

OCCUPANCY SCHEMES ASSOCIATED TO YULE PROCESSES

PHILIPPE ROBERT * ** AND
FLORIAN SIMATOS, * *** INRIA

Abstract

An occupancy problem with an infinite number of bins and a random probability vector for the locations of the balls is considered. The respective sizes of the bins are related to the split times of a Yule process. The asymptotic behavior of the landscape of the first empty bins, i.e. the set of corresponding indices represented by point processes, is analyzed and convergences in distribution to mixed Poisson processes are established. Additionally, the influence of the random environment, the random probability vector, is analyzed. It is represented by two main components: an independent, identically distributed sequence and a fixed random variable. Each of these components has a specific impact on the qualitative behavior of the stochastic model. It is shown in particular that, for some values of the parameters, some rare events, which are identified, determine the asymptotic behavior of the average values of the number of empty bins in some regions.

Keywords: Occupancy scheme; Poisson process; convergence of point processes; random environment; rare events

2000 Mathematics Subject Classification: Primary 60K20

Secondary 60K37; 60J80

1. Introduction

Occupancy schemes in terms of bins and balls offer a very flexible and elegant way to formulate various problems in computer science, biology, and applied mathematics for example. One of the earliest models investigated in the literature consists of throwing m balls at random into n identical bins. Asymptotic behavior of occupancy variables have been analyzed when n grows to ∞ , with different scalings in n for the variable m . The books by Johnson and Kotz [12] and Kolchin *et al.* [15] are classical references on this topic. See also Chapter 6 of Barbour *et al.* [3] for a recent presentation of these problems.

An extension of these models is when there is an infinite number of bins and a probability vector (p_n) on \mathbb{N} describing the way balls are sent: for $n \geq 0$, p_n is the probability that a ball is sent to the n th bin. In one of the first studies in this setting, Karlin [13] analyzed the asymptotic behavior of the number of occupied bins. More recently, Hwang and Janson [11] proved in a quite general framework central limit results for these quantities. In this setting, some additional variables are also of interest such as the sets of indices of occupied or empty bins, adding a geometric component to these problems. For specific probability vectors (p_n) , Csáki and Földes [4] and Flajolet and Martin [6] investigated the index of the first empty bin. See the recent survey by Gnedin *et al.* [8] for more references on the occupancy problem with infinitely many bins.

Received 17 October 2008; revision received 19 March 2009.

* Postal address: INRIA Paris-Rocquencourt, Domaine de Voluceau, BP 105, 78153 Le Chesnay, France.

** Email address: philippe.robert@inria.fr

*** Email address: florian.simatos@inria.fr

A further extension of these stochastic models is to consider *random* probability vectors. Gnedin [7] (and subsequent papers) analyzed the case where (p_n) decays geometrically fast according to some random variables, i.e. for $n \geq 1$, $p_n = \prod_{i=1}^{n-1} Y_i(1 - Y_n)$, where (Y_i) are independent and identically distributed (i.i.d.) random variables on $(0, 1)$. Various asymptotic results on the number of occupied bins in this case have been obtained. The random vector can be seen as a ‘random environment’ for the bins and balls problem, significantly complicating the asymptotic results in some cases. In particular, the indices of the urns in which the balls fall are no longer independent random variables as in the deterministic case.

The general goal of this paper is to investigate in detail the impact of this randomness for a bins and balls problem associated to a Yule process; see [2, p. 109] for the definition of a Yule process. This (quite natural) stochastic model has its origin in network modeling; see [23] for a detailed presentation. It can be described as follows. The nondecreasing sequence (t_n) of split times of the Yule process defines the bins, with the n th bin, $n \geq 1$, being the interval $(t_{n-1}, t_n]$. The locations of the balls are represented by independent exponential random variables with parameter ρ . The main problem investigated here concerns the asymptotic description of the set of indices of the first empty bins when the number of balls goes to ∞ . Mathematically, it is formulated as a convergence in distribution of rescaled point processes having Dirac masses at the indices of empty bins.

For $n \geq 1$, if P_n is the probability that a ball falls into the n th bin, it is easily seen that, for a large n , P_n has a power law decay; it can be expressed as $V E_n/n^{\rho+1}$, where (E_n) are i.i.d. exponential random variables with parameter 1 and V is some independent random variable related to the limit of a martingale. The randomness of the probability vector (P_n) has two components: one which is a part of an i.i.d. sequence, changing from one bin to another, and the other being ‘fixed once and for all’, inducing a dependency structure. As will be seen, the two components have separately a significant impact on the qualitative behavior of this model.

1.1. Convergence in distribution and rare events

Because the variables (E_n) can be arbitrarily small with positive probability, empty bins are likely to be created earlier (i.e. with smaller indices) than for a deterministic probability vector with the same power law decay. It is shown in fact that, for the convergence in distribution, the first empty bins occur around indices of the order of $n^{1/(\rho+2)}$ instead of $(n/\log n)^{1/(1+\rho)}$ in the deterministic case.

The variable V has a more subtle impact; when $\rho > 1$, it is shown that, owing to some heavy tail property of V^{-1} , rare events affect the asymptotic behavior of *averages* of some of the characteristics. For $\alpha \in [1/(2\rho + 1), 1/(\rho + 2))$, despite the fact that the number of empty bins with indices of order n^α converges *in distribution* to 0, the corresponding average converges to $+\infty$. When $\rho < 1$, the average converges to 0. A phase transition phenomenon at $\rho = 1$ has been identified through simulations in a related context (communication networks) in [22]. It is not apparent as far as convergence in distribution is concerned, but it shows up when the average quantities are considered. This phenomenon is due to rare events related to the total size of the first $\lfloor \rho \rfloor$ bins (where $\lfloor \cdot \rfloor$ is the integer part). On these events, the indices of the first empty bins are of the order $n^{1/(2\rho+1)} \ll n^{1/(\rho+2)}$ and a lot of them are created at this occasion. See Proposition 7 and Corollary 2 for a precise statement of this result. Concerning the generality of the results obtained, it is believed that some of them hold in a more general setting, e.g. for the underlying branching process; see Section 6.

1.2. Point processes

Technically, we mainly use point processes on \mathbb{R}_+ to describe the asymptotic behavior of the indices of the first empty bins, not just the index of the first empty bin (or the subsequent bins), as is usually the case in the literature. It turns out that it is quite appropriate in our setting to obtain a full description of the set of the first empty bins and, moreover, it reduces the technicalities of some of the proofs. One of the arguments used in the proofs of the convergence results is a simple convergence result of two-dimensional point processes to a Poisson point process with some intensity measure. A one-dimensional equivalent of this point of view is implicit in most of the literature; see [9] and [11] in particular.

The paper is organized as follows. In Section 2 we introduce the stochastic model investigated. The main results concerning convergence of related point processes in \mathbb{R}_+^2 are presented in Section 3. Convergence results for the indices of empty bins are proved in Section 4. In Section 5 we investigate in detail the case in which $\rho \geq 1$. In Section 6 we present some possible extensions.

2. A bins and balls problem in a random environment

The stochastic model is described in detail and some notation is introduced.

2.1. The bins

Let (E_i) be a sequence of i.i.d. exponential random variables with parameter 1. Define the nondecreasing sequence (t_n) by $t_0 = 0$ and, for $n \geq 1$,

$$t_n = \sum_{i=1}^n \frac{1}{i} E_i.$$

It is easy to check that, for $x \geq 0$,

$$P(t_n \leq x) = P(\max(E_1, E_2, \dots, E_n) \leq x) = (1 - e^{-x})^n. \tag{1}$$

The n th bin will be identified by the interval $(t_{n-1}, t_n]$.

If $H_n = 1 + \frac{1}{2} + \dots + 1/n$ is the n th harmonic number, since $(t_n - H_n)$ is a square-integrable martingale whose increasing process is given by

$$E[(t_n - H_n)^2] = \sum_{i=1}^n \frac{1}{i^2},$$

then $(M_n) := (t_n - \log(n + 1))$ is almost surely converging to some finite random variable M_∞ . See [18] or [24]. By using (1), it is not difficult to show that the distribution of M_∞ is given by

$$P(M_\infty \leq x) = \exp(-e^{-x}), \quad x \in \mathbb{R}. \tag{2}$$

An alternative description of the sequence (t_n) is provided by the split times of a Yule (branching) process starting with one individual. See [2].

2.2. The balls

The locations of the balls are given by an independent sequence (B_j) of i.i.d. exponential random variables with parameter ρ for some $\rho > 0$.

Conditionally on the point process (t_n) associated with the locations of the bins, the probability that a given ball falls into the n th bin $(t_{n-1}, t_n]$ is given by

$$\begin{aligned} P_n &= P(B_1 \in (t_{n-1}, t_n] \mid (t_n)) \\ &= \exp(-\rho t_{n-1}) - \exp(-\rho t_n) \\ &= \exp(-\rho t_{n-1}) \left(1 - \exp\left(-\frac{\rho E_n}{n}\right) \right). \end{aligned}$$

This quantity can be rewritten as

$$P_n = \frac{1}{n^{\rho+1}} W_n^\rho Z_n \quad \text{with} \quad Z_n = n \left(1 - \exp\left(-\frac{\rho E_n}{n}\right) \right) \text{ and } W_n = \exp(-M_{n-1}). \quad (3)$$

The variables W_n and Z_n are independent random variables with different behaviors.

- (a) The variables (Z_n) are independent and converge in distribution to an exponentially distributed random variable with parameter $1/\rho$.
- (b) The random variables (W_n) converge almost surely to the finite random variable $W_\infty = \exp(-M_\infty)$, which is exponentially distributed with parameter 1.

This suggests an asymptotic representation of the sequence (P_n) as

$$\left(\frac{1}{n^{\rho+1}} W_\infty^\rho F_n \right),$$

where (F_n) is an i.i.d. sequence of exponential random variables with mean ρ independent of W_∞ . The sequence (P_n) has a power law decay with a random coefficient consisting of the product of two terms: a fixed random variable W_∞^ρ and an element of an i.i.d. sequence. As will be seen, these two terms have a significant impact on the bins and balls problem studied in this paper.

3. Convergence of point processes

One of the main results, Theorem 2 in Section 4, which establishes convergence results for the indices of the first empty bins is closely related to the asymptotic behavior of the point process $\{(i/n^{1/(2+\rho)}, nP_i), i \geq 1\}$ on \mathbb{R}_+^2 . For this reason, some results on the convergence of point processes in \mathbb{R}_+^2 are first proved. The point process associated to the sequence (nP_i) appears quite naturally, especially in view of the Poisson transform used in the proof of Theorem 2. This is also a central variable in [11] in some cases.

