Article

A Compound Poisson Perspective of Ewens–Pitman Sampling Model

Emanuele Dolera [1,2,3,*] and Stefano Favaro [2,3,4]

1 Department of Mathematics, University of Pavia, Via Adolfo Ferrata 5, 27100 Pavia, Italy
2 Collegio Carlo Alberto, Piazza V. Arbarello 8, 10122 Torino, Italy; stefano.favaro@unito.it
3 IMATI-CNR "Enrico Magenes", 27100 Pavia, Italy
4 Department of Economic and Social Sciences, Mathematics and Statistics, University of Torino, Corso Unione Sovietica 218/bis, 10134 Torino, Italy
* Correspondence: emanuele.dolera@unipv.it

Abstract: The Ewens–Pitman sampling model (EP-SM) is a distribution for random partitions of the set $\{1, \dots, n\}$, with $n \in \mathbb{N}$, which is indexed by real parameters α and θ such that either $\alpha \in [0, 1)$ and $\theta > -\alpha$, or $\alpha < 0$ and $\theta = -m\alpha$ for some $m \in \mathbb{N}$. For $\alpha = 0$, the EP-SM is reduced to the Ewens sampling model (E-SM), which admits a well-known compound Poisson perspective in terms of the log-series compound Poisson sampling model (LS-CPSM). In this paper, we consider a generalisation of the LS-CPSM, referred to as the negative Binomial compound Poisson sampling model (NB-CPSM), and we show that it leads to an extension of the compound Poisson perspective of the E-SM to the more general EP-SM for either $\alpha \in (0, 1)$, or $\alpha < 0$. The interplay between the NB-CPSM and the EP-SM is then applied to the study of the large n asymptotic behaviour of the number of blocks in the corresponding random partitions—leading to a new proof of Pitman's α diversity. We discuss the proposed results and conjecture that analogous compound Poisson representations may hold for the class of α-stable Poisson–Kingman sampling models—of which the EP-SM is a noteworthy special case.

Keywords: Berry–Esseen type theorem; Ewens–Pitman sampling model; exchangeable random partitions; log-series compound poisson sampling model; Mittag–Leffler distribution function; negative binomial compound poisson sampling model; Pitman's α-diversity; wright distribution function

1. Introduction

The Pitman–Yor process is a discrete random probability measure indexed by real parameters α and θ such that either $\alpha \in [0, 1)$ and $\theta > -\alpha$, or $\alpha < 0$ and $\theta = -m\alpha$ for some $m \in \mathbb{N}$—as can be seen in, e.g., Perman et al. [1], Pitman [2] and Pitman and Yor [3]. Let $\{V_i\}_{i \geq 1}$ be independent random variables such that V_i is distributed as a Beta distribution with parameter $(1 - \alpha, \theta + i\alpha)$, for $i \geq 1$, with the convention for $\alpha < 0$ that $V_m = 1$ and V_i is undefined for $i > m$. If $P_1 := V_1$ and $P_i := V_i \prod_{1 \leq j \leq i-1} (1 - V_j)$ for $i \geq 2$, such that almost definitely $\sum_{i \geq 1} P_i = 1$, then the Pitman–Yor is the random probability measure $\tilde{p}_{\alpha, \theta}$ on $(\mathbb{N}, 2^{\mathbb{N}})$ such that $\tilde{p}_{\alpha, \theta}(\{i\}) = P_i$ for $i \geq 1$. The Dirichlet process (Ferguson [4]) arises for $\alpha = 0$. Because of the discreteness of $\tilde{p}_{\alpha, \theta}$, a random sample $(X_1, \dots, X_n)$ induces a random partition Π_n of $\{1, \dots, n\}$ by means of the equivalence $i \sim j \iff X_i = X_j$ (Pitman [5]). Let $K_n(\alpha, \theta) := K_n(X_1, \dots, X_n) \leq n$ be the number of blocks of Π_n and let $M_{r,n}(\alpha, \theta) := M_{r,n}(X_1, \dots, X_n)$, for $r = 1, \dots, n$, be the number of blocks with frequency r of Π_n with $\sum_{1 \leq r \leq n} M_{r,n} = K_n$ and $\sum_{1 \leq r \leq n} r M_{r,n} = n$. Pitman [2] showed that:

$$\Pr[(M_{1,n}(\alpha, \theta), \dots, M_{n,n}(\alpha, \theta)) = (x_1, \dots, x_n)] = n! \frac{\left(\frac{\theta}{\alpha}\right)_{(\sum_{i=1}^n x_i)}}{(\theta)_{(n)}} \prod_{i=1}^n \frac{\left(\frac{\alpha(1-\alpha)_{(i-1)}}{i!}\right)^{x_i}}{x_i!}, \quad (1)$$

with $(x)_{(n)}$ being the ascending factorial of x of order n, i.e., $(x)_{(n)} := \prod_{0 \leq i \leq n-1}(x+i)$. The distribution (1) is referred to as the Ewens–Pitman sampling model (EP-SM), and for $\alpha = 0$, it reduces to the Ewens sampling model (E-SM) in Ewens [6]. The Pitman–Yor process plays a critical role in a variety of research areas, such as mathematical population genetics, Bayesian nonparametrics, machine learning, excursion theory, combinatorics and statistical physics. See Pitman [5] and Crane [7] for a comprehensive treatment of this subject.

The E-SM admits a well-known compound Poisson perspective in terms of the log-series compound Poisson sampling model (LS-CPSM). See Charalambides [8] and the references therein for an overview of compound Poisson models. We consider a population of individuals with a random number K of distinct types, and let K be distributed as a Poisson distribution with parameter $\lambda = -z \log(1-q)$ for $q \in (0,1)$ and $z > 0$. For $i \in \mathbb{N}$, let N_i denote the random number of individuals of type i in the population, and let the N_i's be independent of K and independent from each other, with the same distribution:

$$\Pr[N_1 = x] = -\frac{1}{x \log(1-q)} q^x \tag{2}$$

for $x \in \mathbb{N}$. Let $S = \sum_{1 \leq i \leq K} N_i$ and let $M_r = \sum_{1 \leq i \leq K} \mathbb{1}_{\{N_i = r\}}$ for $r = 1, \ldots, S$, that is, M_r is the random number of N_i equal to r such that $\sum_{r \geq 1} M_r = K$ and $\sum_{r \geq 1} r M_r = S$. If $(M_1(z,n), \ldots, M_n(z,n))$ denotes a random variable whose distribution coincides with the conditional distribution of $(M_1, \ldots, M_S)$ given $S = n$, then (Section 3, Charalambides [8]) it holds:

$$\Pr[(M_1(z,n), \ldots, M_n(z,n)) = (x_1, \ldots, x_n)] = \frac{n!}{(z)_{(n)}} \prod_{i=1}^{n} \frac{\left(\frac{z}{i}\right)^{x_i}}{x_i!}. \tag{3}$$

The distribution (3) is referred to as the LS-CPSM, and it is equivalent to the E-SM. That is, the distribution (3) coincides with the distribution (1) with $\alpha = 0$. Therefore, the distributions of $K(z,n) = \sum_{1 \leq r \leq n} M_r(z,n)$ and $M_r(z,n)$ coincide with the distributions of $K_n(0,z)$ and $M_{r,n}(0,z)$, respectively. Let $\xrightarrow{w}$ denote the weak convergence. From Korwar and Hollander [9], $K(z,n)/\log n \xrightarrow{w} z$ as $n \to +\infty$, whereas from Ewens [6], it follows that $M_r(z,n) \xrightarrow{w} P_{z/r}$ as $n \to +\infty$, where P_z is a Poisson random variable with parameter z.

