From (2.8) we can compute

$$\mathbb{E}(K^q(\alpha, r)) = \frac{1}{\Gamma(r)\Gamma^r(1-\alpha)} \times \int_{x=0}^{\infty} \int_{\lambda>0} x^{r+q-1} f_{Y_x(\lambda)}(1) \exp\left(-\frac{x(\lambda^{-\alpha} \vee 1)}{\Gamma(1-\alpha)}\right) \lambda^{-\alpha r-1} d\lambda dx. \tag{6.12}$$

Then, following a similar path as in the proof of Lemma 4.1, we split the λ -integral in (6.12) into two components: one over $\lambda \in (0, 1]$ and the other over $\lambda > 1$. For the component over $\lambda > 1$, the λ -integral gives $1/\alpha r$, as before. The x -integral for this component can again be computed in terms of gamma functions. The result is

$$\frac{\Gamma(r+q)\Gamma^q(1-\alpha)}{\alpha r \Gamma(r)} \times g_{\alpha, r+q}(1) = \frac{\Gamma(r+q)\Gamma^q(1-\alpha)}{\Gamma(r)} \times \alpha r \times c_{\alpha, r+q+1}, \tag{6.13}$$

where (5.13) was used in the last equality.

For the component over $\lambda \in (0, 1]$, using similar computations as in the proof of Lemma 4.1, we arrive at the expression

$$\frac{\Gamma(r+q)\Gamma^q(1-\alpha)}{\Gamma(r)} \int_1^{\infty} \lambda^{-\alpha q} g_{\alpha, r+q}(\lambda) d\lambda.$$

We can write this as

$$\frac{\Gamma(r+q)\Gamma^q(1-\alpha)}{\Gamma(r)} \mathbb{E}((\alpha, r+q)T)^{-\alpha q} - \frac{\Gamma(r+q)\Gamma^q(1-\alpha)}{\Gamma(r)} \int_0^1 \lambda^{-\alpha q} g_{\alpha, r+q}(\lambda) d\lambda, \tag{6.14}$$

in which the second component is, by (5.13),

$$\begin{aligned} & \frac{\Gamma(r+q)\Gamma^q(1-\alpha)}{\Gamma(r)} \int_0^1 \lambda^{-\alpha q} \times (\alpha r)^2 \times c_{\alpha, r+q+1} \times \lambda^{\alpha(r+q)-1} d\lambda \\ &= \frac{\Gamma(r+q)\Gamma^q(1-\alpha)}{\Gamma(r)} \times (\alpha r)^2 \times c_{\alpha, r+q+1} \times \frac{1}{\alpha r}. \end{aligned}$$

This equals the right-hand side of (6.13), so we have cancellation in (6.14), thereby proving (6.10).

For (6.11), we use (4.2) and the identity

$$\mathbb{E}(((\alpha, r)T)^{-s}) = \frac{1}{\Gamma(s)} \int_{\tau=0}^{\infty} \tau^{s-1} \int_{x=0}^{\infty} e^{-\tau x} g_{\alpha, r}(x) dx d\tau.$$

This completes the proof of Proposition 6.1. □

Tables 1 and 2 in Appendix B show the mean and variance of $K(\alpha, r)$ calculated from (6.10) for some values of α, r using the R package [25]. Some graphs of the density corresponding to (6.9) are also in Appendix B.

6.3. Concluding comments and some related literature

- (i) Zhou, Favaro and Walker in [30] consider a different generalised negative binomial process model constructed from a gamma mixture of Poisson processes. Their emphasis is on fitting it to an observed ‘frequency of frequencies’ sample (i.e., the observed \mathbf{M}_n , in our notation).
- (ii) The (generalised) Dickman function occurs in the calculation and simulation of perpetuities [7], and thereby with sorting routines in computer science [4], [19]. See also [1, Chapter 4] and [2] for relationships with size-biased random variables. In a recent paper [15] we considered size-biased sampling from the Dickman subordinator.

