A Limit Theorem for a Nested Infinite Occupancy Scheme in Random Environment

Oksana Braganets
Taras Shevchenko
National University of Kyiv

Alexander Iksanov
Taras Shevchenko
National University of Kyiv

Abstract

We investigate an infinite balls-in-boxes scheme, in which boxes are arranged in nested hierarchy and random probabilities of boxes are defined in terms of iterated fragmentation of a unit mass. Gnedin and Iksanov (2020) obtained a multivariate functional central limit theorem with centering for the cumulative occupancy counts as the number of balls becomes large. We prove a counterpart of their result, in which centering is not needed and the limit processes are not Gaussian. An application is given to the scheme generated by a residual allocation model.

Keywords: functional limit theorem, infinite occupancy, nested hierarchy, residual allocation model.

1. Introduction

The infinite occupancy scheme is an urn model that has numerous applications to statistics, combinatorics, and computer science. It is often depicted as a balls-in-boxes model. One throws balls successively and independently into an infinite array of boxes 1, 2, ..., so that each ball hits box k with positive probability p_k, and $\sum_{k \in \mathbb{N}} p_k = 1$. This classical model is usually called *Karlin's occupancy scheme* because of Karlin's seminal contribution (Karlin 1967). Features of the occupancy pattern emerging after the first n balls are thrown have been intensively studied, see Gnedin, Hansen, and Pitman (2007) and Iksanov (2016) for surveys.

There is also a randomized version of the classical infinite occupancy scheme, in which 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 (a.s.). We consider here a variant of this occupancy scheme, which corresponds to a nested family of boxes. The construction is conveniently described in terms of the genealogical structure of populations. Let $\mathcal{I}_0 := \{\varnothing\}$ be the initial ancestor and $\mathcal{I}_1 := \{1, 2, \ldots\}$ be the set of the first generation boxes with some random hitting probabilities $P_1, P_2, \ldots$. Divide now each box i into subboxes $i1, i2, \ldots$ and define the hitting probabilities of the subboxes by

$$P(ik) = P_i P_k^{(i)} \quad \text{for } k \in \mathbb{N},$$

where $(P_k^{(i)})_{k \in \mathbb{N}}$ is an independent copy of $(P_k)_{k \in \mathbb{N}}$. These subboxes are interpreted as the

second generation boxes which form the set $\mathcal{I}_2$. We repeat this procedure for boxes of each generation until an ∞-ary tree of nested boxes $\cup_{k\in\mathbb{N}_0}\mathcal{I}_k$ has been constructed. Here, $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. Note that the hitting probabilities of each generation boxes sum up to one by construction, that is, $\sum_{v\in\mathcal{I}_j} P(v) = 1$ a.s. for each $j \in \mathbb{N}$.

We assume that the random probabilities of boxes and the outcome of throwing balls are defined on a common probability space. For $n, j, r \in \mathbb{N}$, denote by $K_{n,j,r}$ the number of the jth generation boxes $v \in \mathcal{I}_j$ containing exactly r out of n first balls, and let

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

be a cumulative count of occupied boxes, where $\lceil\cdot\rceil$ is the ceiling function. With probability one, the random function $u \mapsto K_{n,j}(u)$ is non-decreasing and right-continuous, hence belongs to $D[0,1]$. Here and hereafter, for an interval $I \subseteq [0,\infty)$ we denote by $D(I)$ the Skorokhod space on I, that is, the set of all càdlàg functions defined on I.

2. Related literature

The nested occupancy scheme in random environment is introduced in Bertoin (2008). Among other things, it is proved there that, with both n and j diverging to infinity in such a way that $j \sim a \log n$, where a is some positive constant, $K_{n,j}(1)$ the number of occupied boxes in the jth generation, properly normalized, satisfies a central limit theorem with *random centering* as the number n of balls thrown tends to ∞. Let $H_{n,j}$ denote the index of the first generation in which all the boxes have less then j balls, when n balls have been thrown and $G_{n,j}$ denote the index of the first generation in which there exists a box containing less then j balls. In Joseph (2011) the a.s. asymptotic behaviour of these two random variables is investigated when $n \to \infty$. It turns out that $H_{n,j}$ and $G_{n,j}$ are of the order $\log n$ as $n \to \infty$ and their asymptotics depend on j, when j is smaller than the critical value j^*, and do not depend on j, when j is larger than j^*. Businger (2017) is concerned with similar problems in a more general setting (the underlying branching process is a multitype weighted branching process rather than a weighted branching process). The three aforementioned articles investigated *late generations* of the occupancy scheme in random environment, that is, generations whose indices j are of the order $\log n$. Iksanov and Mallein (2022) provide a full classification of regimes of the a.s. convergence for the number of occupied boxes in the late generations, thereby complementing the earlier results obtained in Bertoin (2008).

In Gnedin and Iksanov (2020) *early generations* (whose indices j do not depend on n) of the nested occupancy scheme in random environment are treated. In particular, a multivariate functional central limit theorem for the process $((K_{n,1}(u))_{u\in[0,1]}, (K_{n,2}(u))_{u\in[0,1]}, \ldots)$, properly scaled, centred and normalized, is proved, with the scaling limit of each coordinate being an a.s. continuous Gaussian process. Under the assumption that the hitting probabilities of boxes are given by a residual allocation model (a.k.a. stick-breaking) weak convergence of finite-dimensional distributions of $(K_{n,\lfloor j_n t\rfloor}(1))_{t>0}$, properly normalized and centered, is obtained in Iksanov, Marynych, and Samoilenko (2022d) for $j_n \to \infty$ and $j_n = o((\log n)^{1/2})$ as $n \to \infty$. A weaker version with $j_n = o((\log n)^{1/3})$ as $n \to \infty$ appeared earlier in Buraczewski, Dovgay, and Iksanov (2020).