In this paper, we consider a generalisation of the LS-CPSM referred to as the negative binomial compound Poisson sampling model (NB-CPSM). The NB-CPSM is indexed by real parameters α and z such that either $\alpha \in (0,1)$ and $z > 0$ or $\alpha < 0$ and $z < 0$. The LS-CPSM is recovered by letting $\alpha \to 0$ and $z > 0$. We show that the NB-CPSM leads to extend the compound Poisson perspective of the E-SM to the more general EP-SM for either $\alpha \in (0,1)$, or $\alpha < 0$. That is, we show that: (i) for $\alpha \in (0,1)$, the EP-SM (1) admits a representation as a randomised NB-CPSM with $\alpha \in (0,1)$ and $z > 0$, where the randomisation acts on z with respect a scale mixture between a Gamma and a scaled Mittag–Leffler distribution (Pitman [5]); (ii) for $\alpha < 0$ the NB-CPSM admits a representation in terms of a randomised EP-SM with $\alpha < 0$ and $\theta = -m\alpha$ for some $m \in \mathbb{N}$, where the randomisation acts on m with respect to a tilted Poisson distribution arising from the Wright function (Wright [10]). The interplay between the NB-CPSM and the EP-SM is then applied to the large n asymptotic behaviour of the number of distinct blocks in the corresponding random partitions. In particular, by combining the randomised representation in (i) with the large n asymptotic behaviour or the number of distinct blocks under the NB-CPSM, we present a new proof of Pitman's α-diversity (Pitman [5]), namely the large n asymptotic behaviour of $K_n(\alpha, \theta)$ under the EP-SM for $\alpha \in (0,1)$.

2. A Compound Poisson Perspective of EP-SM

To introduce the NB-CPSM, we considered a population of individuals with a random number K of types and let K be distributed as a Poisson distribution with parameter $\lambda = z[1-(1-q)^{\alpha}]$ such that either $q \in (0,1)$, $\alpha \in (0,1)$ and $z > 0$, or $q \in (0,1)$, $\alpha < 0$ and

$z < 0$. For $i \in \mathbb{N}$, let N_i be the random number of individuals of type i in the population, and let the N_i be independent of K and independent from each other with the same distribution:

$$\Pr[N_1 = x] = -\frac{1}{[1 - (1-q)^\alpha]} \binom{\alpha}{x} (-q)^x \tag{4}$$

for $x \in \mathbb{N}$. Let $S = \sum_{1 \le i \le K} N_i$ and $M_r = \sum_{1 \le i \le K} \mathbb{1}_{\{N_i = r\}}$ for $r = 1, \ldots, S$, that is, M_r is the random number of N_i equal to r such that $\sum_{r \ge 1} M_r = K$ and $\sum_{r \ge 1} r M_r = S$. If $(M_1(\alpha, z, n), \ldots, M_n(\alpha, z, n))$ is a random variable whose distribution coincides with the conditional distribution of $(M_1, \ldots, M_S)$, given $S = n$, then it holds (Section 3, Charalambides [8]):

$$\Pr[(M_1(\alpha, z, n), \ldots, M_n(\alpha, z, n)) = (x_1, \ldots, x_n)] = \frac{n!}{\sum_{j=0}^{n} \mathscr{C}(n, j; \alpha) z^j} \prod_{i=1}^{n} \frac{\left[z \frac{\alpha(1-\alpha)_{(i-1)}}{i!} \right]^{x_i}}{x_i!}, \tag{5}$$

where $\mathscr{C}(n, j; \alpha) = \frac{1}{j!} \sum_{0 \le i \le j} \binom{j}{i} (-1)^i (-i\alpha)_{(n)}$ is the generalised factorial coefficient (Charalambides [11]), with the proviso $\mathscr{C}(n, 0, \alpha) = 0$ for all $n \in \mathbb{N}$, $\mathscr{C}(n, j, \alpha) = 0$ for all $j > n$ and $\mathscr{C}(0, 0, \alpha) = 1$. The distribution (5) is referred to as the NB-CPSM. As $\alpha \to 0$, the distribution (4) reduces to the distribution (2), and hence the NB-CPSM (5) is reduced to the LS-CPSM (3). The next theorem states the large n asymptotic behaviour of the counting statistics $K(\alpha, z, n) = \sum_{1 \le r \le n} M_r(\alpha, z, n)$ and $M_r(\alpha, z, n)$ arising from the NB-CPSM.

Theorem 1. *Let P_λ denote a Poisson random variable with the parameter $\lambda > 0$. As $n \to +\infty$,*
(i)   for $\alpha \in (0, 1)$ and $z > 0$:

$$K(\alpha, z, n) \xrightarrow{w} 1 + P_z \tag{6}$$

and:

$$M_r(\alpha, z, n) \xrightarrow{w} P_{\frac{\alpha(1-\alpha)_{(r-1)}}{r!} z}; \tag{7}$$

(ii)   for $\alpha < 0$ and $z < 0$:

$$\frac{K(\alpha, z, n)}{n^{\frac{-\alpha}{1-\alpha}}} \xrightarrow{w} \frac{(\alpha z)^{\frac{1}{1-\alpha}}}{-\alpha} \tag{8}$$

and:

$$M_r(\alpha, z, n) \xrightarrow{w} P_{\frac{\alpha(1-\alpha)_{(r-1)}}{r!} z}. \tag{9}$$

Proof. As regards the proof of (6), we start by recalling that the probability generating function $G(\cdot; \lambda)$ of P_λ is $G(s; \lambda) = \exp\{-\lambda(s-1)\}$ for any $s > 0$. Now, let $G(\cdot; \alpha, z, n)$ be the probability generating function of $K(\alpha, z, n)$. The distribution of $K(\alpha, z, n)$ follows by combining the NB-CPSM (5) with Theorem 2.15 of Charalambides [11]. In particular, it follows that:

$$G(s; \alpha, z, n) = \frac{\sum_{j=1}^{n} \mathscr{C}(n, j; \alpha)(sz)^j}{\sum_{j=1}^{n} \mathscr{C}(n, j; \alpha) z^j}.$$

Hereafter, we show that $G(s; \alpha, z, n) \to s \exp\{z(s-1)\}$ as $n \to +\infty$, for any $s > 0$, which implies (6). In particular, by the direct application of the definition of $\mathscr{C}(n, k; \alpha)$, we write the following:

$$\sum_{j=1}^{n} \mathscr{C}(n, j; \alpha) z^j = \sum_{i=1}^{n} (-1)^i (-i\alpha)_{(n)} \sum_{k=i}^{n} \frac{1}{k!} \binom{k}{i} z^k = \sum_{i=1}^{n} (-1)^i (-i\alpha)_{(n)} e^z z^i \frac{\Gamma(n-i+1, z)}{i! \Gamma(n-i+1)},$$

where $\Gamma(a, x) := \int_x^{+\infty} t^{a-1} e^{-t} \mathrm{d}t$ denotes the incomplete gamma function for $a, x > 0$ and $\Gamma(a) := \int_0^{+\infty} t^{a-1} e^{-t} \mathrm{d}t$ denotes the Gamma function for $a > 0$. Accordingly, we write the identity:

$$G(s; \alpha, z, n) = e^{z(s-1)} \frac{-zs \frac{\Gamma(n,zs)}{\Gamma(n)} + \sum_{i=2}^{n}(-1)^i \frac{(-i\alpha)_{(n)}}{(-\alpha)_{(n)}} (zs)^i \frac{\Gamma(n-i+1,zs)}{i!\Gamma(n-i+1)}}{-z \frac{\Gamma(n,z)}{\Gamma(n)} + \sum_{i=2}^{n}(-1)^i \frac{(-i\alpha)_{(n)}}{(-\alpha)_{(n)}} z^i \frac{\Gamma(n-i+1,z)}{i!\Gamma(n-i+1)}}.$$

