ISSN 2279-9362

Collegio Carlo Alberto

Asymptotic behavior of the number of distinct values in a sample from the geometric stick-breaking process

Pierpaolo De Blasi
Ramsés H. Mena
Igor Prünster

No. 640
February 2021

Carlo Alberto Notebooks
www.carloalberto.org/research/working-papers

Asymptotic behavior of the number of distinct values in a sample from the geometric stick-breaking process

Pierpaolo De Blasi*[1], Ramsés H. Mena[2], and Igor Prünster[3]

[1]University of Torino and Carlo Alberto, Torino, Italy
[2]Universidad Nacional Autónoma de Mexico, Mexico
[3]Bocconi University and BIDSA, Milano, Italy

January 19, 2021

Abstract

Discrete random probability measures are a key ingredient of Bayesian nonparametric inferential procedures. A sample generates ties with positive probability and a fundamental object of both theoretical and applied interest is the corresponding random number of distinct values. The growth rate can be determined from the rate of decay of the small frequencies implying that, when the decreasingly ordered frequencies admit a tractable form, the asymptotics of the number of distinct values can be conveniently assessed. We focus on the geometric stick-breaking process and we investigate the effect of the choice of the distribution for the success probability on the asymptotic behavior of the number of distinct values. We show that a whole range of logarithmic behaviors are obtained by appropriately tuning the prior. We also derive a two-term expansion and illustrate its use in a comparison with a larger family of discrete random probability measures having an additional parameter given by the scale of the negative binomial distribution.

Keywords: Bayesian Nonparametrics; random probability measure; geometric stick-breaking process; asymptotic growth rate; occupancy problem.

1 Introduction

Discrete random probability measures can be represented by random frequencies at random locations as

$$\tilde{p}(\mathrm{d}x) = \sum_{j\geq 1} w_j \delta_{x_j}(\mathrm{d}x). \tag{1}$$

The frequencies $(w_j)_{j\geq 1}$ are $(0,1)$-valued variables such that $\sum_{j\geq 1} w_j = 1$ almost surely (a.s.), and the locations $(x_j)_{j\geq 1}$ are draws from some distribution on a Polish space $\mathbb{X}$ endowed with the corresponding Borel σ-field. Discrete measures like $\tilde{p}$ are naturally suited to describe the

*pierpaolo.deblasi@unito.it

structure a population made of potentially infinite different species or types, labeled by x_j, with certain random proportions modeled through w_j. Clearly, a sample drawn from $\tilde{p}$ will exhibit ties with positive probability and thus the random number of distinct values in a sample of size n, here denoted by K_n, is of great interest. From a Bayesian nonparametric perspective the law of $\tilde{p}$ represents the prior distribution. Inference is carried out by predicting the number of new distinct values in an additional sample, conditional on an observed sample. See Lijoi et al. (2007b). According to the applied context at issue the distinct values or species are interpreted as distinct genes (Lijoi et al., 2007a), words (Teh, 2006), economic agents (Lijoi et al., 2016) etc. Another important statistical use of discrete random probality measures is in mixture modeling, when a layer is added to model the data distribution as in

$$Y_i \sim f(Y_i|X_i), \quad X_1, X_2, \ldots |\tilde{p} \overset{\text{iid}}{\sim} \tilde{p}$$

for some probability kernel $f(y|x)$. Here $\tilde{p}$ acts as mixing distribution and K_n represents the random number of mixture components, thus providing a flexible way to model unobserved heterogeneity in the population. The mixture is characterized by the component distribution $f(y|x_j)$, usually referred as the jth mixture component, and the mixing weights w_j. See De Blasi et al. (2015) for a recent review on the inferential implications of different choices of $\tilde{p}$.

In probability theory, distributional properties of K_n are of prime interest in combinatorial stochastic processes; see e.g. Arratia et al. (2003), Pitman (2006), Gnedin (2010), Gnedin et al. (2007). The techniques employed to study the law of K_n depend on the construction of the random frequencies (w_j). Karlin (1967) studied the case of fixed frequencies and derived a key result, which forms the basis to establish general strong laws for K_n: it states that the growth of K_n is ultimately determined by how small the small frequencies are, which can be conveniently expressed by the *tail behavior* of (w_j) once decreasingly ordered. In particular, the faster the decay to zero, the slower K_n diverges to infinity as n increases. There exist essentially two regimes, logarithmic and polynomial growth. Notable examples are, respectively, the Dirichlet process (Ferguson, 1973) and its two parameter extension known as Pitman-Yor process (Pitman and Yor, 1997). The associated distributions of the frequencies in decreasing order, termed Poisson-Dirichlet and the two-parameter Poisson-Dirichlet, respectively, are not tractable enough for a direct application of Karlin's theory. Instead, the distribution of K_n is derived from the Ewens and the Pitman-Ewens sampling formulae, cf. Pitman (2006). In the former case K_n is asymptotically normal, with both mean and variance of the order $\log n$. In the latter case the scale of K_n is n^α, where $\alpha \in (0, 1)$ is the discount parameter of the Pitman-Yor process. The logarithmic growth of the Dirichlet process was first pointed out in Korwar and Hollander (1973). Within the logarithmic regime growth behaviors of K_n slower than the logarithm, e.g. $(\log n)^\alpha$ with $\alpha < 1$ or even iterated logarithms, can be achieved with so-called hierarchical processes Camerlenghi et al. (2019); see also Argiento et al. (2020); Bassetti et al. (2020). In this paper we are able to identify models leading to growth rates of the type $(\log n)^\beta$ with $\beta > 1$, specifically we establish a growth rate $(\log n)^{m+2}$, for m a nonnegative integer, for a class of tractable class of discrete random probability measures. We stress that from a modeling perspective it is crucial to have tractable models, which cover the whole range of possible growth rates. See Lijoi et al. (2007c); De Blasi et al. (2015); Dahl et al. (2017); Caron and Fox (2017); Ayed et al. (2019); Di Benedetto et al. (2020) for motivation and discussion of these issues in diverse application contexts also beyond exchangeability. Note that a power logarithmic growth of $\mathrm{E}(K_n)$ can be obtained by means of a Dirichlet prior with a somehow artificial sample size-dependent specification of the total mass parameter; in particular, one needs the total mass parameter to grow with n, which leads to an increasingly informative prior as more data becomes available, an unnatural scenario.

The asymptotic evaluation K_n, together with its limiting distribution, has been also object of extensive research in the context of regenerative composition structures; see Gnedin (2010) for a survey. In this setting the frequencies (w_j) are constructed from the range of a multiplicative subordinator, that is from the exponential transform $1-\exp\{S(t)\}$ of a subordinator $S(t)$. The logarithmic and polynomial regimes can be recovered from Karlin's theory according to the variation at zero of the right tail of the Lévy measure of $S(t)$. The $\log n$ regime corresponds to finite Lévy measures, that is when $S(t)$ is a compound Poisson process. In this case, the frequencies (w_j) can be conveniently defined in terms of a stick-breaking procedure, or residual allocation scheme, with

$$w_j = W_j \prod_{\ell<j}(1-W_\ell) \tag{2}$$

for $(W_\ell)_{\ell\geq 1}$ independent and identically distributed (iid) $(0,1)$-valued random variables with distribution determined by the Lévy measure. Exploiting the renewal representation of the composition structure, aymptotics for the moments of K_n and a central limit theorem can be derived; cf. Gnedin (2004), Gnedin et al. (2009). Gnedin et al. (2006a) show that when the right tail of the Lévy measure is regularly varying at zero with index $-1<\alpha<0$, the scale of K_n is n^α and the partition structure induced by the Pitman-Yor process can be recovered (Gnedin and Pitman, 2005). In contrast, when the right tail diverges at zero like a slowly varying function, e.g. for $S(t)$ a gamma subordinator, a central limit theorem with mean of the order $(\log n)^2$ and variance of the order $(\log n)^3$ is obtained, cf. Gnedin et al. (2006b).

Discrete random probability measures with w_j as in (2), not necessarily with identically distributed $(W_\ell)_{\ell\geq 1}$, have been proposed in Ishwaran and James (2001) as a Bayesian nonparametric model and termed stick-breaking priors. The Dirichlet and the Pitman-Yor processes belong to this class, their distinctive property being that the law of $(w_j)_{j\geq 1}$ is invariant under size-biased permutation. In this setting, the distribution of W_1 is called the *structural distribution* of (w_j), and the limiting behavior of K_n in the Dirichlet and the Pitman-Yor process cases can be also derived using Karlin's theory from the variation at zero of the structural distribution; see Gnedin et al. (2007).

