

# FROM INFINITE URN SCHEMES TO SELF-SIMILAR STABLE PROCESSES

OLIVIER DURIEU, GENNADY SAMORODNITSKY, AND YIZAO WANG

**ABSTRACT.** We investigate the randomized Karlin model with parameter  $\beta \in (0, 1)$ , which is based on an infinite urn scheme. It has been shown before that when the randomization is bounded, the so-called odd-occupancy process scales to a fractional Brownian motion with Hurst index  $\beta/2 \in (0, 1/2)$ . We show here that when the randomization is heavy-tailed with index  $\alpha \in (0, 2)$ , then the odd-occupancy process scales to a  $(\beta/\alpha)$ -self-similar symmetric  $\alpha$ -stable process with stationary increments.

## 1. INTRODUCTION AND MAIN RESULTS

Consider the following infinite urn scheme. Suppose there is an infinite number of urns labeled by  $\mathbb{N} = \{1, 2, \dots\}$ , all initially empty. Balls are thrown into the urns randomly one after another. At each round, a ball is thrown independently into the urn with label  $k$  with probability  $p_k$ , with  $\sum_{k \geq 1} p_k = 1$ . This random sampling strategy dates back to at least the 60s [3, 13]. The urns may represent different species in a population of interest, and in various applications an interesting question is to infer the population frequencies  $(p_k)_{k \geq 1}$ ; see [11] and references therein. This urn scheme has been extensively investigated in the literature on combinatorial stochastic processes as it induces the so-called *paintbox partition* of  $\mathbb{N}$ , an infinite exchangeable random partition; see for example [19].

Asymptotic results for many statistics of this urn scheme, and in particular of the random partition it induces, have been investigated in the literature (e.g. [11] and references therein), including in particular the so-called *odd-occupancy process*. In the sequel, let  $Y_n$  denote the label of the urn that the  $n$ -th ball falls into, so  $\mathbb{P}(Y_n = k) = p_k$ . Then  $(Y_n)_{n \in \mathbb{N}}$  are i.i.d. random variables, and we let  $Y_{n,k} := \sum_{i=1}^n \mathbb{1}_{\{Y_i=k\}}$  denote the number of balls in the urn with label  $k$  after first  $n$  rounds. The odd-occupancy process is then defined as

$$U_n^* = \sum_{k=1}^{\infty} \mathbb{1}_{\{Y_{n,k} \text{ odd}\}}.$$

This process counts the number of urns that contain an odd number of balls after the first  $n$  rounds. An interpretation of this process due to Spitzer [25] is as follows. One may associate to each urn a lightbulb, and start the sampling procedure with all lightbulbs off. Each time a ball falls in an urn, the corresponding lightbulb changes its status (from off to on or from on to off). The process  $U_n^*$  then represents the total number of lightbulbs that are on after the first  $n$  rounds.

---

*Date:* July 24, 2018.

*2010 Mathematics Subject Classification.* Primary, 60F17; Secondary, 60G18, 60G52.

*Key words and phrases.* Infinite urn scheme, regular variation, functional central limit theorem, self-similar process, stable process.

Karlin [13] proposed and investigated the aforementioned model under the following assumptions on  $(p_k)_{k \geq 1}$ :  $p_k$  is decreasing in  $k$  and

$$(1) \quad \max\{k \geq 1 \mid p_k \geq 1/t\} = t^\beta L(t), \quad t \geq 0, \text{ for some } \beta \in (0, 1),$$

where  $L$  is a slowly varying function at infinity. Among many results, Karlin proved that

$$\frac{U_n^* - \mathbb{E}U_n^*}{\sigma_n} \Rightarrow \mathcal{N}(0, 1)$$

as  $n \rightarrow \infty$  with  $\sigma_n = 2^{\beta-1}(\Gamma(1-\beta)n^\beta L(n))^{1/2}$ . Here and in the sequel  $\Rightarrow$  denotes weak convergence and  $\mathcal{N}(0, 1)$  stands for the standard normal distribution.

