On nested infinite occupancy scheme in random environment

Alexander Gnedin* and Alexander Iksanov†

August 1, 2018

Abstract

We consider an infinite balls-in-boxes occupancy scheme with boxes organised in nested hierarchy, and random probabilities of boxes defined in terms of iterated fragmentation of a unit mass. We obtain a multivariate functional limit theorem for the cumulative occupancy counts as the number of balls approaches infinity. In the case of fragmentation driven by a homogeneous residual allocation model our result generalises the functional central limit theorem for the block counts in Ewens' and more general regenerative partitions.

Key words: Bernoulli sieve; Ewens' partition; functional limit theorem; infinite occupancy; nested hierarchy
2000 Mathematics Subject Classification: Primary: 60F17, 60J80
Secondary: 60C05

1 Introduction

In the infinite multinomial occupancy scheme balls are thrown independently in a series of boxes, so that each ball hits box $k = 1, 2, \dots$ with probability p_k, where $p_k > 0$ and $\sum_{k\in\mathbb{N}} p_k = 1$. This classical model is sometimes named after Karlin due to his seminal contribution [28]. Features of the occupancy pattern emerging after the first n balls are thrown have been intensely studied, see [5, 18, 24] for survey and references and [6, 11, 12, 13] for recent advances. Statistics in focus of most of the previous work, and also relevant to the subject of this paper, are not sensitive to the labelling of boxes but rather only depend on the integer partition of n comprised of nonzero occupancy numbers.

In the infinite occupancy scheme in a random environment the (hitting) probabilities of boxes are positive random variables $(P_k)_{k\in\mathbb{N}}$ with an arbitrary joint distribution satisfying $\sum_{k\in\mathbb{N}} P_k = 1$ almost surely. Conditionally on $(P_k)_{k\in\mathbb{N}}$, balls are thrown independently, with probability P_k of hitting box k. Instances of this general setup have received considerable attention within the circle of questions around exchangeable partitions, discrete random measures and their applications to population genetics, Bayesian statistics and computer science. In the most studied and analytically best tractable case the probabilities of boxes are representable as the residual allocation (or stick-breaking) model

$$P_k = U_1 U_2 \cdots U_{k-1}(1 - U_k), \quad k \in \mathbb{N}, \tag{1}$$

*School of Mathematical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, UK; e-mail: a.gnedin@qmul.ac.uk

†Faculty of Computer Science and Cybernetics, Taras Shevchenko National University of Kyiv, 01601 Kyiv, Ukraine; e-mail: iksan@univ.kiev.ua

where the U_i's are independent with beta$(\theta, 1)$ distribution[1] on $(0, 1)$ and $\theta > 0$. In this case the distribution of the sequence $(P_k)_{k\in\mathbb{N}}$ is known as the Griffiths-Engen-McCloskey (GEM) distribution with parameter θ. The sequence of the P_k's arranged in decreasing order has the Poisson-Dirichlet (PD) distribution with parameter θ, and the induced exchangeable partition on the set of n balls follows the celebrated Ewens sampling formula [3, 29, 31, 32]. Generalisations have been proposed in various directions. The two-parameter extension due to Pitman and Yor [29] involves probabilities of form (1) with independent but not identically distributed U_i's, where the distribution of U_i is beta$(\theta + \alpha i, 1 - \alpha)$ (with $0 < \alpha < 1$ and $\theta > -\alpha$). Residual allocation models with other choices of parameters for the U_i's with different beta distributions are found in [26, 33]. Much effort has been devoted to the occupancy scheme, known as the Bernoulli sieve, which is based on a *homogeneous* residual allocation model (1), that is, with independent and identically distributed (iid) factors U_i having arbitrary distribution on $(0, 1)$, see [2, 14, 20, 24, 25, 30]. The homogeneous model has a multiplicative regenerative property, also inherited by the partition of the set of balls.

In more sophisticated constructions of random environments probabilities $(P_k)_{k\in\mathbb{N}}$ are identified with some arrangement in sequence of masses of a purely atomic random probability measure. A widely explored possibility is to define a random cumulative distribution function F by transforming the path of an increasing drift-free Lévy process (subordinator) $(X(t))_{t\geq 0}$. In particular, in the Poisson-Kingman model $F(t) = X(t)/X(1)$ for a measure supported by $[0, 1]$, see [16, 29]. In the regenerative model $F(t) = 1 - e^{-X(t)},\ t \geq 0$, called in the statistical literature neutral-to-the right prior [16], see [4, 19, 21, 22].

Following [7, 10, 27] we shall study a nested infinite occupancy scheme in random environment. In this context we regard $(P_k)_{k\in\mathbb{N}}$ as a random *fragmentation law* (with $P_k > 0$ and $\sum_{k\in\mathbb{N}} P_k = 1$ a.s.). To introduce hierarchy of boxes, for each $j \in \mathbb{N}_0$ let $\mathbb{W}_j$ be the set of words of length j over $\mathbb{N}$, where $\mathbb{W}_0 := \{\varnothing\}$. The set $\mathbb{W} = \bigcup_{j\in\mathbb{N}_0} \mathbb{W}_j$ of all finite words has the natural structure of a ∞-storey tree with root $\varnothing$ and ∞-ary branching at every node, where $w1, w2, \dots \in \mathbb{W}_{j+1}$ are the immediate followers of $w \in \mathbb{W}_j$. Let $\{(P_k^{(w)})_{k\in\mathbb{N}},\ w \in \mathbb{W}\}$ be a family of independent copies of $(P_k)_{k\in\mathbb{N}}$. With each $w \in \mathbb{W}$ we associate a box divided in sub-boxes $w1, w2, \dots$ of the next level. The probabilities of boxes are defined recursively by

$$P(\varnothing) = 1, \quad P(wk) = P(w)P_k^{(w)} \quad \text{for } w \in \mathbb{W}, k \in \mathbb{N} \tag{2}$$

(note that the factors $P(w)$ and $P_k^{(w)}$ are independent). Given $(P(w))_{w\in\mathbb{W}}$, balls are thrown independently, with probability $P(w)$ of hitting box w. Since $\sum_{w\in\mathbb{W}_j} P(w) = 1$ the allocation of balls in boxes of level j occurs according to the ordinary Karlin's occupancy scheme.

Recursion (2) defines a discrete-time mass-fragmentation process, where the generic mass splits in proportions according to the same fragmentation law, independently of the history and masses of the co-existing fragments. The nested occupancy scheme can be seen as a combinatorial version of this fragmentation process. Initially all balls are placed in box $\varnothing$, and at each consecutive step $j + 1$ each ball in box $w \in \mathbb{W}_j$ is placed in sub-box wk with probability $P_k^{(w)}$. The inclusion relation on the hierarchy of boxes induces a combinatorial structure on the (labelled) set of balls called total partition, that is a sequence of refinements from the trivial one-block partition down to the partition in singletons. The paper [15] highlights the role of exchangeability and gives the general de Finetti-style connection between mass-fragmentations and total partitions.

[1] Recall that a random variable X has a beta distribution with parameters $\alpha > 0$ and $\beta > 0$ if $\mathbb{P}\{X \in dx\} = (1/\mathrm{B}(\alpha, \beta))x^{\alpha-1}(1 - x)^{\beta-1}\, \mathbb{1}_{(0,1)}(x)dx$. Here, $\mathrm{B}(\cdot, \cdot)$ is the beta function.

We consider the random probabilities of the hierarchy of boxes and the outcome of throwing infinitely many balls all defined on the same underlying probability space. For $j, r \in \mathbb{N}$, denote by $K_{n,j,r}$ the number of boxes $w \in \mathbb{W}_j$ of the jth level that contain exactly r out of n first balls, and let

$$K_{n,j}(s) := \sum_{r=\lceil n^{1-s} \rceil}^{n} K_{n,j,r}, \quad s \in [0,1], \tag{3}$$

be a cumulative count of occupied boxes, where $\lceil \cdot \rceil$ is the integer ceiling function. With probability one the random function $s \mapsto K_{n,j}(s)$ is nondecreasing and right-continuous, hence belongs to the Skorokhod space $D[0,1]$. Also observe that $K_{n,j}(0) = K_{n,j,n}$ is zero unless all balls fall in the same box and that $K_{n,j}(1)$ is the number of occupied boxes in the jth level. In [7] a central limit theorem with random centering was proved for $K_{n,j}(1)$ for j growing with n at certain rate. Our focus is different. We are interested in the joint weak convergence of $((K_{n,j_1}(s), \dots, K_{n,j_m}(s)))_{s\in[0,1]}$, properly normalised and centered, for any finite collection of occupancy levels $1 \le j_1 < \dots < j_m$ as the number of balls n tends to ∞. As far as we know, this question has not been addressed so far. We prove a multivariate functional limit theorem (Theorem 2.1) applicable to the fragmentation laws representable by homogeneous residual allocations models (including the GEM/PD distribution) and some other models where the sequence of P_k's arranged in decreasing order approaches zero sufficiently fast. A univariate functional limit for $(K_{n,1}(s))_{s\in[0,1]}$ in the case of Bernoulli sieve was previously obtained in [2].

2 Main result

For given fragmentation law $(P_k)_{k\in\mathbb{N}}$, let $\rho(s) := \#\{k \in \mathbb{N} : P_k \ge 1/s\}$ for $s > 0$, and $N(t) := \rho(e^t), V(t) := \mathbb{E}N(t)$ for $t \in \mathbb{R}$. The joint distribution of $K_{n,j,r}$'s is completely determined by the probability law of the random function $\rho(\cdot)$, which captures the fragmentation law up to re-arrangement of P_k's. For our purposes therefore we can make no difference between fragmentation laws with the same $\rho(\cdot)$.

Similarly, using probabilities of boxes in level $j \in \mathbb{N}$ define $\rho_j(s) := \#\{v \in \mathbb{V}_j : P(v) \ge 1/s\}$ for $s > 0$, and $N_j(t) := \rho_j(e^t), V_j(t) := \mathbb{E}N_j(t)$ for $t \in \mathbb{R}$. Note that $N_j(t) = 0$ for $t \le 0$. Since $\sum_{w\in\mathbb{W}_j} P(w) = 1$ a.s. we have $\rho_j(s) \le s$, whence $N_j(t) \le e^t$ a.s. and $V_j(t) < e^t$.

Let $T_k := -\log P_k$. Here is a basic decomposition of principal importance for what follows:

$$N_j(t) = \sum_{k\in\mathbb{N}} N_{j-1}^{(k)}(t - T_k), \quad t \in \mathbb{R}, \tag{4}$$

where $(N_{j-1}^{(k)}(t))_{t\ge0}$ for $k \in \mathbb{N}$ are independent copies of $N_{j-1}(\cdot)$ which are also independent of $T_1, T_2, \dots$ An immediate consequence of (4) is a recursion for the expectations

$$V_j(t) = \int_{[0,\,t]} V_{j-1}(t-y)\mathrm{d}V(y), \quad t \ge 0,\ j \ge 2, \tag{5}$$

which shows that $V_j(\cdot)$ is the jth convolution power of $V(\cdot)$.