Since $\lim_{n \to +\infty} \frac{\Gamma(n,x)}{\Gamma(n)} = 1$ for any $x > 0$, the proof (6) is completed by showing that, for any $t > 0$:

$$\lim_{n \to +\infty} \sum_{i=2}^{n} (-1)^i \frac{(-i\alpha)_{(n)}}{(-\alpha)_{(n)}} \frac{\Gamma(n-i+1,t)}{\Gamma(n-i+1)} \frac{t^i}{i!} = 0. \tag{10}$$

By the definition of ascending factorials and the reflection formula of the Gamma function, it holds:

$$\frac{(-i\alpha)_{(n)}}{(-\alpha)_{(n)}} = \frac{\Gamma(n - i\alpha)}{\Gamma(n - \alpha)} \frac{\sin i\pi\alpha}{\pi} \Gamma(i\alpha + 1)\Gamma(-\alpha).$$

In particular, by means of the monotonicity of the function $[1, +\infty) \ni z \mapsto \Gamma(z)$, we can write:

$$\frac{1}{i!} \left| \frac{(-i\alpha)_{(n)}}{(-\alpha)_{(n)}} \right| \leq \frac{|\Gamma(-\alpha)|}{\pi} \frac{\Gamma(n - 2\alpha)}{\Gamma(n - \alpha)} \frac{\Gamma(i\alpha + 1)}{i!} \tag{11}$$

for any $n \in \mathbb{N}$ such that $n > 1/(1 - \alpha)$, and $i \in \{2, \dots, n\}$. Note that $\frac{\Gamma(n,x)}{\Gamma(n)} \leq 1$. Then, we apply (11) to obtain:

$$\left| \sum_{i=2}^{n} (-1)^i \frac{(-i\alpha)_{(n)}}{(-\alpha)_{(n)}} \frac{\Gamma(n-i+1,t)}{\Gamma(n-i+1)} \frac{t^i}{i!} \right| \leq \sum_{i=2}^{n} \frac{t^i}{i!} \left| \frac{(-i\alpha)_{(n)}}{(-\alpha)_{(n)}} \right|$$

$$\leq \frac{|\Gamma(-\alpha)|}{\pi} \frac{\Gamma(n - 2\alpha)}{\Gamma(n - \alpha)} \sum_{i \geq 0} t^i \frac{\Gamma(i\alpha + 1)}{i!}.$$

Now, by means of Stirling approximation, it holds $\frac{\Gamma(n-2\alpha)}{\Gamma(n-\alpha)} \sim \frac{1}{n^\alpha}$ as $n \to +\infty$. Moreover, we have:

$$\sum_{i \geq 0} t^i \frac{\Gamma(i\alpha + 1)}{i!} = \int_0^{+\infty} e^{tz^\alpha - z} \mathrm{d}z < +\infty$$

where the finiteness of the integral follows, for any fixed $t > 0$, from the fact that $tz^\alpha < \frac{1}{2}z$ if $z > (2t)^{\frac{1}{1-\alpha}}$. This completes the proof of (10) and hence the proof of (6). As regards the proof of (7), we make use of the falling factorial moments of $M_r(\alpha, z, n)$, which follows by combining the NB-CPSM (5) with Theorem 2.15 of Charalambides [11]. Let $(a)_{[n]}$ be the falling factorial of a of order n, i.e., $(a)_{[n]} = \prod_{0 \leq i \leq n-1}(a - i)$, for any $a \in \mathbb{R}^+$ and $n \in \mathbb{N}_0$ with the proviso $(a)_{[0]} = 1$. Then, we write:

$$\mathbb{E}[(M_r(\alpha, z, n))_{[s]}]$$

$$= (-1)^{rs} (n)_{[rs]} \binom{\alpha}{r}^s (-z)^s \frac{\sum_{j=0}^{n-rs} \mathscr{C}(n - rs, j; \alpha) z^j}{\sum_{j=0}^{n} \mathscr{C}(n, j; \alpha) z^j}$$

$$= (-1)^{rs} (n)_{[rs]} \binom{\alpha}{r}^s (-z)^s \frac{(-z)\frac{\Gamma(n-rs,z)}{\Gamma(n-rs)} + \sum_{i=2}^{n-rs}(-1)^i \frac{(-i\alpha)_{(n-rs)}}{(-\alpha)_{(n-rs)}} (z)^i \frac{\Gamma(n-rs-i+1,z)}{i!\Gamma(n-rs-i+1)}}{(-z)\frac{\Gamma(n,z)}{\Gamma(n)} + \sum_{i=2}^{n}(-1)^i \frac{(-i\alpha)_{(n)}}{(-\alpha)_{(n)}} (z)^i \frac{\Gamma(n-i+1,z)}{\Gamma(n-i+1)}}$$

$$= (-1)^{rs} (n)_{[rs]} \binom{\alpha}{r}^s (-z)^s$$

$$\times \frac{(-\alpha)_{(n-rs)}}{(-\alpha)_{(n)}} \frac{(-z)\frac{\Gamma(n-rs,z)}{\Gamma(n-rs)} + \sum_{i=2}^{n-rs}(-1)^i \frac{(-i\alpha)_{(n-rs)}}{(-\alpha)_{(n-lr)}}(z)^i \frac{\Gamma(n-rs-i+1,z)}{i!\Gamma(n-rs-i+1)}}{(-z)\frac{\Gamma(n,z)}{\Gamma(n)} + \sum_{i=2}^{n}(-1)^i \frac{(-i\alpha)_{(n)}}{(-\alpha)_{(n)}}(z)^i \frac{\Gamma(n-i+1,z)}{\Gamma(n-i+1)}}.$$

Now, by means of the same argument applied in the proof of statement (6), it holds true that:

$$\lim_{n\to+\infty} \frac{(-z)\frac{\Gamma(n-rs,z)}{\Gamma(n-rs)} + \sum_{i=2}^{n-rs}(-1)^i \frac{(-i\alpha)_{(n-rs)}}{(-\alpha)_{(n-lr)}}(z)^i \frac{\Gamma(n-rs-i+1,z)}{i!\Gamma(n-rs-i+1)}}{(-z)\frac{\Gamma(n,z)}{\Gamma(n)} + \sum_{i=2}^{n}(-1)^i \frac{(-i\alpha)_{(n)}}{(-\alpha)_{(n)}}(z)^i \frac{\Gamma(n-i+1,z)}{\Gamma(n-i+1)}} = 1.$$

Then:

$$\lim_{n\to+\infty} \mathbb{E}[(M_r(\alpha,z,n))_{[s]}] = (-1)^{rs}\binom{\alpha}{r}^s (-z)^s = \left[\frac{\alpha(1-\alpha)_{(r-1)}}{r!}z\right]^s.$$

follows from the fact that $(n)_{[rs]} \sim \frac{(-\alpha)_{(n-rs)}}{(-\alpha)_{(n)}}$ as $n \to +\infty$. The proof of the large n asymptotics (7) is completed by recalling that the falling factorial moment of order s of P_λ is $\mathbb{E}[(P_\lambda)_{[s]}] = \lambda^s$.