Observe that the sequence $((-\log P(v))_{v\in\mathcal{I}_j})_{j\in\mathbb{N}}$ forms a branching random walk. In general, the definition of a branching random walk does not require that $\sum_{v\in\mathcal{I}_1} P(v) = 1$ a.s. A branching random walk with $(-\log P(v))_{v\in\mathcal{I}_1}$ being a globally perturbed random walk (see Section 6 for the definition) or a standard random walk is called an iterated perturbed random walk or an iterated standard random walk. Closely related to the present setting are the articles Bohun, Iksanov, Marynych, and Rashytov (2022), Iksanov, Rashytov, and Samoilenko (2023), Iksanov, Marynych, and Rashytov (2022c) and Iksanov, Kabluchko, and Kotelnikova

(2022b), in which various aspects of iterated perturbed random walks and iterated standard random walks are investigated.

Multivariate functional central limit theorems for the number of boxes containing at least 1 ball, 2 balls, ... and the number of boxes containing 1 ball, 2 balls, ..., properly scaled, normalized and centered, are proved in Iksanov, Kabluchko, and Kotelnikova (2022a) and Iksanov and Kotelnikova (2022), respectively for a nested occupancy scheme with deterministic probabilities.

3. Main result

Our purpose is to prove a functional limit theorem for $((K_{n,1}(u))_{u\in[0,1]}, (K_{n,2}(u))_{u\in[0,1]}, \ldots)$, properly scaled and normalized *without centering*. One consequence of our assumptions is that the scaling limit of each component is an a.s. nondecreasing process, which particularly cannot be Gaussian. Thus, even though our setting and that of Gnedin and Iksanov (2020) look similar, the ideas exploited here and in Gnedin and Iksanov (2020) are quite different at places.

For the given fragmentation law $(P_k)_{k\in\mathbb{N}}$, put $T_k := -\log P_k$ for $k \in \mathbb{N}$, $\rho(t) := \sum_{k\geq 1} \mathbb{1}_{\{P_k \geq 1/t\}}$ for $t > 0$, $N(t) := \rho(e^t)$ and $V(t) := \mathbb{E}N(t)$ for $t \in \mathbb{R}$. Let $j \in \mathbb{N}$. Similarly, for the jth generation put $\rho_j(t) := \sum_{v\in\mathcal{I}_j} \mathbb{1}_{\{P(v)\geq 1/t\}}$ for $t > 0$, $N_j(t) := \rho_j(e^t)$ and $V_j(t) := \mathbb{E}N_j(t)$ for $t \in \mathbb{R}$. Observe that $\rho_1 = \rho$, $N_1 = N$ and $V_1 = V$. Here is a basic recursive decomposition, which will be essentially used throughout the paper:

$$N_j(t) = \sum_{k\in\mathbb{N}} N_{j-1}^{(k)}(t - T_k), \quad t \in \mathbb{R}, \ j \geq 2, \tag{1}$$

where $(N_{j-1}^{(1)}(t))_{t\geq 0}$, $(N_{j-1}^{(2)}(t))_{t\geq 0}, \ldots$ are independent copies of $(N_{j-1}(t))_{t\geq 0}$, which are also independent of $T_1, T_2, \ldots$. Passing in (1) to the expectations yields

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

which shows that V_j is the j-fold Lebesgue-Stieltjes convolution of V with itself.

Throughout the paper we write $\Rightarrow$ for weak convergence in a function space. Now we formulate the assumptions of our main result:

$$V(t) \sim t^\alpha \ell(t), \quad t \to \infty, \tag{3}$$

for some $\alpha \geq 0$ and some ℓ slowly varying at ∞;

$$\sup_{t\geq 1} \frac{\mathbb{E}(N(t))^2}{(V(t))^2} < \infty \tag{4}$$

and

$$\left(\frac{N(ut)}{V(t)}\right)_{u\geq 0} \Rightarrow (W(u))_{u\geq 0}, \quad t \to \infty \tag{5}$$

in the J_1-topology on $D[0,\infty)$, where $(W(u))_{u\geq 0}$ is an a.s. locally Hölder continuous process with exponent $\beta \in (0,1]$. The latter means that, for every $T > 0$ there exists an a.s. finite random variable M_T such that, for all $x, y \in [0, T]$,

$$|W(x) - W(y)| \leq M_T |x - y|^\beta \quad \text{a.s.}$$

Put

$$W_j(u) := \int_{[0,\,u]} (u - y)^{\alpha(j-1)}\mathrm{d}W(y), \quad u \geq 0, j \in \mathbb{N},$$

where the integral exists as a pathwise Lebesgue-Stieltjes integral. Observe that $W_1 = W$.

We are ready to state our main result.

Theorem 1. *Suppose* (3), (4) *and* (5). *Then*

$$\left(\left(\frac{K_{n,j}(u)}{(\log n)^{\alpha j}(\ell(\log n))^j}\right)_{u\in[0,\,1]}\right)_{j\in\mathbb{N}} \Rightarrow \left(c_{j-1}(W_j(u))_{u\in[0,\,1]}\right)_{j\in\mathbb{N}}, \quad n\to\infty$$

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

$$c_j := \frac{(\Gamma(1+\alpha))^j}{\Gamma(1+\alpha j)}, \quad j\in\mathbb{N}_0 \tag{6}$$

and Γ is the Euler gamma function.