We are interested in the *randomized version* of the odd-occupancy process, defined as

$$(2) \quad U_n = \sum_{k=1}^{\infty} \varepsilon_k \mathbb{1}_{\{Y_{n,k} \text{ odd}\}},$$

where  $(\varepsilon_k)_{k \in \mathbb{N}}$  are i.i.d. symmetric random variables independent of  $(Y_n)_{n \in \mathbb{N}}$ . This randomization was recently introduced in [8], and it was shown in Corollary 2.8 therein that with  $(\varepsilon_k)_{k \in \mathbb{N}}$  being a sequence of independent Rademacher ( $\pm 1$ -valued and symmetric) random variables, with the same normalization  $\sigma_n$ ,

$$(3) \quad \left( \frac{U_{\lfloor nt \rfloor}}{\sigma_n} \right)_{t \in [0,1]} \Rightarrow 2^{-(\beta-1)/2} \left( \mathbb{B}_t^{\beta/2} \right)_{t \in [0,1]}$$

in  $D([0, 1])$ , where  $\mathbb{B}^H$  is a standard fractional Brownian motion with Hurst index  $H \in (0, 1)$ , a centered Gaussian process with covariance function

$$\text{Cov}(\mathbb{B}_s^H, \mathbb{B}_t^H) = \frac{1}{2} (s^{2H} + t^{2H} - |t-s|^{2H}), \quad s, t \geq 0.$$

The following aspect of the randomization and the resulting functional central limit theorem is particularly interesting. For an arbitrary symmetric distribution of  $\varepsilon_1$ , the randomized odd-occupancy process  $U_n$  is the partial-sum process for a stationary sequence,

$$(4) \quad U_n = X_1 + \cdots + X_n \quad \text{with} \quad X_i = -\varepsilon_{Y_i}(-1)^{Y_i}, \quad i \in \mathbb{N}, n \in \mathbb{N}.$$

Therefore, the infinite urn scheme provides a specific way to generate a stationary sequence of random variables whose marginal distribution is the given symmetric law, and whose partial-sum process is the randomized odd-occupancy process. Moreover, at least for the Rademacher marginal distribution, this stationary sequence  $(X_n)_{n \geq 1}$  exhibits, in view of the limiting result (3), anomalous behavior. Namely, the normalization  $\sigma_n$  has an order of magnitude different from the “usual”  $\sqrt{n}$  normalization needed for partial sums of i.i.d. random variables with the same marginal distribution. Such a behavior indicates *long-range dependence* in the stationary sequence [5, 17, 23]. The long-range dependence in this case is due to the underlying random partition. In particular, the covariance function of  $X$  is determined by the law of the random partition. In general, when  $\varepsilon_1$  is symmetric and in the domain of attraction of the normal distribution, it is expected that a fractional Brownian motion still arises in the limit with the same order of scaling as in (3).

In this paper, we are interested in the randomized odd-occupancy process when  $\varepsilon_1$  has a heavy-tailed distribution. Specifically, we will assume that  $\varepsilon_1$  has infinite variance and, even more specifically, is in the domain of attraction of a non-Gaussian stable law. The stationary-process representation (4) is, clearly, still valid. However, in a functional central limit theorem one expects now a symmetric stable process that is self-similar with stationary increments (we abbreviate the latter two properties as the sssi property). The only sssi Gaussian process is the fractional Brownian motion; however there are many different sssi

symmetric stable processes, which often arise in limit theorems for the partial sums of stationary sequences with long-range dependence [18, 23, 24]. While many sssi stable processes have been well investigated in the literature, new processes in this family are still being discovered [16]. One way to classify different sssi stable processes is via the flow representation, introduced by Rosiński [21] and developed by Samorodnitsky [22], of the corresponding increment processes which is, by necessity, stationary. Certain ergodic-theoretical properties of these flows are invariants for each stationary stable process, and processes corresponding to different types of flows, namely *positive*, *conservative null* and *dissipative*, have drastically different properties.

The main result of this paper is to show that for the Karlin model with a heavy-tailed randomization, the scaling limit of the randomized odd-occupancy process is a new self-similar symmetric stable process with stationary increments. We continue to assume that  $(\varepsilon_k)$  are i.i.d. and symmetric. For simplicity we will also assume that they are in the normal domain of attraction of a symmetric stable law and in particular, for some  $\alpha \in (0, 2)$ ,

$$(5) \quad \lim_{x \rightarrow \infty} \frac{\mathbb{P}(|\varepsilon_1| > x)}{x^{-\alpha}} = C_\varepsilon \in (0, \infty).$$