An important tool to study point processes in \mathbb{R}_+^d for some $d \geq 1$ is the Laplace transform. If $\mathcal{N} = \{t_n, n \geq 1\}$ is a point process and f is a function in $C_c^+(\mathbb{R}_+^d)$, the set of nonnegative continuous functions with a compact support, it is defined as $E[\exp(-\mathcal{N}(f))]$, where

$$\mathcal{N}(f) := - \sum_{n \geq 1} f(t_n).$$

This functional uniquely determines the distribution of \mathcal{N} and it is a key tool to establish convergence results. See [5] and [19] for a comprehensive presentation of this framework. In the following, the quantity $\mathcal{N}(A)$ denotes the number of t_n s in the subset A of \mathbb{R}_+^d .

The main results of this section establish convergence in distribution to mixed Poisson point processes, i.e. distributed as a Poisson point process with a parameter which is a random variable. A natural tool in this domain is the Chen–Stein approach which gives the convergence in distribution and, generally, quite good bounds on the convergence rate. See Chapter 10 of [3] for example. This has been used in [23], when the probability vector is deterministic. For some of the results of this section, this approach can probably also be used. Unfortunately, owing to the almost surely converging sequence (W_n) creating a dependency structure, it does not seem that the main convergence result, Theorem 1, below, can be proved in a simple way by using Chen–Stein’s method. The main problem being that of conditioning on the variable W_∞ and, at the same time, ensuring that the upper bounds on the total variation distance converge to 0.

Condition 1. *A sequence of independent random variables (X_i) satisfies Condition 1 if there exist some $\alpha > 0$ and $\eta > 0$ such that, for all $i \geq 1$,*

$$|\mathbb{P}(X_i \leq x) - \alpha x| \leq Cx^2 \quad \text{when } 0 \leq x \leq \eta. \tag{4}$$

The following proposition is a preliminary result that will be used to prove the main convergence results for the indices of the first empty bins.

Proposition 1. (Convergence to a Poisson process.) *For $\delta > 0$ and $n \geq 1$, let \mathcal{P}_n be the point process on \mathbb{R}_+^2 defined by*

$$\mathcal{P}_n := \left\{ \left(\frac{i}{n^{1/(\delta+1)}}, \frac{n}{i^\delta} X_i \right), i \geq 1 \right\},$$

where (X_i) is a sequence of nonnegative, independent random variables satisfying Condition 1. Then the sequence of point processes (\mathcal{P}_n) converges in distribution to a Poisson point process \mathcal{P} in \mathbb{R}_+^2 with intensity measure $x^\delta dx dy$ on \mathbb{R}_+^2 . In particular, its Laplace transform is given by

$$\mathbb{E}[e^{-\mathcal{P}(f)}] = \exp\left(-\alpha \int_{\mathbb{R}_+^2} (1 - e^{-f(x,y)})x^\delta dx dy\right), \quad f \in C_c^+(\mathbb{R}_+^2). \tag{5}$$

See [20] for the definition and the main properties of Poisson processes in general state spaces.

Proof of Proposition 1. There exists some $\eta_0 > 0$ such that $\mathbb{P}(X_i \leq x) \leq 2\alpha x$ for $0 \leq x \leq \eta_0$ and all $i \geq 1$. Let $f \in C_c^+(\mathbb{R}_+^2)$ be such that f is differentiable with respect to the second variable. There exists some $K > 0$ such that the support of f is included in $[0, K] \times [0, K]$. Define $g(x, y) = 1 - e^{-f(x,y)}$. Then, by independence of the variables X_i , $i \geq 1$,

$$\log \mathbb{E}[\exp(-\mathcal{P}_n(f))] = \sum_{i=1}^{+\infty} \log\left(1 - \mathbb{E}\left[g\left(\frac{i}{n^{1/(\delta+1)}}, \frac{n}{i^\delta} X_i\right)\right]\right).$$

Since

$$\mathbb{E}\left[g\left(\frac{i}{n^{1/(\delta+1)}}, \frac{n}{i^\delta} X_i\right)\right] \leq \mathbb{P}\left(X_i \leq K \frac{i^\delta}{n}\right) \mathbf{1}_{\{i \leq Kn^{1/(\delta+1)}\}},$$

the elementary inequality $|\log(1-y) + y| \leq 3y^2/2$ valid for $0 \leq y \leq \frac{1}{2}$ shows that there exists

some $n_0 \geq 1$ such that

$$\begin{aligned} \left| \log E[\exp(-\mathcal{P}_n(f))] + \sum_{i=1}^{+\infty} E \left[g \left(\frac{i}{n^{1/(\delta+1)}}, \frac{n}{i^\delta} X_i \right) \right] \right| &\leq \frac{6(\alpha K)^2}{n^2} \sum_{i=1}^{\lfloor Kn^{1/(\delta+1)} \rfloor} i^{2\delta} \\ &\leq 6\alpha^2 K^{2\delta+3} \frac{1}{n^{1/(\delta+1)}} \end{aligned}$$

holds for $n \geq n_0$. It is therefore enough to study the asymptotics of the series on the left-hand side of the above inequality. For $x \geq 0$, using Fubini's theorem, we obtain

$$E \left[g \left(x, \frac{n}{i^\delta} X_i \right) \right] = - \int_0^{+\infty} \frac{\partial g}{\partial y} (x, y) P \left(X_i \leq \frac{yi^\delta}{n} \right) dy.$$

Again, by using Condition 1 we find that the log of the Laplace transform of \mathcal{P}_n has the same asymptotic behavior as

$$-\alpha \frac{1}{n^{1/(\delta+1)}} \sum_{i=1}^{+\infty} \int_0^{+\infty} \frac{\partial g}{\partial y} \left(\frac{i}{n^{1/(\delta+1)}}, y \right) y \left(\frac{i}{n^{1/(\delta+1)}} \right)^\delta dy,$$

which is a Riemann sum converging to

$$-\alpha \int_{\mathbb{R}_+^2} \frac{\partial g}{\partial y} (x, y) y x^\delta dx dy = \alpha \int_{\mathbb{R}_+^2} (1 - e^{-f(x,y)}) x^\delta dx dy.$$

This shows in particular that, for any compact set H of \mathbb{R}_+^2 ,

$$\sup_{n \geq 1} E[\mathcal{P}_n(H)] < +\infty;$$

the sequence (\mathcal{P}_n) is therefore tight for the weak topology; see [5].

By the convergence result, if \mathcal{P} is any limiting point of the sequence (\mathcal{P}_n) , for any function $f \in C_c^+(\mathbb{R}_+^2)$ such that $y \rightarrow f(x, y)$ is differentiable, then the Laplace transform of \mathcal{P} at f is given by the right-hand side of (5). Since the differentiable functions in $C_c^+(\mathbb{R}_+^2)$ are dense in $C_c^+(\mathbb{R}_+^2)$ for the uniform topology, this implies that \mathcal{P} is indeed a Poisson point process with intensity measure $x^\delta dx dy$ on \mathbb{R}_+^2 . This completes the proof.

The above result can be (roughly) restated as follows: for the indices of the order of $n^{1/(\delta+1)}$, the points nX_i/i^δ lying in some finite fixed interval converge to a homogeneous Poisson point process. The following proposition gives an asymptotic description of the indices of the points nX_i/i^δ , but for indices of the order of n^κ with $1/(\delta + 1) < \kappa < 1/\delta$. It shows that, on finite intervals, these points accumulate at rate $n^{(1+\delta)\kappa-1}$ according to the Lebesgue measure with some density.

Proposition 2. (Law of large numbers.) *If, for $\kappa \in (1/(1 + \delta), 1/\delta)$ and $n \in \mathbb{N}$, \mathcal{P}_n^κ is the point process on \mathbb{R}_+^2 defined by*

$$\mathcal{P}_n^\kappa(f) = \frac{1}{n^{(1+\delta)\kappa-1}} \sum_{i=1}^{+\infty} f \left(\frac{i}{n^\kappa}, \frac{n}{i^\delta} X_i \right), \quad f \in C_c^+(\mathbb{R}_+^2),$$

where (X_i) is a sequence of nonnegative, independent random variables satisfying Condition 1, then the sequence (\mathcal{P}_n^κ) converges in distribution to the deterministic measure $\mathcal{P}_\infty^\kappa$ defined by

$$\mathcal{P}_\infty^\kappa(f) = \alpha \int_{\mathbb{R}_+^2} f(x, y)x^\delta \, dx \, dy, \quad f \in C_c^+(\mathbb{R}_+^2).$$

Proof. Let $f \in C_c^+(\mathbb{R}_+^2)$ be such that f is differentiable with respect to the second variable. As before, the convergence result is proved for such a function f ; the generalization to an arbitrary function $f \in C_c^+(\mathbb{R}_+^2)$ follows the same lines as the previous proof (relative compactness argument and identification of the limit). Let $K > 0$ be such that the support of f is included in $[0, K] \times [0, K]$. We have

$$E[\mathcal{P}_n^\kappa(f)] = -\frac{1}{n^{(1+\delta)\kappa-1}} \sum_{i=1}^{+\infty} \int_0^{+\infty} \frac{\partial f}{\partial y}\left(\frac{i}{n^\kappa}, y\right) P\left(X_i \leq \frac{yi^\delta}{n}\right) dy,$$

as in the previous proof. By using condition (4) and the fact that $\kappa < 1/\delta$, we obtain the equivalence

$$E[\mathcal{P}_n^\kappa(f)] \sim -\alpha \frac{1}{n^\kappa} \sum_{i=1}^{+\infty} \int_0^{+\infty} \frac{\partial f}{\partial y}\left(\frac{i}{n^\kappa}, y\right) y\left(\frac{i}{n^\kappa}\right)^\delta dy;$$

therefore,

$$\lim_{n \rightarrow +\infty} E[\mathcal{P}_n^\kappa(f)] = -\alpha \int_0^{+\infty} \int_0^{+\infty} \frac{\partial f}{\partial y}(x, y)yx^\delta \, dx \, dy = \alpha \int_{\mathbb{R}_+^2} f(x, y)x^\delta \, dx \, dy.$$