The assumptions on fragmentation law and the functional limit will involve a centered Gaussian process $W := (W(s))_{s\ge0}$ which is a.s. locally Hölder continuous with exponent $\beta > 0$ and satisfy $W(0) = 0$. In particular,

$$|W(x) - W(y)| \le M(x-y)^\beta, \quad 0 \le y < x \le 1 \tag{6}$$

for some a.s. finite random variable M. We set further

$$R_1(s) := W(s), \quad R_j(s) := \int_{[0,\,s]} (s-y)^{j-1} \mathrm{d}W(y), \quad s \geq 0,\ j \geq 2.$$

Alternatively, the process R_j can be defined via repeated integration by parts

$$R_j(s) = (j-1)! \int_0^{s_1} \int_0^{s_2} \ldots \int_0^{s_{j-1}} W(y) \mathrm{d}y \mathrm{d}s_{j-1} \ldots \mathrm{d}s_2, \quad s \geq 0,\ j \geq 2,$$

where $s_1 = s$.

Throughout the paper $D := D[0, \infty)$ denotes the standard Skorokhod space. Here is our main result.

Theorem 2.1. *Assume the following conditions hold:*

(i)
$$b_0 + b_1 t^{\omega - \varepsilon_1} \leq V(t) - ct^{\omega} \leq a_0 + a_1 t^{\omega - \varepsilon_2} \tag{7}$$

for all $t \geq 0$ *and some constants* $c, \omega, a_0, a_1 > 0$, $0 < \varepsilon_1, \varepsilon_2 \leq \omega$ *and* $b_0, b_1 \in \mathbb{R}$,

(ii)
$$\lim_{t \to \infty} \frac{N(t)}{t^{\omega}} = c \quad \text{a.s.}, \tag{8}$$

(iii)
$$\frac{N(t\,\cdot) - c(t\,\cdot)^{\omega}}{at^{\gamma}} \;\Rightarrow\; W(\cdot), \quad t \to \infty \tag{9}$$

in the J_1*-topology on* D *for some* $a > 0$ *and* $\gamma \in (\omega - \min(1, \varepsilon_1, \varepsilon_2), \omega)$.

Then

$$\left(\frac{K_{n,j}(\cdot) - c_j (\log n(\cdot))^{\omega j}}{a c_{j-1} (\log n)^{\gamma + \omega(j-1)}} \right)_{j \in \mathbb{N}} \;\Rightarrow\; (R_j(\cdot))_{j \in \mathbb{N}}, \quad n \to \infty \tag{10}$$

in the J_1*-topology on* $D[0,1]^{\mathbb{N}}$, *where* $\Gamma(\cdot)$ *is the gamma function and*

$$c_j := \frac{(c\Gamma(\omega+1))^j}{\Gamma(\omega j + 1)}, \quad j \geq 0. \tag{11}$$

Remark 2.2. The assumption $0 < \varepsilon_1, \varepsilon_2 \leq \omega$ ensures that $\gamma > 0$. Furthermore, in view of (7) and the choice of γ relation (9) is equivalent to

$$\frac{N(t\,\cdot) - V(t\,\cdot)}{at^{\gamma}} \;\Rightarrow\; W(\cdot), \quad t \to \infty \tag{12}$$

in the J_1-topology on D. Similarly, in view of (13) given below relation (10) is equivalent to

$$\left(\frac{K_{n,j}(\cdot) - V_j(\log n(\cdot))}{a c_{j-1} (\log n)^{\gamma + \omega(j-1)}} \right)_{j \in \mathbb{N}} \;\Rightarrow\; (R_j(\cdot))_{j \in \mathbb{N}}, \quad n \to \infty$$

in the J_1-topology on $D[0,1]^{\mathbb{N}}$.

3 Proof of Theorem 2.1

3.1 Auxiliary results

Lemma 3.1. (a) *Condition* (7) *ensures that, for* $j \in \mathbb{N}$ *and* $t \geq 0$,

$$b_{0,j} + b_{1,j} t^{\omega j - \varepsilon_1} \leq V_j(t) - c_j t^{\omega j} \leq a_{0,j} + a_{1,j} t^{\omega j - \varepsilon_2}, \tag{13}$$

where c_j *is given by* (11), $a_{0,j}, a_{1,j} > 0$ *and* $b_{0,j}, b_{1,j} \in \mathbb{R}$ *are constants with* $a_{0,1} := a_0$, $a_{1,1} := a_1$, $b_{0,1} := b_0$ *and* $b_{1,1} := b_1$. *In particular, for* $j \in \mathbb{N}$,

$$V_j(t) \ \sim \ c_j t^{\omega j}, \quad t \to \infty \tag{14}$$

and, for $j \in \mathbb{N}$ *and* $u, v \geq 0$,

$$\begin{aligned} V_j(u+v) - V_j(v) \ &\leq \ c_j (\mathbb{1}_{\{\omega j \in (0,1]\}} u^{\omega j} + \mathbb{1}_{\{\omega j > 1\}} \omega j (u+v)^{\omega j - 1} u) \\ &+ \ a_{0,j} + a_{1,j} (u+v)^{\omega j - \varepsilon_2} - b_{0,j} - b_{1,j} v^{\omega j - \varepsilon_1}. \end{aligned} \tag{15}$$

(b) *Suppose* (7) *and* (9). *Then, for* $j \in \mathbb{N}$ *and* $h > 0$,

$$t^{-\gamma - \omega(j-1)} \sup_{y \in [0,\,1]} (N_j(yt+h) - N_j(yt)) \ \overset{\mathrm{P}}{\to} \ 0, \quad t \to \infty. \tag{16}$$

Proof. (a) We only prove the second inequality in (13). To this end, we first check that for any $b > 0$

$$\int_{[0,\,t]} (t-y)^b \mathrm{d}V(y) \leq a_0 t^b + b a_1 \mathrm{B}(b, 1+\omega-\varepsilon) t^{\omega - \varepsilon + b} + bc \mathrm{B}(b, 1+\omega) t^{\omega + b},$$

where $\mathrm{B}(\cdot,\cdot)$ is the beta function, and we write ε for ε_2 to ease notation. Indeed, using (7) we obtain

$$\begin{aligned} \int_{[0,\,t]} (t-y)^b \mathrm{d}V(y) \ &= \ b \int_0^t (V(t-y) - c(t-y)^\omega) y^{b-1} \mathrm{d}y + bc \int_0^t (t-y)^\omega y^{b-1} \mathrm{d}y \\ &\leq \ b a_0 \int_0^t y^{b-1} \mathrm{d}y + b a_1 \int_0^t (t-y)^{\omega - \varepsilon} y^{b-1} \mathrm{d}y + bc \int_0^t (t-y)^\omega y^{b-1} \mathrm{d}y \\ &= \ a_0 t^b + b a_1 \mathrm{B}(b, 1+\omega-\varepsilon) t^{\omega-\varepsilon+b} + bc \mathrm{B}(b, 1+\omega) t^{\omega+b}. \end{aligned}$$

To prove the second inequality in (13) we use induction. The case $j = 1$ is covered by (7). Assume the inequality holds for $j = k-1$. Then, for $t \geq 0$ recalling (5) we obtain

$$\begin{aligned} V_k(t) \ &= \ \int_{[0,\,t]} (V_{k-1}(t-y) - c_{k-1}(t-y)^{\omega(k-1)}) \mathrm{d}V(y) + c_{k-1} \int_{[0,\,t]} (t-y)^{\omega(k-1)} \mathrm{d}V(y) \\ &\leq \ a_{0,k-1} V(t) + a_{1,k-1} \int_{[0,\,t]} (t-y)^{\omega(k-1)-\varepsilon} \mathrm{d}V(y) + c_{k-1} \int_{[0,\,t]} (t-y)^{\omega(k-1)} \mathrm{d}V(y) \\ &\leq \ a_{0,k-1} V(t) + a_{1,k-1} \big(a_0 t^{\omega(k-1)-\varepsilon} + (\omega(k-1)-\varepsilon) a_1 \mathrm{B}(\omega(k-1)-\varepsilon, 1+\omega-\varepsilon) t^{\omega k - 2\varepsilon} \\ &+ \ (\omega(k-1)-\varepsilon) c \mathrm{B}(\omega(k-1)-\varepsilon, 1+\omega) t^{\omega k - \varepsilon} \big) \\ &+ \ c_{k-1} \big(a_0 t^{\omega(k-1)} + \omega(k-1) a_1 \mathrm{B}(\omega(k-1), 1+\omega-\varepsilon) t^{\omega k - \varepsilon} \\ &+ \ \omega(k-1) c \mathrm{B}(\omega(k-1), 1+\omega) t^{\omega k} \big) \leq c_k t^{\omega k} + a_{0,k} + a_{1,k} t^{\omega k - \varepsilon} \end{aligned}$$

for appropriate positive $a_{0,k}$ and $a_{1,k}$, where we used $c_k = c_{k-1} \omega(k-1) c \mathrm{B}(\omega(k-1), 1+\omega)$.

Further, (14) is an immediate consequence of (13). To prove (15), we use (13) to obtain, for $j \in \mathbb{N}$ and $u, v \geq 0$,

$$V_j(u+v) - V_j(v) \leq c_j((u+v)^{\omega j} - v^{\omega j}) + a_{0,j} + a_{1,j}(u+v)^{\omega j - \varepsilon_2} - b_{0,j} - b_{1,j} v^{\omega j - \varepsilon_1}.$$

If $\omega j \in (0,1]$, we have $(u+v)^{\omega j} - v^{\omega j} \leq u^{\omega j}$ by subadditivity. If $\omega j > 1$, we have $(u+v)^{\omega j} - v^{\omega j} \leq \omega j (u+v)^{\omega j - 1} u$ by the mean value theorem and monotonicity. This completes the proof of (15).

(b) Since W is a.s. continuous we conclude that, for any $h > 0$,

$$\Big(\frac{N(t\cdot) - c(t\cdot)^\omega}{a t^\gamma}, \frac{N(t\cdot + h) - c(t\cdot + h)^\omega}{a t^\gamma}\Big) \ \Rightarrow\ (W(\cdot), W(\cdot)), \quad t \to \infty$$

in the J_1-topology on $D \times D$, whence

$$t^{-\gamma} \sup_{y \in [0,1]} (N(ty+h) - N(ty) - c((ty+h)^\omega - (ty)^\omega)) \ \overset{\mathrm{P}}{\to}\ 0, \quad t \to \infty.$$

Using

$$\sup_{y \in [0,1]} ((ty+h)^\omega - (ty)^\omega) \ \leq\ \mathbb{1}_{\{\omega \in (0,1]\}} h^\omega + \mathbb{1}_{\{\omega > 1\}} \omega (t+h)^{\omega - 1} h$$

we conclude that the right-hand side is $o(t^\gamma)$ as $t \to \infty$ because $\gamma > \omega - 1$ by assumption. Thus, we have shown that

$$t^{-\gamma} \sup_{y \in [0,1]} (N(yt+h) - N(yt)) \ \overset{\mathrm{P}}{\to}\ 0, \quad t \to \infty. \tag{17}$$

To prove (16) we again use induction. When $j = 1$, the result holds according to (17). Assume that (16) holds for $j = \ell$. Then using a decomposition analogous to (4) we obtain

$$\begin{aligned}
\sup_{y \in [0,t]} (N_{\ell+1}(y+h) - N_{\ell+1}(y)) \ &= \ \sup_{y \in [0,t]} \Big(\sum_{k \in \mathbb{N}} (N_\ell^{(k)}(y+h-T_k) - N_\ell^{(k)}(y - T_k)) \, \mathbb{1}_{\{T_k \leq y\}} \\
&+ \ \sum_{k \in \mathbb{N}} N_\ell^{(k)}(y+h-T_k) \, \mathbb{1}_{\{y < T_k \leq y+h\}} \Big) \\
&\leq \ \sum_{k \in \mathbb{N}} \sup_{y \in [0,t]} (N_\ell^{(k)}(y+h) - N_\ell^{(k)}(y)) \, \mathbb{1}_{\{T_k \leq t\}} \\
&+ \ \sum_{k \in \mathbb{N}} N_\ell^{(k)}(h) \, \mathbb{1}_{\{T_k \leq t+h\}} \, .
\end{aligned}$$