Appendix A. Proof of Lemma 3.1

Proof of Lemma 3.1: The inversion formula for the characteristic function of a discrete random variable [10, p. 233] gives

$$\begin{aligned}
 n\mathbb{P}\left(\sum_{i=1}^{\lfloor xn^\alpha \rfloor} X_{1n}(\lambda n) = n\right) &= \frac{n}{2\pi} \int_{-\pi}^{\pi} e^{-in\theta} \phi_{\lambda n}^{\lfloor xn^\alpha \rfloor}(\theta) d\theta \\
 &= \frac{1}{2\pi} \int_{-n\pi}^{n\pi} e^{-i\theta} \phi_{\lambda n}^{\lfloor xn^\alpha \rfloor}(\theta/n) d\theta.
 \end{aligned}
 \tag{A.1}$$

Formally taking the limit under the integral and using (3.37), we can write

$$\begin{aligned}
 \lim_{n \rightarrow \infty} n\mathbb{P}\left(\sum_{i=1}^{\lfloor xn^\alpha \rfloor} X_{1n}(\lambda n) = n\right) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\theta} (\mathbb{E}e^{i\theta Y_1(\lambda)})^x d\theta \\
 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\theta} \mathbb{E}e^{i\theta Y_x(\lambda)} d\theta.
 \end{aligned}
 \tag{A.2}$$

The last integral is absolutely convergent for each $x > 0$ and $\lambda > 0$, which can be checked as follows. By (4.6)

$$\left| \mathbb{E}e^{i\theta Y_x(\lambda)} \right| = \exp\left(-\frac{x}{\Gamma(1-\alpha)} \int_0^{\lambda \wedge 1} (1 - \cos \theta y) \alpha y^{-\alpha-1} dy\right).$$

For a given $\lambda > 0$ choose $|\theta| > 1/(\lambda \wedge 1) = 1/\lambda \vee 1$. Then since $1 - \cos y \geq y^2/4$ for $|y| \leq 1$, we get

$$\frac{x}{\Gamma(1-\alpha)} \int_0^{\lambda \wedge 1} (1 - \cos \theta y) \alpha y^{-\alpha-1} dy \geq \frac{x\theta^2}{4\Gamma(1-\alpha)} \int_0^{1/|\theta|} \alpha y^{1-\alpha} dy = \frac{\alpha x |\theta|^\alpha}{4(2-\alpha)}.$$

When $|\theta| \leq 1/\lambda \vee 1$, we have $|\mathbb{E}e^{i\theta Y_x(\lambda)}| \leq 1$. Hence for some $c_\alpha > 0$,

$$\int_{-\infty}^{\infty} |\mathbb{E}e^{i\theta Y_x(\lambda)}| d\theta \leq \int_{-\infty}^{\infty} \left(\mathbf{1}_{\{|\theta| \leq 1/\lambda \vee 1\}} + e^{-c_\alpha x |\theta|^\alpha} \mathbf{1}_{\{|\theta| > 1/\lambda \vee 1\}}\right) d\theta < \infty,$$

thus establishing the absolute convergence in (A.2). It follows that the right-hand side of (A.2) is the inverse to the characteristic function of a random variable $Y_x(\lambda)$ with a continuous bounded density and thus equals the right-hand side of (3.29).

To justify the limiting procedure which produces (A.2) we need a bound on the characteristic function in (A.1). For this, split the integral into three components,

$$I_1 = \int_{|\theta| \leq A}, \quad I_2 = \int_{A < |\theta| \leq \varepsilon n}, \quad I_3 = \int_{\varepsilon n < |\theta| \leq \pi n},$$

where $A > 0$ and $0 < \varepsilon < 1/4$ will be chosen large and small, respectively, later. The first component is, by (3.37),

$$I_1 = \frac{1}{2\pi} \int_{-A}^A e^{-i\theta} \phi_{\lambda n}^{\lfloor \lambda n^\alpha \rfloor}(\theta/n) d\theta,$$

and by dominated convergence it has limit

$$\frac{1}{2\pi} \int_{-A}^A e^{-i\theta} \mathbb{E} e^{i\theta Y_x(\lambda)} d\theta.$$

This is arbitrarily close to the integral on the right-hand side of (A.2) once A is large enough, depending on x and λ .