As regards the proof of statement (8), let $\alpha = -\sigma$ for any $\sigma > 0$ and let $z = -\zeta$ for any $\zeta > 0$. Then, by direct application of Equation (2.27) of Charalambides [11], we write the following identity:

$$\sum_{j=0}^{n} \mathscr{C}(n,j;-\sigma)(-\zeta)^j = (-1)^n \sum_{v=0}^{n} s(n,v)(-\sigma)^v \sum_{j=0}^{v} \zeta^j S(v,j),$$

where $S(v,j)$ is the Stirling number of that second type. Now, note that $\sum_{0\le j\le v}^{v} \zeta^j S(v,j)$ is the moment of order v of a Poisson random variable with parameter $\zeta > 0$. Then, we write:

$$\sum_{j=0}^{n} \mathscr{C}(n,j;-\sigma)(-\zeta)^j = \sum_{v=0}^{n} |s(n,v)|\sigma^v \sum_{j\ge0} j^v e^{-\zeta}\frac{\zeta^j}{j!} = \sum_{j\ge0} e^{-\zeta}\frac{\zeta^j}{j!} \int_0^{+\infty} x^n f_{G_{\sigma j,1}}(x)\mathrm{d}x. \quad (12)$$

That is:

$$B_n(w) = \mathbb{E}[(G_{\sigma P_w,1})^n], \quad (13)$$

where $G_{a,1}$ and P_w are independent random variables such that $G_{a,1}$ is a Gamma random variable with a shape parameter $a > 0$ and a scale parameter 1, and P_w is a Poisson random variable with a parameter w. Accordingly, the distribution of $G_{\sigma P_w,1}$, say $\mu_{\sigma,w}$, is the following:

$$\mu_{\sigma,w}(\mathrm{d}t) = e^{-w}\delta_0(\mathrm{d}t) + \left(\sum_{j\ge1} \frac{e^{-w}w^j}{j!}\frac{1}{\Gamma(j\sigma)}e^{-t}t^{j\sigma-1}\right)\mathrm{d}t$$

for $t > 0$. The discrete component of $\mu_{\sigma,w}$ does not contribute to the expectation (13) so that we focus on the absolutely continuous component, whose density can be written as follows:

$$\sum_{j\ge1} \frac{e^{-w}w^j}{j!}\frac{1}{\Gamma(j\sigma)}e^{-t}t^{j\sigma-1} = \frac{e^{-(w+t)}}{t}W_{\sigma,0}(wt^\sigma),$$

where $W_{\sigma,\tau}(y) := \sum_{j\ge0} \frac{y^j}{j!\Gamma(j\sigma+\tau)}$ is the Wright function (Wright [10]). In particular, for $\tau = 0$:

$$B_n(w) = \int_0^{+\infty} t^n \frac{e^{-(w+t)}}{t}W_{\sigma,0}(wt^\sigma)\mathrm{d}t. \quad (14)$$

If we split the integral as $\int_0^M + \int_M^{+\infty}$ for any $M > 0$, the contribution of the latter integral is overwhelming with respect to the contribution of the former. Then, $W_{\sigma,0}$ can be equivalently replaced by the asymptotics $W_{\sigma,0}(y) \sim c(\sigma) y^{\frac{1}{2(1+\sigma)}} \exp\{\sigma^{-1}(\sigma+1)(\sigma y)^{\frac{1}{1+\sigma}}\}$, as $y \to +\infty$, for some constant $c(\sigma)$ solely depending on σ. See Theorem 2 in Wright [10]. Hence:

$$B_n(w) \sim c(\sigma) \int_0^{+\infty} t^{n-1} e^{-(w+t)} (wt^\sigma)^{\frac{1}{2(1+\sigma)}} \exp\left\{ \frac{\sigma+1}{\sigma}(\sigma wt^\sigma)^{\frac{1}{1+\sigma}} \right\} \mathrm{d}t$$

$$= c(\sigma) e^{-w} w^{\frac{1}{2(1+\sigma)}} \int_0^{+\infty} t^{n + \frac{\sigma}{2(1+\sigma)} - 1} \exp\{A(w,\sigma) t^{\frac{\sigma}{1+\sigma}} - t\} \mathrm{d}t,$$

where $A(w,\sigma) := \frac{\sigma+1}{\sigma}(\sigma w)^{\frac{1}{1+\sigma}}$. Then, the problem is reduced to an integral whose asymptotic behaviour is described in Berg [12]. From Equation (31) of the Berg [12] and Stirling approximation, we have:

$$B_n(w) \sim c(\sigma) e^{-w} w^{\frac{1}{2(1+\sigma)}} \Gamma(n) \exp\left\{ A(w,\sigma) n^{\frac{\sigma}{1+\sigma}} \right\}. \tag{15}$$

This last asymptotic expansion leads directly to (8). Indeed, let $G(\cdot; -\sigma, -\zeta, n)$ be the probability generating function of the random variable $K(-\sigma, -\zeta, n)$, which reads as $G(s; -\sigma, -\zeta, n) = B_n(s\zeta)/B_n(\zeta)$ for $s > 0$. Then, by means of (15), for any fixed $s > 0$ we write:

$$G(s; -\sigma, -\zeta, n) \sim e^{-w(s-1)} s^{\frac{1}{2(1+\sigma)}} \exp\left\{ n^{\frac{\sigma}{1+\sigma}} \frac{\sigma+1}{\sigma}(\sigma\zeta)^{\frac{1}{1+\sigma}} [s^{\frac{1}{1+\sigma}} - 1] \right\}. \tag{16}$$

Since (15) holds uniformly in w in a compact set, we consider the function $G(s; -\sigma, -\zeta, n)$ evaluated at some point s_n and extend the validity of (16) with s_n in the place of s, as long as $\{s_n\}_{n\geq 1}$ varies in a compact subset of $[0, +\infty)$. Thus, we can choose $s_n = s^{\beta(n)}$ and $\beta(n) = \frac{1}{n^{\frac{\sigma}{1+\sigma}}}$ and notice that $\beta(n) \to 0$ as $n \to +\infty$. Thus, $s_n \simeq 1 + \beta(n) \log s \to 1$ and we have:

$$n^{\frac{\sigma}{1+\sigma}} \frac{\sigma+1}{\sigma}(\sigma w)^{\frac{1}{1+\sigma}} [s_n^{\frac{1}{1+\sigma}} - 1] \to \frac{(\sigma\zeta)^{\frac{1}{1+\sigma}}}{\sigma} \log s,$$

which implies that $K(-\sigma, -\zeta, n) \to \frac{(\sigma\zeta)^{\frac{1}{1+\sigma}}}{\sigma}$ as $n \to +\infty$. This completes the proof of (8). As regards the proof (9), let $\alpha = -\sigma$ for any $\sigma > 0$ and let $z = -\zeta$ for any $\zeta > 0$. Similarly to the proof of (7), here we make use of the falling factorial moments of $M_r(-\sigma, -\zeta, n)$, that is:

$$\mathbb{E}[(M_r(-\sigma, \zeta, n))_{[s]}] = (-1)^{rs} (n)_{[rs]} \binom{-\sigma}{r}^s \zeta^s \frac{\sum_{j=0}^{n-rs} \mathscr{C}(n-rs, j; -\sigma)(-\zeta)^j}{\sum_{j=0}^n \mathscr{C}(n, j; -\sigma)(-\zeta)^j}.$$

At this point, we make use of the same large n arguments applied in the proof of statement (7). In particular, by means of the large n asymptotic (15), as $n \to +\infty$, it holds true that:

$$\frac{\sum_{j=0}^{n-rs} \mathscr{C}(n-rs, j; -\sigma)(-\zeta)^j}{\sum_{j=0}^n \mathscr{C}(n, j; -\sigma)(-\zeta)^j} \sim n^{-rs}.$$

Then:

$$\lim_{n\to+\infty} \mathbb{E}[(M_r(-\sigma, -\zeta, n))_{[s]}] = (-1)^{rs} \binom{-\sigma}{r}^s \zeta^s = \left[\frac{-\sigma(1+\sigma)_{(r-1)}}{r!} (-\zeta) \right]^s$$

it follows from the fact that $(n)_{[rs]} \sim n^{rs}$ as $n \to +\infty$. The proof of the large n asymptotics (9) is completed by recalling that the falling factorial moment of order s of P_λ is $\mathbb{E}[(P_\lambda)_{[s]}] = \lambda^s$. $\square$