The random variables $\sup_{y \in [0,t]} \big(N_\ell^{(k)}(y+h) - N_\ell^{(k)}(y)\big)$ for $k \in \mathbb{N}$ are independent and identically distributed. Then, for any $\varepsilon > 0$ and $\delta \in (0, c)$

$$\begin{aligned}
&\mathbb{P}\Big\{ \sum_{k \in \mathbb{N}} \sup_{y \in [0,t]} (N_\ell^{(k)}(y+h) - N_\ell^{(k)}(y)) \, \mathbb{1}_{\{T_k \leq t\}} > \varepsilon t^{\gamma + \omega \ell} \Big\} \\
\leq \ &\mathbb{P}\{N(t) > (c+\delta) t^\omega\} \\
+ \ &\mathbb{P}\Big\{ \sum_{k \in \mathbb{N}} \sup_{y \in [0,t]} (N_\ell^{(k)}(y+h) - N_\ell^{(k)}(y)) \, \mathbb{1}_{\{T_k \leq t\}} > \varepsilon t^{\gamma + \omega \ell}, \, N(t) \leq (c+\delta) t^\omega \Big\} \\
\leq \ &\mathbb{P}\{N(t) > (c+\delta) t^\omega\} \\
+ \ &\mathbb{P}\{ \sup_{y \in [0,t]} (N_\ell(y+h) - N_\ell(y)) > \varepsilon t^{\gamma + \omega \ell} \lfloor (c+\delta) t^\omega \rfloor^{-1} \} \ \to \ 0 \quad \text{as } t \to \infty.
\end{aligned}$$

While the second probability on the right-hand side converges to zero by the induction hypothesis, the first does due to (8). A similar but much simpler argument leads to the conclusion that

$$t^{-\gamma-\omega\ell}\sum_{k\in\mathbb{N}}N_\ell^{(k)}(h)\,\mathbb{1}_{\{T_k\le t+h\}}\ \overset{\mathrm{P}}{\to}\ 0,\quad t\to\infty.$$

The proof of Lemma 3.1 is complete. □

3.2 Connecting two ways of box-counting

We retreat for a while from our main theme to focus on Karlin's occupancy scheme with deterministic probabilities $(p_k)_{k\in\mathbb{N}}$. By the law of large numbers a box of probability p gets occupied by about np balls, provided np is big enough. This suggests to relate counting the boxes occupied by at least n^{1-s} balls to the number of boxes with probability at least n^{-s}. Let $\bar\rho(t) := \#\{k\in\mathbb{N} : p_k\ge 1/t\}$ for $t>0$, and let $\bar K_{n,r}$ be the number of boxes containing exactly r out of n balls. We shall estimate uniformly the difference between

$$\bar K_n(s) := \sum_{r=\lceil n^{1-s}\rceil}^{n}\bar K_{n,r}\,,\quad s\in[0,1],$$

and $(\bar\rho(n^s))_{s\in[0,1]}$. The following result is very close to Proposition 4.1 in [2]. However, we did not succeed to apply the cited proposition directly and will combine the estimates obtained in its proof.

Proposition 3.2. *The following universal estimate holds for each $n\in\mathbb{N}$*

$$\begin{aligned}\mathbb{E}\sup_{s\in[0,1]}\big|\bar K_n(s)-\bar\rho(n^s)\big| \ \le\ & 4\big(\bar\rho(n)-\bar\rho(y_0n(\log n)^{-2})\big)+2\bar\rho(n)(\log n)^{-1} \\ & + \int_1^\infty x^{-2}(\bar\rho(nx)-\bar\rho(n))\mathrm{d}x+2\sup_{s\in[0,1]}(\bar\rho(en^s)-\bar\rho(e^{-1}n^s)),\end{aligned}\tag{18}$$

where $y_0\in(0,1)$ is a constant which does not depend on n, nor on $(p_k)_{k\in\mathbb{N}}$.

Proof. For $k\in\mathbb{N}$, denote by $\bar Z_{n,k}$ the number of balls falling in the kth box, so that

$$\bar K_n(s)=\sum_{k\in\mathbb{N}}\mathbb{1}_{\{n^{1-s}\le\bar Z_{n,k}\le n\}},\ s\in[0,1].$$

Then, for $n\in\mathbb{N}$ and $s\in[0,1]$,

$$\begin{aligned}|\bar K_n(1-s)-\bar\rho(n^{1-s})| \ \le\ & \sum_{k\in\mathbb{N}}\mathbb{1}_{\{\bar Z_{n,k}\ge n^s,\,1\le np_k\le n^s\}}+\sum_{k\in\mathbb{N}}\mathbb{1}_{\{\bar Z_{n,k}\ge n^s,\,np_k<1\}}+\sum_{k\in\mathbb{N}}\mathbb{1}_{\{\bar Z_{n,k}\le n^s,\,np_k\ge n^s\}} \\ :=\ & A_n(s)+B_n(s)+C_n(s).\end{aligned}$$

In [2] it was shown that, for $n\in\mathbb{N}$,

$$\mathbb{E}\sup_{s\in[0,1]}A_n(s)\le 2(\bar\rho(n)-\bar\rho(y_0n(\log n)^{-2}))+\frac{\bar\rho(n)}{\log n}+\sup_{s\in[0,1]}(\bar\rho(en^s)-\bar\rho(n^s))$$

(see [2], pp. 1004–1005) and

$$\mathbb{E}\sup_{s\in[0,1]}C_n(s)\le 2(\bar\rho(n)-\bar\rho(y_0n(\log n)^{-2}))+\frac{\bar\rho(n)}{\log n}+\sup_{s\in[0,1]}(\bar\rho(n^s)-\bar\rho(e^{-1}n^s))$$

(see [2], p. 1006). Finally, for $n \in \mathbb{N}$,

$$\begin{aligned}\mathbb{E}\sup_{s\in[0,1]} B_n(s) &= \mathbb{E}\sum_{k\in\mathbb{N}} \mathbb{1}_{\{\bar{Z}_{n,k}\geq 1,\, np_k<1\}} = \sum_{k\in\mathbb{N}}(1-(1-p_k)^n)\,\mathbb{1}_{\{np_k<1\}} \leq \sum_{k\in\mathbb{N}} np_k\,\mathbb{1}_{\{np_k<1\}} \\ &= \int_{(1,\infty)} \frac{1}{x}\mathrm{d}(\bar{\rho}(nx)-\bar{\rho}(n)) = \int_1^\infty \frac{\bar{\rho}(nx)-\bar{\rho}(n)}{x^2}\mathrm{d}x.\end{aligned}$$

Combining the estimates we arrive at (18) because

$$\sup_{s\in[0,1]}\big|\bar{K}_n(s)-\bar{\rho}(n^s)\big| = \sup_{s\in[0,1]}\big|\bar{K}_n(1-s)-\bar{\rho}(n^{1-s})\big|.$$

□

We apply next Proposition 3.2 to the setting of Theorem 2.1. This result shows that (10) is equivalent to the analogous limit relation with $\rho_j(n^t) = N_j(t\log n)$ replacing $K_{n,j}(t)$.

Proposition 3.3. *Suppose* (7) *and* (9). *Then, for each* $j \in \mathbb{N}$,

$$\frac{\sup_{s\in[0,1]}\big|K_{n,j}(s)-\rho_j(n^s)\big|}{(\log n)^{\gamma+\omega(j-1)}} \overset{\mathrm{P}}{\to} 0, \quad n\to\infty. \tag{19}$$

Proof. Fix any $j \in \mathbb{N}$. By Proposition 3.2, for $n \in \mathbb{N}$,

$$\begin{aligned}&\mathbb{E}\Big(\sup_{s\in[0,1]}\big|K_{n,j}(s)-\rho_j(n^s)\big|\Big|(P_k)_{k\in\mathbb{N}}\Big) \qquad (20)\\ \leq\;& 4\big(\rho_j(n)-\rho_j(y_0 n(\log n)^{-2})\big)+2\rho_j(n)(\log n)^{-1}\\ +\;& \int_1^\infty x^{-2}(\rho_j(nx)-\rho_j(n))\mathrm{d}x + 2\sup_{s\in[0,1]}(\rho_j(en^s)-\rho_j(e^{-1}n^s)).\end{aligned}$$

Recall the notation

$$c_j = \frac{(c\Gamma(\omega+1))^j}{\Gamma(\omega j+1)}, \quad j\in\mathbb{N}$$

and our choice of $\gamma > \omega - \min(1,\varepsilon_1,\varepsilon_2)$. In view of (14),

$$\frac{\mathbb{E}\rho_j(n)}{\log n} = \frac{V_j(\log n)}{\log n} \sim c_j(\log n)^{\omega j-1} = o((\log n)^{\gamma+\omega(j-1)}), \quad n\to\infty. \tag{21}$$

The next step is to show that

$$\mathbb{E}\int_1^\infty x^{-2}(\rho_j(nx)-\rho_j(n))\mathrm{d}x = o((\log n)^{\gamma+\omega(j-1)}), \quad n\to\infty. \tag{22}$$

As a preparation for the proof of (22) we first note that according to (15)

$$\begin{aligned}\mathbb{E}(\rho_j(nx)-\rho_j(n)) &= V_j(\log n+\log x)-V_j(\log n)\\ &\leq c_j(\mathbb{1}_{\{\omega j\in(0,1]\}}(\log x)^{\omega j}+\mathbb{1}_{\{\omega j>1\}}\,\omega j(\log n+\log x)^{\omega j-1}\log x)\\ &+ a_{0,j}+a_{1,j}(\log n+\log x)^{\omega j-\varepsilon_2}-b_{0,j}+|b_{1,j}|(\log n)^{\omega j-\varepsilon_1}\end{aligned}$$

for $n\in\mathbb{N}$ and $x\geq 1$. Further, using the inequality $(u+v)^\alpha \leq (2^{\alpha-1}\wedge 1)(u^\alpha+v^\alpha)$ which holds for $\alpha>0$ and $u,v\geq 0$ yields

$$\int_1^\infty x^{-2}(\log n+\log x)^{\omega j-\varepsilon_2}\mathrm{d}x = O((\log n)^{\omega j-\varepsilon_2}), \quad n\to\infty$$

and

$$\int_1^\infty x^{-2}(\log n+\log x)^{\omega j-1}\mathrm{d}x=O((\log n)^{\omega j-1}),\quad n\to\infty,$$

and (22) follows.