In this paper we further broaden the realm of application of the fundamental result of Karlin in order to derive a two-term expansion of the mean of K_n. The expansion relies on de Haan's regular variation theory and requires a precise assessment of the tail behavior of (w_j) together with a deconditioning argument, cf. Theorem 1. To illustrate the applicability of this technique, we consider the geometric stick-breaking process, first proposed in Fuentes-García et al. (2010), which gained quite some popularity in Bayesian applications (Mena et al., 2011; Gutiérrez et al., 2014; Hatjispyros et al., 2018). It is a discrete random probability measure (1) with locations independent of the frequencies and, importantly, $(w_j)_{j\geq 1}$ naturally arranged in decreasing order, which facilitates the evaluation of the tail behavior of the sequence. Specifically the frequencies are of geometric type,

$$w_j = p(1-p)^{j-1}, \quad j=1,2,\dots \tag{3}$$

with p, the probability of success, random and endowed with a (prior) distribution $\pi(p)$ on $(0,1)$. In Theorem 2 we derive a two-term expansion for a choice of $\pi(p)$, the key technical tool being the regular variation of fractional integrals. As anticipated, the leading term shows that the mean of K_n can covers the whole range of logarithmic behaviors $(\log n)^{m+2}$, for m a nonnegative integer, upon setting $\pi(p)$ as an exponential transform of the gamma distribution of shape parameter

m. From a practical perspective this result widens the range of achievable asymptotic behaviors by means of tractable models and also allows a principled prior elicitation. To illustrate the importance of the second order term in the expansion, we also consider an extension of the geometric stick-breaking process, which has an additional parameter s corresponding to the scale of the negative binomial distribution. Such a construction reduces to the geometric stick-breaking process when $s = 2$ and was exploited by De Blasi et al. (2020) within a mixture model, to which the present study provides further theoretical support. The frequencies $(w_j)_{j\geq 1}$ are still decreasingly ordered and are available in closed form for any integer $s \geq 2$. The parameter s determines the tail behavior of $(w_j)_{j\geq 1}$, the larger s the slower the decay to zero. In order to single out the effect of s on K_n, we set $s = 3$ and compare the asymptotic behavior of the mean of K_n with that of the geometric stick-breaking case, while keeping $\pi(p)$ to be uniform. It turns out that K_n grows faster for $s = 3$, as predicted by Karlin's theory, the difference however emerging only in the second order term of the expansion, cf. Proposition 2. We conjecture that similar conclusions apply also for s an integer larger than 3 and other prior specifications of $\pi(p)$, although we do not pursue it here. It would be of interest to investigate the asymptotics of higher order moments like the variance and whether a central limit theorem holds. These are left for future research.

Layout of the paper. In Section 2 we review Karlin's theory and establish a general two-term expansion of the mean of K_n. In Section 3 we introduce the geometric stick-breaking process and investigate the impact of the choice of prior $\pi(p)$ on the asymptotic behavior of K_n. In Section 4 we deal with the negative binomial extension and apply the asymptotic expansion of Section 2 to show that the scale parameter s enters in the second order term. Some proofs are deferred to the Appendix.

Notation. Let $F(x)$ be a positive nondecreasing function on $\mathbb{R}$ with $F(x) = 0$ for $x \leq 0$ and $\alpha \geq 0$. The fractional integral of order α of $F(x)$ is given by

$$ {}_{\alpha}F(x) = \frac{1}{\Gamma(\alpha+1)} \int_0^x (x-t)^{\alpha} f(t) \mathrm{d}t. $$

We use $f \sim g$ for $f/g \to 1$, the limit being clear from the context. When either f or g is random, the notation $f \sim_{\text{a.s.}} g$ means that the asymptotic relation holds with probability one. For x a real number, $\lfloor x \rfloor$ is the integer part of x.

2 Occupancy problem and regular variation

Let $\tilde{p}$ be a discrete random probability measure (1). Assume $(w_j)_{j\geq 1}$ and $(x_j)_{j\geq 1}$ are independent with $(x_j)_{j\geq 1}$ independent and identically distributed form a non atomic distribution. Then $\tilde{p}$ is a *species sampling model* (Pitman, 1995). The partition induced by a sample from $\tilde{p}$ depends only on the random frequencies $(w_j)_{j\geq 1}$ and can be studied in terms of a multinomial occupancy problem. The theory is well established and dates back to the seminal paper Karlin (1967). The main tools are a Poissonization argument and regular variation theory. We provide a concise overview taking the set-up from Gnedin et al. (2007).

The multinomial occupancy problem can be described as the experiment of throwing balls independently at a fixed infinite series of boxes, with probability w_j of hitting the jth box. First consider the case of fixed, or non random, frequencies. As n balls are thrown, their allocation is captured by the array $X_n = (X_{n,j})_{j\geq 1}$ where $X_{n,j}$ is the number of balls out of the first n that fall in box j. K_n, the number of occupied boxes, is then given by $K_n = \sum_{j\geq 1} \mathbb{1}(X_{n,j} > 0)$ with mean

$$\mathrm{E}(K_n) = \sum\nolimits_{j\geq 1}(1-(1-w_j)^n).$$

In general, it is difficult to work with $\mathrm{E}(K_n)$ since the indicators in K_n are not independent. In the Poissonized version of the problem the balls are thrown in continuous time at epochs of a unit rate Poisson process $(P(t), t \geq 0)$, which is independent of $(X_n, n = 1, 2, \ldots)$. The balls then fall in the boxes according to independent Poisson processes $(X_j(t))_{t\geq 0}$, at rate w_j for box j. Hence $K(t) := K_{P(t)} = \sum_{j\geq 1} \mathbb{1}(X_j(t) > 0)$ and

$$\Phi(t) := \mathrm{E}(K(t)) = \sum\nolimits_{j\geq 1}(1-\mathrm{e}^{-tw_j}).$$

Encoding the frequencies into the counting measure $\nu(\mathrm{d}x) = \sum_{j\geq 1} \delta_{w_j}(\mathrm{d}x)$ and integrating by parts,

$$\Phi(t) = \int_0^1 (1-\mathrm{e}^{-tx})\nu(\mathrm{d}x) = t\int_0^1 \mathrm{e}^{-tx}\overrightarrow{\nu}(x)\mathrm{d}x,$$

where $\overrightarrow{\nu}(x) = \nu([x, 1))$, the right tail of ν, represents the number of frequencies w_j not smaller than x. $\Phi(t)$ provides an approximation of $\mathrm{E}(K_n)$ for n large according to

$$|\mathrm{E}(K_n) - \Phi(n)| \leq \tfrac{2}{n}\Phi(n) \to 0 \tag{4}$$

cf. (Gnedin et al., 2007, Lemma 1). The convenience of working with $\Phi(t)$ is that, being $\Phi(t)$ the Laplace-Stieltjes transform of $\overrightarrow{\nu}(x)$, its behavior as $t \to \infty$ is determined by the behavior of $\overrightarrow{\nu}(x)$ as $x \to 0$ by an application of the Tauberian theorem; see Bingham et al. (1987) for a full account on Abel-Tauberian theorems for Laplace transforms. Hence, ultimately, by regular variation theory the growth of $\mathrm{E}(K_n)$, as $n \to \infty$, is determined by the behavior of $\overrightarrow{\nu}(x)$ at zero. In the case of random frequencies, the same result holds with the counting measure $\nu(\mathrm{d}x)$ being replaced by its mean measure, and correspondingly adapting the meaning of $\overrightarrow{\nu}(x)$. See (Gnedin et al., 2007, Section 7, Page 162) and Section 3 for an illustration.

Here we work under the hypothesis that $\overrightarrow{\nu}(x)$ is slowly varying at zero, that is $\lim_{x\to 0}\overrightarrow{\nu}(\lambda x)/\overrightarrow{\nu}(x) = 1$ for all $\lambda > 0$. According to (Bingham et al., 1987, Theorems 1.7.1' and 1.7.6) (see also (Gnedin et al., 2007, Proposition 19)), $\Phi(1/x) \sim \overrightarrow{\nu}(x)$ as $x \to 0$, so that via (4)

$$\mathrm{E}(K_n) \sim \overrightarrow{\nu}\big(\tfrac{1}{n}\big) \quad \text{as } n \to \infty,$$

cf. (Karlin, 1967, Theorem 1'). In Theorem 1 we derive a two term expansion of $\mathrm{E}(K_n)$ under the hypothesis that $\overrightarrow{\nu}(x)$ is a de Haan slowly varying function at zero, that is for a constant c and a slowly varying function $\ell(x)$ at zero, called the auxiliary function of $\overrightarrow{\nu}(x)$,

$$\frac{\overrightarrow{\nu}(\lambda x) - \overrightarrow{\nu}(x)}{\ell(x)} \to c\log\lambda, \quad \text{as } x \to 0. \tag{5}$$

Theorem 1. *If $\ell(x)$ is slowly varying at zero and $c \geq 0$ satisfy (5) for all $\lambda > 0$, then*

$$\mathrm{E}(K_n) = \overrightarrow{\nu}(1/n) - c\gamma\ell(1/n) + o(\ell(1/n)), \quad \textit{as } n \to \infty$$

where γ is the Euler-Mascheroni constant.