In the rest of the section, we make use of the NB-CPSM (5) to introduce a compound Poisson perspective of the EP-SM. In particular, our result extends the well-known compound Poisson perspective of the E-SM to the EP-SM for either $\alpha \in (0,1)$, or $\alpha < 0$. For $\alpha \in (0,1)$ let f_α denote the density function of a positive α-stable random variable X_α, that is X_α is a random variable for which $\mathbb{E}[\exp\{-tX_\alpha\}] = \exp\{-t^\alpha\}$ for any $t > 0$. For $\alpha \in (0,1)$ and $\theta > -\alpha$, let $S_{\alpha,\theta}$ be a positive random variable with the density function:

$$f_{S_{\alpha,\theta}}(s) = \frac{\Gamma(\theta+1)}{\alpha\Gamma(\theta/\alpha+1)}s^{\frac{\theta-1}{\alpha}-1}f_\alpha(s^{-\frac{1}{\alpha}}).$$

That is, $S_{\alpha,\theta}$ is a scaled Mittag–Leffler random variable (Chapter 1, Pitman [5]). Let $G_{a,b}$ be a Gamma random variable with the scale parameter $b > 0$ and shape parameter $a > 0$, and let us assume that $G_{a,b}$ is independent of $S_{\alpha,\theta}$. Then, for $\alpha \in (0,1)$, $\theta > -\alpha$ and $n \in \mathbb{N}$ let:

$$\bar{X}_{\alpha,\theta,n} \stackrel{d}{=} G^\alpha_{\theta+n,1}S_{\alpha,\theta}. \tag{17}$$

Finally, for $\alpha < 0$, $z < 0$ and $n \in \mathbb{N}$, let $\tilde{X}_{\alpha,z,n}$ be a random variable on $\mathbb{N}$ whose distribution is a tilted Poisson distribution arising from the identity (12). Precisely, for any $x \in \mathbb{N}$:

$$\Pr[\tilde{X}_{\alpha,z,n} = x] = \frac{1}{\sum_{j=1}^{n}\mathscr{C}(n,j;\alpha)z^j}\frac{e^z(-z)^x\Gamma(-x\alpha+n)}{x!\Gamma(-x\alpha)}. \tag{18}$$

The next theorem makes use of $\bar{X}_{\alpha,\theta,n}$ and $\tilde{X}_{\alpha,z,n}$ to set an interplay between NB-CPSM (5) and EP-SM (1). This extends the compound Poisson perspective of the E-SM.

Theorem 2. *Let $(M_{1,n}(\alpha,\theta),\ldots,M_{n,n}(\alpha,\theta))$ be distributed as the EP-SM (1) and let $\bar{X}_{\alpha,\theta,n}$ be the random variable defined in (17), which is independent of $(M_{1,n}(\alpha,\theta),\ldots,M_{n,n}(\alpha,\theta))$. Moreover, let $(M_1(\alpha,z,n),\ldots,M_n(\alpha,z,n))$ be distributed as the NB-CPSM (5), and let $\tilde{X}_{\alpha,z,n}$ be the random variable defined in (18), which is independent of $(M_1(\alpha,z,n),\ldots,M_n(\alpha,z,n))$. Then:*

(i) *for $\alpha \in (0,1)$ and $\theta > -\alpha$:*

$$(M_{1,n}(\alpha,\theta),\ldots,M_{n,n}(\alpha,\theta)) \stackrel{d}{=} (M_1(\alpha,\bar{X}_{\alpha,\theta,n},n),\ldots,M_n(\alpha,\bar{X}_{\alpha,\theta,n},n));$$

(ii) *for $\alpha < 0$ and $z < 0$:*

$$(M_1(\alpha,z,n),\ldots,M_n(\alpha,z,n)) \stackrel{d}{=} (M_{1,n}(\alpha,-\tilde{X}_{\alpha,z,n}\alpha),\ldots,M_{n,n}(\alpha,-\tilde{X}_{\alpha,z,n}\alpha)).$$

Proof. As regards the proof of statement (i), it relies on the classical integral representation of the Gamma function. That is, by applying the integral representation of $\Gamma(\theta/\alpha+k)$ to the EP-SM (1), for $x_1,\ldots,x_n \in \{0,\ldots,n\}$ with $\sum_{i=1}^{n}x_i = k$ and $\sum_{i=1}^{n}ix_i = n$, we can write that:

$$\Pr[(M_{1,n}(\alpha,\theta),\ldots,M_{n,n}(\alpha,\theta)) = (x_1,\ldots,x_n)]$$

$$= n!\frac{\alpha^k}{\Gamma(\theta+n)}\prod_{i=1}^{n}\frac{\left(\frac{(1-\alpha)_{(i-1)}}{i!}\right)^{x_i}}{x_i!}\frac{\Gamma(\theta+1)}{\alpha\Gamma(\theta/\alpha+1)}$$

$$\times \int_0^{+\infty}z^{\theta/\alpha-1}e^{-z}\frac{z^k}{\sum_{j=1}^{n}\mathscr{C}(n,j;\alpha)z^j}\left(\sum_{j=1}^{n}\mathscr{C}(n,j;\alpha)z^j\right)dz$$

By Equation (13) of Favaro et al. [13]:

$$= n!\frac{\alpha^k}{\Gamma(\theta+n)}\prod_{i=1}^{n}\frac{\left(\frac{(1-\alpha)_{(i-1)}}{i!}\right)^{x_i}}{x_i!}\frac{\Gamma(\theta+1)}{\alpha\Gamma(\theta/\alpha+1)}$$

$$\times \int_0^{+\infty} z^{\theta/\alpha-1}\mathrm{e}^{-z} \frac{z^k}{\sum_{j=1}^n \mathscr{C}(n,j;\alpha)z^j} \left(\mathrm{e}^z z^{n/\alpha} \int_0^{+\infty} y^n \mathrm{e}^{-yz^{1/\alpha}} f_\alpha(y)\mathrm{d}y \right) \mathrm{d}z$$

$$= \int_0^{+\infty} \frac{n!}{\sum_{j=0}^n \mathscr{C}(n,j,\alpha)z^j} \prod_{i=1}^n \frac{\left(z \frac{\alpha(1-\alpha)_{(i-1)}}{i!} \right)^{x_i}}{x_i!}$$

$$\times \frac{\Gamma(\theta+1)}{\alpha\Gamma(\theta+n)\Gamma(\theta/\alpha+1)} z^{\theta/\alpha+n/\alpha-1} \int_0^{+\infty} y^n \mathrm{e}^{-yz^{1/\alpha}} f_\alpha(y)\mathrm{d}y\mathrm{d}z$$

$$= \int_0^{+\infty} \Pr[(M_1(\alpha,x,n),\ldots,M_n(\alpha,x,n)) = (x_1,\ldots,x_n)]$$

$$\times \frac{\Gamma(\theta+1)}{\alpha\Gamma(\theta+n)\Gamma(\theta/\alpha+1)} z^{\theta/\alpha+n/\alpha-1} \int_0^{+\infty} y^n \mathrm{e}^{-yz^{1/\alpha}} f_\alpha(y)\mathrm{d}y\mathrm{d}z$$

By the distribution of $\bar{X}_{\alpha,\theta,n}$:

$$= \int_0^{+\infty} \Pr[(M_1(\alpha,z,n),\ldots,M_n(\alpha,z,n)) = (x_1,\ldots,x_n)] f_{\bar{X}_{\alpha,\theta,n}}(z)\mathrm{d}z,$$

where $f_{\bar{X}_{\alpha,\theta,n}}$ is the density function of the random variable $\bar{X}_{\alpha,\theta,n}$. This completes the proof of (i).