To deal with I_2 , we use an inequality of the form

$$|\phi|^{n^\alpha} = e^{\frac{1}{2}n^\alpha \log |\phi|^2} = e^{\frac{1}{2}n^\alpha \log(1-(1-|\phi|^2))} \leq e^{-\frac{1}{2}n^\alpha(1-|\phi|^2)}. \tag{A.3}$$

We see that we need a lower bound for $1 - |\phi_{\lambda n}^{\lfloor \lambda n^\alpha \rfloor}(\theta/n)|^2$. Let $X_{1n}^s(\lambda)$ denote a symmetrised version of $X_{1n}(\lambda)$, obtained by subtracting an independent copy, and having probability mass function $p_{nj}^s(\lambda)$, $-n \leq j \leq n$. Then, for $\theta \neq 0$,

$$\begin{aligned} 1 - |\phi_{\lambda n}^{\lfloor \lambda n^\alpha \rfloor}(\theta/n)|^2 &= \mathbb{E}(1 - \cos(\theta X_{1n}^s(\lambda n)/n)) \\ &= \sum_{j=-n}^n (1 - \cos(\theta j/n)) p_{nj}^s(\lambda n) \geq \frac{\theta^2}{4n^2} \sum_{|\theta j/n| \leq 1} j^2 p_{nj}^s(\lambda n) \\ &= \frac{\theta^2}{4n^2} \mathbb{E}((X_{1n}^s(\lambda n))^2 \mathbf{1}_{\{|X_{1n}^s(\lambda n)| < n/|\theta|\}}). \end{aligned} \tag{A.4}$$

To replace X_{1n}^s by X_{1n} in this we use the inequality

$$\begin{aligned} &\mathbb{E}((X_{1n}^s(\lambda n))^2 \mathbf{1}_{\{|X_{1n}^s(\lambda n)| < n/|\theta|\}}) \\ &\geq 2\mathbb{E}((X_{1n}(\lambda n))^2 \mathbf{1}_{\{X_{1n}(\lambda n) < n/|\theta|\}}) \mathbb{P}(X_{1n}(\lambda n) < n/|\theta|\}) \\ &\quad - 2\mathbb{E}^2(X_{1n}(\lambda n) \mathbf{1}_{\{X_{1n}(\lambda n) < n/|\theta|\}}). \end{aligned} \tag{A.5}$$

This follows from a general inequality proved as follows:

$$\begin{aligned} \int_{|X-X'| \leq a} (X-X')^2 dP &\geq \int_{X \leq a, X' \leq a} (X^2 - 2XX' + (X')^2) dP \\ &= 2\mathbb{E}(X^2 \mathbf{1}_{\{X < a\}}) \mathbb{P}(X < a) - 2\mathbb{E}^2(X \mathbf{1}_{\{X < a\}}), \end{aligned}$$

where X and X' are any two nonnegative i.i.d. random variables. (Here $a > 0$, and in the inequality we used that $0 < X \leq a$ and $0 < X' \leq a$ imply $X - X' \leq a$ and $X - X' \geq -a$, so that

$|X - X'| \leq a$.) Applying this to $X_{1n}^s(\lambda n)$ and $X_{1n}(\lambda n)$ with $a = n/|\theta|$ we get (A.5). Referring to (A.4) and (A.5), we need a lower bound for

$$\begin{aligned} \frac{\theta^2}{n^2} \mathbb{E}((X_{1n}(\lambda n))^2 \mathbf{1}_{\{X_{1n}(\lambda n) < n/|\theta|\}}) &= \frac{\theta^2}{n^2} \sum_{j=1}^{\lfloor n/|\theta| \rfloor} j^2 p_{nj}(\lambda n) \\ &\geq \frac{\alpha \theta^2}{n^2 \Gamma(1 - \alpha)} \int_0^{\lambda n} \sum_{j=1}^{\lfloor n/|\theta| \rfloor} \frac{j^2 z^j}{j!} z^{-\alpha-1} e^{-z} dz \\ &\geq \frac{\alpha \theta^2}{n^\alpha \Gamma(1 - \alpha)} \int_0^\lambda \sum_{j=2}^{\lfloor n/|\theta| \rfloor} \frac{(nz)^{j-2}}{(j-2)!} z^{1-\alpha} e^{-nz} dz. \end{aligned} \tag{A.6}$$