The remainder of the paper is structured as follows. Section 4 collects a number of preparatory results, which are then used in Section 5 for the proof of Theorem 1. An application of Theorem 1 to the occupancy scheme in random environment generated by a residual allocation model is given in Section 6.

4. Auxiliary results

In this section we collect several auxiliary results.

Lemma 2. *Suppose* (3). *Then, for each $j\in\mathbb{N}$,*

$$V_j(t) \sim c_j t^{\alpha j}(\ell(t))^j, \quad t\to\infty \tag{7}$$

with c_j given in (6).

Proof. Let $j\geq 2$ and $\hat{V}(s) := \int_{[0,\infty)} e^{-st}\mathrm{d}V(t)$ for $s>0$. Then in view of (2)

$$(\hat{V}(s))^j = \int_{[0,\infty)} e^{-st}\mathrm{d}V_j(t), \quad s>0. \tag{8}$$

By Karamata's Tauberian theorem (Theorem 1.7.1 in Bingham, Goldie, and Teugels (1987)) (3) is equivalent to

$$\hat{V}(s) \sim \Gamma(1+\alpha)s^{-\alpha}\ell(1/s), \quad s\to 0+,$$

whence

$$(\hat{V}(s))^j \sim (\Gamma(1+\alpha))^j s^{-\alpha j}(\ell(1/s))^j, \quad s\to 0+.$$

Observe that ℓ^j is a slowly varying function. In view of (8), another application of Theorem 1.7.1 in Bingham, Goldie, and Teugels (1987) yields (7).

$\square$

Lemma 3. *Suppose* (3) *and* (4). *Then, for each $j\in\mathbb{N}$,*

$$\sup_{t\geq 1}\frac{\mathbb{E}(N_j(t))^2}{(V_j(t))^2} < \infty. \tag{9}$$

Proof. We use mathematical induction. According to (4) inequality (9) holds with $j=1$. Assuming its validity for $j=i-1$ we shall show that (9) also holds with $j=i$.

In view of $\mathbb{E}(N_i(t))^2 = \mathbb{E}\big(N_i(t) - V_i(t)\big)^2 + (V_i(t))^2$ it suffices to prove

$$\sup_{t\geq 1}\frac{\mathbb{E}(N_i(t) - V_i(t))^2}{(V_i(t))^2} < \infty. \tag{10}$$

We first provide a convenient representation for $I(t) := \mathbb{E}\big(N_i(t) - V_i(t)\big)^2$. Using (1) we obtain

$$I(t) = \mathbb{E}\left(\sum_{k\geq 1}\big(N_{i-1}^{(k)}(t-T_k) - \mathbb{E}V_{i-1}(t-T_k)\big)\right)^2 = \sum_{k\geq 1}\mathbb{E}\big(N_{i-1}^{(k)}(t-T_k) - \mathbb{E}V_{i-1}(t-T_k)\big)^2. \tag{11}$$

The last equality is a consequence of

$$\mathbb{E}\big(\big(N_{i-1}^{(r)}(t-T_r) - \mathbb{E}V_{i-1}(t-T_r)\big)\big(N_{i-1}^{(s)}(t-T_s) - \mathbb{E}V_{i-1}(t-T_s)\big)|\mathcal{G}_s\big) = 0 \quad \text{a.s.}$$

for $r < s$, where $\mathcal{G}_k$ is the σ-algebra generated by $N_{i-1}^{(1)}, \ldots, N_{i-1}^{(k-1)}$ and $(T_l)_{l\in\mathbb{N}}$, $k \geq 2$. Here, while the expression in the first parantheses is $\mathcal{G}_s$-measurable, the conditional expectation of the expression in the second parantheses is equal to 0 a.s., for T_s is $\mathcal{G}_s$-measurable and $N_{i-1}^{(s)}$ is independent of $\mathcal{G}_s$.

Further,

$$I(t) \leq \sum_{k\geq 1} \mathbb{E}(N_{i-1}^{(k)}(t-T_k))^2 = \int_{[0,\,t]} \mathbb{E}(N_{i-1}(t-x))^2 \mathrm{d}V(x).$$

Inequality (9) with $j = i-1$ implies that $\mathbb{E}(N_{i-1}(t))^2 \leq C(V_{i-1}(t))^2$ for all $t \geq 1$ and some constant $C \in (0, \infty)$. Hence,

$$\int_{[0,\,t-1]} \mathbb{E}(N_{i-1}(t-x))^2 \mathrm{d}V(x) \leq C \int_{[0,\,t]} (V_{i-1}(t-x))^2 \mathrm{d}V(x) \leq CV_{i-1}(t)V_i(t)$$
$$= o((V_i(t))^2), \quad t \to \infty$$

having utilized monotonicity of V_{i-1} for the last inequality and Lemma 2 for the asymptotic relation. Using monotonicity of $y \mapsto \mathbb{E}(N_{i-1}(y))^2$ and Lemma 2 we infer

$$\int_{(t-1,\,t]} \mathbb{E}(N_{i-1}(t-x))^2 \mathrm{d}V(x) \leq \mathbb{E}(N_{i-1}(1))^2(V(t) - V(t-1)) = o((V_i(t))^2)$$

as $t \to \infty$. This completes the proof of (10). $\qquad\square$

Let $(Y_k)_{k\in\mathbb{N}_0}$ be a sequence of nonnegative random variables. Put $M(t) = \sum_{k\geq 0} \mathbb{1}_{\{Y_k \leq t\}}$ for $t \geq 0$ and assume that $M(t) < \infty$ a.s. for $t \geq 0$. For a function $h \in D[0, \infty)$, put

$$X(t) = \int_{[0,\,t]} h(t-y) \mathrm{d}M(y), \quad t \geq 0.$$

In the terminology of the article Iksanov and Rashytov (2021) $(X(t))_{t\geq 0}$ is a *general shot noise process* with the response function h and the counting process M.