As regards the proof of statement (ii), for any $\alpha < 0$, $m \in \mathbb{N}$, $k \leq m$ and $n \in \mathbb{N}$, we define the function $m \mapsto A(m;k,\alpha,n) = \frac{m!}{(m-k)!}\frac{\Gamma(-m\alpha)}{\Gamma(-m\alpha+n)}$, and then consider the following identity:

$$\frac{(-z)^k}{\sum_{j=1}^n \mathscr{C}(n,j;\alpha)z^j} = \sum_{m\geq k} A(m;k,\alpha,n)\Pr[\tilde{X}_{\alpha,z,n} = m]. \tag{19}$$

By applying (19) to the NB-CPSM (5), for $x_1,\ldots,x_n \in \{0,\ldots,n\}$ with $\sum_{i=1}^n x_i = k$ and $\sum_{i=1}^n ix_i = n$, we write:

$$\Pr[(M_1(\alpha,z,n),\ldots,M_n(\alpha,z,n)) = (x_1,\ldots,x_n)]$$

$$= \sum_{m\geq k} n!(-1)^k A(m;k,\alpha,n)\Pr[\tilde{X}_{\alpha,z,n} = m] \prod_{i=1}^n \frac{\left(\frac{\alpha(1-\alpha)_{(i-1)}}{i!} \right)^{x_i}}{x_i!}$$

$$= \sum_{m\geq k} n!(-1)^k \frac{m!}{(m-k)!}\frac{\Gamma(-m\alpha)}{\Gamma(-m\alpha+n)}\Pr[\tilde{X}_{\alpha,z,n} = m] \prod_{i=1}^n \frac{\left(\frac{\alpha(1-\alpha)_{(i-1)}}{i!} \right)^{x_i}}{x_i!}$$

$$= \sum_{m\geq k} n! \frac{\left(\frac{-m\alpha}{\alpha} \right)_{(k)}}{(-m\alpha)_{(n)}} \prod_{i=1}^n \frac{\left(\frac{\alpha(1-\alpha)_{(i-1)}}{i!} \right)^{x_i}}{x_i!}\Pr[\tilde{X}_{\alpha,z,n} = m]$$

$$= \sum_{m\geq k} \Pr[(M_1(\alpha,-m\alpha),\ldots,M_n(\alpha,-m\alpha)) = (x_1,\ldots,x_n)]\Pr[\tilde{X}_{\alpha,z,n} = m].$$

This completes the proof of (ii). □

Theorem 2 presents a compound Poisson perspective of the EP-SM in terms of the NB-CPSM, thus extending the well-known compound Poisson perspective of the E-SM in terms of the LS-CPSM. Statement (i) of Theorem 2 shows that for $\alpha \in (0,1)$ and $\theta > -\alpha$, the EP-SM admits a representation in terms of the NB-CPSM with $\alpha \in (0,1)$ and $z > 0$, where the randomisation acts on the parameter z with respect to the distribution (17). Precisely, this is a compound mixed Poisson sampling model. That is, a compound sampling model in which the distribution of the random number K of distinct types in the population is a mixture of Poisson distributions with respect to the law of $\bar{X}_{\alpha,\theta,n}$. Statement (ii) of Theorem 2 shows that for $\alpha < 0$ and $z < 0$, the NB-CPSM admits a representation in

terms of a randomised EP-SM with $\alpha < 0$ and $\theta = -m\alpha$ for some $m \in \mathbb{N}$, where the randomisation acts on the parameter m with respect to the distribution (17).

Remark 1. *The randomisation procedure introduced in Theorem 2 is somehow reminiscent of a class of Gibbs-type sampling models introduced in Gnedin and Pitman [14]. This class is defined from the EP-SM with $\alpha < 0$ and $\theta = -m\alpha$, for some $m \in \mathbb{N}$, and then it assumes that the parameter m is distributed according to an arbitrary distribution on $\mathbb{N}$. This can be seen in Theorem 12 of Gnedin and Pitman [14] and Gnedin [15] for example. However, differently from the definition of Gnedin and Pitman [14], in our context, the distribution of m depends on the sample size n.*

For $\alpha \in (0,1)$ and $\theta > -\alpha$, Pitman [5] first studied the large n asymptotic behaviour of $K_n(\alpha, \theta)$. This can also be seen in Gnedin and Pitman [14] and the references therein. Let $\xrightarrow{a.s.}$ denote the almost sure convergence, and let $S_{\alpha,\theta}$ be the scaled Mittag–Leffler random variable defined above. Theorem 3.8 of Pitman [5] exploited a martingale convergence argument to show that:

$$\frac{K_n(\alpha, \theta)}{n^\alpha} \xrightarrow{a.s.} S_{\alpha,\theta} \tag{20}$$

as $n \to +\infty$. The random variable $S_{\alpha,\theta}$ is referred to as Pitman's α-diversity. For $\alpha < 0$ and $\theta = -m\alpha$ for some $m \in \mathbb{N}$, the large n asymptotic behaviour of $K_n(\alpha, \theta)$ is trivial, that is:

$$K_n(\alpha, \theta) \xrightarrow{w} m \tag{21}$$

as $n \to +\infty$. We refer to Dolera and Favaro [16,17] for Berry–Esseen type refinements of (20) and to Favaro et al. [18,19] and Favaro and James [13] for generalisations of (20) with applications to Bayesian nonparametrics. This can also be seen in Pitman [5] (Chapter 4) for a general treatment of (20). According to Theorem 2, it is natural to ask whether there exists an interplay between Theorem 1 and the large n asymptotic behaviours (20) and (21). Hereafter, we show that: (i) (20), with the almost sure convergence replaced by the convergence in distribution, arises by combining (6) with (i) of Theorem 2; (ii) (8) arises by combining (21) with (ii) of Theorem 2. This provides an alternative proof of Pitman's α-diversity.

Theorem 3. *Let $K_n(\alpha, \theta)$ and $K(\alpha, z, n)$ under the EP-SM and the NB-CPSM, respectively. As $n \to +\infty$:*

(i) *For $\alpha \in (0,1)$ and $\theta > -\alpha$:*

$$\frac{K_n(\alpha, \theta)}{n^\alpha} \xrightarrow{w} S_{\alpha,\theta}. \tag{22}$$

(ii) *For $\alpha < 0$ and $z < 0$:*

$$\frac{K(\alpha, z, n)}{n^{\frac{-\alpha}{1-\alpha}}} \xrightarrow{w} \frac{(\alpha z)^{\frac{1}{1-\alpha}}}{-\alpha}. \tag{23}$$

Proof. We show that (22) arises by combining (6) with statement (i) of Theorem 2. For any pair of $\mathbb{N}$-valued random variables U and V, let $d_{TV}(U; V)$ be the total variation distance between the distribution of U and the distribution of V. Furthermore, let P_c denote a Poisson random variable with parameter $c > 0$. For any $\alpha \in (0,1)$ and $t > 0$, we show that as $n \to +\infty$:

$$d_{TV}(K(\alpha, tn^\alpha, n); 1 + P_{tn^\alpha}) \to 0. \tag{24}$$

This implies (22). The proof of (24) requires a careful analysis of the probability generating function of $K(\alpha, tn^\alpha, n)$. In particular, let us define $\omega(t; n, \alpha) := tn^\alpha + \frac{tM'_\alpha(t)}{M_\alpha(t)}$, where $M_\alpha(t) := \frac{1}{\pi} \sum_{m=1}^\infty \frac{(-t)^{m-1}}{(m-1)!} \Gamma(\alpha m) \sin(\pi \alpha m)$ is the Wright–Mainardi function (Mainardi et al. [20]). Then, we apply Corollary 2 of Dolera and Favaro [16] to conclude that