We will define the limiting sssi symmetric  $\alpha$ -stable (SaS) process in terms of stochastic integrals with respect to an SaS random measure as follows. Let  $(\Omega', \mathcal{F}', \mathbb{P}')$  be a probability space and  $N'$  a standard Poisson process defined on this space. Let  $\mathcal{M}_{\alpha, \beta}$  be a SaS random measure on  $\mathbb{R}_+ \times \Omega'$  with control measure  $\beta r^{-\beta-1} dr \mathbb{P}'(d\omega')$ ; we refer the reader to [24] for detailed information on SaS random measures and stochastic integrals with respect to these measures. The random measure  $\mathcal{M}_{\alpha, \beta}$  is itself defined on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , the same probability space on which the limiting process  $\mathbb{U}^{\alpha, \beta}$  is defined as

$$(6) \quad \mathbb{U}_t^{\alpha, \beta} := \int_{\mathbb{R}_+ \times \Omega'} \mathbb{1}_{\{N'(tr)(\omega') \text{ odd}\}} \mathcal{M}_{\alpha, \beta}(dr, d\omega'), \quad t \geq 0.$$

(See (7) below for the characteristic function of finite-dimensional distributions of  $\mathbb{U}^{\alpha, \beta}$ .)

The process  $\mathbb{U}^{\alpha, \beta}$  is, to the best of our knowledge, a new class of sssi SaS processes, with self-similarity index  $\beta/\alpha$ . In particular, we shall show that its increment process is driven by a positive flow [21, 22]. The following is the main result of the paper; we use the notation  $\xrightarrow{f.d.d.}$  for convergence in finite-dimensional distributions.

**Theorem 1.** *Under the assumptions (1) and (5), with  $b_n = (n^\beta L(n))^{1/\alpha}$ ,*

$$\left( \frac{U_{[nt]}}{b_n} \right)_{t \in [0, 1]} \xrightarrow{f.d.d.} \sigma_\varepsilon \left( \mathbb{U}_t^{\alpha, \beta} \right)_{t \in [0, 1]},$$

where  $\sigma_\varepsilon^\alpha = C_\varepsilon \int_0^\infty x^{-\alpha} \sin x \, dx$ . If, in addition,  $\alpha \in (0, 1)$ , then convergence in distribution in the Skorohod  $J_1$  topology on  $D([0, 1])$  also holds.

We will prove this theorem by first conditioning on the urn sampling sequence  $(Y_n)_{n \geq 1}$ . It turns out that the characteristic function of finite-dimensional distributions of  $U$  can be expressed in terms of certain statistics of that sequence. The same idea can be used in the case of a bounded  $\varepsilon_1$  (although the proof in [8] was different and actually more involved as the results are stronger; see also [9] for the same idea applied to a generalization of Karlin model). Indeed, in this case the limit process is Gaussian, so one addresses the convergence of the covariance function by essentially examining the joint even/odd-occupancies *at two different time points*. In the present case the limit is a stable process and, hence, no longer characterized by bivariate distributions. Therefore, as an intermediate

step, we have to examine the joint even/odd-occupancies *at multiple time points*  $(n_1, \dots, n_d)$ . For this purpose we investigate the following multiparameter process

$$M_{n_1, \dots, n_d}^{\delta_1, \dots, \delta_d} = \sum_{k \geq 1} \mathbb{1}_{\{Y_{n_1, k} = \delta_1 \bmod 2, \dots, Y_{n_d, k} = \delta_d \bmod 2\}},$$

with  $n_1, \dots, n_d \in \mathbb{N}_0 = \{0\} \cup \mathbb{N}$  and  $(\delta_1, \dots, \delta_d) \in \{0, 1\}^d \setminus \{(0, \dots, 0)\}$ . We refer to this process as the *multiparameter even/odd-occupancy process*. A weak law of large numbers for the process  $M$  will turn out to be sufficient to prove the first part of Theorem 1. However, we will establish a functional central limit theorem for the multiparameter even/odd-occupancy process (Theorem 2 below). The limit in that result can be viewed as a multiparameter generalization of the bi-fractional Brownian motion [12] and, hence, is of interest on its own. There is a huge literature on limit theorems for various counting statistics of the Karlin model (see e.g. [2, 11] and references therein). However, the investigation of multiparameter even/odd-occupancy process above seems to be new.

The paper is organized as follows. Section 2 reviews the background on flow representation of stationary stable processes and shows that the increment process of  $\mathbb{U}^{\alpha, \beta}$  is driven by a positive flow. Section 3 establishes limit theorems for the multiparameter even/odd-occupancy process  $M$ . Section 4 presents the proof of main result.