The proof consists in an adaptation to the present setting of (Bingham et al., 1987, Theorem 3.9.1) for the study of the remainder of Tauberian theorem, $\Phi(1/x) - \overrightarrow{\nu}(x)$, as $x \to 0$, combined with an application of (4). The proof is reported in the Appendix. In order to apply this result, one needs to establish the variation of $\overrightarrow{\nu}(x)$ at 0, so some explicit or at least tractable form of $\overrightarrow{\nu}(x)$ is in order. In the next two sections we apply the asymptotic expansion of Theorem 1 to species sampling priors that features stochastically decreasing frequencies for which $\overrightarrow{\nu}(x)$ is tractable enough.

3 Geometric stick-breaking process

The geometric stick breaking process is a species sampling model with random frequencies $(w_j)_{j\geq 1}$ of geometric type,

$$w_j = p(1-p)^{j-1}, \quad j = 1, 2, \dots$$

with random success probability p. The number of frequencies w_j not smaller than x, $\max\{j : p(1-p)^{j-1} \geq x\}$, can be explicitly found as the solution in j to the equation $p(1-p)^{j-1} = x$. By direct calculation

$$\overrightarrow{\nu}(x,p) = \left\lfloor \frac{\log(x/p)}{\log(1-p)} + 1 \right\rfloor \mathbb{1}_{(p\geq x)},$$

where the notation $\overrightarrow{\nu}(x,p)$ makes the dependence on p explicit. The case of fixed p provides an illustration of Theorem 1.

Example 1. Let K_n be the number of distinct values among n iid draws from the geometric distribution with success probability p. Accurate formulae for the mean and the variance of K_n are given in Archibald et al. (2006). Since $\overrightarrow{\nu}(x,p) \sim \log x / \log(1-p)$ as $x \to 0$, $\overrightarrow{\nu}(x,p)$ is a de Haan slowly varying function with auxiliary function $\ell(x) = 1$ and $c = 1/\log(1-p)$, cf. (5). Hence Theorem 1 yields

$$\mathrm{E}(K_n) = \left\lfloor \frac{\log(np)}{|\log(1-p)|} + 1 \right\rfloor + \frac{\gamma}{|\log(1-p)|} + o(1) \quad \text{as } n \to \infty,$$

in accordance with the expansion of (Archibald et al., 2006, Theorem 1).

Now return to the random case with $\pi(p)$ on $(0,1)$ denoting the (prior) distribution of the success probability p in (3). The results about the expected value of K_n now hold with $\nu(\mathrm{d}x)$ being the mean measure of the counting measure $\sum_{j\geq 1} \delta_{w_j}$ and $\overrightarrow{\nu}(x)$ obtained by averaging the number of frequencies w_j not smaller than x with respect to $\pi(p)$:

$$\overrightarrow{\nu}(x) = \int_0^1 \overrightarrow{\nu}(x,p)\pi(p)\mathrm{d}p.$$

In the sequel it is convenient to work with

$$m(x) = \int_x^1 \frac{\log x - \log p}{\log(1-p)} \pi(p)\mathrm{d}p,$$

since $m(x) \leq \overrightarrow{\nu}(x) \leq m(x) + 1$. The variation of $\overrightarrow{\nu}(x)$ in zero can then be studied in terms of $m(x)$. By the change of variable $t = \log 1/p$,

$$m(x) = \int_0^{\log 1/x} \left(\log \frac{1}{x} - t \right) \pi(\mathrm{e}^{-t}) f(t)\mathrm{d}t, \quad f(t) = \frac{\mathrm{e}^{-t}}{-\log(1-\mathrm{e}^{-t})} \tag{6}$$

Properties of $f(t)$ in (6) are collected in Lemma 1, whose proof is deferred to the Appendix.

Lemma 1. *The function $f(t)$ defined in (6) is nondecreasing on $\mathbb{R}_+$ with $\lim_{t\to 0} f(t) = 0$, $\lim_{t\to\infty} f(t) = 1$ and $1 - f(t) \sim \mathrm{e}^{-t}/2$ as $t \to \infty$. Moreover $\int_0^\infty (1 - f(t))\mathrm{d}t = \gamma$, with $\gamma = -\Gamma'(1) - \int_0^\infty (\log x)\mathrm{e}^{-x}\mathrm{d}x$ the Euler-Mascheroni constant.*

The variation at zero of $m(x)$ is determined by $f(t)$ and the success probability distribution $\pi(p)$. First consider p uniformly distributed on the unit interval.

Proposition 1. *Let p in (3) be uniformly distributed on $(0, 1)$. Then*

$$\begin{aligned} \overrightarrow{\nu}(x) &= \frac{1}{2}\big(\log 1/x\big)^2 - \gamma \log 1/x + O(1), \quad x \to 0 \\ \mathrm{E}(K_n) &= \frac{1}{2}(\log n)^2 + o(\log n), \quad n \to \infty. \end{aligned}$$

Proof. For $\pi(p) = \mathbb{1}_{(0,1)}(p)$, $m(x)$ in (6) is given by

$$m(x) = \int_0^{\log 1/x} (\log 1/x - t) f(t)\mathrm{d}t = {}_1F(\log 1/x),$$

where $F(t) = \int_0^t f(s)\mathrm{d}s$ and ${}_1F(x) = \int_0^x (x-t)f(t)\mathrm{d}t$ is the fractional integral of order one of F. To prove the first statement, it is sufficient to prove it for $m(x)$ in place of $\overrightarrow{\nu}(x)$. Integrating by parts, ${}_1F(x) = \int_0^x F(t)\mathrm{d}t$. Hence, since $\log 1/x \to \infty$ as $x \to 0$, we derive an asymptotic expansion of $F(x)$ as $x \to \infty$. According to Lemma 1, $f(x)$ is a distribution function on $\mathbb{R}_+$. Moreover, given $1 - f(t) \sim \mathrm{e}^{-t}/2$ as $t \to \infty$, the distribution function $f(x)$ has moments of any order and, in particular, the first moment is equal to the Euler-Mascheroni constant γ. Then, $F(x)$ is regularly varying at infinity with exponent $\beta = 1$ and, as $x \to \infty$,

$$F(x) = x - \int_0^x (1 - f(t))\mathrm{d}t = x - \gamma + \int_x^\infty (1 - f(t))\mathrm{d}t = x - \gamma + O(\mathrm{e}^{-x}). \tag{7}$$

Computing $\int_0^x F(t)\mathrm{d}t$ with the asymptotic expansion $F(x) \sim x - \gamma + O(\mathrm{e}^{-x})$ leads to

$${}_1F(x) = \int_0^x F(t)\mathrm{d}t = \frac{(x-\gamma)^2}{2} + O(1), \quad x \to \infty.$$

Substituting x for $\log 1/x$ yields the first statement. In view of the application of Theorem 1, note that, as $x \to 0$,

$$\begin{aligned} m(\lambda x) - m(x) &= \frac{1}{2}\big((\log(\lambda x))^2 - (\log x)^2\big) + \gamma\big(\log(\lambda x) - \log x\big) + O(1) \\ &= \frac{1}{2}\big((\log x)^2 + 2\log\lambda \log x - (\log x)^2\big) + O(1) = \log\lambda \log x + O(1) \end{aligned}$$

so that, as $x \to 0$, $(m(\lambda x) - m(x))/\log x \to \log\lambda$. Hence $\overrightarrow{\nu}(x)$ is a de Haan slowly varying function at zero with auxiliary function $\ell(x) = \log x$ and $c = 1$, cf. (5). An application of Theorem 1 yields the second statement. $\square$

Remark 1. Using only the leading term of the expansion of $m(x)$ in the application of Theorem 1, we would get the second order term in the asymptotic expansion of $\mathrm{E}(K_n)$ wrong, i.e. differing by $\gamma \log n$. Hence, in this case, an application of Karamata's Theorem to the evaluation of $\int_0^x (x-t)f(t)\mathrm{d}t$ would be not precise enough, as the latter would yield $\int_0^x (x-t)f(t)\mathrm{d}t \sim \frac{1}{2}xF(x)$ and, in turn, $m(x) \sim \frac{1}{2}(\log \frac{1}{x})^2$.

Next we tackle the case of the success probability distribution $\pi(p)$ chosen such that the integrand in (6) behaves like $t^m f(t)$ for m a positive integer. This is obtained by setting $p = \mathrm{e}^{-X}$ for $X \sim \mathrm{ga}(m+1, 1)$, a gamma distributed random variable with shape $m+1$ and unit rate. Note $m = 0$ yields the uniform distribution considered in Proposition 1. As detailed in Theorem 2 below, this choice makes $\mathrm{E}(K_n)$ grow as a power of $\log n$ with exponent determined by m. Before that, we first provide an illustration of how the arguments used in Proposition 1 can be adapted to the case $m = 1$, paving the way for the techniques used in the general case.