$\mathrm{d}_{TV}(K(\alpha, tn^\alpha, n); 1 + P_{\omega(t;n,\alpha)}) \to 0$ as $n \to +\infty$. Finally, we applied inequality (2.2) in Adell and Jodrá [21] to obtain:

$$\mathrm{d}_{TV}(1 + P_{tn^\alpha}; 1 + P_{\omega(t;n,\alpha)}) = \mathrm{d}_{TV}(P_{tn^\alpha}; P_{\omega(t;n,\alpha)}) \le \frac{tM'_\alpha(t)}{M_\alpha(t)} \min\left\{1, \frac{\sqrt{(2/e)}}{\sqrt{\omega(t;n,\alpha)} + \sqrt{tn^\alpha}}\right\}$$

So that $\mathrm{d}_{TV}(1 + P_{tn^\alpha}; 1 + P_{\omega(t;n,\alpha)}) \to 0$ as $n \to +\infty$, and (24) follows. Now, keeping α and t fixed as above, we show that (24) entails (22). To this aim, we introduced the Kolmogorov distance d_K which, for any pair of $\mathbb{R}_+$-valued random variables U and V, is defined by $\mathrm{d}_K(U;V) := \sup_{x \ge 0}|\Pr[U \le x] - \Pr[V \le x]|$. The claim to be proven is equivalent to:

$$\mathrm{d}_K(K_n(\alpha, \theta)/n^\alpha; S_{\alpha,\theta}) \to 0$$

as $n \to +\infty$. We exploit statement (i) of Theorem 2. This leads to the distributional identity $K_n(\alpha, \theta) \overset{d}{=} K(\alpha, \bar{X}_{\alpha,\theta,n}, n)$. Thus, in view of the basic properties of the Kolmogorov distance:

$$\begin{aligned} \mathrm{d}_K(K_n(\alpha, \theta)/n^\alpha; S_{\alpha,\theta}) &\le \mathrm{d}_K(K_n(\alpha, \theta); K(\alpha, n^\alpha S_{\alpha,\theta}, n)) \\ &\quad + \mathrm{d}_K(K(\alpha, n^\alpha S_{\alpha,\theta}, n); 1 + P_{n^\alpha S_{\alpha,\theta}}) \\ &\quad + \mathrm{d}_K([1 + P_{n^\alpha S_{\alpha,\theta}}]/n^\alpha; S_{\alpha,\theta}), \end{aligned} \tag{25}$$

where the $\{P_\lambda\}_{\lambda \ge 0}$ is thought of here as a homogeneous Poisson process with a rate of 1, independent of $S_{\alpha,\theta}$. The desired conclusion will be reached as soon as we will prove that all the three summands on the right-hand side of (25) go to zero as $n \to +\infty$. Before proceeding, we recall that $\mathrm{d}_K(U;V) \le \mathrm{d}_{TV}(U;V)$. Therefore, for the first of these terms, we write:

$$\begin{aligned} &\mathrm{d}_K(K_n(\alpha, \theta); K(\alpha, n^\alpha S_{\alpha,\theta}, n)) \\ &\le \frac{1}{2}\sum_{k=1}^n \left|\mathscr{C}(n,k;\alpha)\frac{\Gamma(k + \theta/\alpha)}{\alpha\Gamma(\theta/\alpha + 1)}\frac{\Gamma(\theta + 1)}{\Gamma(n + \theta)} - \int_0^{+\infty}\frac{\mathscr{C}(n,k;\alpha)(tn^\alpha)^k}{d_n(t)}f_{S_{\alpha,\theta}}(t)\mathrm{d}t\right| \end{aligned}$$

with $d_n(t) := \sum_{j=1}^n \mathscr{C}(n,j;\alpha)(tn^\alpha)^j$. Now, let us define $d_n^*(t) := e^{tn^\alpha}(n-1)!\frac{1}{t^{1/\alpha}}f_\alpha\left(\frac{1}{t^{1/\alpha}}\right)$. Accordingly, we can make the above right-hand side major by means of the following quantity:

$$\begin{aligned} &\frac{1}{2}\sum_{k=1}^n \left|\mathscr{C}(n,k;\alpha)\frac{\Gamma(k + \theta/\alpha)}{\alpha\Gamma(\theta/\alpha + 1)}\frac{\Gamma(\theta + 1)}{\Gamma(n + \theta)} - \int_0^{+\infty}\frac{\mathscr{C}(n,k;\alpha)(tn^\alpha)^k}{d_n^*(t)}f_{S_{\alpha,\theta}}(t)\mathrm{d}t\right| \\ &+ \frac{1}{2}\int_0^{+\infty}\frac{|d_n^*(t) - d_n(t)|}{d_n^*(t)}f_{S_{\alpha,\theta}}(t)\mathrm{d}t \, . \end{aligned}$$

Then, by exploiting the identity $\int_0^{+\infty}\frac{(tn^\alpha)^k}{d_n^*(t)}f_{S_{\alpha,\theta}}(t)\mathrm{d}t = \frac{1}{(n-1)!}\frac{\Gamma(k+\theta/\alpha)}{n^\theta}\frac{\Gamma(\theta+1)}{\alpha\Gamma(\theta/\alpha+1)}$, we can write:

$$\sum_{k=1}^n \left|\mathscr{C}(n,k;\alpha)\frac{\Gamma(k + \theta/\alpha)}{\alpha\Gamma(\theta/\alpha + 1)}\frac{\Gamma(\theta + 1)}{\Gamma(n + \theta)} - \int_0^{+\infty}\frac{\mathscr{C}(n,k;\alpha)(tn^\alpha)^k}{d_n^*(t)}f_{S_{\alpha,\theta}}(t)\mathrm{d}t\right| = \left|1 - \frac{\Gamma(n + \theta)}{\Gamma(n)n^\theta}\right|$$

which goes to zero as $n \to +\infty$ for any $\theta > -\alpha$, by Stirling's approximation. To show that the integral $\int_0^{+\infty}\frac{|d_n^*(t) - d_n(t)|}{d_n^*(t)}f_{S_{\alpha,\theta}}(t)\mathrm{d}t$ also goes to zero as $n \to +\infty$, we may resort to identities (13)–(14) of Dolera and Favaro [16], as well as Lemma 3 therein. In particular, let $\Delta : (0, +\infty) \to (0, +\infty)$ denote a suitable continuous function independent of n, and such that $\Delta(z) = O(1)$ as $z \to 0$ and $\Delta(z)f_\alpha(1/z) = O(z^{-\infty})$ as $z \to +\infty$. Then, we write that:

$$\int_0^{+\infty}\frac{|d_n^*(t) - d_n(t)|}{d_n^*(t)}f_{S_{\alpha,\theta}}(t)\mathrm{d}t$$

$$\leq \left| \frac{(n/e)^n \sqrt{2\pi n}}{n!} - 1 \right| + \left(\frac{(n/e)^n \sqrt{2\pi n}}{n!} \right) \frac{1}{n} \int_0^{+\infty} \Delta(t^{1/\alpha}) f_{S_{\alpha,\theta}}(t) \mathrm{d}t \ .$$

Since $\int_0^{+\infty} \Delta(t^{1/\alpha}) f_{S_{\alpha,\theta}}(t) \mathrm{d}t < +\infty$ by Lemma 3 of Dolera and Favaro [16], both the summands on the above right-hand side go to zero as $n \to +\infty$, again by Stirling's approximation. Thus, the first summand on the right-hand side of (25) goes to zero as $n \to +\infty$. As for the second summand on the right-hand side of (25), it can be bounded by

$$\int_0^{+\infty} \mathrm{d}_{TV}(K(\alpha, tn^\alpha, n); 1 + P_{tn^\alpha}) f_{S_{\alpha,\theta}}(t) \mathrm{d}t \ .$$

By a dominated convergence argument, this quantity goes to zero as $n \to +\infty$ as a consequence of (24). Finally, for the third summand on the right-hand side of (25), we can resort to a conditioning argument in order to reduce the problem to a direct application of the law of large numbers for renewal processes (Section 10.2 , Grimmett and Stirzaker [22]). In particular, this leads to $n^{-\alpha} P_{tn^\alpha} \xrightarrow{a.s.} t$ for any $t > 0$, which entails that $n^{-\alpha} P_{n^\alpha S_{\alpha,\theta}} \xrightarrow{a.s.} S_{\alpha,\theta}$ as $n \to +\infty$. Thus, this third term also goes to zero as $n \to +\infty$ and (22) follows.