An appeal to (13) enables us to conclude that for large enough n

$$\begin{aligned}
& \mathbb{E}(\rho_j(n)-\mathbb{E}\rho_j(y_0 n(\log n)^{-2}))\\
= \;& V_j(\log n)-V_j(\log n+\log y_0-2\log\log n)\\
\leq \;& c_j(\log n)^{\omega j}\Big(1-\Big(1-\frac{2\log\log n-\log y_0}{\log n}\Big)^{\omega j}\Big)\\
+ \;& a_{0,j}+a_{1,j}(\log n)^{\omega j-\varepsilon_2}-b_{0,j}-b_{1,j}(\log n+\log y_0-2\log\log n)^{\omega j-\varepsilon_1}\\
\leq \;& 4\omega j c_j(\log n)^{\omega j-1}\log\log n+a_{0,j}+a_{1,j}(\log n)^{\omega j-\varepsilon_2}\\
- \;& b_{0,j}+|b_{1,j}|(\log n+\log y_0-2\log\log n)^{\omega j-\varepsilon_1}.
\end{aligned}$$

Hence,

$$\mathbb{E}(\rho_j(n)-\rho_j(y_0 n(\log n)^{-2}))=o((\log n)^{\gamma+\omega(j-1)}),\quad n\to\infty \tag{23}$$

by the same reasoning as above. Finally,

$$\frac{\sup_{s\in[0,1]}(\rho_j(en^s)-\rho_j(e^{-1}n^s))}{(\log n)^{\gamma+\omega(j-1)}}=\frac{\sup_{s\in[0,1]}(N_j(s\log n+1)-N_j(s\log n-1))}{(\log n)^{\gamma+\omega(j-1)}}\overset{\mathrm{P}}{\to}0,\quad n\to\infty \tag{24}$$

by Lemma 3.1(b). Using (21), (22), (23) and (24) in combination with Markov's inequality (applied to the first three terms on the right-hand side of (20)) shows that the left-hand side of (20) converges to zero in probability as $n\to\infty$. Now (19) follows by another application of Markov's inequality and the dominated convergence theorem. □

Our main result, Theorem 2.1, is an immediate consequence of Proposition 3.3 and the next Theorem 3.4 which in turn follows from Propositions 3.5 and 3.6.

Theorem 3.4. *Suppose* (7) *and* (9). *Then*

$$\Big(\frac{N_j(t\cdot)-V_j(t\cdot)}{ac_{j-1}t^{\gamma+\omega(j-1)}}\Big)_{j\in\mathbb{N}}\ \Rightarrow\ (R_j(\cdot))_{j\in\mathbb{N}} \tag{25}$$

in the J_1-topology on $D^{\mathbb{N}}$.

Recalling (4), write, for $j\geq 2$ and $t\geq 0$,

$$\begin{aligned}
N_j(t)-V_j(t) \;&=\; \sum_{k\in\mathbb{N}}\big(N_{j-1}^{(k)}(t-T_k)-V_{j-1}(t-T_k)\big)\\
&+\;\Big(\sum_{k\in\mathbb{N}}V_{j-1}(t-T_k)-V_j(t)\Big)=:X_j(t)+Y_j(t).
\end{aligned} \tag{26}$$

Proposition 3.5. *Suppose* (7) *and* (9). *Then*

$$\Big(\frac{N_1(t\cdot)-V_1(t\cdot)}{at^\gamma},\Big(\frac{Y_j(t\cdot)}{ac_{j-1}t^{\gamma+\omega(j-1)}}\Big)_{j\geq 2}\Big)\ \Rightarrow\ (R_j(\cdot))_{j\in\mathbb{N}},\quad t\to\infty, \tag{27}$$

in the J_1-topology on $D^{\mathbb{N}}$.

Proposition 3.6. *Suppose* (7), (8) *and* (9). *Then, for each $j\geq 2$ and each $T>0$,*

$$t^{-(\gamma+\omega(j-1))}\sup_{y\in[0,\,T]}X_j(ty)\ \overset{\mathrm{P}}{\to}\ 0,\quad t\to\infty. \tag{28}$$

3.3 Proof of Proposition 3.5

We shall use an integral representation similar to (5)

$$Y_j(t) = \sum_{k\in\mathbb{N}} V_{j-1}(t-T_k) - V_j(t) = \int_{[0,\,t]} V_{j-1}(t-y)\mathrm{d}(N_1(y)-V_1(y)) \tag{29}$$

for $j \geq 2$ and $t \geq 0$.

In view of (12) Skorokhod's representation theorem ensures that there exist versions $\widehat{N}_1$ and $\widehat{W}$ such that

$$\lim_{t\to\infty} \sup_{y\in[0,\,T]} \left| \frac{\widehat{N}_1(ty)-V_1(ty)}{at^\gamma} - \widehat{W}(y) \right| = 0 \quad \text{a.s.} \tag{30}$$

for all $T > 0$. This implies that (27) is equivalent to

$$\left(\widehat{W}(\cdot), \left(\frac{\widehat{Z}_j(t,\cdot)}{c_{j-1}t^{\omega(j-1)}} \right)_{j\geq 2} \right) \ \Rightarrow\ (R_j(\cdot))_{j\in\mathbb{N}}, \quad t\to\infty, \tag{31}$$

where we set $\widehat{Z}_j(t,x) := \int_{(0,\,x]} \widehat{W}(y)\mathrm{d}_y(-V_{j-1}(t(x-y))$ for $j \geq 2$ and $t, x \geq 0$. As far as the first coordinate is concerned the equivalence is an immediate consequence of (30). As for the other coordinates, integration by parts yields, for $s > 0$ fixed and $j \geq 2$

$$\begin{aligned} \int_{[0,\,st]} \frac{V_{j-1}(st-y)}{c_{j-1}t^{\omega(j-1)}} \mathrm{d}_y \frac{\widehat{N}_1(y)-V_1(y)}{at^\gamma} &= \int_{(0,\,s]} \left(\frac{\widehat{N}_1(ty)-V_1(ty)}{at^\gamma} - \widehat{W}(y) \right) \mathrm{d}_y \frac{-V_{j-1}(t(s-y))}{c_{j-1}t^{\omega(j-1)}} \\ &+ \int_{(0,\,s]} \widehat{W}(y)\mathrm{d}_y \frac{-V_{j-1}(t(s-y))}{c_{j-1}t^{\omega(j-1)}}. \end{aligned}$$

Observe that the first term is a counterpart of (29) in which $\widehat{N}_1$ replaces N_1. Denoting by $L(t)$ the first term on the right-hand side, we infer

$$|L(t)| \leq \sup_{0\leq y\leq s} \left| \frac{\widehat{N}_1(ty)-V_1(ty)}{at^\gamma} - \widehat{W}(y) \right| \left(\left(c_{j-1}t^{\omega(j-1)}\right)^{-1} V_{j-1}(st) \right) \to 0 \ \text{ a.s. for } t\to\infty$$

in view of (14) which implies that

$$\lim_{t\to\infty} \left(c_{j-1}t^{\omega(j-1)}\right)^{-1} V_{j-1}(st) = s^{\omega(j-1)} \tag{32}$$

and (30).

For $j \geq 2$ and $t, x \geq 0$, set $Z_j(t,x) := \int_{(0,\,x]} W(y)\mathrm{d}_y(-V_{j-1}(t(x-y))$ and note that (31) is equivalent to

$$\left(W(\cdot), \left(\frac{Z_j(t,\cdot)}{c_{j-1}t^{\omega(j-1)}} \right)_{j\geq 2} \right) \ \Rightarrow\ (R_j(\cdot))_{j\in\mathbb{N}}, \quad t\to\infty \tag{33}$$

because the left-hand sides of (31) and (33) have the same distribution.

It remains to check two properties:
(a) weak convergence of finite-dimensional distributions, i.e. that for all $n \in \mathbb{N}$, all $0 \leq s_1 < s_2 < \ldots < s_n < \infty$ and all integer $\ell \geq 2$

$$\left(W(s_i), \left(\frac{Z_j(t,s_i)}{c_{j-1}t^{\omega(j-1)}} \right)_{2\leq j\leq \ell} \right)_{1\leq i\leq n} \ \stackrel{\mathrm{d}}{\to}\ (R_j(s_i))_{1\leq j\leq \ell,\,1\leq i\leq n} \tag{34}$$

as $t\to\infty$;
(b) tightness of the distributions of coordinates in (33), excluding the first one.
PROOF OF (34). If $s_1=0$, we have $W(s_1)=Z_j(t,s_1)=R_k(s_1)=0$ a.s. for $j\geq 2$ and $k\in\mathbb{N}$. Hence, in what follows we consider the case $s_1>0$. Both the limit and the converging vectors in (34) are Gaussian. In view of this it suffices to prove that

$$\lim_{t\to\infty} t^{-\omega(k+j-2)}\mathbb{E}[Z_k(t,s)Z_j(t,u)] = c_{k-1}c_{j-1}\mathbb{E}[R_k(s)R_j(u)] \tag{35}$$
$$= \begin{cases} c_{k-1}c_{j-1}\int_0^s\int_0^u r(s-y,u-z)\mathrm{d}y^{k-1}\mathrm{d}z^{j-1}, & \text{if } k,j\geq 2,\\ c_{j-1}\int_0^u r(s,u-z)\mathrm{d}z^{j-1}, & \text{if } k=1, j\geq 2\end{cases}$$

for $k,j\in\mathbb{N}$, $k+j\geq 3$ and $s,u>0$, where we set $Z_1(t,\cdot)=W(\cdot)$ and $r(x,y):=\mathbb{E}[W(x)W(y)]$ for $x,y\geq 0$. We only consider the case where $k,j\geq 2$, the complementary case being similar and simpler.

To prove (35) we need some preparation. For each $t>0$ denote by $\theta_{k,t}$ and $\theta_{j,t}$ independent random variables with the distribution functions $\mathbb{P}\{\theta_{k,t}\leq y\}=V_{k-1}(ty)/V_{k-1}(ts)$ on $[0,s]$ and $\mathbb{P}\{\theta_{j,t}\leq y\}=V_{j-1}(ty)/V_{j-1}(tu)$ on $[0,u]$, respectively. Further, let θ_k and θ_j denote independent random variables with the distribution functions $\mathbb{P}\{\theta_k\leq y\}=(y/s)^{k-1}$ on $[0,s]$ and $\mathbb{P}\{\theta_j\leq y\}=(y/u)^{j-1}$ on $[0,u]$, respectively. According to (14), $(\theta_{k,t},\theta_{j,t})\overset{\mathrm{d}}{\to}(\theta_k,\theta_j)$ as $t\to\infty$. Now observe that the function $r(x,y)=\mathbb{E}[W(x)W(y)]$ is continuous, hence bounded, on $[0,T]\times[0,T]$ for every $T>0$. This follows from the assumed a.s. continuity of W, the dominated convergence theorem in combination with $\mathbb{E}(\sup_{z\in[0,T]}W(z))^2<\infty$ for every $T>0$ (for the latter, see Theorem 3.2 on p. 63 in [1]). As a result, $r(s-\theta_{k,t},u-\theta_{j,t})\overset{\mathrm{d}}{\to}r(s-\theta_k,u-\theta_j)$ as $t\to\infty$ and thereupon

$$\lim_{t\to\infty}\mathbb{E}r(s-\theta_{k,t},u-\theta_{j,t})=\mathbb{E}r(s-\theta_k,u-\theta_j)$$

by the dominated convergence theorem.