Example 2. By direct calculation, the density function of $p \overset{d}{=} \mathrm{e}^{-X}$, for $X \sim \mathrm{ga}(2, 1)$, is $\pi(p) = -\log p$. Then $m(x)$ in (6) is given by

$$m(x) = \int_0^{\log 1/x} (\log 1/x - t)tf(t)\mathrm{d}t = \int_0^{\log 1/x} \int_0^t sf(s)\mathrm{d}s\,\mathrm{d}t.$$

where the second equality follows by integration by parts. We first derive an asymptotic expansion for $\int_0^x tf(t)\mathrm{d}t$ as $x \to \infty$. Since

$$\int_0^x tf(t)\mathrm{d}t = xF(x) - \int_0^x F(t)\mathrm{d}t = xF(x) - {}_1F(x)$$

using the asymptotic expansion $F(x) = x - \gamma + O(\mathrm{e}^{-x})$ in (7) we find

$$\int_0^x tf(t)\mathrm{d}t = x(x-\gamma) + O(x\mathrm{e}^{-x}) - \frac{(x-\gamma)^2}{2} + O(1) = \frac{x^2}{2} + O(1)$$

so that

$$\int_0^x \int_0^t sf(s)\mathrm{d}s\,\mathrm{d}t = \int_0^x \left(\frac{t^2}{2} + O(1)\right)\mathrm{d}x = \frac{x^3}{6} + O(x)$$

and, in turn,

$$m(x) = \frac{1}{6}\left(\log\frac{1}{x}\right)^3 + O(\log x).$$

We look now for the auxiliary function $\ell(x)$ of the slowly varying function $m(x)$. We have

$$\begin{aligned} 6\big(m(\lambda x) - m(x)\big) &= -(\log x + \log \lambda)^3 + (\log x)^3 + O(\log x) \\ &= -3\log\lambda(\log x)^2 + O(\log x) \end{aligned}$$

so that, as $x \to 0$

$$\frac{m(\lambda x) - m(x)}{(\log x)^2} \to -\frac{1}{2}\log\lambda.$$

Hence, the auxiliary function of $\overrightarrow{\nu}(x)$ is found to be $\ell(x) = (\log x)^2$ with $c = -\frac{1}{2}$, cf. (5). Note that, because of the cancellation of the $(\log 1/x)^2$ term in the expansion of $m(x)$ for $x \to \infty$, the derivation of $\ell(x)$ only requires the leading term of $m(x)$. The latter can be alternatively obtained by using regular variation theory. Since $\int_0^x tf(t)\mathrm{d}t$ is regularly varying at infinity with index 2, Karamata's theorem yields $\int_0^x (x-t)tf(t)\mathrm{d}t/x\int_0^x tf(t)\mathrm{d}t \to \frac{1}{3}$ as $x\to\infty$. A second application of Karamata's Theorem yields $\int_0^x tf(t)\mathrm{d}t/x^2 f(x) \to \frac{1}{2}$ to conclude that, as $x\to\infty$,

$$\frac{\int_0^x (x-t)tf(t)\mathrm{d}t}{x^3 f(x)} \to \frac{1}{2}\frac{1}{3} = \frac{1}{6}$$

Since $f(x)\to 1$ as $x\to\infty$, we get $m(x)\sim \frac{1}{6}(\log 1/x)^3$. Finally, by applying Theorem 1 we conclude that

$$\mathrm{E}(K_n) = \frac{1}{6}(\log n)^3 + \frac{1}{2}\gamma(\log n)^2 + o(\log^2 n)$$

It is worth stressing that $\pi(p) = -\log p$ yields $\mathrm{E}(p) = 1/4$, i.e. a mass shift to lower values of p compared to $\pi(p) = 1$. This, according to $\overrightarrow{\nu}(x,p)\sim \log x/\log(1-p)$, favors larger values of $\mathrm{E}(K_n|p)$, which explain a faster growth of $\mathrm{E}(K_n)$.

The asymptotic expansion of $\mathrm{E}(K_n)$, for m any positive integer, is derived in the following theorem. The key ingredient consists in expressing $\int_0^x t^m f(t)\mathrm{d}t$ in terms of fractional integrals of $F(x)$.

Theorem 2. *Let p in (3) have distribution $\pi(p)$ defined by $p \overset{d}{=} \mathrm{e}^{-X}$ with $X\sim \mathrm{ga}(m+1,1)$ and m a positive integer. Then*

$$\overrightarrow{\nu}(x) = \frac{(\log 1/x)^{m+2}}{(m+2)!} + O\big((\log x)^m\big), \quad x\to 0$$
$$\mathrm{E}(K_n) = \frac{(\log n)^{m+2}}{(m+2)!} + \gamma\frac{(\log n)^{m+1}}{(m+1)!} + o((\log n)^{m+1}), \quad n\to\infty.$$

Proof. Since $\pi(p) = (-\log p)^m/m!$, $m(x)$ in (6) becomes

$$m(x) = \int_0^{\log 1/x} (\log 1/x - t)\frac{t^m}{m!}f(t)\mathrm{d}t = \int_0^{\log 1/x}\int_0^t \frac{s^m}{m!}f(s)\mathrm{d}s\,\mathrm{d}t. \tag{8}$$

where the second equality follows again by integration by parts. Recall that, for an integer-valued index, fractional integrals correspond to "higher order" primitives of $f(x)$: $F(x) = {}_0F(x)$, ${}_1F(x) = \int_0^x F(t)\mathrm{d}t$ and

$${}_{(k+1)}F(x) = \int_0^x (x-t)\mathrm{d}\,{}_kF(t) = \int_0^x {}_kF(t)\mathrm{d}t, \quad k = 1,2,\dots$$

By repeated integration by parts the inner integral in (8) is found to be

$$\int_0^x \frac{t^m}{m!}f(t)\mathrm{d}t = \sum_{k=0}^m \frac{(-1)^k}{(m-k)!}x^{m-k}{}_kF(x).$$

Next, we exploit the asymptotic evaluation (7), $F(x) = x-\gamma+O(\mathrm{e}^{-x})$ as $x\to\infty$, to find

$${}_1F(x) = \frac{(x-\gamma)^2}{2} + O(1) = \frac{x^2-2\gamma x}{2} + O(1)$$

$$
\begin{aligned}
{}_2F(x) &= \frac{(x-\gamma)^3}{3!} + O(x) = \frac{x^3 - 3\gamma x^2}{3!} + O(x) \\
{}_kF(x) &= \frac{(x-\gamma)^{k+1}}{(k+1)!} + O(x^{k-1}) = \frac{x^{k+1} - (k+1)\gamma x^k}{(k+1)!} + O(x^{k-1}).
\end{aligned}
$$

We get

$$
\begin{aligned}
\int_0^x \frac{t^m}{m!} f(t)\mathrm{d}t &= \sum_{k=0}^{m} \frac{(-1)^k}{(m-k)!} x^{m-k} \left(\frac{x^{k+1} - (k+1)\gamma x^k}{(k+1)!} + O(x^{k-1}) \right) \\
&= x^{m+1} \sum_{k=0}^{m} \frac{(-1)^k}{(m-k)!(k+1)!} + \gamma x^m \sum_{k=0}^{m} \frac{(-1)^{k+1}}{(m-k)!k!} + O(x^{m-1}).
\end{aligned}
$$

As for the x^{m+1}-term we obtain

$$
\begin{aligned}
\sum_{k=0}^{m} \frac{(-1)^k}{(m-k)!(k+1)!} &= \sum_{k=1}^{m+1} \frac{(-1)^{k-1}}{(m+1-k)!(k)!} = \frac{1}{(m+1)!} \sum_{k=1}^{m+1} \binom{m+1}{k} (-1)^{k-1} \\
&= -\frac{1}{(m+1)!} \left(\sum_{k=0}^{m+1} \binom{m+1}{k} (-1)^k - 1 \right) = \frac{1}{(m+1)!},
\end{aligned}
$$

where in the last step we used

$$
\sum_{k=0}^{m+1} \binom{m+1}{k} (-1)^k = \sum_{k=0}^{m+1} \binom{m+1}{k} (-1)^k (+1)^{m+1-k} = (-1+1)^{m+1} = 0.
$$

A similar application of the binomial formula shows that the x^m-term is zero, namely

$$
\sum_{k=0}^{m} \frac{(-1)^{k+1}}{(m-k)!k!} = -\sum_{k=0}^{m} \frac{(-1)^k}{(m-k)!k!} = -\frac{1}{m!}(-1+1)^m = 0
$$