Now, we consider (23), showing that it arises by combining (21) with statement (ii) of Theorem 2. In particular, by an obvious conditioning argument, we can write that as $n \to +\infty$:

$$\frac{K_n(\alpha, \tilde{X}_{\alpha,z,n} | \alpha|)}{\tilde{X}_{\alpha,z,n}} \xrightarrow{a.s.} 1.$$

At this stage, we consider the probability generating function of $\tilde{X}_{\alpha,z,n}$ and we immediately obtain $\mathbb{E}[s^{\tilde{X}_{\alpha,z,n}}] := B_n(-sz)/B_n(-z)$ for $n \in \mathbb{N}$ and $s \in [0,1]$ with the same B_n as in (13) and (14). Therefore, the asymptotic expansion we already provided in (15) entails:

$$\frac{\tilde{X}_{\alpha,z,n}}{n^{\frac{-\alpha}{1-\alpha}}} \xrightarrow{w} \frac{(\alpha z)^{\frac{1}{1-\alpha}}}{-\alpha} \tag{26}$$

as $n \to +\infty$. In particular, (26) follows by applying exactly the same arguments used to prove (8). Now, since:

$$\frac{K_n(\alpha, \tilde{X}_{\alpha,z,n} | \alpha|)}{n^{\frac{-\alpha}{1-\alpha}}} \overset{d}{=} \frac{K_n(\alpha, \tilde{X}_{\alpha,z,n} | \alpha|)}{\tilde{X}_{\alpha,z,n}} \frac{\tilde{X}_{\alpha,z,n}}{n^{\frac{-\alpha}{1-\alpha}}},$$

the claim follows from a direct application of Slutsky's theorem. This completes the proof. $\square$

3. Discussion

The NB-CPSM is a compound Poisson sampling model generalising the popular LS-CMSM. In this paper, we introduced a compound Poisson perspective of the EP-SM in terms of the NB-CPSM, thus extending the well-known compound Poisson perspective of the E-SM in terms of the LS-CPSM. We conjecture that an analogous perspective holds true for the class of α-stable Poisson–Kingman sampling models (Pitman [23] and Pitman [5]), of which the EP-SM is a noteworthy special case. That is, for $\alpha \in (0,1)$, we conjecture that an α-stable Poisson–Kingman sampling model admits a representation as a randomised NB-CPSM with $\alpha \in (0,1)$ and $z > 0$, where the randomisation acts on z with respect to a scale mixture between a Gamma and a suitable transformation of the Mittag–Leffler distribution. We believe that such a compound Poisson representation would be critical in order to introduce Berry–Esseen type refinements of the large n asymptotic behaviour of K_n under α-stable Poisson–Kingman sampling models. This can be seen in Section 6.1 of Pitman [23] and the references therein. Such a line of research aims to extend the preliminary works of Dolera and Favaro [16,17] on Berry–Esseen type theorems under the EP-SM. Work on this, and on the more general settings induced by normalised random measures (Regazzini et al. [24]) and Poisson–Kingman models (Pitman [23]), is ongoing.

References

1. Perman, M.; Pitman, J.; Yor, M. Size-biased sampling of Poisson point processes and excursions. *Probab. Theory Relat. Fields* **1992**, *92*, 21–39.
2. Pitman, J. Exchangeable and partially exchangeable random partitions. *Probab. Theory Relat. Fields* **1995**, *102*, 145–158.
3. Pitman, J.; Yor, M. The two parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.* **1997**, *25*, 855–900.
4. Ferguson, T.S. A Bayesian analysis of some nonparametric problems. *Ann. Stat.* **1973**, *1*, 209–230.
5. Pitman, J. *Combinatorial Stochastic Processes*; Lecture Notes in Mathematics; Springer: Berlin/Heidelberg, Germany, 2006.
6. Ewens, W. The sampling theory or selectively neutral alleles. *Theor. Popul. Biol.* **1972**, *3*, 87–112.
7. Crane, H. The ubiquitous Ewens sampling formula. *Stat. Sci.* **2016**, *31*, 1–19.
8. Charalambides, C.A. Distributions of random partitions and their applications. *Methodol. Comput. Appl. Probab.* **2007**, *9*, 163–193.
9. Korwar, R.M.; Hollander, M. Contributions to the theory of Dirichlet processes. *Ann. Stat.* **1973**, *1*, 705–711.
10. Wright, E.M. The asymptotic expansion of the generalized Bessel function. *Proc. Lond. Math. Soc.* **1935**, *38*, 257–270.
11. Charalambides, C.A. *Combinatorial Methods in Discrete Distributions*; Wiley: Hoboken, NJ, USA, 2005.
12. Berg, L. Asymptotische darstellungen für integrale und reihen mit anwendungen. *Math. Nachrichten* **1958**, *17*, 101–135.
13. Favaro, S.; James, L.F. A note on nonparametric inference for species variety with Gibbs-type priors. *Electron. J. Stat.* **2015**, *9*, 2884–2902.
14. Gnedin, A.; Pitman, J. Exchangeable Gibbs partitions and Stirling triangles. *J. Math. Sci.* **2006**, *138*, 5674–5685.
15. Gnedin, A. A species sampling model with finitely many types. *Electron. Commun. Probab.* **2010**, *8*, 79–88.
16. Dolera, E.; Favaro, S. A Berry—Esseen theorem for Pitman's α—Diversity. *Ann. Appl. Probab.* **2020**, *30*, 847–869.
17. Dolera, E.; Favaro, S. Rates of convergence in de Finetti's representation theorem, and Hausdorff moment problem. *Bernoulli* **2020**, *26*, 1294–1322.
18. Favaro, S.; Lijoi, A.; Prünster, I. Asymptotics for a Bayesian nonparametric estimator of species richness. *Bernoulli* **2012**, *18*, 1267–1283
19. Favaro, S.; Lijoi, A.; Mena, R.H.; Prünster, I. Bayesian nonparametric inference for species variety with a two parameter Poisson-Dirichlet process prior. *J. R. Stat. Soc. Ser. B* **2009**, *71*, 992–1008.
20. Mainardi, F.; Mura, A.; Pagnini, G. The M-Wright function in time-fractional diffusion processes: A tutorial survey. *Int. J. Differ. Equat.* **2010**, 104505. doi: 10.1155/2010/104505.
21. Adell, J.A.; Jodrá, P. Exact Kolmogorov and total variation distances between some familiar discrete distributions. *J. Inequalities Appl.* **2006**, 64307. doi. 10.1155/JIA/2006/64307.
22. Grimmett, G.; Stirzaker, D. *Probability and Random Processes*; Oxford University Press: Oxford, UK, 2001.
23. Pitman, J. Poisson-Kingman partitions. In *Science and Statistics: A Festschrift for Terry Speed*; Goldstein, D.R., Eds.; Institute of Mathematical Statistics: Tachikawa, Tokyo, Japan, 2003.
24. Regazzini, E.; Lijoi, A.; Prünster, I. Distributional results for means of normalized random measures with independent increments. *Ann. Stat.* **2003**, *31*, 560–585.