This together with (32) leads to formula (35):

$$\begin{aligned}
&\mathbb{E}[t^{-\omega(k+j-2)}Z_k(t,s)Z_j(t,u)]\\
=\ &\frac{V_{k-1}(ts)}{t^{\omega(k-1)}}\frac{V_{j-1}(tu)}{t^{\omega(j-1)}}\int_0^s\int_0^u r(s-y,u-z)\mathrm{d}_y\Big(\frac{V_{k-1}(ty)}{V_{k-1}(ts)}\Big)\mathrm{d}_z\Big(\frac{V_{j-1}(tz)}{V_{j-1}(tu)}\Big)\\
=\ &\frac{V_{k-1}(ts)}{t^{\omega(k-1)}}\frac{V_{j-1}(tu)}{t^{\omega(j-1)}}\mathbb{E}r(s-\theta_{k,t},u-\theta_{j,t})\ \to\ c_{k-1}s^{k-1}c_{j-1}s^{j-1}\mathbb{E}r(s-\theta_k,u-\theta_j)\\
=\ &c_{k-1}c_{j-1}\int_0^s\int_0^u r(s-y,u-z)\mathrm{d}y^{k-1}\mathrm{d}z^{j-1}
\end{aligned}$$

as $t\to\infty$.
PROOF OF TIGHTNESS. Choose $j\geq 2$. We intend to prove tightness of $(t^{-\omega(j-1)}Z_j(t,u))_{u\geq 0}$ on $D[0,T]$ for all $T>0$. Since the function $t\mapsto t^{-\omega(j-1)}$ is regularly varying at ∞ it is enough to investigate the case $T=1$ only. By Theorem 15.5 in [8] it suffices to show that for any $\kappa_1>0$ and $\kappa_2>0$ there exist $t_0>0$ and $\delta>0$ such that

$$\mathbb{P}\Big\{\sup_{0\leq u,v\leq 1,|u-v|\leq\delta}|Z_j(t,u)-Z_j(t,v)|>\kappa_1 t^{-\omega(j-1)}\Big\}\leq\kappa_2 \tag{36}$$

for all $t\geq t_0$. We only analyze the case where $0\leq v<u\leq 1$, the complementary case being analogous.

Set $W(x)=0$ for $x<0$. The basic observation for the subsequent proof is that (6) extends to

$$|W(x)-W(y)|\le M(x-y)^\beta \tag{37}$$

whenever $-\infty<y<x\le 1$ for the same positive random variable M as in (6). This is trivial when $x\le 0$ and a consequence of (6) when $y\ge 0$. Assume that $y\le 0<x$. Then $|W(x)-W(y)|=|W(x)|\le Mx^\beta\le M(x-y)^\beta$, where the first inequality follows from (6) with $y=0$.

Let $0\le v<u\le 1$ and $u-v\le\delta$ for some $\delta\in(0,1]$. Using (37) and (14) we obtain

$$\begin{aligned} t^{-\omega(j-1)}|Z_j(t,u)-Z_j(t,v)| &= t^{-\omega(j-1)}\Big|\int_{[0,\,u)}\big(W(u-y)-W(v-y)\big)\mathrm{d}V_{j-1}(ty)\Big| \\ &\le M(u-v)^\beta(t^{-\omega(j-1)}V_{j-1}(t))\le M\delta^\beta\lambda \end{aligned}$$

for large enough t and a positive constant λ. This proves (36).

3.4 Proof of Proposition 3.6

It suffices to check that for $j\ge 2$

$$\lim_{t\to\infty}t^{-(\gamma+\omega(j-1))}X_j(t)=0\quad\text{a.s.} \tag{38}$$

To this end, we pick $\delta\in(1,\min(1,\varepsilon_1,\varepsilon_2)/(\omega-\gamma))$ which is possible because $\min(1,\varepsilon_1,\varepsilon_2)>\omega-\gamma>0$ by assumption. Note that for each $t\ge 0$, there exists $\ell\in\mathbb{N}_0$ such that $t\in[\ell^\delta,(\ell+1)^\delta)$ and use a.s. monotonicity of N_{j-1} and monotonicity of V_{j-1} to obtain

$$\begin{aligned} t^{-(\gamma+\omega(j-1))}X_j(t) &\le \ell^{-\delta(\gamma+\omega(j-1))}\sum_{k\in\mathbb{N}}\big(N_{j-1}^{(k)}((\ell+1)^\delta-T_k)-V_{j-1}((\ell+1)^\delta-T_k)\big)\,\mathbb{1}_{\{T_k\le(\ell+1)^\delta\}} \\ &+ \ell^{-\delta(\gamma+\omega(j-1))}\sum_{k\in\mathbb{N}}\big(V_{j-1}((\ell+1)^\delta-T_k)-V_{j-1}(\ell^\delta-T_k)\big)\,\mathbb{1}_{\{T_k\le\ell^\delta\}} \\ &+ \ell^{-\delta(\gamma+\omega(j-1))}\sum_{k\in\mathbb{N}}V_{j-1}((\ell+1)^\delta-T_k)\,\mathbb{1}_{\{\ell^\delta<T_k\le(\ell+1)^\delta\}} \\ &\le \ell^{-\delta(\gamma+\omega(j-1))}X_j((\ell+1)^\delta) \\ &+ \ell^{-\delta(\gamma+\omega(j-1))}\Big(\sum_{k\in\mathbb{N}}\big(V_{j-1}((\ell+1)^\delta-T_k)-V_{j-1}(\ell^\delta-T_k)\big)\,\mathbb{1}_{\{T_k\le\ell^\delta\}} \\ &+ V_{j-1}((\ell+1)^\delta-\ell^\delta)N((\ell+1)^\delta)\Big). \end{aligned} \tag{39}$$

Similarly,

$$\begin{aligned} t^{-(\gamma+\omega(j-1))}X_j(t) &\ge (\ell+1)^{-\delta(\gamma+\omega(j-1))}X_j(\ell^\delta) \\ &- (\ell+1)^{-\delta(\gamma+\omega(j-1))}\Big(\sum_{k\in\mathbb{N}}\big(V_{j-1}((\ell+1)^\delta-T_k)-V_{j-1}(\ell^\delta-T_k)\big)\,\mathbb{1}_{\{T_k\le\ell^\delta\}} \\ &+ V_{j-1}((\ell+1)^\delta-\ell^\delta)N((\ell+1)^\delta)\Big). \end{aligned} \tag{40}$$

We intend to show that the second and the third summands on the right-hand side of (39) (hence, of (40)) converge to zero a.s. as $\ell\to\infty$.

THE 3RD SUMMAND. Recall the notation

$$c_k = \frac{(c\Gamma(\omega+1))^k}{\Gamma(\omega k+1)}, \quad k \in \mathbb{N}_0.$$

By (8) and (14), $N(\ell) \sim c_1 \ell^{\omega}$ a.s. and, for $j \in \mathbb{N}$, $V_j(\ell) \sim c_j \ell^{\omega j}$ as $\ell \to \infty$, respectively, whence, as $\ell \to \infty$,

$$\ell^{-\delta(\gamma+\omega(j-1))} V_{j-1}((\ell+1)^{\delta} - \ell^{\delta}) N((\ell+1)^{\delta}) \ \sim\ \frac{c_1 c_{j-1} \delta^{\omega(j-1)}}{\ell^{\omega j-(\omega+(\omega-\gamma)\delta)}} \quad \text{a.s.} \tag{41}$$

This converges to zero as $\ell \to \infty$ because, for $j \geq 2$, $\omega j - (\omega + (\omega+\gamma)\delta) \geq \omega - (\omega-\gamma)\delta > 0$. The positivity is a consequence of $\delta < \min(1, \varepsilon_1, \varepsilon_2)/(\omega-\gamma) \leq \varepsilon_1/(\omega-\gamma) \leq \omega/(\omega-\gamma)$, where the last inequality holds by our choice of ω in Theorem 2.1.

THE 2ND SUMMAND. Using (15) with $u = (\ell+1)^{\delta} - \ell^{\delta}$ and $v = \ell^{\delta} - T_k$ on the event $\{T_k \leq \ell^{\delta}\}$ yields

$$\begin{aligned}
& \sum_{k\in\mathbb{N}} \big(V_{j-1}((\ell+1)^{\delta} - T_k) - V_{j-1}(\ell^{\delta} - T_k)\big) \mathbb{1}_{\{T_k \leq \ell^{\delta}\}} \\
\leq\ & \sum_{k\in\mathbb{N}} \big(c_{j-1}(\mathbb{1}_{\{\omega(j-1)\in(0,1]\}}((\ell+1)^{\delta} - \ell^{\delta})^{\omega(j-1)} \\
+\ & \mathbb{1}_{\{\omega(j-1)>1\}}\, \omega(j-1)((\ell+1)^{\delta} - T_k)^{\omega(j-1)-1}((\ell+1)^{\delta} - \ell^{\delta})) \\
+\ & a_{0,j-1} + a_{1,j-1}((\ell+1)^{\delta} - T_k)^{\omega(j-1)-\varepsilon_2} - b_{0,j-1} - b_{1,j-1}(\ell^{\delta} - T_r)^{\omega(j-1)-\varepsilon_1}\big) \mathbb{1}_{\{T_k \leq \ell^{\delta}\}} \\
\leq\ & N(\ell^{\delta})\big(\, \mathbb{1}_{\{\omega(j-1)\in(0,1]\}}\, c_{j-1}(((\ell+1)^{\delta} - \ell^{\delta})^{\omega(j-1)} \\
+\ & \mathbb{1}_{\{\omega(j-1)>1\}}\, \omega(j-1)(\ell+1)^{\delta(\omega(j-1)-1)}((\ell+1)^{\delta} - \ell^{\delta})) \\
+\ & a_{0,j-1} + a_{1,j-1}(\ell+1)^{\delta\omega(j-1)-\varepsilon_2} - b_{0,j-1} + |b_{1,j-1}|\ell^{\delta\omega(j-1)-\varepsilon_1}\big).
\end{aligned}$$

As was explained right after formula (41) the right-hand side of

$$\ell^{-\delta(\gamma+\omega(j-1))} N(\ell^{\delta})((\ell+1)^{\delta} - \ell^{\delta})^{\omega(j-1)} \ \sim\ \frac{c_1 \delta^{\omega(j-1)}}{\ell^{\omega j-(\omega+(\omega-\gamma)\delta)}} \quad \text{a.s.}$$

converges to zero as $\ell \to \infty$. Further, for $i = 1, 2$,

$$\ell^{-\delta(\gamma+\omega(j-1))} N(\ell^{\delta}) \ell^{\delta\omega(j-1)-\varepsilon_i} \ \sim\ \frac{c_1}{\ell^{\varepsilon_i - \delta(\omega-\gamma)}} \quad \text{a.s.}$$

and

$$\ell^{-\delta(\gamma+\omega(j-1))} N(\ell^{\delta})(\ell+1)^{\delta(\omega(j-1)-1)}((\ell+1)^{\delta} - \ell^{\delta}) \ \sim\ \frac{c_1 \delta}{\ell^{1-\delta(\omega-\gamma)}}.$$

The right-hand sides of the last two centered formulas converge to zero as $\ell \to \infty$ by the choice of δ. Thus, we have proved that (38) is a consequence of

$$\lim_{\ell\to\infty} \ell^{-\delta(\gamma+\omega(j-1))} X_j(\ell^{\delta}) \ =\ 0 \quad \text{a.s.} \tag{42}$$

Relation (42) will be proved by induction in two steps.