We state a slightly corrected version of Theorem 1 in Iksanov and Rashytov (2021). The formulation in Iksanov and Rashytov (2021) contains an extra assumption, which is omitted here.

Proposition 4. *Fix any $\alpha > 0$, $\lambda \in (0, 1]$ and $\beta \geq 0$. Let $h : [0, \infty) \to [0, \infty)$ be a nondecreasing function which varies regularly at ∞ of index β and $a : [0, \infty) \to [0, \infty)$ a nonincreasing function which varies regularly at ∞ of index $-\alpha$. Assume that $(a(t)M(ut))_{u\geq 0} \Rightarrow (V_\lambda(u))_{u\geq 0}$ as $t \to \infty$ in the J_1-topology on $D[0, \infty)$, where $(V_\lambda(u))_{u\geq 0}$ is an a.s. nondecreasing random process, which is a.s. locally Hölder continuous with exponent λ and satisfies $V_\lambda(0) = 0$ a.s. Then*

$$\Big(\frac{a(t)}{h(t)}X(ut)\Big)_{u\geq 0} \Rightarrow \Big(\int_{[0,\,u]} (u-y)^\beta \mathrm{d}V_\lambda(y)\Big)_{u\geq 0}, \quad t \to \infty$$

in the J_1-topology on $D[0, \infty)$.

5. Proof of the main result

Theorem 1 is an immediate consequence of the following two results.

Theorem 5. *Suppose* (3), (4) *and* (5). *Then*

$$\left(\left(\frac{N_j(ut)}{V_j(t)}\right)_{u\geq 0}\right)_{j\in\mathbb{N}} \ \Rightarrow\ \left(\frac{c_{j-1}}{c_j}\left(W_j(u)\right)_{u\geq 0}\right)_{j\in\mathbb{N}}, \quad n\to\infty$$

in the product J_1-topology on $(D[0,\infty))^{\mathbb{N}}$, with c_j as given in (6).

Proposition 6. *Suppose* (3), (4) *and* (5). *Then, for each $j\in\mathbb{N}$,*

$$\sup_{u\in[0,1]}\frac{|K_{n,j}(u)-\rho_j(n^u)|}{V_j(\log n)} \ \xrightarrow{\mathbb{P}}\ 0, \quad n\to\infty,$$

where $\xrightarrow{\mathbb{P}}$ denotes convergence in probability.

Proof of Theorem 5. We claim it suffices to prove weak convergence of finite-dimensional distributions. According to the Cramér-Wold device, the task amounts to showing that, for any $i,m\in\mathbb{N}$, any nonnegative $\beta_{r,k}$, $r,k\in\mathbb{N}$, $r\leq i$, $k\leq m$ and any nonnegative $u_{r,k}$, $r,k\in\mathbb{N}$, $r\leq i$, $k\leq m$,

$$\sum_{r=1}^{i}\sum_{k=1}^{m}\beta_{r,k}\frac{N_r(u_{r,k}t)}{V_r(t)} \ \xrightarrow{\text{d}}\ \sum_{r=1}^{i}\sum_{k=1}^{m}\beta_{r,k}\frac{c_{r-1}}{c_r}W_r(u_{r,k}), \quad t\to\infty, \tag{12}$$

where $\xrightarrow{\text{d}}$ denotes one-dimensional distributional convergence. Since the product J_1-topology on $(D[0,\infty))^{\mathbb{N}}$ is used, tightness of the distributions of the converging processes is secured by tightness of the distributions of their coordinates. The latter is ensured by Remark 2.1 in Yamazato (2009), because, for each $j\in\mathbb{N}$ and each $t>0$, the random function $u\mapsto N_j(ut)$ is a.s. nondecreasing on $[0,\infty)$ and the limit process W_j is a.s. continuous (this can be easily checked).

Using (1), write, for $t\in\mathbb{R}$ and $j\geq 2$,

$$N_j(t)=\sum_{k\geq 1}N_{j-1}^{(k)}(t-T_k)=\sum_{k\geq 1}(N_{j-1}^{(k)}(t-T_k)-V_{j-1}(t-T_k))+\sum_{k\geq 1}V_{j-1}(t-T_k)=:X_j(t)+Y_j(t).$$

We shall prove that

$$\sum_{k=1}^{m}\beta_{1,k}\frac{N_1(u_{1,k}t)}{V_1(t)}+\sum_{r=2}^{i}\sum_{k=1}^{m}\beta_{r,k}\frac{Y_r(u_{r,k}t)}{V_r(t)} \ \xrightarrow{\text{d}}\ \sum_{r=1}^{i}\sum_{k=1}^{m}\beta_{r,k}\frac{c_{r-1}}{c_r}W_r(u_{r,k}), \quad t\to\infty, \tag{13}$$

and, for each $j\in\mathbb{N}$,

$$\frac{X_j(t)}{V_j(t)} \ \xrightarrow{\mathbb{P}}\ 0, \quad t\to\infty. \tag{14}$$