Hence, we have

$$
\int_0^x \frac{t^m}{m!} f(t)\mathrm{d}t = \frac{x^{m+1}}{(m+1)!} + O(x^{m-1}),
$$

which yields

$$
\int_0^x \int_0^t \frac{s^m}{m!} f(s)\mathrm{d}s\,\mathrm{d}t = \int_0^x \left(\frac{t^{m+1}}{(m+1)!} + O(t^{m-1}) \right)\mathrm{d}t = \frac{x^{m+2}}{(m+2)!} + O(x^m)
$$

to conclude that, as $x \to 0$,

$$
m(x) = \frac{(\log 1/x)^{m+2}}{(m+2)!} + O\big((\log x)^m\big).
$$

In view of $m(x) \le \overrightarrow{\nu}(x) \le m(x) + 1$, the first statement is proved. We now proceed to derive the auxiliary function $\ell(x)$ of $m(x)$ and, in turn, of $\overrightarrow{\nu}(x)$. We have

$$
(m+2)!\big(m(\lambda x) - m(x)\big) = (-\log(\lambda x))^{m+2} - (-\log x)^{m+2} + O\big((\log x)^m\big)
$$

$$
\begin{aligned}
&= (-1)^{m+2}\big((\log x + \log\lambda)^{m+2} - (\log x)^{m+2}\big) + O\big((\log x)^m\big)\\
&= (-1)^{m+2}(m+2)\log\lambda(\log x)^{m+1} + O\big((\log x)^m\big)
\end{aligned}
$$

so that, as $x \to 0$,

$$\frac{m(\lambda x) - m(x)}{(\log x)^{m+1}} \to \frac{(-1)^{m+2}}{(m+1)!}\log\lambda.$$

So we find $\ell(x) = (\log(x))^{m+1}$ and $c = (-1)^{m+2}/(m+1)!$ in (5). Note that

$$c\ell(1/n) = \frac{(-1)^{m+2}}{(m+1)!}(-\log n)^{m+1} = \frac{(-1)^{m+2+m+1}}{(m+1)!}(\log n)^{m+1} = -\frac{(\log n)^{m+1}}{(m+1)!}.$$

Finally, an application of Theorem 1 yields the second statement. □

Remark 2. The phenomenon observed in Example 2 for $m = 1$, namely the cancellation of the term $(\log 1/x)^2$ in the expansion of $m(x)$, applies to any $m \geq 1$ meaning that the term $(\log 1/x)^{m+1}$ cancels out. Since the auxiliary function $\ell(x)$ is found to be of the order $(\log 1/x)^{m+1}$, we conclude that for any $m \geq 1$ the leading term in the expansion of $m(x)$ is sufficient for the derivation of the second order term in the expansion of $\mathrm{E}(K_n)$ according to Theorem 1. As observed in Example 2, a double application of Karamata's Theorem yields the leading term of $m(x)$ as it yields, for $x \to \infty$,

$$\frac{\int_0^x (x-t)t^m f(t)\mathrm{d}t}{x^{m+2}f(x)} \to \frac{1}{m+1}\frac{1}{m+2}.$$

These calculations can be easily extended to the case of $p \overset{d}{=} \mathrm{e}^{-X}$ for $X \sim \mathrm{ga}(1+\rho, 1)$ with $\rho > -1$. Since $\pi(\mathrm{e}^{-t})f(t)$ is regularly varying at infinity with index $\rho + 1 > 0$,

$$\frac{\int_0^x (x-t)\pi(\mathrm{e}^{-t})f(t)\mathrm{d}t}{x^{\rho+2}f(x)} \to \frac{1}{\Gamma(\rho+1)}\frac{1}{\rho+1}\frac{1}{\rho+2} = \frac{1}{\Gamma(\rho+3)}$$

so we obtain

$$\mathrm{E}(K_n) \sim \frac{1}{\Gamma(\rho+3)}(\log n)^{\rho+2}.$$

However, a more accurate approximation of $\int_0^x (x-t)\pi(\mathrm{e}^{-t})f(t)\mathrm{d}t$ is necessary in order to apply Theorem 1 and obtain the second order term in the expansion of $\mathrm{E}(K_n)$.

4 Negative binomial extension

In the following we use the notation $w_j(p)$ for the jth frequency as a function of the parameter p. Recall that the asymptotic behavior of $\mathrm{E}(K_n)$ depends on the behavior at zero of the tail mean measure $\overrightarrow{\nu}(x) = \int_0^1 \overrightarrow{\nu}(x,p)\pi(p)\mathrm{d}p$, where $\pi(p)$ is the success probability distribution, and $\overrightarrow{\nu}(x,p) = \#\{j : w_j(p) \geq x\}$ is the number of frequencies larger than a threshold $x \in [0,1]$. When the frequencies $w_j(p)$ are decreasing, $\overrightarrow{\nu}(x,p) = \sup\{j : w_j(p) \geq x\}$, that is $\overrightarrow{\nu}(x,p)$ is obtained in terms of the inverse of $w_j(p)$ with respect to j. In the case of geometric frequencies the inverse is explicitly found to be $\log(x/p)/\log(1-p)+1$, thus we have $\overrightarrow{\nu}(x,p) = \lfloor \log(x/p)/\log(1-p)+1 \rfloor$

for $p \geq x$ or, equivalently, for $w_1(p) \geq x$. In Section 3, the behavior in zero of $\overrightarrow{\nu}(x)$ was studied in terms of

$$m(x) = \int_0^1 \frac{\log(x/p)}{\log(1-p)} \mathbb{1}_{(w_1(p) \geq x)} \pi(p) \mathrm{d}p,$$

based on the fact that $m(x) \leq \overrightarrow{\nu}(x) \leq m(x) + 1$.

In this section we consider a different model for $w_j(p)$ that can be seen as an extension of the geometric case. To this aim, we resort to a derivation of the geometric weights $w_j(p) = p(1-p)^{j-1}$ as a special case of a general construction of distributions on the positive integers with decreasing frequencies. Let $\phi(r;p)$ be a probability function for $r = 1, 2, \dots$ with parameter $p \in (0,1)$. Then

$$w_j(p) = \sum_{r \geq j} \frac{\phi(r;p)}{r}, \qquad j = 1, 2, \dots,$$

form a decreasing sequence, $w_j(p) > w_{j+1}(p)$, of positive numbers summing up to one. As such, $(w_j(p))_{j \geq 1}$ defines a new distribution on the positive integers parametrized by p. An interesting instance of $\phi(r;p)$ is given by

$$\phi(r;p) = \phi(r;s,p) = \binom{r+s-2}{r-1} p^s (1-p)^{r-1}, \quad r = 1, 2, \dots$$

that is the negative binomial distribution shifted by one. The geometric frequencies are obtained by taking the scale parameter $s = 2$. In fact

$$\begin{aligned} w_j(p) &= \sum_{r \geq j} p^2 (1-p)^{r-1} = p^2 (1-p)^{j-1} \sum_{i \geq 1} (1-p)^{i-1} \\ &= p^2 (1-p)^{j-1} \sum_{i \geq 0} (1-p)^i = p(1-p)^{j-1}. \end{aligned}$$

An explicit expression for $w_j(p)$ can be found for any integer $s \geq 2$. When $s = 3$, differentiating with respect to the geometric series one finds

$$\begin{aligned} \frac{2(1-p)}{p^3} w_j(p) &= \sum_{r \geq j} (r+1)(1-p)^r = -\frac{\mathrm{d}}{\mathrm{d}p} \sum_{r \geq j} (1-p)^{r+1} = -\frac{\mathrm{d}}{\mathrm{d}p} \frac{(1-p)^{j+1}}{p} \\ &= \frac{(1-p)^j}{p^2} \big((j+1)p + (1-p) \big) = \frac{(1-p)^j}{p^2} \big(1 + jp \big) \end{aligned}$$

to conclude that

$$w_j(p) = p(1-p)^{j-1} \frac{1+jp}{2}, \quad j = 1, 2, \dots \tag{9}$$

Similar formulae are derived for $s > 3$: one finds that $w_j(p)$ is proportional to the geometric probability $p(1-p)^{j-1}$ multiplied by a polynomial in (j, p) of order determined by s. Details are omitted. For instance, $s = 4$ yields

$$w_j(p) = p(1-p)^{j-1} \frac{2 + 2jp + jp^2 + j^2p^2}{6}, \quad j = 1, 2, \dots$$

With a closed form expression of $w_j(p)$, an asymptotic evaluation of $\overrightarrow{\nu}(x, p)$ can be derived for $x \to 0$, and in turn, for $\overrightarrow{\nu}(x)$, so that the asymptotics of $\mathrm{E}(K_n)$ is obtained through Theorem

1. Let us restrict attention to $s = 3$ with $w_j(p)$ as in (9) and $\pi(p)$ the uniform distribution. Our goal is to investigate the asymptotics of $\mathrm{E}(K_n)$ in comparison with the geometric case of Proposition 1. It is reasonable to expect that $\mathrm{E}(K_n)$ grows at a faster rate: in fact, the larger s, the larger the mean of the negative binomial distribution $\phi(r; s, p)$, so $w_j(p)$ decrease slower in j for $s = 3$ compared to $s = 2$, which corresponds to the geometric case. This implies that $\overrightarrow{\nu}(x, p)$ grows slower as $x \to 0$ for $s = 3$ and, in turn, via a Tauberian theorem $\mathrm{E}(K_n)$ grows faster as $n \to \infty$. In Proposition 2 we establish that the asymptotic behavior of $\mathrm{E}(K_n)$ differs from the one found in Proposition 1 only in the second order term of the expansion.