The first inequality in (A.6) follows because the denominator of $p_{nj}(\lambda n)$, by (3.13) and (3.14), is less than $\Gamma(1 - \alpha)$, and the second inequality follows just because $j^2 \geq j(j - 1)$. The right-hand side of (A.6) equals

$$\frac{\alpha \theta^2}{n^\alpha \Gamma(1 - \alpha)} \int_0^\lambda \left(1 - \mathbb{P} \left(\text{Pois}(nz) \geq \frac{n}{|\theta|} - 2 \right) \right) z^{1-\alpha} dz, \tag{A.7}$$

in which, by Markov’s inequality, and because $|\theta| \leq \varepsilon n$ in I_2 ,

$$\mathbb{P} \left(\text{Pois}(nz) \geq \frac{n}{|\theta|} - 2 \right) \leq \frac{nz}{n/|\theta| - 2} \leq \frac{z|\theta|}{1 - 2\varepsilon}.$$

Now choose $A > 1/\lambda$ for the given $\lambda > 0$. Then since $A < |\theta| \leq \varepsilon n$ in I_2 , we have $1/|\theta| < \lambda$. It follows that the integral in (A.7) is no smaller than

$$\begin{aligned} &\frac{\alpha \theta^2}{n^\alpha \Gamma(1 - \alpha)} \int_0^{1/2|\theta|} \left(1 - \frac{z|\theta|}{1 - 2\varepsilon} \right) z^{1-\alpha} dz \\ &\geq \frac{1 - 4\varepsilon}{2(1 - 2\varepsilon)} \times \frac{\alpha |\theta|^\alpha}{(2 - \alpha) 2^{2-\alpha} n^\alpha \Gamma(1 - \alpha)} =: \frac{C_1(\varepsilon, \alpha) |\theta|^\alpha}{n^\alpha}. \end{aligned}$$

Returning to (A.5), we also need a lower bound for the probability term. By (3.13) and (3.14),

$$\begin{aligned} \mathbb{P}(X_{1n}(\lambda n) < n/|\theta|) &\geq \frac{\alpha}{\Gamma(1 - \alpha)} \int_0^{\lambda n} \sum_{j=1}^{\lfloor n/|\theta| \rfloor} \frac{z^j}{j!} z^{-\alpha-1} e^{-z} dz \\ &= \frac{\alpha}{n^\alpha \Gamma(1 - \alpha)} \int_0^\lambda \sum_{j=1}^{\lfloor n/|\theta| \rfloor} \frac{(nz)^j}{j!} e^{-nz} z^{-\alpha-1} dz \\ &= \frac{\alpha}{n^\alpha \Gamma(1 - \alpha)} \int_0^\lambda \mathbb{P} \left(1 \leq \text{Pois}(nz) \leq \frac{n}{|\theta|} \right) z^{-\alpha-1} dz. \end{aligned} \tag{A.8}$$

Again since $A > 1/\lambda$ and $A < |\theta| \leq \varepsilon n$ in I_2 , we have $1/|\theta| < \lambda$, so the last expression is no smaller than

$$\frac{\alpha}{n^\alpha \Gamma(1 - \alpha)} \int_{1/n}^{1/2|\theta|} \mathbb{P} \left(1 \leq \text{Pois}(nz) \leq \frac{n}{|\theta|} \right) z^{-\alpha-1} dz. \tag{A.9}$$

The probability in the integrand is, by Markov’s inequality,

$$1 - \mathbb{P}\left(\text{Poiss}(nz) > \frac{n}{|\theta|}\right) - \mathbb{P}(\text{Poiss}(nz) = 0) \geq 1 - \frac{nz|\theta|}{n} - e^{-nz} \geq \frac{1}{10},$$

so we obtain

$$\mathbb{P}\left(X_{1n}(\lambda n) < \frac{n}{|\theta|}\right) \geq \frac{n^\alpha - (2|\theta|)^\alpha}{10n^\alpha \Gamma(1 - \alpha)} = \frac{1}{10\Gamma(1 - \alpha)} \left(1 - \frac{(2|\theta|)^\alpha}{n^\alpha}\right). \tag{A.10}$$

Since $|\theta| < \varepsilon n$, the term in brackets is no smaller than $(1 - (2\varepsilon)^\alpha) > 0$ for a small enough choice of ε . So for the first term on the right-hand side of (A.5) we have the lower bound

$$2\mathbb{E}\left((X_{1n}(\lambda n))^2 \mathbf{1}_{\{X_{1n}(\lambda n) < n/|\theta|\}}\right) \mathbb{P}(X_{1n}(\lambda n) < n/|\theta|) \geq \frac{C_2(\varepsilon, \alpha) |\theta|^\alpha}{n^\alpha}, \tag{A.11}$$

where $C_2(\varepsilon, \alpha) := 2C_1(1 - (2\varepsilon)^\alpha)/10\Gamma(1 - \alpha)$.