These limit relations entail (12). Note that, in view of the regular variation of V_j, (14) guarantees that, for each $u\geq 0$,

$$\frac{X_j(ut)}{V_j(t)} \ \xrightarrow{\mathbb{P}}\ 0, \quad t\to\infty.$$

PROOF OF (13). Since

$$Y_r(u_{r,k}t)=\int_{[0,u_{r,k}]}V_{r-1}(t(u_{r,k}-y))\mathrm{d}_y N(ty), \quad t\geq 0$$

and

$$W(u_{r,k})=\int_{[0,u_{r,k}]}(u_{r,k}-y)^{\alpha(r-1)}\mathrm{d}W(y),$$

(13) is equivalent to

$$\sum_{k=1}^{m}\beta_{1,k}\frac{N_1(u_{1,k}t)}{V_1(t)} + \sum_{r=2}^{i}\sum_{k=1}^{m}\beta_{r,k}\frac{V_{r-1}(t)V(t)}{V_r(t)}\int_{[0,u_{r,k}]}\frac{V_{r-1}(t(u_{r,k}-y))}{V_{r-1}(t)}\mathrm{d}\frac{N(ty)}{V(t)}$$

$$\xrightarrow{\mathrm{d}}\sum_{k=1}^{m}\beta_{1,k}W(u_{1,k}) + \sum_{r=2}^{i}\sum_{k=1}^{m}\beta_{r,k}\frac{c_{r-1}}{c_r}\int_{[0,u_{r,k}]}(u_{r,k}-y)^{\alpha(r-1)}\mathrm{d}W(y), \quad t\to\infty. \quad (15)$$

Let $(t_n)_{n\in\mathbb{N}}$ be an arbitrary sequence of positive numbers satisfying $\lim_{n\to\infty}t_n=\infty$. According to the Skorokhod representation theorem there exist processes $((\tilde{N}_{t_n}(y))_{y\geq 0})_{n\in\mathbb{N}}$ having the same distribution as $((N(t_n y)/V(t_n))_{y\geq 0})_{n\in\mathbb{N}}$ and a process $\tilde{W}$ having the same distribution as W, such that

$$\lim_{n\to\infty}\tilde{N}_{t_n}(y) = \tilde{W}(y) \quad \text{a.s. on } D[0,\infty).$$

The distribution of the left- (right-) hand side of (15) does not change upon replacing $N(t_n y)/V(t_n)$ with $\tilde{N}_{t_n}(y)$ (W with $\tilde{W}$). Hence, (15) follows if we can show that

$$\lim_{n\to\infty}\Big(\sum_{k=1}^{m}\beta_{1,k}\tilde{N}_{t_n}(u_{1,k}) + \sum_{r=2}^{i}\sum_{k=1}^{m}\beta_{r,k}\frac{V_{r-1}(t_n)V(t_n)}{V_r(t_n)}\int_{[0,u_{r,k}]}\frac{V_{r-1}(t_n(u_{r,k}-y))}{V_{r-1}(t_n)}\mathrm{d}\tilde{N}_{t_n}(y)\Big)$$

$$= \sum_{k=1}^{m}\beta_{1,k}\tilde{W}(u_{1,k}) + \sum_{r=2}^{i}\sum_{k=1}^{m}\beta_{r,k}\frac{c_{r-1}}{c_r}\int_{[0,u_{r,k}]}(u_{r,k}-y)^{\alpha(r-1)}\mathrm{d}\tilde{W}(y) \quad \text{a.s.} \quad (16)$$

According to (5), the first sum on the left-hand side converges a.s. to the first sum on the right-hand side. For $r\geq 2$, the process Y_r is an instance of a general shot noise process with the response function V_{r-1} and the counting process N, see the paragraph preceding Proposition 4 for the definition. The function V_{r-1} is nondecreasing and, by Lemma 2, regularly varying at ∞ of index $\alpha(r-1)$. The function $1/V$ is nonincreasing and, by assumption, is regularly varying at ∞ of index $-\alpha$. The process W is an a.s. locally Hölder continuous with exponent $\beta\in(0,1]$. Thus, the process Y_r satisfies all the assumptions of Proposition 4, and an application of this result yields

$$\Big(\frac{Y_r(ut)}{V_{r-1}(t)V(t)}\Big)_{u\geq 0} \Rightarrow (W_{r-1}(u))_{u\geq 0}, \quad t\to\infty$$

in the J_1-topology on $D[0,\infty)$ or equivalently

$$\Big(\frac{Y_r(ut)}{V_r(t)}\Big)_{u\geq 0} \Rightarrow \frac{c_{r-1}}{c_r}(W_{r-1}(u))_{u\geq 0}, \quad t\to\infty$$

because

$$V_{r-1}(t)V(t) \sim (c_{r-1}/c_r)V_r(t), \quad t\to\infty$$

by Lemma 2. This entails that each term on the left-hand side of (16) which corresponds to a single $r=2,3,...,i$ converges a.s. to the corresponding term on the right-hand side of (16). This proves (16) and thereupon (13).