STEP 1. Assume (42), hence (38) and (28) hold for $j = 2, \ldots, k$. We claim that

$$\left(\frac{N_j(t\cdot) - V_j(t\cdot)}{a c_{j-1} t^{\gamma+\omega(j-1)}}\right)_{j=1,\ldots,k} \ \Rightarrow\ (R_j(\cdot))_{j=1,\ldots k} \tag{43}$$

in the J_1-topology on D^k. Indeed, in view of (26) and the induction hypothesis relation (43) is equivalent to

$$\left(\frac{N_1(t\cdot)-V_1(t\cdot)}{at^\gamma},\left(\frac{Y_j(t\cdot)}{ac_{j-1}t^{\gamma+\omega(j-1)}}\right)_{j=2,\ldots,k}\right) \Rightarrow (R_j(\cdot))_{j=1,\ldots,k}, \quad t\to\infty. \tag{44}$$

The latter holds by Proposition 3.5.

Step 2. Using

$$\frac{N_k(t\cdot)-V_k(t\cdot)}{ac_{k-1}t^{\gamma+\omega(k-1)}} \Rightarrow R_k(\cdot) \tag{45}$$

in the J_1-topology on D which is a consequence of (43) we shall prove that (42) holds with $j=k+1$.

In view of (45) and the fact that R_k is a.s. continuous Skorokhod's representation theorem ensures that there exists a probability space which is rich enough to accomodate

- independent random processes $\widehat N_k^{(1)}$, $\widehat N_k^{(2)},\ldots$ which are versions of N_k;
- independent random processes $\widehat R_k^{(1)}$, $\widehat R_k^{(2)},\ldots$ which are versions of R_k;
- random variables $\widehat T_1$, $\widehat T_2,\ldots$ which are versions of T_1, $T_2,\ldots$ independent of $(\widehat N_k^{(1)},\widehat R_k^{(1)})$, $(\widehat N_k^{(2)},\widehat R_k^{(2)}),\ldots$

Furthermore,

$$\lim_{t\to\infty}\sup_{y\in[0,\,T]}\left|\frac{\widehat N_k^{(r)}(ty)-V_k(ty)}{ac_{k-1}t^{\gamma+\omega(k-1)}}-\widehat R_k^{(r)}(y)\right|=0 \quad \text{a.s.} \tag{46}$$

for all $T>0$ and $r\in\mathbb{N}$.

Set

$$\widehat X_{k+1}(t):=\sum_{r\in\mathbb{N}}\big(\widehat N_k^{(r)}(t-\widehat T_r)-V_k(t-\widehat T_r)\big)\,\mathbb{1}_{\{\widehat T_r\le t\}}, \quad t\ge 0.$$

The process $(\widehat X_{k+1}(t))_{t\ge0}$ has the same distribution as $(X_{k+1}(t))_{t\ge0}$. Therefore, (42) with $j=k+1$ is equivalent to

$$\lim_{\ell\to\infty}\ell^{-\delta(\gamma+\omega k)}\widehat X_{k+1}(\ell^\delta) \;=\; 0 \quad \text{a.s.} \tag{47}$$

To prove this, write

$$\begin{aligned}\ell^{-\delta(\gamma+\omega k)}\widehat X_{k+1}(\ell^\delta) \;&=\; \ell^{-\delta\omega}\sum_{r\in\mathbb{N}}\Big(\frac{\widehat N_k^{(r)}(\ell^\delta-\widehat T_r)-V_k(\ell^\delta-\widehat T_r)}{ac_{k-1}\ell^{\delta(\gamma+\omega(k-1))}}-\widehat R_k^{(r)}(1-\ell^{-\delta}\widehat T_r)\Big)\,\mathbb{1}_{\{\widehat T_r\le\ell^\delta\}}\\ &+\; \ell^{-\delta\omega}\sum_{r\in\mathbb{N}}\widehat R_k^{(r)}(1-\ell^{-\delta}\widehat T_r)\,\mathbb{1}_{\{\widehat T_r\le\ell^\delta\}} =: Z_1(\ell)+Z_2(\ell).\end{aligned}$$

For all $T>0$,

$$\begin{aligned}&t^{-\omega}\sup_{y\in[0,\,T]}\Big|\sum_{r\in\mathbb{N}}\Big(\frac{\widehat N_k^{(r)}(ty-\widehat T_r)-V_k(ty-\widehat T_r)}{ac_{k-1}t^{\gamma+\omega(k-1)}}-\widehat R_k^{(r)}(y-T_r/t)\Big)\,\mathbb{1}_{\{\widehat T_r\le ty\}}\Big|\\ \le\;& t^{-\omega}\sum_{r\in\mathbb{N}}\sup_{y\in[0,\,T]}\Big|\frac{\widehat N_k^{(r)}(ty)-V_k(ty)}{ac_{k-1}t^{\gamma+\omega(k-1)}}-\widehat R_k^{(r)}(y)\Big|\,\mathbb{1}_{\{\widehat T_r\le tT\}}\,.\end{aligned} \tag{48}$$

In view of (8) and (46) we can argue as in the proof of Lemma 3.1 (b) to conclude that the right-hand side of (48) converges to zero in probability as $t \to \infty$, whence $\lim_{\ell\to\infty} Z_1(\ell) = 0$ a.s.

Pick now $n \in \mathbb{N}$ such that $n\omega\delta > 1$. To avoid considering separately a simpler case $n = 1$ we assume in what follows that $n \geq 2$. Since R_k is a centered Gaussian process which is self-similar of exponent $\gamma + \omega(k-1)$ we infer, for $j \in \mathbb{N}$, $\mathbb{E}[R_k(y)]^{2j} = \mathbb{E}[R_k(1)]^{2j} y^{2j(\gamma+\omega(k-1))} =: m_{2j} y^{2j(\gamma+\omega(k-1))}$. By Rosenthal's inequality (Theorem 3 in [34])

$$\begin{aligned}
&\mathbb{E}\Big(\Big(\sum_{r\in\mathbb{N}} \widehat{R}_k^{(r)}(1-\ell^{-\delta}\widehat{T}_r)\,\mathbb{1}_{\{\widehat{T}_r \leq \ell^{\delta}\}}\Big)^{2n}\Big|(\widehat{T}_r)_{r\in\mathbb{N}}\Big) \\
\leq\;& C\Big(\Big(m_2\int_{[0,\,\ell^{\delta}]}(1-\ell^{-\delta}x)^{2(\gamma+\omega(k-1))}\mathrm{d}N(x)\Big)^{n} \\
+\;& m_{2n}\int_{[0,\,\ell^{\delta}]}(1-\ell^{-\delta}x)^{2n(\gamma+\omega(k-1))}\mathrm{d}N(x)\Big) \leq C(m_2^n N^n(\ell^{\delta}) + m_{2n}N(\ell^{\delta}))
\end{aligned}$$

for a positive constant C. By Markov's inequality, for all $\kappa > 0$,

$$\sum_{\ell\in\mathbb{N}} \mathbb{P}\{Z_2(\ell) > \kappa | (\widehat{T}_k)_{k\in\mathbb{N}}\} \leq C\kappa^{-2n}\sum_{\ell\in\mathbb{N}} \ell^{-2n\omega\delta}(m_2^n N^n(\ell^{\delta}) + m_{2n}N(\ell^{\delta})).$$

In view of (8), the general term of the series is $O(\ell^{-n\omega\delta})$ a.s. Hence, our choice of n ensures that the series converges a.s. thereby proving (with the help of the Borel-Cantelli lemma) that $\lim_{\ell\to\infty} Z_2(\ell) = 0$ a.s. conditionally on $(\widehat{T}_k)_{k\in\mathbb{N}}$, hence, also unconditionally. The proof of (47) is complete.

4 The case of homogeneous residual allocation model

In this section we apply Theorem 2.1 to the case of fragmentation law given by homogeneous residual allocation model (1). Let $B := (B(s))_{s\geq 0}$ be a standard Brownian motion (BM) and for $q \geq 0$ let

$$B_q(s) := \int_{[0,\,s]} (s-y)^q \mathrm{d}B(y), \quad s \geq 0.$$

The process $B_q := (B_q(s))_{s\geq 0}$ is a centered Gaussian process called the fractionally integrated BM or the Riemann-Liouville process. Clearly $B = B_0$, and for $q \in \mathbb{N}$ the process can be obtained as a repeated integral of the BM. It is known that B_q is locally Hölder continuous with any exponent $\beta < q + 1/2$ [23].

Theorem 4.1. *Let $(P_k)_{k\in\mathbb{N}}$ be given by (1) with iid U_i's such that*

$$\mu := \mathbb{E}|\log U_1| < \infty,\; \sigma^2 := \mathrm{Var}(\log U_1) \in (0,\infty)$$

and $\mathbb{E}|\log(1-U_1)|^d < \infty$ for some $d > 1/2$. Then

$$\Big(\frac{(j-1)!(K_{n,j}(\cdot) - (j!)^{-1}(\mu^{-1}\log n(\cdot))^j)}{\sqrt{\sigma^2\mu^{-2j-1}(\log n)^{2j-1}}}\Big)_{j\in\mathbb{N}} \;\Rightarrow\; (B_{j-1}(\cdot))_{j\in\mathbb{N}}, \quad n\to\infty$$

in the product J_1-topology on $D[0,1]^{\mathbb{N}}$.

Proof. Let $(\xi_k, \eta_k)_{k\in\mathbb{N}}$ be independent copies of a random vector (ξ, η) with positive arbitrarily dependent components. Denote by $(S_k)_{k\in\mathbb{N}_0}$ the zero-delayed ordinary random walk with increments ξ_k, that is, $S_0 := 0$ and $S_k := \xi_1 + \ldots + \xi_k$ for $k \in \mathbb{N}$. Consider a *perturbed* random walk

$$\tilde{T}_k := S_{k-1} + \eta_k, \quad k \in \mathbb{N} \tag{49}$$

and then define $\tilde{N}(t) := \#\{k \in \mathbb{N} : \tilde{T}_k \le t\}$ and $\tilde{V}(t) := \mathbb{E}\tilde{N}(t)$ for $t \ge 0$. It is clear that

$$\tilde{V}(t) = \mathbb{E}U((t-\eta)^+) = \int_{[0,\,t]} U(t-y)\mathrm{d}\tilde{G}(y), \quad t \ge 0 \tag{50}$$

where, for $t \ge 0$, $U(t) := \sum_{k\ge 0} \mathbb{P}\{S_k \le t\}$ is the renewal function and $\tilde{G}(t) = \mathbb{P}\{\eta \le t\}$.

For P_k written as (1), $T_k = -\log P_k$ becomes

$$T_k = |\log U_1| + \ldots + |\log U_{k-1}| + |\log(1-U_k)|, \quad k \in \mathbb{N}$$

which is a particular case of (49) with $(\xi, \eta) = (|\log U_1|, |\log(1-U_1)|)$. In view of this and Lemma 4.2 given below, the conditions of Theorem 2.1 hold with $\omega = 1$, $\varepsilon_2 = 1$, $\varepsilon_1 = \min(s, 1)$, $\gamma = 1/2$, $c = \mu^{-1}$, $W = B$ and $R_j = B_{j-1}$. $\square$

Lemma 4.2. (a) *Assume that* $\mathtt{m} := \mathbb{E}\xi < \infty$, $\mathtt{s}^2 := \mathrm{Var}\,\xi < \infty$ *and* $\mathbb{E}\eta^d < \infty$ *for some* $d > 1/2$. *Then*

$$b_1 t^{1-\min(d,1)} \le \tilde{V}(t) - \mathtt{m}^{-1}t \le a_0, \quad t \ge 0 \tag{51}$$

for some constants $b_1 < 0$ *and* $a_0 > 0$. *Also,*

$$\frac{\tilde{N}(t\cdot) - \mathtt{m}^{-1}(t\cdot)}{(\mathtt{s}^2\mathtt{m}^{-3}t)^{1/2}} \;\Rightarrow\; B(\cdot), \quad t \to \infty$$

in the J_1*-topology on* D.