Proposition 2. *Let $w_j(p)$ be defined as in (9) and p be uniformly distributed on $(0, 1)$. Then*

$$\overrightarrow{\nu}(x) = \frac{1}{2}\Big(\log\frac{1}{x}\Big)^2 + \log\frac{1}{x}\log\log\frac{1}{x} - \gamma\log\frac{1}{x} - (1+\log 2)\log\frac{1}{x} + O\Big(\log\log\frac{1}{x}\Big), \quad x \to 0$$

$$\mathrm{E}(K_n) = \frac{1}{2}\big(\log n\big)^2 + \log n\log(\log n) - (1+\log 2)\log n + o(\log n), \quad n \to \infty.$$

Proof. Let $m(x, p) \geq 0$ be defined by

$$p(1-p)^{m(x,p)}\frac{1}{2}\big(1 + p + p\,m(x, p)\big) = x, \tag{10}$$

which corresponds to the solution in m to the equation $w_{m+1}(p) = x$. Note that $m(x, p) \geq 0$ when $w_1(p) \geq x$, $\overrightarrow{\nu}(x, p) = \lfloor m(x, p) + 1\rfloor \mathbb{1}_{(w_1(p)\geq x)}$ and, as in the geometric case, $m(x) \leq \overrightarrow{\nu}(x) \leq m(x) + 1$ for

$$m(x) = \int_0^1 m(x, p)\mathbb{1}_{(w_1(p)\geq x)}\mathrm{d}p.$$

Equation (11) provides an asymptotic expansion of $m(x, p)$ as $x \to 0$. The proof is reported in the Appendix and involves the Lambert W function (Corless et al., 1996).

$$\begin{aligned} m(x,p) = {} & \frac{\log x/p}{\log(1-p)} \\ & - \frac{1}{\log(1-p)}\log\left(\frac{1}{2}\left(1 + p + p\frac{\log x/p}{\log(1-p)} + \frac{p}{\log(1-p)}\log\frac{-2\log(1-p)}{p}\right)\right) \\ & + \frac{1}{\log(1-p)}O\left(\frac{\log(\log 1/x)}{\log 1/x}\right). \end{aligned} \tag{11}$$

An heuristic derivation inspired by (Barndorff-Nielsen and Cox, 1989, Example 3.13) is as follows. In equation (10), we have that for small x, $m(x, p)$ will be large and the term $p(1-p)^{m(x,p)}$ is thus dominant. Rewrite the equation after taking the log and keeping $m(x, p)$ on the left hand side,

$$m(x,p) = \frac{\log x/p}{\log(1-p)} - \frac{1}{\log(1-p)}\log\left(\frac{1}{2}\big(1 + p + pm(x, p)\big)\right).$$

It defines a convergent iterative scheme via

$$m_{(k)}(x,p) = \frac{\log x/p}{\log(1-p)} - \frac{1}{\log(1-p)}\log\left(\frac{1}{2}\big(1 + p + pm_{(k-1)}(x, p)\big)\right)$$

with $m_{(1)}(x,p)$ solution to $p(1-p)^{m(x,p)} = x$, that is

$$m_{(1)}(x,p) = \frac{\log x/p}{\log(1-p)}.$$

For $k=2$ we get

$$m_{(2)}(x,p) = \frac{\log x/p}{\log(1-p)} - \frac{1}{\log(1-p)} \log\left(\frac{1}{2}\Big(1+p+p\frac{\log x/p}{\log(1-p)}\Big)\right),$$

which nearly matches the expansion in (11). Now we use it to evaluate the behavior of $m(x)$ and, in turn, $\overrightarrow{\nu}(x)$ as $x \to 0$. Note that

$$m(x) = \int_0^1 m(x,p)\mathbb{1}_{(w_1(p)\geq x)}\mathrm{d}p = \int_{\tilde{x}}^1 m(x,p)\mathrm{d}p,$$

where $\tilde{x}$ is defined by $w_1(\tilde{x}) = x$, that is $\tilde{x}(1+\tilde{x})/2 = x$. It is easy to check that $x \leq \tilde{x} \leq 2x$ for any x and $\tilde{x} \sim 2x$ as $x \to 0$. In order to exploit the derivation of the asymptotic expansion of $m(x)$ as $x \to 0^+$ in the geometric case, cf. proof of Proposition 1, we use (11) as follows:

$$m(x) = \int_x^1 \frac{\log x/p}{\log(1-p)}\mathrm{d}p - \int_x^{\tilde{x}} \frac{\log x/p}{\log(1-p)}\mathrm{d}p + \int_{\tilde{x}}^1 \left(m(x,p) - \frac{\log x/p}{\log(1-p)}\right)\mathrm{d}p.$$

From Proposition 1, as $x \to 0$

$$\int_x^1 \frac{\log x/p}{\log(1-p)}\mathrm{d}p = \frac{1}{2}\Big(\log\frac{1}{x}\Big)^2 - \gamma\log\frac{1}{x} + O(1).$$

The first statement of the thesis about the behavior of $\overrightarrow{\nu}(x)$ as $x \to 0$ is then implied by $m(x) \leq \overrightarrow{\nu}(x) \leq m(x)+1$ and by showing that, as $x \to 0$,

$$\int_x^{\tilde{x}} \frac{\log x/p}{\log(1-p)}\mathrm{d}p = O(1) \tag{12}$$

$$\int_{\tilde{x}}^1 \left(m(x,p) - \frac{\log x/p}{\log(1-p)}\right)\mathrm{d}p = \log\frac{1}{x}\log\log\frac{1}{x} - (1+\log 2)\log\frac{1}{x} + O\Big(\log\log\frac{1}{x}\Big). \tag{13}$$

As for (12), the maximum of $(\log x/p)/\log(1-p)$ is attained at $p = p(x)$, where $p(x)$, the solution to the first order equation in p $-(1-p)\log(1-p)+p\log x/p = 0$, goes to zero as $x \to 0$. It can be shown that that $p(x) \geq 2x$ for $x \leq 1/4$, that is $(\log x/p)/\log(1-p)$ is increasing for $x \leq p \leq 2x$ and x sufficiently small. Since $2x \geq \tilde{x}$,

$$\int_x^{\tilde{x}} \frac{\log x/p}{\log(1-p)}\mathrm{d}p \leq \int_x^{2x} \frac{\log x/p}{\log(1-p)}\mathrm{d}p \leq x\frac{\log x/(2x)}{\log(1-2x)} = -x\log 2/\log(1-2x) \leq \log 2/2$$

so (12) follows. As for (13), note that by the change of variable $t = \log 1/p$,

$$\int_{\tilde{x}}^1 -\frac{1}{\log(1-p)}\mathrm{d}p = \int_0^{\log 1/\tilde{x}} f(t)\mathrm{d}t = F(\log 1/\tilde{x}),$$

for $f(t)$ in (6) and $F(t)$ the primitive of $f(t)$. By equation (11), (7) and $\tilde{x} \sim 2x$ as $x \to 0$, we have that, as $x \to 0$,

$$\int_{\tilde{x}}^1 \left(m(x,p) - \frac{\log x/p}{\log(1-p)}\right)\mathrm{d}p = -\log 2 \log 1/x + O(\log(\log 1/x))$$
$$+ \int_{\tilde{x}}^1 -\frac{1}{\log(1-p)} \log\left(1+p+p\frac{\log x/p}{\log(1-p)} + \frac{p}{\log(1-p)}\log\frac{-2\log(1-p)}{p}\right)\mathrm{d}p.$$

In studying the asymptotic behavior of the integral in the last display, it is sufficient to focus on

$$\int_{\tilde{x}}^1 -\frac{1}{\log(1-p)} \log\left(1+p\frac{\log x/p}{\log(1-p)}\right)\mathrm{d}p,$$

since the extra terms inside the logarithm satisfy

$$-2\mathrm{e}^{-1} \le p + \frac{p}{\log(1-p)}\log\frac{-2\log(1-p)}{p} \le 1$$

for $0 \le p \le 1$. By the change of variable $t = \log 1/p$

$$\int_{\tilde{x}}^1 -\frac{1}{\log(1-p)} \log\left(1+p\frac{\log x/p}{\log(1-p)}\right)\mathrm{d}p = \int_0^{\log 1/\tilde{x}} \log\Big(1+\big(\log 1/\tilde{x} - t\big)f(t)\Big)\, f(t)\mathrm{d}t,$$

for $f(t)$ defined in (6). Hence, (13) is implied by

$$r(x) = \int_0^x \log\big(1+\big(x-t\big)f(t)\big)\, f(t)\mathrm{d}t = x\log(x) - x + O(\log x), \tag{14}$$

as $x \to \infty$. We have

$$\begin{aligned} r(x) &= \int_0^x \left(\log x + \log f(t) + \log\frac{1+\big(x-t\big)f(t)}{xf(t)}\right) f(t)\mathrm{d}t \\ &= \log(x)F(x) + \int_0^x f(t)\log f(t)\mathrm{d}t + \int_0^x \log\left(1+\frac{1-tf(t)}{xf(t)}\right)f(t)\mathrm{d}t. \end{aligned}$$