By independence of the X_i s, the second moment of the difference

$$\begin{aligned} & \mathcal{P}_n^\kappa(f) - E[\mathcal{P}_n^\kappa(f)] \\ &= -\frac{1}{n^{(1+\delta)\kappa-1}} \sum_{i=1}^{+\infty} \int_0^{+\infty} \frac{\partial f}{\partial y}\left(\frac{i}{n^\kappa}, y\right) \left(\mathbf{1}_{\{X_i \leq yi^\delta/n\}} - P\left(X_i \leq \frac{yi^\delta}{n}\right)\right) dy \end{aligned}$$

can be expressed as

$$\begin{aligned} & n^{2((1+\delta)\kappa-1)} E[(\mathcal{P}_n^\kappa(f) - E[\mathcal{P}_n^\kappa(f)])^2] \\ &= \sum_{i=1}^{+\infty} E\left[\left(\int_0^{+\infty} \frac{\partial f}{\partial y}\left(\frac{i}{n^\kappa}, y\right) \left(\mathbf{1}_{\{X_i \leq yi^\delta/n\}} - P\left(X_i \leq \frac{yi^\delta}{n}\right)\right) dy\right)^2\right] \\ &\leq K \sum_{i=1}^{+\infty} \int_0^{+\infty} \left(\frac{\partial f}{\partial y}\left(\frac{i}{n^\kappa}, y\right)\right)^2 E\left[\left(\mathbf{1}_{\{X_i \leq yi^\delta/n\}} - P\left(X_i \leq \frac{yi^\delta}{n}\right)\right)^2\right] dy \\ &\leq K \sum_{i=1}^{+\infty} \int_0^{+\infty} \left(\frac{\partial f}{\partial y}\left(\frac{i}{n^\kappa}, y\right)\right)^2 P\left(X_i \leq \frac{yi^\delta}{n}\right) dy, \end{aligned}$$

by Cauchy–Shwartz’s inequality. The last term is, with the same arguments as for the asymptotics of $E[\mathcal{P}_n^\kappa(f)]$, equivalent to

$$Kn^{(1+\delta)\kappa-1} \int_{\mathbb{R}_+^2} \left(\frac{\partial f}{\partial y}(x, y)\right)^2 yx^\delta \, dx \, dy.$$

In particular, the sequence $(\mathcal{P}_n^\kappa(f))$ converges in L_2 (and, therefore, in distribution) to $\mathcal{P}_\infty^\kappa(f)$. This completes the proof.

The main convergence result can now be established.

Theorem 1. *If, for $n \geq 1$, \mathcal{P}_n is the point process on \mathbb{R}_+^2 defined by*

$$\mathcal{P}_n = \left\{ \left(\frac{i}{n^{1/(\rho+2)}}, nP_i \right), i \geq 1 \right\},$$

where the random variables $P_i, i \geq 1$, are defined in (3), then the sequence (\mathcal{P}_n) converges in distribution and the relation

$$\lim_{n \rightarrow +\infty} E[\exp(-\mathcal{P}_n(f))] = E \left[\exp \left(-\frac{W_\infty^{-\rho}}{\rho} \int_{\mathbb{R}_+^2} (1 - e^{-f(x,y)}) x^{\rho+1} dx dy \right) \right] \quad (6)$$

holds for any $f \in C_c^+(\mathbb{R}_+^2)$.

In other words the point process \mathcal{P}_n converges in distribution to a mixed Poisson point process: conditionally on W_∞ , it is a Poisson process with intensity measure $W_\infty^{-\rho} x^{\rho+1} dx dy / \rho$.

Proof of Theorem 1. The proof proceeds in several steps. The main objective of these steps is to decouple the sequences (W_i) and (Z_i) defining the (P_i) , and then to apply Proposition 1.

Step 1. We define the sequences

$$P_i^1 = \frac{1}{i^{\rho+1}} \tilde{W}_\infty^\rho \tilde{Z}_i, \quad i \geq 1, \quad P_i^2 = \frac{1}{i^{\rho+1}} \tilde{W}_{\beta_n}^\rho \tilde{Z}_i, \quad i \geq 1,$$

where (β_n) is some sequence of integers converging to $+\infty$. Note that (P_i^2) depends on n . The sequences of random variables $(\tilde{W}_i, 1 \leq i \leq +\infty)$ and (\tilde{Z}_i) are assumed to be independent and to respectively have the same distribution as $(W_i, 1 \leq i \leq +\infty)$ and (Z_i) defined in (3). Recall that the sequence (\tilde{W}_i) converges almost surely to \tilde{W}_∞ . These sequences define point processes in the following way: for $j = 1$ and 2,

$$\mathcal{P}_n^j = \left\{ \left(\frac{i}{n^{1/(\rho+2)}}, nP_i^j \right), i \geq 1 \right\}.$$

If f is a nonnegative continuous function with compact support on \mathbb{R}_+^2 , because, conditionally on \tilde{W}_∞ , the variables $(\tilde{W}_\infty Z_i)$ satisfy Condition 1 with $\alpha = \tilde{W}_\infty^{-\rho} / \rho$, Proposition 1, with $\delta = \rho + 1$, shows that

$$\lim_{n \rightarrow +\infty} E[\exp(-\mathcal{P}_n^1(f)) \mid \tilde{W}_\infty] = \exp \left(-\frac{\tilde{W}_\infty^{-\rho}}{\rho} \int_{\mathbb{R}_+^2} (1 - e^{-f(x,y)}) x^{\rho+1} dx dy \right).$$

Because of the boundedness of these quantities, by Lebesgue’s theorem, the same result holds for the expected values. Therefore, the sequence (\mathcal{P}_n^1) converges in distribution to the point process \mathcal{P} on \mathbb{R}_+^2 whose Laplace transform is given by (6).

Let $K \geq 2$ be such that the support of f is a subset of $[0, K]^2$, and let $\varepsilon > 0$. Since the limiting point process \mathcal{P} is almost surely a Radon measure, there exists some $m \in \mathbb{N}$ such that $P(\mathcal{P}_n^1([0, 2K]^2) \geq m) \leq \varepsilon$ for all $n \geq 1$. By uniform continuity, there exists $0 < \eta < \frac{1}{2}$ such that $|f(u) - f(v)| \leq \varepsilon/m$ for $u, v \in \mathbb{R}_+^2$ such that $\|u - v\| \leq \eta$. For $n \geq 1$, if

$$\mathcal{A} = \left\{ \left| \frac{\tilde{W}_{\beta_n}^\rho}{\tilde{W}_\infty^\rho} - 1 \right| \geq \frac{\eta}{2K} \right\} \cup \{ \mathcal{P}_n^1([0, 2K]^2) \geq m \}$$

then

$$\begin{aligned} & |E[\exp(-\mathcal{P}_n^2(f))] - E[\exp(-\mathcal{P}_n^1(f))]| \\ & \leq P(\mathcal{A}) + E\left[\left(\exp\left(\sum_{i \geq 1} \left| f\left(\frac{i}{n^{1/(\rho+2)}}, \frac{\tilde{W}_{\beta_n}^\rho}{\tilde{W}_\infty^\rho} n P_i^1\right) - f\left(\frac{i}{n^{1/(\rho+2)}}, n P_i^1\right) \right| \right) - 1\right) \mathbf{1}_{\mathcal{A}^c}\right] \\ & \leq P\left(\left|\frac{W_{\beta_n}^\rho}{W_\infty^\rho} - 1\right| \geq \frac{\eta}{2K}\right) + (1 + e^\varepsilon)\varepsilon; \end{aligned}$$

hence, by the almost-sure convergence of (W_n) to W_∞ , the right-hand side of the last relation can be arbitrarily small as n goes to ∞ . We conclude that the sequence (\mathcal{P}_n^2) also converges in distribution to the point process \mathcal{P} .

Step 2. For $n \geq 1$, define $\beta_n = \lfloor n^{1/(\rho+2)} / \log n \rfloor$. Then it will be shown that the point processes

$$\mathcal{Q}_n = \left\{ \left(\frac{i}{n^{1/(\rho+2)}}, n P_i \right), 1 \leq i \leq \beta_n \right\}$$

converge to the measure identically null. It is sufficient to prove that, for any $f \in C_c^+(\mathbb{R}_+)$, the sequence $(\mathcal{Q}_n(f))$ converges in distribution to 0. For a fixed i , the sequence $(n P_i)$ converges in distribution to ∞ , since f is continuous with compact support and, therefore, bounded, we find that, in the definition of \mathcal{Q}_n , it can be assumed that the indices i are restricted to the set $\{\lceil \rho \rceil, \dots, \beta_n\}$.

Let K be such that the support of f is included in $[0, K]^2$. If $u_n = \log \log n$, for $i \geq \lceil \rho \rceil$,

$$E\left[f\left(\frac{i}{n^{1/(\rho+2)}}, n P_i\right) \mathbf{1}_{\{t_{\lceil \rho \rceil} \leq u_n\}} \right] \leq \|f\|_\infty P(t_{\lceil \rho \rceil} \leq u_n, n P_i \leq K).$$

Since $P_i = \exp(-\rho t_{\lceil \rho \rceil}) \exp(-\rho(t_{i-1} - t_{\lceil \rho \rceil}))(1 - \exp(-\rho E_i/i))$,

$$\begin{aligned} & E\left[f\left(\frac{i}{n^{1/(\rho+2)}}, n P_i\right) \mathbf{1}_{\{t_{\lceil \rho \rceil} \leq u_n\}} \right] \\ & \leq \|f\|_\infty P\left(\left(\frac{1 - \exp(-\rho E_i/i)}{\rho/i}\right) \leq \frac{i K \exp(\rho u_n) \exp(\rho(t_{i-1} - t_{\lceil \rho \rceil}))}{n}\right). \end{aligned}$$

By using the elementary inequality, if E_1 is exponentially distributed with mean 1,

$$P\left(\frac{1}{y}(1 - \exp(-yE_1)) \leq x\right) \leq e(1 - e^{-x}), \quad y \leq 1, x \geq 0,$$

we obtain, for $i > \rho$,

$$\begin{aligned} & E\left[f\left(\frac{i}{n^{1/(\rho+2)}}, n P_i\right) \mathbf{1}_{\{t_{\lceil \rho \rceil} \leq u_n\}} \right] \\ & \leq e \|f\|_\infty E\left[1 - \exp\left(-\frac{i}{n\rho} K \exp(\rho u_n) \exp(\rho(t_{i-1} - t_{\lceil \rho \rceil}))\right) \right] \\ & \leq eK \|f\|_\infty \frac{i \exp(\rho u_n)}{n\rho} E[\exp(\rho(t_{i-1} - t_{\lceil \rho \rceil}))] \\ & = eK \|f\|_\infty \frac{i \exp(\rho u_n)}{n\rho} \exp\left(\rho \sum_{k=\lceil \rho \rceil}^{i-1} \frac{1}{k}\right) \exp\left(\sum_{k=\lceil \rho \rceil}^{i-1} \left(-\log\left(1 - \frac{\rho}{k}\right) - \frac{\rho}{k}\right)\right). \end{aligned}$$

Thus, there exists some finite constant C such that, for $i > \rho$,

$$E\left[f\left(\frac{i}{n^{1/(\rho+2)}}, nP_i\right) \mathbf{1}_{\{t_{\lfloor \rho \rfloor} \leq u_n\}}\right] \leq C \frac{i^{\rho+1} \exp(\rho u_n)}{n} = C \frac{i^{\rho+1} (\log n)^\rho}{n};$$

consequently,

$$E[\mathcal{Q}_n(f) \mathbf{1}_{\{t_{\lfloor \rho \rfloor} \leq u_n\}}] \leq C \frac{\beta_n^{\rho+2} (\log n)^\rho}{n} \leq C \frac{1}{(\log n)^2}.$$

This relation and the inequality

$$E[1 - \exp(-\mathcal{Q}_n(f))] \leq P(t_{\lfloor \rho \rfloor} > u_n) + E[\mathcal{Q}_n(f) \mathbf{1}_{\{t_{\lfloor \rho \rfloor} \leq u_n\}}]$$

give the desired result.