(b) *If* $\mathtt{m} < \infty$ *then* $\lim_{t\to\infty}(\tilde{N}(t)/t) = 1/\mathtt{m}$ *a.s.*

Proof. (a) A standard result of the renewal theory tells us that

$$0 \le U(t) - \mathtt{m}^{-1}t \le a_0, \tag{52}$$

where a_0 is a known positive constant. The second inequality in combination with $\tilde{V}(t) \le U(t)$ proves the second inequality in (51). Using the first inequality in (52) yields

$$\begin{aligned}\tilde{V}(t) - \mathtt{m}^{-1}t &= \int_{[0,\,t]} (U(t-y) - \mathtt{m}^{-1}(t-y))\mathrm{d}\tilde{G}(y) \\ &- \mathtt{m}^{-1}\int_0^t (1-\tilde{G}(y))\mathrm{d}y \ge -\mathtt{m}^{-1}\int_0^t (1-\tilde{G}(y))\mathrm{d}y.\end{aligned}$$

To obtain the first inequality in (51) it remains to note that $\lim_{t\to\infty} t^{\min(d,1)}(1-\tilde{G}(t)) = 0$ by assumption. For a proof of weak convergence see Theorem 3.2 in [2].

(b) Set $\nu(t) := \sum_{k\ge 0} \mathbb{1}_{\{S_k \le t\}}$ for $t \ge 0$. For $t > 0$ and $y \in (0, t)$, the following inequalities hold with probability one

$$\nu(t-y) - \sum_{k=1}^{\nu(t)} \mathbb{1}_{\{\eta_k > y\}} = \sum_{k=1}^{\nu(t)} \mathbb{1}_{\{S_{k-1} \le t-y\}} - \sum_{k=1}^{\nu(t)} \mathbb{1}_{\{\eta_k > y\}} \le \sum_{k=1}^{\nu(t)} \mathbb{1}_{\{S_{k-1}+\eta_k \le t\}} = \tilde{N}(t) \le \nu(t). \tag{53}$$

By the strong law of large numbers for ordinary random walks $\lim_{n\to\infty} n^{-1}\sum_{k=1}^{n} \mathbb{1}_{\{\eta_k>y\}} = \mathbb{E}\,\mathbb{1}_{\{\eta>y\}} = \mathbb{P}\{\eta > y\}$ a.s. Since $\lim_{t\to\infty} \nu(t) = \infty$ a.s., it follows that $\lim_{t\to\infty} \sum_{k=1}^{\nu(t)} \mathbb{1}_{\{\eta_k>y\}} / \nu(t) = \mathbb{P}\{\eta > y\}$ a.s. Recall that $\lim_{t\to\infty} t^{-1}\nu(t) = \mathtt{m}^{-1}$ a.s. by the strong law of large numbers for renewal processes, whence

$$\frac{\sum_{k=1}^{\nu(t)} \mathbb{1}_{\{\eta_k>y\}}}{t} = \frac{\sum_{k=1}^{\nu(t)} \mathbb{1}_{\{\eta_k>y\}}}{\nu(t)} \frac{\nu(t)}{t} \to \frac{\mathbb{P}\{\eta > y\}}{\mathtt{m}} \quad \text{a.s.}$$

as $t \to \infty$. Hence, using (53) we infer that

$$\mathtt{m}^{-1} - \mathtt{m}^{-1}\mathbb{P}\{\eta > y\} \le \liminf_{t\to\infty} t^{-1}\tilde{N}(t) \le \limsup_{t\to\infty} t^{-1}\tilde{N}(t) \le \mathtt{m}^{-1} \quad \text{a.s.}$$

Letting $y \to \infty$ gives $\lim_{t\to\infty} t^{-1}\tilde{N}(t) = \mathtt{m}^{-1}$ a.s.

□

Recall that $(P_k)_{k\in\mathbb{N}}$ follows the GEM distribution with parameter $\theta > 0$ when U_i's in (1) are beta$(\theta, 1)$-distributed, in which case $\mu = \mathbb{E}|\log U_1| = \theta^{-1}, \sigma^2 = \mathrm{Var}(\log U_1) = \theta^{-2}$ and we can take $d = 1$.

Corollary 4.3. *For $\theta > 0$ let $(P_k)_{k\in\mathbb{N}}$ be* GEM*-distributed with parameter θ, or any random sequence such that the sequence of P_k's arranged in decreasing order follows the* PD *distribution with parameter θ. Then*

$$\left(\frac{(j-1)!(K_{n,j}(\cdot) - (j!)^{-1}(\theta\log n(\cdot))^j)}{\sqrt{(\theta\log n)^{2j-1}}}\right)_{j\in\mathbb{N}} \Rightarrow (B_{j-1}(\cdot))_{j\in\mathbb{N}}, \quad n \to \infty. \tag{54}$$

in the product J_1-topology on $D[0,1]^{\mathbb{N}}$.

5 Some regenerative models

For $(X(t))_{t\ge0}$ a drift-free subordinator with $X(0) = 0$ and a nonzero Lévy measure ν supported by $(0, \infty)$ let

$$\Delta X(t) = X(t) - X(t-),\ t \ge 0,$$

be the associated process of jumps. The process $\Delta X(\cdot)$ assumes nonzero values on a countable set, which is dense in case $\nu(0, \infty) = \infty$. The transformed process (multiplicative subordinator) $F(t) = 1 - e^{-X(t)}$, $t \ge 0$, has the associated process of jumps

$$\Delta F(t) = e^{-X(t-)}(1 - e^{-\Delta X(t)}),\ t \ge 0.$$

In this section we identify the fragmentation law $(P_k)_{k\in\mathbb{N}}$ with nonzero jumps $\Delta F(\cdot)$arranged in some order (for instance by decrease). Note that mutlypling the Lévy measure by a positive factor corresponds to a time-change for F, hence does not affect the derived fragmentation law.

We shall assume that the Lévy measure ν is infinite and has the right tail $\nu([x, \infty))$ satisfying

$$\beta_0 + \beta_1|\log x|^{q-r_2} \le \nu([x, \infty)) - c_0|\log x|^q \le \alpha_0 + \alpha_1|\log x|^{q-r_1} \tag{55}$$

for small enough $x > 0$ and some $q, c_0, \alpha_0, \alpha_1 > 0$, $1/2 < r_1, r_2 \le q + 1$ and $\beta_0, \beta_1 < 0$.

Theorem 5.1. *Assume that* (55) *holds and*

$$m := \mathbb{E}X(1) = \int_{[0,\infty)} x\nu(\mathrm{d}x) < \infty, \quad s^2 := \mathrm{Var}\, X(1) = \int_{[0,\infty)} x^2\nu(\mathrm{d}x) < \infty.$$

Then

$$\Big(\frac{K_{n,j}(\cdot) - c_j^*(\log n(\cdot))^{(q+1)j}}{q\mathrm{B}(q,j)sm^{-3/2}c_{j-1}^*(\log n)^{(q+1)j-1/2}}\Big)_{j\in\mathbb{N}} \ \Rightarrow \ (B_{j+q-1}(\cdot))_{j\in\mathbb{N}}, \quad n\to\infty$$

in the product J_1-topology on $D[0,1]^{\mathbb{N}}$, where

$$c_j^* := \frac{c_0\Gamma(q+2)}{m(q+1)\Gamma((q+1)j+1)}, \quad j\in\mathbb{N}.$$

Theorem 5.1 applies to the gamma subordinator with the Lévy measure

$$\nu(\mathrm{d}x) = \theta x^{-1}e^{-\lambda x}\, \mathbb{1}_{(0,\infty)}(x)\mathrm{d}x$$

and to the subordinator with

$$\nu(\mathrm{d}x) = \theta(1-e^{-x})^{-1}e^{-\lambda x}\, \mathbb{1}_{(0,\infty)}(x)\mathrm{d}x,$$

where $\theta,\lambda>0$. In both cases $s^2<\infty$ and (55) holds with $c_0=q=r_1=r_2=1$.

Theorem 5.1 is a consequence of Theorem 2.1, the easily checked formula

$$\int_{[0,\,u]}(u-y)^{j-1}\mathrm{d}B_q(y) = q\mathrm{B}(q,j)\int_{[0,\,u]}(u-y)^{j+q-1}\mathrm{d}B(y), \quad u\geq 0,\ j\in\mathbb{N},\ q>0$$

and the next lemma.

Lemma 5.2. *Assume that* (55) *holds and $s^2<\infty$. Then the following is true:*

(a)

$$b_0 + b_1t^{q-\min(r_2-1,0)} \leq V(t) - c_0(m(q+1))^{-1}t^{q+1} \leq a_0 + a_1t^{q-\min(r_1-1,0)}, \quad t>0 \quad (56)$$

for some constants a_0, $a_1>0$ and b_0, $b_1\leq 0$, where $m=\mathbb{E}X(1)<\infty$,

(b)

$$\lim_{t\to\infty}\frac{N(t)}{t^{q+1}} = \frac{c_0}{m(q+1)} \quad \text{a.s.},$$

(c)

$$\frac{N(t\cdot) - c_0(m(q+1))^{-1}(t\cdot)^{q+1}}{sm^{-3/2}t^{q+1/2}} \ \Rightarrow \ B_q(\cdot), \quad t\to\infty$$

in the J_1-topology on D.

Proof. (a) Set $f(x) := \nu([-\log(1-e^{-x}),\infty))$ for $x\geq 0$. Inequality (55) in combination with $\lim_{x\to\infty}\nu([x,\infty))=0$ entails

$$\beta_0+\beta_1x^{q-r_2} \leq f(x) - c_0x^q \leq \alpha_0+\alpha_1x^{q-r_1} \quad (57)$$

for all $x>0$ and some constants α_0, α_1, β_0 and β_1 which are not necessarily the same as in (55).

Since

$$N(t)=\sum \mathbb{1}_{\{X(s-)-\log(1-e^{-\Delta X(s)})\leq t\}}=\sum \mathbb{1}_{\{\Delta X(s)\geq -\log(1-e^{-(t-X(s-))})\}},$$

where the summation extends to all $s>0$ with $\Delta X(s)>0$, we conclude that $V(x)=\mathbb{E}N(x)=\int_{[0,\,x]}f(x-y)\mathrm{d}U^*(y)$, where $U^*(x):=\int_0^\infty \mathbb{P}\{X(t)\leq x\}\mathrm{d}x=\mathbb{E}T(x)$ is the renewal function and $T(x):=\inf\{t>0: X(t)>x\}$ for $x\geq 0$.