The first term on the right hand side is $\log x(x-\gamma+O(\mathrm{e}^{-x}))$, as $x \to \infty$, due to the asymptotic expansion of $F(x)$ in (7). The second term is easily shown to be bounded in absolute value uniformly in x. As for the third term, we are left to show that

$$\int_0^x \log\left(1+\frac{1-tf(t)}{xf(t)}\right)f(t)\mathrm{d}t = -x + \log x + O(1),$$

as $x \to \infty$. To this aim, it is convenient to split the integral as

$$\int_0^1 \log\left(1+\frac{1-tf(t)}{xf(t)}\right)f(t)\mathrm{d}t + \int_1^x \log\left(1+\frac{1-tf(t)}{xf(t)}\right)f(t)\mathrm{d}t. \tag{15}$$

The first integral in (15) is bounded in x since

$$\int_0^1 \log\left(1+\frac{1-tf(t)}{xf(t)}\right)f(t)\mathrm{d}t \le \frac{1}{x}\int_0^1 \big(1-tf(t)\big)\mathrm{d}t,$$

whereas the last integral is a positive and finite constant. As for the second integral in (15), since $1-f(t) \sim \mathrm{e}^{-t}/2$ for $t \to \infty$, cf. Lemma 1,

$$\int_1^x \log\left(1+\frac{1-tf(t)}{xf(t)}\right)f(t)\mathrm{d}t \sim \int_1^x \log\left(1+\frac{1-t}{x}\right)\mathrm{d}t$$

as $x \to \infty$ and

$$\int_1^x \log\left(1 + \frac{1-t}{x}\right)\mathrm{d}t = -x + \log x + 1$$

by direct calculation. Hence (14), and in turn (13), follow. The proof of the first statement of the proposition is then complete. The second statement about the expansion of $\mathrm{E}(K_n)$, as $n \to \infty$, follows from an application of Theorem 1. $\square$

Appendix

Proof of Theorem 1

The proof follows arguments similar to those of (Bingham et al., 1987, Theorem 3.9.1). It consists in evaluating $\big(\Phi(n) - \overrightarrow{\nu}(1/n)\big)/\ell(1/n)$ in the decomposition

$$\mathrm{E}(K_n) = \overrightarrow{\nu}(1/n) + \frac{\Phi(n) - \overrightarrow{\nu}(1/n)}{\ell(1/n)}\ell(1/n) + \mathrm{E}(K_n) - \Phi(n).$$

Indeed, as $\Phi(n) \sim \overrightarrow{\nu}(1/n)$ and $\ell(1/n)$ are slowly varying, $|\mathrm{E}(K_n) - \Phi(n)| \leq \frac{2}{n}\Phi(n) = o(\ell(1/n))$, cf. (4), so the conclusion follows by showing that $\big(\Phi(n) - \overrightarrow{\nu}(1/n)\big)/\ell(1/n) \to -c\gamma$. To this aim,

$$\begin{aligned}
\frac{\Phi(1/x) - \overrightarrow{\nu}(x)}{\ell(x)} &= \frac{1}{\ell(x)}\left[\int_0^\infty \frac{1}{x}\mathrm{e}^{-y/x}\overrightarrow{\nu}(y)\mathrm{d}y - \int_0^\infty \overrightarrow{\nu}(x)\mathrm{e}^{-\lambda}\mathrm{d}\lambda\right] \\
&= \frac{1}{\ell(x)}\left[\int_0^\infty \mathrm{e}^{-\lambda}\overrightarrow{\nu}(\lambda x)\mathrm{d}\lambda - \int_0^\infty \overrightarrow{\nu}(x)\mathrm{e}^{-\lambda}\mathrm{d}\lambda\right] = \int_0^\infty \frac{\overrightarrow{\nu}(\lambda x) - \overrightarrow{\nu}(x)}{\ell(x)}\mathrm{e}^{-\lambda}\mathrm{d}\lambda \\
&\to \int_0^\infty c(\log\lambda)\mathrm{e}^{-\lambda}\mathrm{d}\lambda = c\Gamma'(1) = -c\gamma, \quad \text{as } x \to 0,
\end{aligned}$$

where in taking the limit we used the dominated convergence theorem, cf. global bounds in (Bingham et al., 1987, Theorem 3.8.6). $\square$

Proof of Lemma 1

We will use the following integral representation of the Euler-Mascheroni constant:

$$\gamma = \int_0^\infty \left(\frac{1}{1-\mathrm{e}^{-x}} - \frac{1}{x}\right)\mathrm{e}^{-x}\mathrm{d}x.$$

By the change of variable $t = -\log(1-\mathrm{e}^{-x})$ so that $\mathrm{d}t = -\frac{\mathrm{e}^{-x}}{1-\mathrm{e}^{-x}}\mathrm{d}x$ and $x = -\log(1-\mathrm{e}^{-t})$, we obtain

$$\begin{aligned}
\gamma &= \int_0^\infty \left(\frac{1}{1-\mathrm{e}^{-x}} - \frac{1}{x}\right)\mathrm{e}^{-x}\mathrm{d}x = \int_0^\infty \left(1 - \frac{1-\mathrm{e}^{-x}}{x}\right)\frac{\mathrm{e}^{-x}}{1-\mathrm{e}^{-x}}\mathrm{d}x \\
&= \int_0^\infty \left(1 - \frac{\mathrm{e}^{-t}}{-\log(1-\mathrm{e}^{-t})}\right)\mathrm{d}t = \int_0^\infty \big(1 - f(t)\big)\mathrm{d}t.
\end{aligned}$$

It is easy to check that $\lim_{t\to 0} f(t) = 0$ and $\lim_{t\to\infty} f(t) = 1$. As for the tail behavior, by the Taylor expansion of $\log(1+x) = x - x^2/2 + O(x^3)$ as $x \to 0$, we find that, as $t \to \infty$,

$$1 - f(t) = 1 - \frac{\mathrm{e}^{-t}}{-\log(1-\mathrm{e}^{-t})} \sim 1 - \frac{\mathrm{e}^{-t}}{\mathrm{e}^{-t} + \mathrm{e}^{-2t}/2} = \frac{\mathrm{e}^{-2t}/2}{\mathrm{e}^{-t} + \mathrm{e}^{-2t}/2} \sim \frac{\mathrm{e}^{-t}}{2}.$$

$\square$

Proof of Equation (11)

Let $W(z)$ be the Lambert function defined by $W(z)\mathrm{e}^{W(z)} = z$, where $W(z)$ is a multivalued function that has, for z a real number, two branches, the principal branch $W_0(z)$ for $W(z) \geq -1$, and the branch $W_{-1}(z)$ for $W(z) < -1$. We have that $\lim_{z\to 0^+} W_0(z) = 0$ while $\lim_{z\to 0^-} W_{-1}(z) = -\infty$. In particular, according to (Corless et al., 1996, Section 4), as $z \to 0^-$

$$W_{-1}(z) = \log(-z) - \log(-\log(-z)) + O\left(\frac{\log(-\log(-z))}{\log(-z)}\right). \tag{16}$$

By algebraic manipulation of (10)

$$\begin{aligned}
&p(1-p)^{m(x,p)}\frac{1}{2}(1+p+p\,m(x,p)) = x; \quad \mathrm{e}^{\log(1-p)m(x,p)}(1+p+p\,m(x,p)) = 2x/p;\\
&(1+p+p\,m(x,p))\log(1-p)\mathrm{e}^{\log(1-p)m(x,p)} = \frac{2x\log(1-p)}{p};\\
&(1+p+p\,m(x,p))\log(1-p)\mathrm{e}^{\log(1-p)(1/p+1+m(x,p)} = \frac{2x\log(1-p)}{p}\mathrm{e}^{(1/p+1)\log(1-p)};\\
&\frac{\log(1-p)}{p}(1+p+p\,m(x,p))\mathrm{e}^{\frac{\log(1-p)}{p}(1+p+p\,m(x,p))} = \frac{2x\log(1-p)}{p^2}\mathrm{e}^{\frac{1+p}{p}\log(1-p)};\\
&\frac{\log(1-p)}{p}(1+p+p\,m(x,p)) = W(z),
\end{aligned}$$

where, in the last display,

$$z = \frac{2x\log(1-p)}{p^2}\exp\left(\frac{1+p}{p}\log(1-p)\right). \tag{17}$$

Solving for $m(x,p)$,

$$m(x,p) = \frac{1}{p\log(1-p)}\left(pW_{-1}(z) - \log(1-p)\right) - 1, \tag{18}$$