For the \mathbb{E}^2 term on the right-hand side of (A.5), use the formula

$$\mathbb{E}(X_{1n}(\lambda n) \mathbf{1}_{\{X_{1n}(\lambda n) < n/|\theta|\}}) = \sum_{j=1}^{\lfloor n/|\theta| \rfloor} j p_{nj}(\lambda n) = \frac{\sum_{j=1}^{\lfloor n/|\theta| \rfloor} j F_j(\lambda n)}{\sum_{j=1}^n F_j(\lambda n)} \tag{A.12}$$

(cf. (3.16)), and for the denominator, use the lower bound

$$\begin{aligned} \sum_{j=1}^n F_j(\lambda n) &= \alpha \int_0^{\lambda n} \sum_{j=1}^n \frac{z^j}{j!} z^{-\alpha-1} e^{-z} dz \\ &\geq \alpha \int_0^{1/\varepsilon} \sum_{j=1}^n \frac{z^j}{j!} z^{-\alpha-1} e^{-z} dz \\ &\rightarrow \alpha \int_0^{1/\varepsilon} z^{-\alpha-1} (1 - e^{-z}) dz \text{ (as } n \rightarrow \infty) \\ &\geq \frac{1}{2} \Gamma(1 - \alpha) \end{aligned} \tag{A.13}$$

(for ε small enough). Thus the lower bound $\sum_{j=1}^n F_j(\lambda n) \geq \Gamma(1 - \alpha)/2$ holds for $\lambda > 1/|\theta|$, $|\theta| \leq \varepsilon n$, n greater than or equal to some $n_0(\varepsilon, \alpha)$, and ε less than or equal to some $\varepsilon_0(\alpha) > 0$. To deal with the numerator in the \mathbb{E}^2 term, write, as in (3.17),

$$\begin{aligned} n^{\alpha-1} \sum_{j=1}^{\lfloor n/|\theta| \rfloor} j F_j(\lambda n) &= \alpha n^{\alpha-1} \int_0^{\lambda n} \sum_{j=1}^{\lfloor n/|\theta| \rfloor} \frac{z^{j-1}}{(j-1)!} z^{-\alpha} e^{-z} dz \\ &= \alpha \int_0^\lambda \sum_{j=1}^{\lfloor n/|\theta| \rfloor} \frac{(nz)^{j-1}}{(j-1)!} z^{-\alpha} e^{-nz} dz \\ &\leq \alpha \int_0^\lambda \mathbb{P}(\text{Poiss}(nz) \leq n/|\theta|) z^{-\alpha} dz. \end{aligned} \tag{A.14}$$

This time we keep $|\theta| > 2/\lambda$ and upper-bound the right-hand side of (A.14) by

$$\begin{aligned} &\alpha \left(\int_0^{2/|\theta|} + \int_{2/|\theta|}^\lambda \right) \mathbb{P}(\text{Poiss}(nz) \leq n/|\theta|) z^{-\alpha} dz \\ &\leq \alpha \int_0^{2/|\theta|} z^{-\alpha} dz + \int_{2/|\theta|}^\lambda \mathbb{P}(\text{Poiss}(nz) \leq n/|\theta|) z^{-\alpha} dz. \end{aligned}$$

Here the first integral equals

$$C_3(\alpha)|\theta|^{\alpha-1}, \quad \text{where } C_3(\alpha) := \alpha 2^{1-\alpha}/(1-\alpha).$$