Similarly to (52) we have

$$0\leq U^*(t)-m^{-1}t\leq a_0^*,\quad t\geq 0, \tag{58}$$

where a_0^* is a known positive constant. Using this and (57) we infer

$$\begin{aligned} V(t)-c_0(m(q+1))^{-1}t^{q+1} &= \int_{[0,\,t]}(U^*(t-y)-m^{-1}(t-y))\mathrm{d}f(y) \\ &+ m^{-1}\int_0^t (f(y)-c_0y^q)\mathrm{d}y\leq a_0^*f(t)+m^{-1}\int_0^t(\alpha_0+\alpha_1y^{q-r_1})\mathrm{d}y \\ &\leq a_0(\alpha_0+\alpha_1t^{q-r_1}+c_0t^q)+m^{-1}(\alpha_0t+\alpha_1(q-r_1+1)^{-1}t^{q-r_1+1}). \end{aligned}$$

This proves the second inequality in (56). Arguing analogously we obtain

$$\begin{aligned} V(t)-c_0(m(q+1))^{-1}t^{q+1} &\geq m^{-1}\int_0^t(f(y)-c_0y^q)\mathrm{d}y\geq m^{-1}\int_0^t(\beta_0+\beta_1y^{q-r_2})\mathrm{d}y \\ &\geq m^{-1}(\beta_0t+\beta_1(q-r_2+1)^{-1}t^{q+1-r_2}), \end{aligned}$$

thereby proving the first inequality in (56).
(b) Write

$$\begin{aligned} N(t) &= \sum \big(\mathbb{1}_{\{\Delta X(s)\geq -\log(1-e^{-(t-X(s-))})\}}-f(t-X(s-))\big)\,\mathbb{1}_{\{X(s-)\leq t\}} \\ &+ \sum f(t-X(s-)))\,\mathbb{1}_{\{X(s-)\leq t\}}=: N_1(t)+N_2(t). \end{aligned} \tag{59}$$

We intend to prove that

$$t^{-q-1}N_2(t)=t^{-q-1}\int_{[0,\,t]}f(t-y)\mathrm{d}T(y)=(m(q+1))^{-1}\quad \text{a.s.} \tag{60}$$

and

$$\lim_{t\to\infty}t^{-q-1/2}N_1(t)=0\quad \text{a.s.} \tag{61}$$

as $t\to\infty$ (this is more than we need for the present proof; however, this limit relation is an essential ingredient in the proof of part (c)).
Proof of (60). Inequality (56) tells us that it suffices to show that, for all $\kappa>0$,

$$\lim_{t\to\infty}t^{-\kappa-1}\int_{[0,\,t]}(t-y)^\kappa\mathrm{d}T(y)=(m(\kappa+1))^{-1}\quad \text{a.s.}$$

By the strong law of large numbers for $T(t)$ and Dini's theorem

$$\lim_{t\to\infty}\sup_{y\in[0,\,1]}|t^{-1}T(yt)-m^{-1}y|=0\quad \text{a.s.}$$

This entails

$$
\begin{aligned}
t^{-\kappa-1}\int_{[0,\,t]}(t-y)^{\kappa}\mathrm{d}T(y) &= \kappa t^{-1}\int_0^1 T(t(1-y))y^{\kappa-1}\mathrm{d}y \\
&\to \kappa m^{-1}\int_0^1 (1-y)y^{\kappa-1}\mathrm{d}y = (m(\kappa+1))^{-1}
\end{aligned}
$$

a.s. as $t\to\infty$.

Proof of (61) is similar to that of Proposition 3.6. To reduce technicalities to a minimum we only consider the case $q>1$. Since $\mathbb{E}[N_1(t)]^2 \le V(t)$ and $V(t)\sim c_0(m(q+1))^{-1}t^{q+1}$ as $t\to\infty$ we conclude that

$$\lim_{\mathbb{N}\ni\ell\to\infty}\ell^{(q+1/2)}N_1(\ell)=0\quad\text{a.s.}$$

by the Borel-Cantelli lemma.

For each $t\ge 0$, there exists $\ell\in\mathbb{N}_0$ such that $t\in[\ell,\ell+1)$. Now we use a.s. monotonicity of $N(t)$ and $N_2(t)$ to obtain

$$
\begin{aligned}
&(\ell+1)^{-(q+1/2)}(N(\ell)-(N_2(\ell+1)-N_2(\ell)))\le t^{-(q+1/2)}N_1(t)\\
\le\;& \ell^{-(q+1/2)}(N(\ell+1)+N_2(\ell+1)-N_2(\ell))\quad\text{a.s.}
\end{aligned}
$$

Thus, it remains to prove that

$$\lim_{\ell\to\infty}\ell^{-(q+1/2)}(N_2(\ell+1)-N_2(\ell))=0,\quad\text{a.s.}$$

In view of (57), f satisfies a counterpart of (15), whence

$$
\begin{aligned}
N_2(\ell+1)-N_2(\ell) &= \int_{[0,\,\ell]}(f(\ell+1-y)-f(\ell-y))\mathrm{d}T(y)+\int_{(\ell,\,\ell+1]}f(\ell+1-y)\mathrm{d}T(y)\\
&\le (c_0(q-1)(\ell+1)^{q-1}+\alpha_0+\alpha_1(\ell+1)^{q-r_1}-\beta_0+|\beta_1|\ell^{q-r_2}+f(1))T(\ell+1)\\
&= O(\ell^{q+\max(1-r_1,1-r_2,0)})=o(\ell^{q+1/2})
\end{aligned}
$$

a.s. as $\ell\to\infty$. For the penultimate equality we have used the strong law of large numbers for $T(y)$. The last equality follows from $r_1,r_2>1/2$.

(c) We use representation (59). Relation (61) entails

$$t^{-q-1/2}\sup_{y\in[0,T]}N_1(ty)\overset{\mathrm{P}}{\to}0,\quad t\to\infty. \tag{62}$$

for each $T>0$. Thus, we are left with showing that

$$\frac{N_2(t\cdot)-c_0(m(q+1))^{-1}(t\cdot)^{q+1}}{sm^{-3/2}t^{q+1/2}}\;\Rightarrow\;B_q(\cdot),\quad t\to\infty$$

in the J_1-topology on D. The proof of this is similar to that of weak convergence of the second coordinates in (27). The only difference is that, instead of (12), we use

$$\frac{T(t\cdot)-m^{-1}(t\cdot)}{sm^{-3/2}t^{1/2}}\;\Rightarrow\;B(\cdot),\quad t\to\infty$$

in the J_1-topology on D, where B is a Brownian motion, see Theorem 2a in [9]. □

References

[1] R. J. Adler, *An introduction to continuity, extrema, and related topics for general Gaussian processes*. Institute of Mathematical Statistics, 1990.

[2] G. Alsmeyer, A. Iksanov and A. Marynych, *Functional limit theorems for the number of occupied boxes in the Bernoulli sieve*. Stoch. Proc. Appl. **127** (2017), 995–1017.

[3] R. Arratia, A. D. Barbour and S. Tavaré, *Logarithmic combinatorial structures: a probabilistic approach*. European Mathematical Society, 2003.

[4] A. D. Barbour and A. Gnedin, *Regenerative compositions in the case of slow variation*. Stoch. Proc. Appl. **116** (2006), 1012–1047.

[5] A. D. Barbour and A. V. Gnedin, *Small counts in the infinite occupancy scheme*. Electron. J. Probab. **14** (2009), 365–384.

[6] A. Ben-Hamou, S. Boucheron and M. I. Ohannessian, *Concentration inequalities in the infinite urn scheme for occupancy counts and the missing mass, with applications*. Bernoulli. **23** (2017), 249–287.

[7] J. Bertoin, *Asymptotic regimes for the occupancy scheme of multiplicative cascades*. Stoch. Proc. Appl. **118** (2008), 1586–1605.

[8] P. Billingsley, *Convergence of probability measures*. Wiley, 1968.

[9] N. H. Bingham, *Limit theorems for regenerative phenomena, recurrent events and renewal theory*. Z. Wahrsch. Verw. Gebiete. **21** (1972), 20–44.

[10] S. Businger, *Asymptotics of the occupancy scheme in a random environment and its applications to tries*. Discrete Mathematics and Theoretical Computer Science. **19** (2017), #22.

[11] M. Chebunin, *On the accuracy of the poissonisation in the infinite occupancy scheme*. Preprint available at `https://arxiv.org/pdf/1712.03487.pdf`

[12] M. Chebunin and A. Kovalevskii, *Functional central limit theorems for certain statistics in an infinite urn scheme*. Statist. Probab. Letters. **119** (2016), 344–348.

[13] O. Durieu, G. Samorodnitsky and Y. Wang, *From infinite urn schemes to self-similar stable processes*. Preprint available at `https://arxiv.org/pdf/1710.08058.pdf`

[14] J.-J. Duchamps, J. Pitman and W. Tang, *Renewal sequences and record chains related to multiple zeta sums*. Preprint available at `https://arxiv.org/pdf/1707.07776.pdf` To appear in Trans. Amer. Math. Soc.

[15] N. Forman, C. Haulk and J. Pitman, *Representation of exchangeable hierarchies by sampling from random real trees*. Probab. Theory Relat. Fields. (2017) (online first)

[16] S. Ghosal and A. van der Vaart, *Fundamentals of nonparametric Bayesian inference*. Cambridge University Press, 2017.

[17] A. V. Gnedin, *The Bernoulli sieve*. Bernoulli **10** (2004), 79–96.

[18] A. Gnedin, A. Hansen and J. Pitman, *Notes on the occupancy problem with infinitely many boxes: general asymptotics and power laws*. Probab. Surv. **4** (2007), 146–171.

[19] A. Gnedin and A. Iksanov, *Regenerative compositions in the case of slow variation: a renewal theory approach*. Electron. J. Probab. **17** (2012), paper no. 77, 19 pp.

[20] A. Gnedin, A. Iksanov, and A. Marynych, *The Bernoulli sieve: an overview*. In Proceedings of the 21st International Meeting on Probabilistic, Combinatorial, and Asymptotic Methods in the Analysis of Algorithms (AofA-2010), Discrete Math. Theor. Comput. Sci. **AM** (2010), 329–341.

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

[22] A. Gnedin, J. Pitman and M. Yor, *Asymptotic laws for regenerative compositions: gamma subordinators and the like*. Probab. Theory Relat. Fields. **135** (2006), 576–602.

[23] Y. Hu, D. Nualart and J. Song, *Fractional martingales and characterization of the fractional Brownian motion*. Ann. Probab. **37** (2009), 2404–2430.

[24] A. Iksanov, *Renewal theory for perturbed random walks and similar processes*. Birkhäuser, 2016.

[25] A. Iksanov, W. Jedidi and F. Bouzeffour, *A law of the iterated logarithm for the number of occupied boxes in the Bernoulli sieve*. Statist. Probab. Letters. **126** (2017), 244–252.

[26] H. Ishwaran and L. F. James, *Gibbs sampling methods for stick-breaking priors*. J. Amer. Stat. Assoc. **96**, no. 453 (2001), 161–173.

[27] A. Joseph, *A phase transition for the heights of a fragmentation tree*. Random Structures and Algorithms. **39** (2011), 247–274.

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

[29] J. Pitman, *Combinatorial stochastic processes*. Springer, 2006.

[30] J. Pitman and W. Tang, *Regenerative random permutations of integers*. Preprint available at `https://arxiv.org/pdf/1704.01166.pdf`. To appear in Ann. Probab.

[31] J. Pitman and Y. Yakubovich, *Extremes and gaps in sampling from a GEM random discrete distribution*. Electron. J. Probab. **22** (2017), no. 44, 1–26.

[32] J. Pitman and Y. Yakubovich, *Ordered and size-biased frequencies in GEM and Gibbs' models for species sampling*. Ann. Appl. Probab. **28** (2018), 1793–1820.

[33] P. Robert and F. Simatos, *Occupancy schemes associated to Yule processes*. Adv. Appl. Probab. **41** (2009), 600–622.

[34] H. P. Rosenthal, *On the subspaces of $L^p (p > 2)$ spanned by sequences of independent random variables*. Israel J. Math. **8** (1970), 273–303.