PROOF OF (14). Fix any integer $j\geq 2$. According to Markov's inequality it is enough to prove that $\mathbb{E}(X_j(t))^2 = o((V_j(t))^2)$ as $t\to\infty$. Invoking (11) we obtain

$$\mathbb{E}(X_j(t))^2 = \int_{[0,t]}\mathrm{Var}\,(N_{j-1}(t-x))\mathrm{d}V(x) = \int_{[0,t-1]}\ldots + \int_{(t-1,t]}\ldots.$$

It follows from (9) that $\mathrm{Var}\,(N_{j-1}(t))\leq \mathbb{E}(N_{j-1}(t))^2 \leq C(V_{j-1}(t))^2$ for some positive constant C and $t\geq 1$. Hence,

$$\mathbb{E}(X_j(t))^2 \leq C\int_{[0,t]}(V_{j-1}(t-x))^2\mathrm{d}V(x) + \sup_{y\in[0,1]}\mathbb{E}(N_{j-1}(y))^2(V(t)-V(t-1))$$

$$\leq CV_{j-1}(t)V_j(t) + \mathbb{E}(N_{j-1}(1))^2 V(t) = o((V_j(t))^2), \quad t\to\infty.$$

Here, the second inequality is justified by monotonicity of V_{j-1} and (2), and the last equality is secured by Lemma 2.

$\square$

Proof of Proposition 6. Fix any $j \in \mathbb{N}$. By Proposition 3.6 in Gnedin and Iksanov (2020), for $n \in \mathbb{N}$,

$$\mathbb{E}\left(\sup_{s \in [0, 1]} |K_{n, j}(s) - \rho_j(n^s)| \Big| (P(v))_{v \in \mathcal{I}_j} \right) \le 4\big(\rho_j(n) - \rho_j(y_0 n (\log n)^{-2})\big) + 2\rho_j(n)(\log n)^{-1}$$

$$+ \int_1^\infty x^{-2}(\rho_j(nx) - \rho_j(n))\mathrm{d}x + 2 \sup_{s \in [0, 1]} (\rho_j(en^s) - \rho_j(e^{-1}n^s)), \quad (17)$$

where $y_0 \in (0, 1)$ is a deterministic constant which does not depend on n, nor on $(P(v))_{v \in \mathcal{I}_j}$. In view of Theorem 5 and Lemma 2,

$$\left(\frac{N_j(s \log n + 1)}{V_j(\log n)}, \frac{N_j(s \log n - 1)}{V_j(\log n)} \right)_{s \ge 0} \Rightarrow ((c_{j-1}/c_j)W_j(s), (c_{j-1}/c_j)W_j(s))_{s \ge 0}, \quad n \to \infty$$

in the J_1-topology on $D[0, \infty) \times D[0, \infty)$, whence

$$\frac{\sup_{s \in [0, 1]}(\rho_j(en^s) - \rho_j(e^{-1}n^s))}{V_j(\log n)} = \frac{\sup_{s \in [0, 1]}(N_j(s \log n + 1) - N_j(s \log n - 1))}{V_j(\log n)} \xrightarrow{\mathbb{P}} 0, \quad n \to \infty.$$

To complete the proof, according to Markov's inequality, it suffices to show that the expectation of each of the first three terms in (17) divided by $V_j(\log n)$ converges to 0 as $n \to \infty$.

For the first term, this follows from the fact, which is a consequence of Lemma 2, that the function $t \mapsto V_j(\log t)$ is slowly varying at ∞. For the second term, this is trivial. For the third term, write

$$\lim_{n \to \infty} \frac{\mathbb{E} \int_1^\infty x^{-2}(\rho_j(nx) - \rho_j(n))\mathrm{d}x}{V_j(\log n)} = \lim_{n \to \infty} \int_1^\infty x^{-2} \frac{V_j(\log(nx))}{V_j(\log n)}\mathrm{d}x - 1$$

$$= \int_1^\infty x^{-2} \lim_{n \to \infty} \frac{V_j(\log(nx))}{V_j(\log n)}\mathrm{d}x - 1 = 0$$

having utilized slow variation of $t \mapsto V_j(\log t)$ for the last equality. The penultimate equality is justified by Lebesgue's dominated convergence theorem in combination with Potter's bound (Theorem 1.5.6(i) in Bingham, Goldie, and Teugels (1987)): for all $x \ge 1$ and large n, $V_j(\log(nx))/V_j(\log n) \le 2x^{1/2}$.

$\square$

6. An application to a residual allocation model

Assume that $(P_k)_{k \in \mathbb{N}}$ follow a residual allocation model

$$P_k = U_1 U_2 \cdot \ldots \cdot U_{k-1}(1 - U_k), \quad k \in \mathbb{N}, \quad (18)$$

where $U_1, U_2, \ldots$ are independent copies of a random variable U taking values in $(0, 1)$ and satisfying

$$\mathbb{P}\{|\log U| > x\} \sim x^{-\rho}L(x), \quad x \to \infty \quad (19)$$

for $\rho \in (0, 1)$ and some L slowly varying at ∞. We intend to apply Theorem 1 to the corresponding infinite occupancy scheme in random environment. To this end, we first check that, under (18) and (19), conditions (3), (4) and (5) hold true with $\alpha = \rho$, $\ell(t) = (\Gamma(1 -$

$\rho)\Gamma(1+\rho)L(t))^{-1}$ and W being a constant multiple of an inverse ρ-stable subordinator $\mathcal{S}_\rho^\leftarrow$, say. The process $\mathcal{S}_\rho^\leftarrow$ is defined by

$$\mathcal{S}_\rho^\leftarrow(u) := \inf\{v \geq 0 : \mathcal{S}_\rho(v) > u\}, \quad u \geq 0,$$

where $(\mathcal{S}_\rho(v))_{v\geq 0}$ is a ρ-stable subordinator with $-\log \mathbb{E}e^{-s\mathcal{S}_\rho(1)} = \Gamma(1-\rho)s^\rho$ for $s \geq 0$.