The second integral can be bounded using Chebyshev’s inequality as

$$\begin{aligned} & \alpha \int_{2/|\theta|}^\lambda \mathbb{P}(\text{Pois}(nz) - nz \leq -n/(z - 1/|\theta|))z^{-\alpha} dz \\ & \leq \alpha \int_{2/|\theta|}^\lambda \frac{nz^{1-\alpha}}{n^2(z - 1/|\theta|)^2} dz \\ & \leq \frac{\alpha|\theta|^\alpha}{n} \int_2^\infty \frac{y^{1-\alpha}}{(y - 1)^2} dy \leq \varepsilon C_4(\alpha)|\theta|^{\alpha-1} \leq C_4(\alpha)|\theta|^{\alpha-1}, \end{aligned}$$

using $|\theta| \leq \varepsilon n, n \geq |\theta|/\varepsilon$, and $\varepsilon \leq 1$ in the last inequality. Combining this with the first integral, we now have a bound for the left-hand side of (A.14) of the form $C_5(\alpha)|\theta|^{\alpha-1}$, where $C_5(\alpha) = C_3(\alpha) + C_4(\alpha)$. This leads to the bound

$$\begin{aligned} \frac{\theta^2}{n^2} \mathbb{E}^2(X_{1n}(\lambda n) \mathbf{1}_{\{X_{1n}(\lambda n) < n/|\theta|\}}) & \leq \frac{4\theta^2}{n^2 \Gamma(1-\alpha)} (C_5(\alpha)n^{1-\alpha}|\theta|^{\alpha-1})^2 \\ & = \frac{C_6(\alpha)}{n^{2\alpha}} |\theta|^{2\alpha} \leq \frac{\varepsilon^\alpha C_6(\alpha)}{n^\alpha} |\theta|^\alpha \end{aligned}$$

for the \mathbb{E}^2 term on the right-hand side of (A.5). This is smaller than the first term on the right-hand side of (A.5), which is bounded below in (A.11), giving a lower bound for the left-hand side of (A.4) as $C_7(\varepsilon, \alpha)|\theta|^\alpha/n^\alpha$, where $C_7(\varepsilon, \alpha) = 2(C_2(\varepsilon, \alpha) - \varepsilon C_6(\varepsilon, \alpha))$. Going back to (A.4), we now have a lower bound for

$$1 - |\phi_{\lambda n}^{\lfloor n^\alpha \rfloor}(\theta/n)|^2$$

of the form $C_7(\varepsilon, \alpha)|\theta|^\alpha/n^\alpha$. Then from (A.3),

$$|\phi_{\lambda n}(\theta/n)|^{n^\alpha} \leq e^{-\frac{1}{2}C_7(\varepsilon, \alpha)|\theta|^\alpha} \tag{A.15}$$

is integrable on $(0, \infty)$ and provides the required upper bound for I_2 .

Finally, to deal with I_3 , we use the fact that $X_{1n}(\lambda n)$ is a lattice variable with span 1 (it takes values $j = 1, 2, \dots, n$ with probabilities $p_{nj}(\lambda) > 0, \sum_{j=1}^n p_{nj}(\lambda) = 1$.) Thus by Corollary 2 to Theorem 5 in Section 14 of [10], for all $\varepsilon > 0$ there is a $c = c(\varepsilon, \lambda)$ such that for $\varepsilon < |\theta| < 2\pi - \varepsilon$ we have $|\phi_{\lambda n}(\theta)| < e^{-c}$. Thus

$$|\phi_{\lambda n}(\theta/n)|^{n^\alpha} < e^{-cn^\alpha}$$

for $\varepsilon n < |\theta| < n\pi$, and so

$$|I_3| \leq \frac{1}{2\pi} \int_{\varepsilon n < |\theta| < n\pi} |\phi_{\lambda n}^{\lfloor xn^\alpha \rfloor}(\theta/n)| d\theta \leq ne^{-xcn^\alpha} \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

This completes the proof of Lemma 3.1. □

TABLE 1. Expected value of $K(\alpha, r)$.