Step 3. The proof of the theorem can now be completed. By (3), for $i \geq 1$, $P_i = W_i^\rho Z_i / i^{\rho+1}$, by using step 2 and the same techniques as in step 1 together with the fact that, for $\eta > 0$, the probability of the event

$$\left\{ \sup\left(\left|\frac{W_i^\rho}{W_{\beta_n}^\rho} - 1\right| : i \geq \beta_n\right) \geq \eta \right\}$$

converges to 0 as n gets large. It is not difficult to show that the sequences of point processes

$$\left\{ \left(\frac{i}{n^{1/(\rho+2)}}, \frac{n}{i^{\rho+1}} W_i^\rho Z_i\right), i \geq 1 \right\} \quad \text{and} \quad \left\{ \left(\frac{i}{n^{1/(\rho+2)}}, \frac{n}{i^{\rho+1}} W_{\beta_n}^\rho Z_i\right), i \geq \beta_n \right\}$$

have the same limit in distribution. Because W_{β_n} is independent of $(Z_i, i \geq \beta_n)$, the last point process has the same distribution as \mathcal{P}_n^2 (up to the first β_n terms which are negligible, similarly as in step 2). By step 1, the convergence in distribution is therefore proved.

The following proposition strengthens the statement of Proposition 1 and it will be used to prove the main asymptotic result on the indices of empty bins.

Proposition 3. *If $f : \mathbb{R}_+^2 \rightarrow \mathbb{R}_+$ is a continuous function such that*

- (a) *there exists K such that $f(x, y) = 0$ for any $x \geq K$ and $y \in \mathbb{R}_+$,*
- (b) *for all $x \in \mathbb{R}_+$, the function $y \rightarrow f(x, y)$ is differentiable and*

$$y \rightarrow y \left\| \frac{\partial f}{\partial y} \right\|_y := y \sup_{x \in \mathbb{R}_+} \left| \frac{\partial f}{\partial y} \right|(x, y)$$

is integrable on \mathbb{R}_+ ,

then convergence (6) also holds for f .

Proof. For $M, L \geq 0$ and $i, n \in \mathbb{N}$, we have

$$\begin{aligned} & E\left[f\left(\frac{i}{n^{\rho+2}}, nP_i\right) \mathbf{1}_{\{nP_i \geq M, t_{\lfloor \rho \rfloor} \leq L\}}\right] \\ &= - \int_0^{+\infty} \frac{\partial f}{\partial y}\left(\frac{i}{n^{\rho+2}}, y\right) P(M \leq nP_i \leq y, t_{\lfloor \rho \rfloor} \leq L) dy. \end{aligned}$$

By using similar arguments as in the end of the proof of the above theorem, we obtain

$$\begin{aligned} & \mathbb{E} \left[f \left(\frac{i}{n^{\rho+2}}, nP_i \right) \mathbf{1}_{\{nP_i \geq M, t_{[\rho]} \leq L\}} \right] \\ & \leq e \int_M^{+\infty} \left\| \frac{\partial f}{\partial y} \right\|_y \mathbb{E} \left[1 - \exp \left(-\frac{i}{n\rho} y e^{\rho L} \exp(\rho(t_{i-1} - t_{[\rho]})) \right) \right] dy \\ & \leq \frac{i e^{\rho L}}{n\rho} e \mathbb{E}[\exp(\rho(t_{i-1} - t_{[\rho]}))] \int_M^{+\infty} y \left\| \frac{\partial f}{\partial y} \right\|_y dy \\ & \leq C \frac{i^{\rho+1} e^{\rho L}}{n} \int_M^{+\infty} y \left\| \frac{\partial f}{\partial y} \right\|_y dy \end{aligned}$$

for some fixed constant C . Define $k_n = \lfloor K n^{1/(\rho+2)} \rfloor$. By summing up these terms, this gives the relation

$$\begin{aligned} & \mathbb{E} \left[\sum_{i \geq 1} f \left(\frac{i}{n^{\rho+2}}, nP_i \right) \mathbf{1}_{\{M \leq nP_i, t_{[\rho]} \leq L\}} \right] \\ & \leq C \frac{k_n^{\rho+2} e^{\rho L}}{n} \int_M^{+\infty} y \left\| \frac{\partial f}{\partial y} \right\|_y dy \tag{7} \\ & \leq CK^{\rho+2} e^{\rho L} \int_M^{+\infty} y \left\| \frac{\partial f}{\partial y} \right\|_y dy. \end{aligned}$$

Define $f_0(x, y) = f(x, y) \mathbf{1}_{\{y \leq M\}}$. By using a convolution kernel on the variable y , there exist sequences (g_p^+) and (g_p^-) in $C_c^+(\mathbb{R}_+)$ converging pointwise to f_0 for all $y \neq M$ such that $g_p^- \leq f_0 \leq g_p^+$. See [21] for example. Proposition 1 gives

$$\begin{aligned} \mathbb{E}[\exp(-\mathcal{P}(g_p^+))] & \leq \liminf_{n \rightarrow +\infty} \mathbb{E}[\exp(-\mathcal{P}_n(f_0))] \\ & \leq \limsup_{n \rightarrow +\infty} \mathbb{E}[\exp(-\mathcal{P}_n(f_0))] \\ & \leq \mathbb{E}[\exp(-\mathcal{P}(g_p^-))], \end{aligned}$$

and expression (5) shows that, as p goes to ∞ , the left- and right-hand side terms of this relation converge to the Laplace transform of \mathcal{P} at f_0 . Therefore, convergence (6) holds at f_0 . Since

$$\begin{aligned} 0 & \leq \mathbb{E}[\exp(-\mathcal{P}_n(f)) - \mathbb{E}[\exp(-\mathcal{P}_n(f_0))] \\ & \leq \mathbb{P}(t_{[\rho]} \geq L) + \mathbb{E}[(1 - \exp(-(\mathcal{P}_n(f) - \mathcal{P}_n(f_0)))) \mathbf{1}_{\{t_{[\rho]} \leq L\}}] \\ & \leq \mathbb{P}(t_{[\rho]} \geq L) + \mathbb{E}[(\mathcal{P}_n(f) - \mathcal{P}_n(f_0)) \mathbf{1}_{\{t_{[\rho]} \leq L\}}], \end{aligned}$$

where the last term is the left-hand side of relation (7), we can choose L and M sufficiently large so that this difference is arbitrarily small. This completes the proof.

4. Asymptotic behavior of the indices of the first empty bins

It is assumed that a large number of balls, n , are thrown in the bins according to the probability distribution (P_i) defined in (3). The purpose of this section is to establish limit theorems to describe the limiting distribution of the set of indices of bins having a fixed number of balls.

Theorem 2. *The point process of rescaled indices of empty bins associated to the probability vector (P_i) when n balls have been used,*

$$\mathcal{N}_n = \left\{ \frac{i}{n^{1/(\rho+2)}} : i \geq 1, \text{ the } i\text{th bin is empty} \right\},$$

converges in distribution as n goes to ∞ to a point process \mathcal{N}_∞ whose distribution is given by

$$E[\exp(-\mathcal{N}_\infty(g))] = E \left[\exp \left(-\frac{W_\infty^{-\rho}}{\rho} \int_{\mathbb{R}_+} (1 - e^{-g(x)}) x^{\rho+1} dx \right) \right] \tag{8}$$

for $g \in C_c^+(\mathbb{R}_+)$. Equivalently, (\mathcal{N}_n) converges in distribution to the point process

$$(W_\infty^{\rho/(\rho+2)} t_i^{1/(\rho+2)}),$$

where (t_i) is a standard Poisson process with parameter $(\rho(\rho + 2))^{-1/(\rho+2)}$.

It can also be shown that the same result holds when the indices of bins containing k balls are considered. If $\mathcal{N}_{k,n}$ is the corresponding point process, the limiting point process does not in fact depend on k and, moreover, the sequence $(\mathcal{N}_{k,n}, k \geq 0)$ converges in distribution to $(\mathcal{N}_{k,\infty}, k \geq 0)$ and, conditionally on W_∞ , the random variables $\mathcal{N}_{k,\infty}, k \geq 0$, are independent with the same distribution.

Proof of Theorem 2. (Poissonization.) A closely related model is first analyzed when U_n balls are used, U_n being an independent Poisson random variable with mean n . Let \mathcal{N}_n^0 denote the corresponding point process. For this model, conditionally on the sequence (P_i) , the number of balls in the bins are independent Poisson random variables with respective parameters (nP_i) . In a first step, the convergence in distribution of the sequence (\mathcal{N}_n^0) of point processes is investigated. Let $g \in C_c^+(\mathbb{R}_+)$. We have

$$E[\exp(-\mathcal{N}_n^0(g))] = E \left[\exp \left(\sum_{i=1}^{+\infty} \log \left(1 - \exp(-nP_i) \left(1 - \exp \left(-g \left(\frac{i}{n^{1/(\rho+2)}} \right) \right) \right) \right) \right) \right].$$

If we define $f(x, y) = -\log(1 - e^{-y}(1 - e^{-g(x)}))$ then

$$E[\exp(-\mathcal{N}_n^0(g))] = E[\exp(-\mathcal{P}_n(f))],$$

where \mathcal{P}_n is the point process defined in Theorem 1. By using Proposition 3, we obtain the relation

$$\begin{aligned} \lim_{n \rightarrow +\infty} E[\exp(-\mathcal{N}_n^0(g))] &= E \left[\exp \left(-\frac{W_\infty^{-\rho}}{\rho} \int_{\mathbb{R}_+^2} (1 - e^{-f(x,y)}) x^{\rho+1} dx dy \right) \right] \\ &= E \left[\exp \left(-\frac{W_\infty^{-\rho}}{\rho} \int_{\mathbb{R}_+} (1 - e^{-g(x)}) x^{\rho+1} dx \right) \right]. \end{aligned}$$