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

$$T_k^* := S_{k-1} + \eta_k, \quad k \in \mathbb{N}.$$

The random sequence $(T_k^*)_{k\in\mathbb{N}}$ is known in the literature as a (globally) perturbed random walk, see Section 1 in Iksanov (2016). With P_k as in (18),

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

that is, $(T_k)_{k\in\mathbb{N}}$ is a particular instance of the perturbed random walk with

$$(\xi, \eta) = (|\log U|, |\log(1 - U)|). \tag{20}$$

Throughout this section, $N(t) = \sum_{k\geq 1} \mathbb{1}_{\{T_k \leq t\}}$ and $V(t) = \mathbb{E}N(t) = \sum_{k\geq 1} \mathbb{P}\{T_k \leq t\}$, $t \in \mathbb{R}$ for the particular $(T_k)_{k\in\mathbb{N}}$ as above.

CONDITION (3). Write, for $s > 0$,

$$\int_{[0,\infty)} e^{-st} d\Big(\sum_{k\geq 1} \mathbb{P}\{T_k^* \leq t\}\Big) = \mathbb{E}e^{-s\eta} \sum_{k\geq 0} \mathbb{E}e^{-sS_k} = \mathbb{E}e^{-s\eta}(1 - \mathbb{E}e^{-s\xi})^{-1}.$$

As a consequence,

$$\int_{[0,\infty)} e^{-st} d\Big(\sum_{k\geq 1} \mathbb{P}\{T_k^* \leq t\}\Big) \sim (1 - \mathbb{E}e^{-s\xi})^{-1}, \quad s \to 0+.$$

We shall apply this limit relation to (ξ, η) as in (20).

By Corollary 8.1.7 in Bingham, Goldie, and Teugels (1987), (19) is equivalent to $1 - \mathbb{E}e^{-s|\log U|} \sim \Gamma(1-\rho)s^\rho L(1/s)$ as $s \to 0+$, whence

$$\int_{[0,\infty)} e^{-st} dV(t) \sim (1 - \mathbb{E}e^{-s|\log U|})^{-1} \sim (\Gamma(1-\rho)s^\rho L(1/s))^{-1}, \quad s \to 0+.$$

Invoking Theorem 1.7.1 Bingham, Goldie, and Teugels (1987) yields

$$V(t) \sim \frac{1}{\Gamma(1-\rho)\Gamma(1+\rho)} \frac{t^\rho}{L(t)}, \quad t \to \infty, \tag{21}$$

that is, condition (3) does indeed hold with $\alpha = \rho$ and $\ell(t) = (\Gamma(1-\rho)\Gamma(1+\rho)L(t))^{-1}$.

CONDITION (4). Since $N(t) \leq 1 + \sum_{k\geq 1} \mathbb{1}_{\{|\log U_1| + \ldots + |\log U_k| \leq t\}} =: \hat{N}(t)$ for $t \geq 0$ a.s. and, under (19), $\sup_{t\geq 1}(\mathbb{E}(\hat{N}(t))^2/(V(t))^2) < \infty$ by Theorem 1.5 in Iksanov, Marynych, and Meiners (2016), condition (4) holds true.

CONDITION (5). By part (B4) of Theorem 3.2 in Alsmeyer, Iksanov, and Marynych (2017),

$$(\mathbb{P}\{|\log U| > t\}N(tu))_{u\geq 0} \Rightarrow (\mathcal{S}_\rho^\leftarrow(u))_{u\geq 0}, \quad t \to \infty$$

in the J_1-topology on $D[0, \infty)$. In view of (21), this is equivalent to

$$\Big(\frac{N(tu)}{V(t)}\Big)_{u\geq 0} \Rightarrow \Gamma(1-\rho)\Gamma(1+\rho)(\mathcal{S}_\rho^\leftarrow(u))_{u\geq 0}, \quad t \to \infty.$$

According to Lemma 3.4 in Owada and Samorodnitsky (2015), the process $\mathcal{S}_\rho^\leftarrow$ is a.s. locally Hölder continuous with exponent smaller than ρ.

An application of Theorem 1 yields the following.

Theorem 7. *Suppose* (18) *and* (19). *Then*

$$\left(\left(\frac{(L(\log n))^j K_{n,j}(u)}{(\log n)^{\rho j}}\right)_{u\in[0,\,1]}\right)_{j\in\mathbb{N}}$$
$$\Rightarrow \left(\frac{1}{(\Gamma(1-\rho))^{j-1}\Gamma(1+\rho(j-1))}\left(\int_{[0,\,u]}(u-y)^{\rho(j-1)}\mathrm{d}\mathcal{S}_\rho^{\leftarrow}(y)\right)_{u\in[0,\,1]}\right)_{j\in\mathbb{N}},\quad n\to\infty$$

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

Acknowledgement

References

Alsmeyer G, Iksanov A, Marynych A (2017). "Functional Limit Theorems for the Number of Occupied Boxes in the Bernoulli Sieve." *Stochastic Processes and Their Applications*, **127**(3), 995–1017. `doi:10.1016/j.spa.2016.07.007`.

Bertoin J (2008). "Asymptotic Regimes for the Occupancy Scheme of the Multiplicative Cascades." *Stochastic Processes and Applications*, **118**(9), 1586–1605. `doi:10.1016/j.spa.2007.10.004`.