	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.7$	$\alpha = 0.9$
$r = 0.5$	1.62	1.56	1.42	1.19
$r = 1$	2.10	1.94	1.66	1.28
$r = 2$	2.99	2.58	2.01	1.39
$r = 4$	4.58	3.57	2.48	1.50
$r = 8$	7.27	5.02	3.07	1.61
$r = 16$	11.73	7.09	3.79	1.73

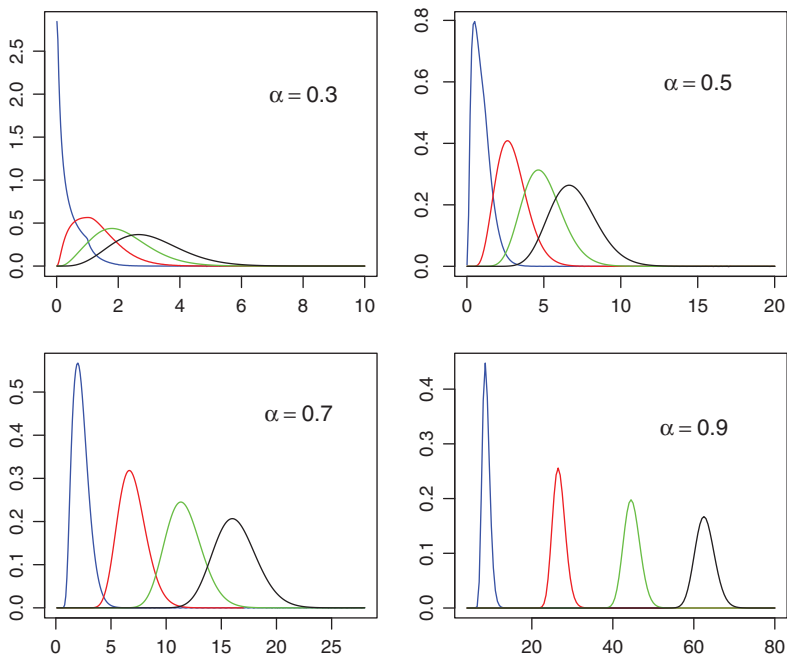


FIGURE 1: Densities of the subordinator \tilde{Y}_t for values of $\alpha = 0.3, 0.5, 0.7, 0.9$ and values of t being (left to right) 1 (blue), 2 (red), 5 (green), and 7 (black). The horizontal axis shows values of x ; the vertical axis shows values of $f_{\tilde{Y}_t}(x)$. \tilde{Y}_t tends to 0 in probability as $t \downarrow 0$, for large t , and tends to normality with mean and variance proportional to t as $t \rightarrow \infty$.

Appendix B. Mean and variance of $K(\alpha, r)$, and density of \tilde{Y}_t

Tables 1 and 2 show the mean and variance of $K(\alpha, r)$ for various values of α and r calculated from (6.10) using the R package [25].

The tables show that the expected value of $K(\alpha, r)$ increases with r for each $\alpha \in (0, 1)$, while the variance of $K(\alpha, r)$ increases with r for $\alpha \leq 1/2$ but, curiously, decreases with r for $\alpha > 1/2$, eventually tending to 0 as $\alpha \uparrow 1$ or $r \rightarrow \infty$.

Figure 1 shows plots of the density of the subordinator \tilde{Y}_t having Laplace transform (6.4) for various values of α and t . The functions were calculated in the package R by evaluating the cumulative distribution function in (6.9), then using R’s numerical differentiation routine.

TABLE 2. Variance of $K(\alpha, r)$.

	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.7$	$\alpha = 0.9$
$r = 0.5$	1.35	0.85	0.37	0.06
$r = 1$	1.65	0.92	0.34	0.04
$r = 2$	2.13	0.98	0.28	0.02
$r = 4$	2.80	1.02	0.22	0.01
$r = 8$	3.70	1.03	0.16	0.01
$r = 16$	4.89	1.04	0.12	0.00