Consequently, the sequence (\mathcal{N}_n^0) converges in distribution to \mathcal{N}_∞ . For $0 < \alpha < 1$, it is not difficult to check that the same convergence result holds when U_{n+n^α} balls are used. If \mathcal{N}_n^1 denotes the associated point process then (\mathcal{N}_n^1) also converges in distribution to \mathcal{N}_∞ . For $x > 0$, the monotonicity property, $\mathcal{N}_a([0, x]) \leq \mathcal{N}_b([0, x])$ for $b \leq a$, gives the relation

$$P(\mathcal{N}_n^1([0, x]) \leq k) \leq P(\mathcal{N}_n([0, x]) \leq k) + P(U_{n+n^\alpha} \leq n).$$

The central limit theorem for Poisson processes shows that the quantity $P(U_{n+n^\alpha} \leq n)$ for $\alpha \in (\frac{1}{2}, 1)$ converges to 0 as n gets large; therefore, if $k \geq 0$,

$$P(\mathcal{N}_\infty([0, x]) \leq k) = \lim_{n \rightarrow +\infty} P(\mathcal{N}_n^1([0, x]) \leq k) \geq \liminf_{n \rightarrow +\infty} P(\mathcal{N}_n([0, x]) \leq k).$$

By using a similar argument with the lim sup, we find that the sequence $(\mathcal{N}_n([0, x]))$ converges in distribution to $\mathcal{N}_\infty([0, x])$. With the same coupling argument, we obtain, for $x_1 \leq x_2 \leq \dots \leq x_p \in \mathbb{R}_+$ and $(k_i, 1 \leq i \leq p) \in \mathbb{N}^p$,

$$\begin{aligned} \lim_{n \rightarrow +\infty} P(\mathcal{N}_n([0, x_1]) \leq k_1, \mathcal{N}_n([0, x_2]) \leq k_2, \dots, \mathcal{N}_n([0, x_p]) \leq k_p) \\ = P(\mathcal{N}_\infty([0, x_1]) \leq k_1, \mathcal{N}_\infty([0, x_2]) \leq k_2, \dots, \mathcal{N}_\infty([0, x_p]) \leq k_p), \end{aligned}$$

and, therefore, the convergence in distribution of (\mathcal{N}_n) to \mathcal{N}_∞ . This completes the proof.

Corollary 1. *If v_n is the index of the first empty bin when n balls are thrown then*

$$\lim_{n \rightarrow +\infty} P\left(\frac{v_n}{n^{1/(\rho+2)}} \geq x\right) = \mathbb{E}\left[\exp\left(-\frac{x^{\rho+2} W_\infty^{-\rho}}{\rho(\rho+2)}\right)\right], \quad x \geq 0.$$

4.1. Comparison with the deterministic power law decay

For $\delta > 1$, we consider the bins and balls problem with the probability vector $Q = (Q_i, i \geq 1) = \alpha/i^\delta$. Note that, for the problems analyzed in this paper, only the asymptotic behavior of the sequence (Q_i) matters. The equivalent of Theorem 2 can be obtained directly from Theorem 1 of [23].

Proposition 4. *As n goes to ∞ , the point process*

$$\left\{ \frac{i(\log n)^{1/\delta+1}}{(\alpha \delta n)^{1/\delta}} - \log n - \frac{1 + \delta}{\delta} \log \log n : \text{the } i\text{th bin is empty} \right\}$$

converges in distribution to a Poisson point process with the intensity measure $(\alpha \delta)^{1/\delta} e^x dx$ on \mathbb{R} .

The probability vector considered in the above theorem has an asymptotic expression of the form $(P_i) = (W_\infty^\rho F_i / i^{\rho+1})$. In this case, empty bins show up for indices of the order of $n^{1/(\rho+2)}$, i.e. much earlier than for the deterministic case where the exponent of n is $1/\delta = 1/(\rho + 1)$ (if we ignore the log). This can be explained simply by the fact that some of the i.i.d. exponential random variables (F_i) can be very small, thereby creating an additional possibility of having empty bins.

In this picture, the variable W_∞ does not seem to have an influence on the qualitative behavior of these occupancy schemes other than creating some dependency structure for the vector (P_i) . In the next section we show that this variable has nevertheless an important role if we look at the averages of the number of empty bins.

5. Rare events

From (6), for $x > 0$, the limiting number (in distribution) of empty bins whose index is less than $x n^{1/(\rho+2)}$ has an average value given by

$$\frac{x^{\rho+2}}{\rho(\rho+2)} \mathbb{E}[W_\infty^{-\rho}] = \frac{x^{\rho+2}}{\rho(\rho+2)} \int_0^{+\infty} \frac{1}{u^\rho} e^{-u} du,$$

by (2) and since $W_\infty = \exp(-M_\infty)$. This quantity is infinite when $\rho \geq 1$. The goal of this section is to investigate this phenomenon, which has a significant impact on the system at the origin of this model (a peer-to-peer architecture). For this purpose, a family of rescaled point processes is introduced.

Definition 1. If $\phi: \mathbb{N} \rightarrow \mathbb{R}_+$ is a nondecreasing function, for $n \geq 1$, \mathcal{N}_n^ϕ denotes the point process defined by

$$\mathcal{N}_n^\phi = \left\{ \frac{i}{\phi(n)} : i \geq 1, \text{ the } i\text{th bin is empty} \right\}.$$

For $i \geq 1$, recall that, by (3), the probability P_i of throwing a ball into the i th urn is, using the notation of Section 2,

$$P_i = \frac{1}{i^{\rho+1}} W_i^\rho Z_i \quad \text{with} \quad Z_i = i \left(1 - \exp\left(-\frac{\rho E_i}{i}\right) \right) \text{ and } W_i = \exp(-M_{i-1}),$$

where the (E_i) are i.i.d. exponential random variables with parameter 1. We define its asymptotic representation as

$$\tilde{P}_i := \frac{1}{i^{\rho+1}} W_\infty^\rho F_i \quad \text{with} \quad W_\infty = \exp(-M_\infty),$$

where W_i and W_∞ are independent of (F_i) , an i.i.d. sequence of exponential random variables with mean ρ .

The following proposition shows that, as in the proof of Theorem 2, for some asymptotics for (\mathcal{N}_n^ϕ) , it is enough to analyze the asymptotic behavior of a functional of the sequence (\tilde{P}_i) .

Proposition 5. For $x > 0$ and a nondecreasing function ϕ ,

$$\lim_{n \rightarrow +\infty} \mathbb{E}[\mathcal{N}_n^\phi([0, x])] \bigg/ \sum_{i=1}^{\lfloor x\phi(n) \rfloor} \mathbb{E}[\exp(-n\tilde{P}_i)] = 1,$$

provided that

$$\lim_{n \rightarrow +\infty} \frac{\phi(n)}{n^{\alpha_0}} = 0 \quad \text{and} \quad \lim_{n \rightarrow +\infty} \sum_{i=1}^{\lfloor x\phi(n) \rfloor} \mathbb{E}[\exp(-n\tilde{P}_i)] = +\infty$$

for some $\alpha_0 < 1/(\rho + 1)$.

Proof. See Appendix A.

For $i > \lfloor \rho \rfloor$, the quantity P_i will be written as $P_i = \exp(-\rho t_{\lfloor \rho \rfloor}) D_i Z_i / i^{\rho+1}$ with

$$D_i := \exp(-\rho[M_{i-1} - M_{\lfloor \rho \rfloor} - \log \lfloor \rho + 1 \rfloor]).$$

The sequence (D_i) converges almost surely to a finite limit D_ρ given by

$$D_\rho := \exp(-\rho[M_\infty - M_{\lfloor \rho \rfloor} - \log \lfloor \rho + 1 \rfloor]), \tag{9}$$

and, since $\exp(\rho E_i / i)$ is integrable for $i > \rho$, a similar result holds for the expected values

$$\lim_{i \rightarrow +\infty} \mathbb{E} \left[\frac{1}{D_i} \right] = \mathbb{E} \left[\frac{1}{D_\rho} \right] < +\infty.$$

With this definition, the sequence (\tilde{P}_i) can be represented as

$$(\tilde{P}_i) = \left(\frac{\exp(-\rho t_{\lfloor \rho \rfloor})}{i^{\rho+1}} D_\rho F_i \right).$$

5.1. Some preliminary estimations

For the moment, $k \in \mathbb{N}$ is fixed. If $n \geq 1$ and $i > \rho$ then

$$\begin{aligned} E[\exp(-n\tilde{P}_i)] &= E \left[\exp \left(- \frac{n\rho D_\rho \exp(-\rho t_{\lfloor \rho \rfloor}) F_1}{i^{\rho+1}} \right) \right] \\ &= E \left[\frac{i^{\rho+1}/n}{i^{\rho+1}/n + \exp(-\rho t_{\lfloor \rho \rfloor}) \rho D_\rho} \right]. \end{aligned}$$

By summing up these terms, if $\varepsilon_{k,n} := k/n^{1/(\rho+1)}$, we obtain

$$\sum_{i=\lfloor \rho \rfloor + 1}^k E[\exp(-n\tilde{P}_i)] = n^{1/(\rho+1)} \int_0^{\varepsilon_{k,n}} E \left[\frac{v^{\rho+1}}{v^{\rho+1} + \exp(-\rho t_{\lfloor \rho \rfloor}) \rho D_\rho} \right] dv + O(\varepsilon_{k,n}),$$

which gives the relation

$$\sum_{i=1}^k E[\exp(-n\tilde{P}_i)] = n^{1/(\rho+1)} \varepsilon_{k,n}^{\rho+2} \int_0^1 E \left[\frac{v^{\rho+1}}{\varepsilon_{k,n}^{\rho+1} v^{\rho+1} + \exp(-\rho t_{\lfloor \rho \rfloor}) \rho D_\rho} \right] dv + O(\varepsilon_{k,n}),$$

with a change of variable. Using (1) and a change of variable again, we obtain the relation

$$\begin{aligned} \sum_{i=1}^k E[\exp(-n\tilde{P}_i)] &= \frac{\lfloor \rho \rfloor}{\rho} n^{1/(\rho+1)} \varepsilon_{k,n}^{(2\rho+1)/\rho} \int_0^{1/\varepsilon_{k,n}} u^{1/\rho-1} (1 - \varepsilon_{k,n}^{(\rho+1)/\rho} u^{1/\rho})^{\lfloor \rho \rfloor - 1} du \\ &\quad \times \int_0^1 E \left[\frac{v^{\rho+1}}{v^{\rho+1} + u\rho D_\rho} \right] dv + O(\varepsilon_{k,n}). \end{aligned} \tag{10}$$

This quantity is now analyzed according to the values of ρ .