Bingham N, Goldie C, Teugels J (1987). *Regular Variation*. Cambridge University Press. `doi:10.1017/CBO9780511721434`. ISBN 9780511721434.

Bohun V, Iksanov A, Marynych A, Rashytov B (2022). "Renewal Theory for Iterated Perturbed Random Walks on a General Branching Process Tree: Intermediate Generations." *Journal of Applied Probability*, **59**(2), 421–446. `doi:10.1017/jpr.2021.60`.

Buraczewski D, Dovgay B, Iksanov A (2020). "On Intermediate Levels of Nested Occupancy Scheme in Random Environment Generated by Stick-Breaking I." *Electronic Journal of Probability*, **25**, 1–24. `doi:10.1214/20-EJP534`.

Businger S (2017). "Asymptotics of the Occupancy Scheme in a Random Environment and Its Applications to Tries." *Discrete Mathematics & Theoretical Computer Science*, **19**(1), 1–22. `doi:10.23638/DMTCS-19-1-22`.

Gnedin A, Hansen A, Pitman J (2007). "Notes on the Occupancy Problem with Infinitely Many Boxes: General Asymptotics and Power Laws." *Probability Surveys*, **4**, 146–171. `doi:10.1214/07-ps092`.

Gnedin A, Iksanov A (2020). "On Nested Infinite Occupancy Scheme in Random Environment." *Probability Theory and Related Fields*, **177**, 855–890. `doi:10.1007/s00440-020-00963-0`.

Iksanov A (2016). *Renewal Theory for Perturbed Random Walks and Similar Processes*. Birkhäuser. `doi:10.1007/978-3-319-49113-4`. ISBN 978-3319840857.

Iksanov A, Kabluchko Z, Kotelnikova V (2022a). "A Functional Limit Theorem for Nested Karlin's Occupancy Scheme Generated by Discrete Weibull-like Distributions." *Journal of Mathematical Analysis and Applications*, **507(2)**, 125798. `doi:10.1016/j.jmaa.2021.125798`.

Iksanov A, Kabluchko Z, Kotelnikova V (2022b). "A Law of the Iterated Logarithm for Iterated Random Walks, With Application to Random Recursive Trees." `doi:10.48550/ARXIV.2212.13441`.

Iksanov A, Kotelnikova V (2022). "Small Counts in Nested Karlin's Occupancy Scheme Generated by Discrete Weibull-like Distributions." *Stochastic Processes and their Applications*, **153**, 283–320. `doi:10.1016/j.spa.2022.08.006`.

Iksanov A, Mallein B (2022). "Late Levels of Nested Occupancy Scheme in Random Environment." *Stochastic Models*, **38**(1), 130–166. `doi:10.1080/15326349.2021.1984948`.

Iksanov A, Marynych A, Meiners M (2016). "Moment Convergence of First-Passage Times in Renewal Theory." *Statistics & Probability Letters*, **119**, 134–143. `doi:10.1016/j.spl.2016.07.019`.

Iksanov A, Marynych A, Rashytov B (2022c). "Stable Fluctuations of Iterated Perturbed Random Walks in Intermediate Generations of a General Branching Process Tree." *Lithuanian Mathematical Journal*, **62**, 447–466. `doi:10.1007/s10986-022-09574-9`.

Iksanov A, Marynych A, Samoilenko I (2022d). "On Intermediate Levels of a Nested Occupancy Scheme in a Random Environment Generated by Stick-Breaking II." *Stochastics*, **94**(7), 1077–1101. `doi:10.1080/17442508.2021.2019739`.

Iksanov A, Rashytov B (2021). "Functional Limit Theorem Without Centering for General Shot-Noise Processes." *Ukrainian Mathematical Journal*, **73**(2), 181–202. `doi:10.37863/umzh.v73i2.6210`.

Iksanov A, Rashytov B, Samoilenko I (2023). "Renewal Theory for Iterated Perturbed Random Walks on a General Branching Process Tree: Early Generations." *Journal of Applied Probability*, **60**(1), 45–67. `doi:10.1017/jpr.2022.26`.

Joseph A (2011). "A Phase Transition for the Heights of the Fragmentation Tree." *Random Structures Algorithms*, **39**(2), 247–274. `doi:10.1002/rsa.20340`.

Karlin S (1967). "Central Limit Theorems for Certain Infinite Urn Schemes." *Journal of Mathematics and Mechanics*, **17**(4), 373–401. `doi:10.1512/IUMJ.1968.17.17020`.

Owada T, Samorodnitsky G (2015). "Functional Central Limit Theorem for Heavy Tailed Stationary Infinitely Divisible Processes Generated by Conservative Flows." *Annals of Probability*, **43**(1), 240–285. `doi:10.1214/13-AOP899`.

Yamazato M (2009). "On a J_1-convergence Theorem for Stochastic Processes on $D[0, \infty)$ Having Monotone Sample Paths and Its Applications." *RIMS Kôkyûroku*, **1620**, 109–118. URL `http://hdl.handle.net/2433/140220`.

Affiliation:

Oksana Braganets
Taras Shevchenko National University of Kyiv
Kyiv-01033, Ukraine
E-mail: oksanabraganets@knu.ua
Alexander Iksanov
Taras Shevchenko National University of Kyiv
Kyiv-01033, Ukraine
E-mail: iksan@univ.kiev.ua
URL: https://do.csc.knu.ua/iksan