References

- [1] ARRATIA, R., BARBOUR, A. AND TAVARÉ, S. (2003). *Logarithmic Combinatorial Structures: a Probabilistic Approach*. European Mathematical Society, Zurich.
- [2] ARRATIA, R. AND BAXENDALE, P. (2015). Bounded size bias coupling: a Gamma function bound, and universal Dickman-function behavior. *Prob. Theory Relat. Fields* **162**, 411–429.
- [3] CARAVENNA, F., SUN, R. AND ZYGOURAS, N. (2019). The Dickman subordinator, renewal theorems, and disordered systems. *Electron. J. Prob.* **24**, paper no. 101.
- [4] DIACONIS, P. (1980). Average running time of the fast Fourier transform. *J. Algorithms* **1**, 187–208.
- [5] DICKMAN, K. (1930). On the frequency of numbers containing primes of a certain relative magnitude. *Ark. Mat. Astron. Fys.* **22**, 1–14.
- [6] EWENS, W. (1972). The sampling theory of selectively neutral alleles. *Theoret. Pop. Biol.* **3**, 87–112.
- [7] FILL, J. A. AND HUBER, M. L. (2010). Perfect simulation of Vervaat perpetuities. *Electron. J. Prob.* **15**, 96–109.
- [8] GIL-PELAEZ, J. (1951). Note on the inversion theorem. *Biometrika* **38**, 481–482.
- [9] GREGOIRE, G. (1984). Negative binomial distributions for point processes. *Stoch. Process. Appl.* **16**, 179–188.
- [10] GNEDENKO, B. V. AND KOLMOGOROV, A. N. (1968). *Limit Distributions for Sums of Independent Random Variables*. Addison-Wesley, Cambridge, MA.
- [11] HANDA, K. (2009). The two-parameter Poisson–Dirichlet point process. *Bernoulli* **15**, 1082–1116.
- [12] IPSEN, Y. F. AND MALLER, R. A. (2017). Negative binomial construction of random discrete distributions on the infinite simplex. *Theory Stoch. Process.* **22**, 34–46.
- [13] IPSEN, Y. F., MALLER, R. A. AND RESNICK, S. (2020). Trimmed Lévy processes and their extremal components. *Stoch. Process. Appl.* **130**, 2228–2249.
- [14] IPSEN, Y. F., MALLER, R. A. AND SHEMEHSAVAR, S. (2019). Limiting distributions of generalised Poisson–Dirichlet distributions based on negative binomial processes. *J. Theoret. Prob.* **33**, 1974–2000.
- [15] IPSEN, Y. F., MALLER, R. A. AND SHEMEHSAVAR, S. (2020). Size biased sampling from the Dickman subordinator. *Stoch. Process. Appl.* **130**, 6880–6900.
- [16] IPSEN, Y. F., SHEMEHSAVAR, S. AND MALLER, R. A. (2018). Species sampling models generated by negative binomial processes. Preprint. Available at <https://arxiv.org/abs/1904.13046>.
- [17] KALLENBERG, O. (2002). *Foundations of Modern Probability*, 2nd edn. Springer, New York.
- [18] KINGMAN, J. F. C. (1975). Random discrete distributions. *J. R. Statist. Soc. B* **37**, 1–22.
- [19] MAHMOUD, H. M., MODARRES, R. E. AND SMYTHE, R. T. (1995). Analysis of quickselect: an algorithm for order statistics. *RAIRO Inf. Théor. Appl.* **29**, 255–276.
- [20] MOREE, P. (2013). Nicolaas Govert de Bruijn, the enchanter of friable integers. *Indag. Math.* **24**, 774–801.
- [21] PENROSE, M. D. AND WADE, A. R. (2004). Random minimal directed spanning trees and Dickman-type distributions. *Adv. Appl. Prob.* **36**, 691–714.
- [22] PINSKY, R. G. (2018). On the strange domain of attraction to generalized Dickman distributions for sums of independent random variables. *Electron. J. Prob.* **23**, 1–17.
- [23] PITMAN, J. (2006). *Combinatorial Stochastic Processes*. Springer, Berlin.
- [24] PITMAN, J. AND YOR, M. (1997). The two-parameter Poisson–Dirichlet distribution derived from a stable subordinator. *Ann. Prob.* **25**, 855–900.
- [25] R Core Team (2013). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.
- [26] SATO, K. I. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press.
- [27] WATTERSON, G. A. (1974). Models for the logarithmic species abundance distributions. *Theoret. Pop. Biol.* **6**, 217–250.

- [28] WATTERSON, G. A. (1976). The stationary distribution of the infinitely-many neutral alleles diffusion model. *J. Appl. Prob.* **13**, 639–651.
- [29] WATTERSON, G. A. AND GUESS, H. A. (1977). Is the most frequent allele the oldest? *Theoret. Pop. Biol.* **11**, 141–160.
- [30] ZHOU, M., FAVARO, S. AND WALKER, S. G. (2017). Frequency of frequencies distributions and size-dependent exchangeable random partitions. *J. Amer. Statist. Assoc.* **112**, 1623–1635.