where we used the branch W_{-1} of $W(z)$ since z in (17) is ≤ 0 and $W(z) \leq -1$. The fact that $W(z) \leq -1$ is easily checked by using $m(x,p) \geq 0$. In fact

$$\begin{aligned}
&\frac{1}{p\log(1-p)}\left(pW(z) - \log(1-p)\right) - 1 \geq 0; \quad pW(z) - \log(1-p) \leq p\log(1-p);\\
&pW(z) \leq \log(1-p)(1+p); \quad W(z) \leq \log(1-p)\frac{1+p}{p}
\end{aligned}$$

and $\frac{1+p}{p}\log(1-p)$ decreases from -1 to $-\infty$ for $p \in (0,1)$. From (17) one finds that $z \to 0^-$ as $x \to 0^+$. In particular, from $w_1(p) > x$, that is $p(1+p)/2 > x$, it follows that

$$\frac{1+p}{p}\log(1-p)\exp\left(\frac{1+p}{p}\log(1-p)\right) \le z \le 0$$

and the lower bound is larger than $-\mathrm{e}^{-1}$ for any $p \in (0,1)$. Hence $\log(-z) < -1$ and $\log(-\log(-z)) > 0$. By direct calculation

$$\log(-z) = \log(1-p)\left(\frac{\log x/p}{\log(1-p)} + \frac{1}{\log(1-p)}\log\frac{-2\log(1-p)}{p} + \frac{1+p}{p}\right)$$

and

$$\frac{1}{p\log(1-p)}\Big(p\log(-z) - \log(1-p)\Big) - 1 = \frac{\log x/p}{\log(1-p)} + \frac{1}{\log(1-p)}\log\frac{-2\log(1-p)}{p}.$$

Substitute in (18) $W_{-1}(z)$ for $\log(-z) - \log(-\log(-z))$ according to the two terms expansion in (16), to find

$$\begin{aligned}
&\frac{1}{p\log(1-p)}\Big(p\Big(\log(-z) - \log(-\log(-z))\Big) - \log(1-p)\Big) - 1 \\
&\quad= \frac{1}{p\log(1-p)}\Big(p\log(-z) - \log(1-p)\Big) - 1 - \frac{1}{\log(1-p)}\log(-\log(-z)) \\
&\quad= \frac{\log x/p}{\log(1-p)} + \frac{1}{\log(1-p)}\log\frac{-2\log(1-p)}{p} - \frac{1}{\log(1-p)}\log(-\log(-z)) \\
&\quad= \frac{\log x/p}{\log(1-p)} - \frac{1}{\log(1-p)}\log\left(\frac{p}{2\log(1-p)}\log(-z)\right) \\
&\quad= \frac{\log x/p}{\log(1-p)} - \frac{1}{\log(1-p)}\log\left(\frac{p}{2}\left(\frac{\log x/p}{\log(1-p)} + \frac{1}{\log(1-p)}\log\frac{-2\log(1-p)}{p} + \frac{1+p}{p}\right)\right) \\
&\quad= \frac{\log x/p}{\log(1-p)} - \frac{1}{\log(1-p)}\log\left(\frac{1}{2}\left(1 + p + p\frac{\log x/p}{\log(1-p)} + \frac{p}{\log(1-p)}\log\frac{-2\log(1-p)}{p}\right)\right).
\end{aligned}$$

The remainder of the expansion is easily found. □

References

Archibald, M., Knopfmacher, A., and Prodinger, H. (2006). The number of distinct values in a geometrically distributed sample. *European Journal of Combinatorics*, 27:1059–1081.

Argiento, R., Cremaschi, A., and Vannucci, M. (2020). Hierarchical normalized completely random measures to cluster grouped data. *Journal of the American Statistical Association*, 115(529):318–333.

Arratia, R., Barbour, A. D., and Tavaré, S. (2003). *Logarithmic Combinatorial Structures: A Probabilistic Approach.* EMS Monographs in Mathematics. European Mathematical Society.

Ayed, F., Lee, J., and Caron, F. (2019). Beyond the Chinese Restaurant and Pitman-Yor processes: Statistical Models with double power-law behavior. In Chaudhuri, K. and Salakhutdinov, R., editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97, pages 395–404. PMLR.

Barndorff-Nielsen, O. E. and Cox, D. R. (1989). *Asymptotic Techniques for Use in Statistics*. Chapman and Hall.

Bassetti, F., Casarin, R., and Rossini, L. (2020). Hierarchical species sampling models. *Bayesian Anal.*, 15(3):809–838.

Bingham, N. H., Goldie, C. M., and Teugels, J. L. (1987). *Regular Variation*. Cambridge University Press.

Camerlenghi, F., Lijoi, A., Orbanz, P., and Prünster, I. (2019). Distribution theory for hierarchical processes. *Ann. Statist.*, 47(1):67–92.

Caron, F. and Fox, E. B. (2017). Sparse graphs using exchangeable random measures. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(5):1295–1366.

Corless, R. M., Gonnet, G. H., Hare, D. E. G., Jeffrey, D. J., and Knuth, D. E. (1996). On the Lambert W function. *Advances in Computational Mathematics*, 5:329–359.

Dahl, D. B., Day, R., and Tsai, J. W. (2017). Random partition distribution indexed by pairwise information. *Journal of the American Statistical Association*, 112(518):721–732.

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

De Blasi, P., Martinez, A. F., Mena, R. H., and Pruenster, I. (2020). On the inferential implications of decreasing weight structures in mixture models. *Computational Statistics and Data Analysis*, 147:106940.

Di Benedetto, G., Caron, F., and Teh, Y. W. (2020). Non-exchangeable random partition models for microclustering. *Annals of Statistics*. (forthcoming).

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.*, 1:209–230.

Fuentes-García, R., Mena, R. H., and Walker, S. G. (2010). A new Bayesian nonparametric mixture model. *Communications in Statistics - Simulation and Computation*, 39(4):669–682.

Gnedin, A. (2004). The Bernoulli sieve. *Bernoulli*, 10:79–96.

Gnedin, A. (2010). Regeneration in random combinatorial structures. *Probability Surveys*, 7:105–156.

Gnedin, A., Hansen, B., and Pitman, J. (2007). Notes on the occupancy problem with infinitely many boxes: general asymptotics and power laws. *Probability Surveys*, 4:146–171.

Gnedin, A., Iksanov, A. M., Pavlo, N., and Uwe, R. (2009). The Bernoulli sieve revisited. *The Annals of Applied Probability*, 19:1634–1655.

Gnedin, A. and Pitman, J. (2005). Regenerative composition structures. *Ann. Probab.*, 33(2):445–479.

Gnedin, A., Pitman, J., and Yor, M. (2006a). Asymptotic laws for compositions derived from transformed subordinators. *Ann. Probab.*, 34(2):468–492.

Gnedin, A., Pitman, J., and Yor, M. (2006b). Asymptotic laws for regenerative compositions: gamma subordinators and the like. *Probability Theory and Related Fields*, 135.

Gutiérrez, L., Gutiérrez-Peña, E., and Mena, R. H. (2014). Bayesian nonparametric classification for spectroscopy data. *Comput. Statist. Data Anal.*, 78:56–68.

Hatjispyros, J., Merkatas, C., Nicoleris, T., and Walker, S. (2018). Dependent mixtures of geometric weights priors. *Comput. Statist. Data Anal.*, 119:1–18.

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

Karlin, S. (1967). Central limit theorems for certain infinite urn schemes. *J. Math. Mech*, 17(24):373–401.

Korwar, R. M. and Hollander, M. (1973). Contributions to the theory of Dirichlet processes. *Ann. Probab.*, 1(4):705–711.

Lijoi, A., Mena, R. H., and Prünster, I. (2007a). A Bayesian nonparametric method for prediction in EST analysis. *BMC Bioinformatics*, 8:339.

Lijoi, A., Mena, R. H., and Prünster, I. (2007b). Bayesian nonparametric estimation of the probability of discovering new species. *Biometrika*, 94(4):769–786.

Lijoi, A., Mena, R. H., and Prünster, I. (2007c). Controlling the reinforcement in Bayesian non-parametric mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(4):715–740.

Lijoi, A., Muliere, P., Prünster, I., and Taddei, F. (2016). Innovation, growth and aggregate volatility from a Bayesian nonparametric perspective. *Electron. J. Statist.*, 10(2):2179–2203.

Mena, R. H., Ruggiero, M., and Walker, S. G. (2011). Geometric stick-breaking processes for continuous-time Bayesian nonparametric modeling. *J. Statist. Plann. Inference*, 141(9):3217–3230.

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

Pitman, J. (2006). *Combinatorial Stochastic Processes*. Springer.

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

Teh, Y. W. (2006). A hierarchical Bayesian language model based on Pitman-Yor processes. *Proceedings of Coling/ACL*, pages 985–992.