Case 1: $\rho > 1$. If $k_n = \lfloor xn^\alpha \rfloor$ with $1/(2\rho + 1) \leq \alpha < 1/(\rho + 1)$ then

$$\varepsilon_{k_n,n} \sim xn^{(\alpha(\rho+1)-1)/(\rho+1)}$$

and, by relation (10),

$$\begin{aligned} \lim_{n \rightarrow +\infty} \frac{1}{n^{((2\rho+1)\alpha-1)/\rho}} \sum_{i=1}^{k_n} E[\exp(-n\tilde{P}_i)] \\ = x^{(2\rho+1)/\rho} \frac{\lfloor \rho \rfloor}{\rho} \int_0^{+\infty} u^{1/\rho-1} du \int_0^1 E \left[\frac{v^{\rho+1}}{v^{\rho+1} + u\rho D_\rho} \right] dv. \end{aligned}$$

Case 2: $\rho = 1$. Equation (10) is for this case

$$\sum_{i=1}^k E[\exp(-n\tilde{P}_i)] = \sqrt{n} \varepsilon_{k,n}^3 \int_0^{1/\varepsilon_{k,n}^2} du \int_0^1 E \left[\frac{v^2}{v^2 + uD_1} \right] dv + O(\varepsilon_{k,n}).$$

If $k_n = \lfloor xn^{1/3} / \log^\beta n \rfloor$ with $\beta \in \mathbb{R}$ then $\varepsilon_{k_n, n} \sim x / (n^{1/6} (\log n)^\beta)$ and, for $\beta \leq \frac{1}{3}$,

$$\lim_{n \rightarrow +\infty} \frac{1}{(\log n)^{(1-3\beta)}} \sum_{i=1}^{k_n} \mathbb{E}[\exp(-n\tilde{P}_i)] = \frac{1}{9} x^3 \mathbb{E}\left[\frac{1}{D_1}\right].$$

The following proposition has therefore been proved.

Proposition 6. (Average of the number of empty bins.) For $\alpha, \beta > 0$ and $n \in \mathbb{N}$, let $p_{\alpha, \beta}(n) = n^\alpha (\log n)^{-\beta}$ and, by convention, let $p_\alpha = p_{\alpha, 0}$.

(a) If $\rho > 1$ and $1/(2\rho + 1) \leq \alpha < 1/(\rho + 1)$,

$$\begin{aligned} & \lim_{n \rightarrow +\infty} \frac{1}{n^{(\alpha(2\rho+1)-1)/\rho}} \mathbb{E}[\mathcal{N}_n^{p_\alpha}([0, x])] \\ &= x^{(2\rho+1)/\rho} \frac{\lfloor \rho \rfloor}{\rho} \int_0^{+\infty} u^{1/\rho-1} du \int_0^1 \mathbb{E}\left[\frac{v^{\rho+1}}{v^{\rho+1} + u\rho D_\rho}\right] dv. \end{aligned}$$

(b) If $\rho = 1$ and $\beta \leq \frac{1}{3}$,

$$\lim_{n \rightarrow +\infty} \frac{1}{(\log n)^{(1-3\beta)}} \mathbb{E}[\mathcal{N}^{p_{1/3, \beta}}([0, x])] = \frac{1}{9} x^3 \mathbb{E}\left[\frac{1}{D_1}\right].$$

5.2. A double threshold

For the convergence in distribution of the sequence of point processes (\mathcal{N}_n^ϕ) , Theorem 2 has shown that the correct scaling ϕ for the order of magnitude of the indices of the first empty bins is given by $\phi(n) = n^{1/(\rho+2)}$, $n \geq 1$. For the average number of points in a finite interval, the above proposition states that, for $\rho > 1$, the correct scaling is in fact $\phi(n) = n^{1/(2\rho+1)} \ll n^{1/(\rho+2)}$.

For $\alpha > 0$, using the notation in the above proposition, we conclude that, under the condition $\rho > 1$ and for $1/(2\rho + 1) < \alpha < 1/(\rho + 2)$, the following limit results hold:

$$\mathcal{N}_n^{p_\alpha} \xrightarrow{D} 0 \quad \text{and} \quad \lim_{n \rightarrow +\infty} \mathbb{E}[\mathcal{N}_n^{p_\alpha}[0, x]] = +\infty \quad \text{for all } x > 0,$$

where ‘ \xrightarrow{D} ’ denotes convergence in distribution. This suggests that in this case, with a high probability, all the bins with index less than $n^{1/(\rho+2)}$ have a large number of balls. But also that there exists some rare event for which a very large number of empty bins with indices of an order slightly greater than $n^{1/(2\rho+1)}$ are created. The following proposition shows that the total size of the first $\lfloor \rho \rfloor$ bins is the key variable to explain this phenomenon. It should be of the order of $\log n$ in order to have sufficiently many empty bins in the appropriate region.

Proposition 7. For $\rho > 1$, and if $p_\alpha(n) = n^\alpha$, $\alpha \in [1/(2\rho + 1), 1/(\rho + 2))$,

$$\delta_0(\alpha) := \frac{1 - \alpha(\rho + 2)}{\rho - 1}, \quad \text{and} \quad \delta_1(\alpha) := \frac{1 - \alpha(\rho + 1)}{\rho},$$

then, for $a \in \mathbb{R}$ and $x > 0$,

(a) if $\delta < \delta_0(\alpha)$ then

$$\lim_{n \rightarrow +\infty} \mathbb{E}[\mathcal{N}_n^{p_\alpha}([0, x]) \mathbf{1}_{\{t_{\lfloor \rho \rfloor} \leq \delta \log n\}}] = 0,$$

(b) if $\delta \in [\delta_0(\alpha), \delta_1(\alpha)]$ then

$$\lim_{n \rightarrow +\infty} \frac{E[\mathcal{N}_n^{P\alpha}([0, x]) \mathbf{1}_{\{t_{\lfloor \rho \rfloor} \leq \delta \log n + a\}}]}{n^{(\rho+2)\alpha + \delta(\rho-1) - 1}} = \frac{x^{\rho+2}}{\rho+2} \frac{\lfloor \rho \rfloor}{\rho-1} E\left[\frac{1}{\rho D_\rho}\right] e^{(\rho-1)a},$$

(c) if $\delta \geq \delta_1(\alpha)$,

$$\begin{aligned} & \lim_{n \rightarrow +\infty} \frac{E[\mathcal{N}_n^{P\alpha}([0, x]) \mathbf{1}_{\{t_{\lfloor \rho \rfloor} \leq \delta \log n + a\}}]}{n^{(2\rho+1)\alpha - 1/\rho}} \\ &= x^{(2\rho+1)/\rho} \frac{\lfloor \rho \rfloor}{\rho} \int_{e^{-\rho a} \mathbf{1}_{\{\delta = \delta_1(\alpha)\}}}^{+\infty} u^{1/\rho-1} du \int_0^1 E\left[\frac{v^{\rho+1}}{v^{\rho+1} + u\rho D_\rho}\right] dv, \end{aligned}$$

where D_ρ is the random variable defined in (9).

Proof. To begin with, it is assumed that $\delta \in [\delta_0(\alpha), \delta_1(\alpha)]$. If $k \geq 1$, $\varepsilon_{k,n} = k/n^{1/(\rho+1)}$, $b > 0$, $k = \lfloor xn^\alpha \rfloor$, and $b = \delta \log n + a$, in the same way as for (10), we obtain

$$\begin{aligned} & \sum_{i=1}^k E[\exp(-n\tilde{P}_i) \mathbf{1}_{\{t_{\lfloor \rho \rfloor} \leq b\}}] \\ &= \frac{\lfloor \rho \rfloor}{\rho} n^{1/(\rho+1)} \varepsilon_{k,n}^{(2\rho+1)/\rho} \int_{e^{-\rho b}/\varepsilon_{k,n}^{\rho+1}}^{1/\varepsilon_{k,n}^{\rho+1}} u^{1/\rho-1} (1 - \varepsilon_{k,n}^{(\rho+1)/\rho} u^{1/\rho})^{\lfloor \rho \rfloor - 1} du \\ & \quad \times \int_0^1 E\left[\frac{v^{\rho+1}}{v^{\rho+1} + u\rho D_\rho}\right] dv + O(\varepsilon_{k,n}) \\ &= \frac{\lfloor \rho \rfloor}{\rho} n^{1/(\rho+1)} \varepsilon_{k,n}^{(2\rho+1)/\rho} \int_{e^{-\rho b}/\varepsilon_{k,n}^{\rho+1}}^{1/\varepsilon_{k,n}^{\rho+1}} u^{1/\rho-2} du \int_0^1 E\left[\frac{uv^{\rho+1}}{v^{\rho+1} + u\rho D_\rho}\right] dv + O(\varepsilon_{k,n}). \end{aligned} \tag{11}$$

Note that

$$\frac{e^{-\rho b}}{\varepsilon_{k,n}^{\rho+1}} \sim n^{1-\rho\delta-\alpha(\rho+1)} e^{-\rho a} \nearrow +\infty;$$

hence, the range of the first integral goes to ∞ as n gets large. Since

$$\int_0^1 E\left[\frac{uv^{\rho+1}}{v^{\rho+1} + u\rho D_\rho} - \frac{v^{\rho+1}}{\rho D_\rho}\right] dv = \int_0^1 E\left[\frac{v^{2(\rho+1)}}{\rho(v^{\rho+1} + u\rho D_\rho)D_\rho}\right] dv,$$

by Lebesgue’s theorem, this integral is arbitrarily small as u gets large, which implies the equivalence

$$\sum_{i=1}^k E[\exp(-n\tilde{P}_i) \mathbf{1}_{\{t_{\lfloor \rho \rfloor} \leq b\}}] \sim \frac{\lfloor \rho \rfloor}{\rho(\rho+2)} E\left[\frac{1}{\rho D_\rho}\right] n^{1/(\rho+1)} \varepsilon_{k,n}^{(2\rho+1)/\rho} \int_{e^{-\rho b}/\varepsilon_{k,n}^{\rho+1}}^{1/\varepsilon_{k,n}^{\rho+1}} u^{1/\rho-2} du.$$

If C is the multiplicative constant on the right-hand side of the above relation then

$$\sum_{i=1}^k E[\exp(-n\tilde{P}_i) \mathbf{1}_{\{t_{\lfloor \rho \rfloor} \leq b\}}] \sim \frac{C\rho}{\rho-1} \frac{k^{\rho+2}}{n} (e^{b(\rho-1)} - 1),$$

which gives the equivalence

$$\sum_{i=1}^k \mathbb{E}[\exp(-n\tilde{P}_i) \mathbf{1}_{\{t_{[\rho]} \leq b\}}] \sim x^{\rho+2} \frac{C\rho}{\rho-1} e^{a(\rho-1)} n^{(\rho+2)\alpha + \delta(\rho-1) - 1}.$$

The proof of this case is completed.

The case in which $\delta \geq \delta_1(\alpha)$ uses (11). The term $e^{-\rho b} / \varepsilon_{k,n}^{\rho+1}$ converges to $e^{-\rho a}$ if $\delta = \delta_1(\alpha)$ and 0 otherwise. This directly gives the desired convergence.

Finally, if $\delta < \delta_0(\alpha)$ for any $a \in \mathbb{R}$, there exists n_0 so that if $n \geq n_0$ then $\delta \log n \leq \delta_0(\alpha) \log n + a$. In particular,

$$\mathbb{E}[\mathcal{N}_n^{P\alpha}([0, x]) \mathbf{1}_{\{t_{[\rho]} \leq \delta \log n\}}] \leq \mathbb{E}[\mathcal{N}_n^{P\alpha}([0, x]) \mathbf{1}_{\{t_{[\rho]} \leq \delta_0(\alpha) \log n + a\}}];$$

hence,

$$\limsup_{n \rightarrow +\infty} \mathbb{E}[\mathcal{N}_n^{P\alpha}([0, x]) \mathbf{1}_{\{t_{[\rho]} \leq \delta \log n\}}] \leq \frac{x^{\rho+2}}{\rho+2} \frac{[\rho]}{\rho-1} \mathbb{E}\left[\frac{1}{\rho D_\rho}\right] e^{(\rho-1)a}.$$

The proof is completed by letting a go to $-\infty$.

As a consequence of the above proposition, for $\alpha \in [1/(2\rho + 1), 1/(\rho + 2))$, the average of the variable $\mathcal{N}_n^{P\alpha}([0, x])$ converges to ∞ only when the total size $t_{[\rho]}$ of the first $[\rho]$ bins is of the order $\delta \log n$ for a sufficiently large δ . The following corollary gives a more precise formulation.

Corollary 2. For $\rho > 1$, and if $p_\alpha(n) = n^\alpha$, $\alpha \in [1/(2\rho + 1), 1/(\rho + 2))$, and

$$\delta_1(\alpha) = \frac{1 - \alpha(\rho + 1)}{\rho},$$

then, for $a, b > 0$,

$$\lim_{n \rightarrow +\infty} \frac{\mathbb{E}[\mathcal{N}_n^{P\alpha}([0, x]) \mathbf{1}_{\{\delta_1(\alpha) \log n - a \leq t_{[\rho]} \leq \delta_1(\alpha) \log n + b\}}]}{\mathbb{E}[\mathcal{N}_n^{P\alpha}([0, x])]} = \psi(-a, b),$$

where, for $y, z \in \mathbb{R}$, $\psi(y, z) = \phi(y, z) / \phi(-\infty, +\infty)$ and

$$\phi(y, z) = x^{(2\rho+1)/\rho} \frac{[\rho]}{\rho} \int_{[e^{-\rho z}, e^{-\rho y}]} u^{1/\rho-1} du \int_0^1 \mathbb{E}\left[\frac{v^{\rho+1}}{v^{\rho+1} + u\rho D_\rho}\right] dv.$$

A rough (nonrigorous) interpretation of this result could be as follows. On the event where ‘most’ (i.e. for the averages) of the empty bins are created in the interval $[0, xn^\alpha]$, the random variable $t_{[\rho]} - \delta_1(\alpha) \log n$ converges in distribution to some random variable X on \mathbb{R} , such that $P(X \leq a) = \psi(-\infty, a)$.

The following analogous result is proved in a similar way for the critical case, $\rho = 1$.

Proposition 8. For $\rho = 1$ and using the notation of Proposition 7, then, for $0 < \beta < \frac{1}{3}$, $x > 0$, and $0 \leq a \leq \frac{1}{3}$,

$$\lim_{n \rightarrow +\infty} \frac{1}{(\log n)^{1-3\beta}} \mathbb{E}[\mathcal{N}_n^{P_{1/3,\beta}}([0, x]) \mathbf{1}_{\{t_{[\rho]} \leq a \log n\}}] = \frac{a}{3} x^3 \mathbb{E}\left[\frac{1}{D_1}\right],$$

and, for $a > \frac{1}{3}$,

$$\lim_{n \rightarrow +\infty} \frac{1}{(\log n)^{1-3\beta}} \mathbb{E}[\mathcal{N}_n^{P_{1/3,\beta}}([0, x]) \mathbf{1}_{\{t_{[\rho]} \leq a \log n\}}] = \frac{1}{9} x^3 \mathbb{E}\left[\frac{1}{D_1}\right],$$

where D_1 is the random variable defined in (9).

6. Generalizations

The problem analyzed in the present paper can be generalized to two directions. On the one hand, the sequence (t_n) can stem from a general branching process instead of the particular Yule process; on the other hand, the locations of the balls can have a general distribution. In this section we discuss these possible extensions.

6.1. Exponential balls and the general branching process

Let (t_n) be the birth instants of a general supercritical branching process $(Z(t))$. See [14] and [17] for example. Let α be the Malthusian parameter, and let W be the almost-sure limit of $(e^{-\alpha t} Z(t))$. Under reasonable technical assumptions, Harnqvist [10] showed the following result.

Theorem 3. *Define the point process Ψ_t^* by*

$$\Psi_t = \sum_{k \geq 1} \mathbf{1}_{\{t \leq t_k\}} \delta_{t_k} e^{\alpha t}.$$

As t becomes large, Ψ_t converges in distribution to a mixed Poisson process whose parameter is distributed as γW for some constant $\gamma > 0$.

From this result, it is possible to prove that the process $(n(t_{n+k} - t_n), k \geq 1)$ converges in distribution, as n goes to ∞ , to a Poisson process: clearly,

$$\Psi_{t_n} = \sum_{k \geq 1} \delta_{(t_{n+k} - t_n)} \exp(\alpha t_n),$$

and provided that, up to a multiplicative constant, $\exp(\alpha t_k)/k$ converges to W , the point process $\sum_{k \geq 1} \delta_{n(t_{n+k} - t_n)}$ should converge to a Poisson random variable with a deterministic parameter. In this case the probability that a ball falls into the n th bin, which is given by

$$P_n = \exp(-\rho t_{n-1})(1 - \exp(-\rho(t_n - t_{n-1}))),$$

therefore has the following asymptotic behavior:

$$P_n \sim n^{-\rho/\alpha - 1} W^{\rho/\alpha} F_n,$$

where the (F_i) are i.i.d. exponential random variables with mean ρ . In the Bellman–Harris case, following [1], it is possible to show that W and (F_i) are independent, so that in this case, the asymptotic behavior of (P_n) is exactly the same as in the case of a Yule process. One can conjecture that this independence property still holds in the general case.

The main obstacle to generalizing the results of this paper, even in the Bellman–Harris case, is that although W and (F_i) are independent, t_{n-1} and $t_n - t_{n-1}$ are not independent. In the proof of Proposition 1, this independence plays a crucial role; therefore, it has to be generalized to variables which are only asymptotically independent. Additionally, since the heavy tail property of the limiting variable $W_\infty^{-\rho}$ is also true in the general case (see, e.g. [16]), a similar rare events phenomenon to the one described in Section 5 is plausible in this case.

6.2. General balls and the Yule process

When the underlying branching process is changed, the above discussion suggests that the asymptotic behavior of the sequence (P_n) remains essentially the same as for a Yule process.

The situation changes significantly when the law of the location X of a ball is changed. In this case, using the same notation as before for the Yule process,

$$P_n = \mathbb{P}\left(t_{n-1} < X \leq t_{n-1} + \frac{E_n}{n}\right).$$

The tail distribution of X then plays a key role. Consider, for instance, a power law, i.e. $\mathbb{P}(X \geq x)$ behaves as $\delta x^{-\beta}$ for some β and $\delta > 0$. Then

$$P_{n+1} \sim t_n^{-\beta} - \left(t_n + \frac{E_{n+1}}{n+1}\right)^{-\beta} \sim \frac{\beta \delta E_{n+1}}{n t_n^{\beta+1}} \sim \frac{\beta \delta E_{n+1}}{n (\log n)^{\beta+1}},$$

and it can be seen that the random variable W_∞ may not play a role anymore in the asymptotic behavior of the system.

Appendix A

This section is devoted to the proof of Proposition 5. The notation of Section 5 is used. For $N \geq 1$, we let $k_n = \lfloor x\phi(n) \rfloor$. It is assumed that $\phi(n) \ll n^{\alpha_0}$ for some $\alpha_0 < 1/(\rho + 1)$.

For $0 \leq p \leq 1$ and $n \geq 1$, the elementary inequality

$$|e^{-np} - (1 - p)^n| \leq \frac{p^2}{2} n e^{-np} \leq \frac{2e^2}{n}$$

and the relation $\mathbb{E}[\mathcal{N}_n^\phi([0, x])] = \mathbb{E}[(1 - P_1)^n] + \dots + \mathbb{E}[(1 - P_{\lfloor x\phi(n) \rfloor})^n]$ directly give the following lemma.

Lemma 1. *For a nondecreasing function ϕ , $x \geq 0$, and $n \geq 1$,*

$$\left| \mathbb{E}[\mathcal{N}_n^\phi([0, x])] - \sum_{i=1}^{k_n} \mathbb{E}[\exp(-nP_i)] \right| \leq 2e^2 \frac{k_n}{n}.$$

Lemma 2. *There exists $\eta > 0$ such that*

$$\mathbb{E}\left[\sup_{n \geq 1} W_n^{-\eta}\right] < +\infty.$$

Proof. For $n \geq 1$, $W_n = \exp(-V_{n-1}) \exp(-(H_{n-1} - \log n))$, where $V_n = t_n - H_n$ and (H_n) is the harmonic sequence. For $0 < \eta < \frac{1}{2}$, it is easy to check that

$$\sup\{\mathbb{E}[\exp(2\eta V_n)]: n \geq 0\} < +\infty.$$

The sequence (V_n) being a martingale, Doob's inequality states that, for any $n \geq 0$,

$$\mathbb{E}\left[\sup_{n \geq 1} \exp(2\eta V_n)\right] \leq 4 \sup_{n \geq 1} \mathbb{E}[\exp(2\eta V_n)].$$

The result is proved.

Lemma 3. *If F is an exponential random variable with mean ρ independent of W_i then*

$$\lim_{n \rightarrow \infty} \left| \mathbb{E}\left[\sum_{i=1}^{k_n} \exp(-ni^{-\rho-1} W_i^\rho Z_i)\right] - \mathbb{E}\left[\sum_{i=1}^{k_n} \exp(-ni^{-\rho-1} W_i^\rho F)\right] \right| = 0.$$

Proof. Let $y > 0$ and $Z = (1 - e^{-yF})/y$. Then, trite calculations with the exponential distribution give, for any $\beta > 0$,

$$0 \leq E[e^{-\beta Z}] - E[e^{-\beta F}] \leq e^{-(1+\rho\beta)/(2y)} + \frac{2y}{(1 + \rho\beta)^2}.$$

For $k \geq 1$ and $1 \leq i \leq k$, since $Z_i = i(1 - \exp(-\rho E_i/i))$, where E_i is an exponential random variable with mean 1, by using this relation we obtain

$$\begin{aligned} 0 &\leq E\left[\sum_{i=1}^k \exp(-ni^{-\rho-1}W_i^\rho Z_i)\right] - E\left[\sum_{i=1}^k \exp(-ni^{-\rho-1}W_i^\rho F)\right] \\ &\leq E\left[\sum_{i=1}^k \exp\left(-\frac{i}{2\rho}(1 + \rho ni^{-\rho-1}W_i^\rho)\right)\right] + E\left[\sum_{i=1}^k \frac{2\rho}{i(1 + \rho ni^{-\rho-1}W_i^\rho)^2}\right]. \end{aligned} \tag{12}$$

Since $(i^{-\rho-1}W_i^\rho)$ is a nonincreasing sequence, then

$$\begin{aligned} \sum_{i=1}^{k_n} \exp\left(-\frac{i}{2\rho}(1 + \rho ni^{-\rho-1}W_i^\rho)\right) &\leq \sum_{i=1}^{k_n} \exp\left(-\frac{i}{2\rho}(1 + \rho nk_n^{-\rho-1}W_{k_n}^\rho)\right) \\ &\leq \frac{\exp(-(1 + \rho nk_n^{-\rho-1}W_{k_n}^\rho)/(2\rho))}{1 - \exp(-(1 + \rho nk_n^{-\rho-1}W_{k_n}^\rho)/(2\rho))}, \end{aligned}$$

and this last term converges almost surely to 0 because $k_n \ll n^{1/(\rho+1)}$. Using the relation

$$\sum_{i=1}^{k_n} \exp\left(-\frac{i}{2\rho}(1 + \rho ni^{-\rho-1}W_i^\rho)\right) \leq \sum_{i=1}^{\infty} \exp\left(-\frac{i}{2\rho}\right) < +\infty,$$

the first term on the right-hand side of (12) vanishes by Lebesgue’s theorem.

Since

$$\sum_{i=1}^{k_n} \frac{1}{i(1 + \rho ni^{-\rho-1}W_i^\rho)^2} \leq \frac{1}{(1 + \rho nk_n^{-\rho-1}W_{k_n}^\rho)^2} \sum_{i=1}^{k_n} \frac{1}{i},$$

the second term on the right-hand side of (12) converges almost surely to 0 because $k_n \ll n^{1/(\rho+1)}$. For $\varepsilon > 0$, again by monotonicity,

$$\sum_{i=1}^{k_n} \frac{1}{i(1 + \rho ni^{-\rho-1}W_i^\rho)^2} = \sum_{i=1}^{k_n} \frac{i^\varepsilon}{i^{1+\varepsilon}(1 + \rho ni^{-\rho-1}W_i^\rho)^2} \leq C_\varepsilon \frac{k_n^\varepsilon}{(1 + \rho nk_n^{-\rho-1}W_{k_n}^\rho)^2},$$

where C_ε is a finite constant. By using some elementary calculations and Lemma 2, it is not difficult to show that one can choose $\varepsilon > 0$ so that this last term is uniformly bounded in n by an integrable random variable. Again, Lebesgue’s theorem shows that the second term on the right-hand side of (12) converges to 0. The lemma is proved.

Combined with the last result, the following lemma completes the proof of Proposition 5.

Lemma 4. *If F is an exponential random variable with mean ρ independent of W_i^ρ then*

$$\lim_{n \rightarrow +\infty} E\left[\sum_{i=1}^{k_n} \exp(-ni^{-\rho-1}W_i^\rho F)\right] / E\left[\sum_{i=1}^{k_n} \exp(-ni^{-\rho-1}W_\infty^\rho F)\right] = 1$$

holds under the assumption of Proposition 5 that

$$\sum_{i=1}^{k_n} \mathbb{E}[\exp(-n\tilde{P}_i)]$$

converges to ∞ as n gets large.

Proof. For any $n \geq 1$,

$$\begin{aligned} & \left| \mathbb{E} \left[\sum_{i=1}^{k_n} \exp(-ni^{-\rho-1} W_i^\rho F) \right] - \mathbb{E} \left[\sum_{i=1}^{k_n} \exp(-ni^{-\rho-1} W_\infty^\rho F) \right] \right| \\ &= \left| \mathbb{E} \left[\sum_{i=1}^{k_n} \frac{1}{1 + \rho ni^{-\rho-1} W_i^\rho} - \sum_{i=1}^{k_n} \frac{1}{1 + \rho ni^{-\rho-1} W_\infty^\rho} \right] \right| \\ &\leq \sum_{i=1}^{k_n} \mathbb{E} \left[\frac{\rho ni^{-\rho-1} |W_\infty^\rho - W_i^\rho|}{(1 + \rho ni^{-\rho-1} W_i^\rho)(1 + \rho ni^{-\rho-1} W_\infty^\rho)} \right] \\ &\leq \sum_{i=1}^{k_n} \mathbb{E} \left[\frac{|W_i^\rho / W_\infty^\rho - 1|}{1 + \rho ni^{-\rho-1} W_i^\rho} \right]; \end{aligned}$$

hence,

$$\begin{aligned} & \left| \mathbb{E} \left[\sum_{i=1}^{k_n} \exp(-ni^{-\rho-1} W_i^\rho F) \right] - \mathbb{E} \left[\sum_{i=1}^{k_n} \exp(-ni^{-\rho-1} W_\infty^\rho F) \right] \right| \\ &\leq \sum_{i=1}^{k_n} \mathbb{E} \left[\frac{1}{1 + \rho ni^{-\rho-1} W_i^\rho} \right] \mathbb{E} \left[\left| \frac{W_i^\rho}{W_\infty^\rho} - 1 \right| \right], \end{aligned}$$

by the independence of W_i^ρ / W_∞^ρ and W_i^ρ . The sequence $(W_i^\rho / W_\infty^\rho)$ being uniformly integrable, it converges to 1 only almost surely and in L_1 ; consequently,

$$\sum_{i=1}^{k_n} \mathbb{E} \left[\frac{1}{1 + \rho ni^{-\rho-1} W_i^\rho} \right] \mathbb{E} \left[\left| \frac{W_i^\rho}{W_\infty^\rho} - 1 \right| \right] = o \left(\sum_{i=1}^{k_n} \mathbb{E} \left[\frac{1}{1 + \rho ni^{-\rho-1} W_i^\rho} \right] \right)$$

since the sequence

$$\left(\sum_{i=1}^{k_n} \mathbb{E}[\exp(-n\tilde{P}_i)] \right) = \left(\sum_{i=1}^{k_n} \mathbb{E} \left[\frac{1}{1 + \rho ni^{-\rho-1} W_i^\rho} \right] \right)$$

converges to ∞ . This completes the proof.

References

- [1] ATHREYA, K. B. AND KAPLAN, N. (1976). Convergence of the age distribution in the one-dimensional supercritical age-dependent branching process. *Ann. Prob.* **4**, 38–50.
- [2] ATHREYA, K. B. AND NEY, P. E. (1972). *Branching Processes*. Springer, New York.
- [3] BARBOUR, A. D., HOLST, L. AND JANSON, S. (1992). *Poisson Approximation* (Oxford. Stud. Prob. 2). Clarendon Press, New York.

- [4] CSÁKI, E. AND FÖLDES, A. (1976). On the first empty cell. *Studia Sci. Math. Hung.* **11**, 373–382.
- [5] DAWSON, D. A. (1993). Measure-valued Markov processes. In *École d'Été de Probabilités de Saint-Flour XXI* (Lecture Notes Math. **1541**), Springer, Berlin, pp. 1–260.
- [6] FLAJOLET, P. AND MARTIN, G. N. (1985). Probabilistic counting algorithms for data base applications. *J. Comput. System Sci.* **21**, 182–209.
- [7] GNEDIN, A. V. (2004). The Bernoulli Sieve. *Bernoulli* **10**, 79–96.
- [8] GNEDIN, A. V., HANSEN, B. AND PITMAN, J. (2007). Notes on the occupancy problem with infinitely many boxes: general asymptotics and power laws. *Prob. Surveys* **4**, 146–171.
- [9] GNEDIN, A., IKSANOV, A. AND ROESLER, U. (2008). Small parts in the Bernoulli sieve. *Discrete Math. Theoret. Computer Sci.* **A1**, 239–246. Available at <http://arxiv.org/abs/0804.3052>.
- [10] HÄRNQVIST, M. (1981). Limit theorems for point processes generated in a general branching process. *Adv. Appl. Prob.* **13**, 650–668.
- [11] HWANG, H.-K. AND JANSON, S. (2008). Local limit theorems for finite and infinite urn models. *Ann. Prob.* **36**, 992–1022.
- [12] JOHNSON, N. L. AND KOTZ, S. (1977). *Urn Models and Their Application*. John Wiley, New York.
- [13] KARLIN, S. (1967). Central limit theorems for certain infinite urn schemes. *J. Math. Mech.* **17**, 373–401.
- [14] KINGMAN, J. F. C. (1975). The first birth problem for an age-dependent branching process. *Ann. Prob.* **3**, 790–801.
- [15] KOLCHIN, V. F., SEVAST'YANOV, B. A. AND CHISTYAKOV, V. P. (1978). *Random Allocations*, V. H. Winston, Washington, D.C.
- [16] LIU, Q. (2001). Asymptotic properties and absolute continuity of laws stable by random weighted mean. *Stoch. Process. Appl.* **95**, 83–107.
- [17] NERMAN, O. (1981). On the convergence of supercritical general (C-M-J) branching processes. *Z. Wahrscheinlichkeitsthe.* **57**, 365–395.
- [18] NEVEU, J. (1972). *Martingales à Temps Discret*. Masson et Cie, éditeurs, Paris.
- [19] NEVEU, J. (1977). Processus ponctuels. In *École d'Été de Probabilités de Saint-Flour VI* (Lecture Notes Math. **598**), Springer, Berlin, pp. 249–445.
- [20] ROBERT, P. (2003). *Stochastic Networks and Queues* (Appl. Math. **52**). Springer, New York.
- [21] RUDIN, W. (1987). *Real and Complex Analysis*, 3rd edn., McGraw-Hill, New York.
- [22] SADDI, W. AND GUILLEMIN, F. (2007). Measurement based modeling of edonkey peer-to-peer file sharing system. In *Proc. Internat. Teletraffic Congress 20* (Ottawa, June 2007), pp. 974–985.
- [23] SIMATOS, F., ROBERT, P. AND GUILLEMIN, F. (2008). A queueing system for modeling a file sharing principle. In *Proc. ACM SIGMETRICS 2008*, ACM Press, New York.
- [24] WILLIAMS, D. (1991). *Probability with Martingales*. Cambridge University Press.