## 2. A NEW CLASS OF SELF-SIMILAR STABLE PROCESSES WITH STATIONARY INCREMENTS

We start by verifying the self-similarity of the process  $\mathbb{U}^{\alpha, \beta}$  introduced in (6). It will also follow from Theorem 1 and the Lamperti theorem (see e.g. [23]), but a direct argument is simple.

**Proposition 1.** *The process  $\mathbb{U}^{\alpha, \beta}$  is  $(\beta/\alpha)$ -self-similar. That is,*

$$\left(\mathbb{U}_{\lambda t}^{\alpha, \beta}\right)_{t \geq 0} \stackrel{f.d.d.}{=} \lambda^{\beta/\alpha} \left(\mathbb{U}_t^{\alpha, \beta}\right)_{t \geq 0} \quad \text{for all } \lambda > 0.$$

*Proof.* Fix  $\lambda > 0$ . For any  $d \in \mathbb{N}$ ,  $t_1, \dots, t_d \geq 0$ ,  $a_1, \dots, a_d \in \mathbb{R}$ ,

$$\begin{aligned} \mathbb{E} \exp \left( i \sum_{k=1}^d a_k \mathbb{U}_{\lambda t_k}^{\alpha, \beta} \right) &= \exp \left( - \int_0^\infty \int_{\Psi'} \sum_{k=1}^d \left( a_k \mathbb{1}_{\{N'(\lambda t_k r)(\omega') \text{ odd}\}} \right) \mathbb{P}'(d\omega') \beta r^{-\beta-1} dr \right) \\ &= \exp \left( - \int_0^\infty \int_{\Psi'} \sum_{k=1}^d \left( a_k \mathbb{1}_{\{N'(t_k r)(\omega') \text{ odd}\}} \right) \mathbb{P}'(d\omega') \beta r^{-\beta-1} dr \right) \\ &= \mathbb{E} \exp \left( -i \sum_{k=1}^d a_k \lambda^{\beta/\alpha} \mathbb{U}_{t_k}^{\alpha, \beta} \right) \end{aligned}$$

as required.  $\square$

We now consider the increment process of  $\mathbb{U}^{\alpha, \beta}$ . We will see that the increment process is stationary (which, of course, will follow from Theorem 1 as well). More importantly, we will classify the flow structure of this process in the spirit of Rosiński [21] and Samorodnitsky [22]. We start with a background on the flow structure of stationary SaS processes. It was shown by Rosiński [21] that, given a stationary SaS process  $X = (X_t)_{t \in \mathbb{R}}$  there exist a measurable space  $(S, \mathcal{S}, \mu)$ , a non-singular flow  $(T_t)_{t \in \mathbb{R}}$  on it, and a function  $f \in L^\alpha(S, \mu)$ , such that

$$(8) \quad (X_t)_{t \in \mathbb{R}} \stackrel{f.d.d.}{=} \int_{\mathbb{S}} \left( c_t(s) f \circ T_t(s) \left( \frac{d\mu \circ T_t}{d\mu}(s) \right)^{1/\alpha} \mathcal{M}_\alpha(ds) \right)_{t \in \mathbb{R}},$$

where  $\mathcal{M}_\alpha$  is an S $\alpha$ S random measure on  $(S, \mathcal{S})$  with control measure  $\mu$ , and  $(c_t)_{t \in \mathbb{R}}$  is a  $\pm 1$ -valued cocycle with respect to  $(T_t)_{t \in \mathbb{R}}$ . Recall that a non-singular flow  $(T_t)_{t \in \mathbb{R}}$  is a group of measurable maps from  $S$  onto  $S$  such that for all  $t \in \mathbb{R}$  the measure  $\mu \circ T$  is equivalent to  $\mu$ . A cocycle  $(c_t)_{t \in \mathbb{R}}$  is a family of measurable functions on  $(S, \mathcal{S})$  such that for all  $t_1, t_2 \in \mathbb{R}$ ,  $c_{t_1+t_2}(s) = c_{t_1}(s)c_{t_2} \circ T_{t_1}(s)$   $\mu$ -almost everywhere. See Krengel [14] and Aaronson [1] for more information. In particular,  $T$  induces a unique decomposition of  $(S, \mathcal{S})$  modulo  $\mu$ ,

$$S = P \cup CN \cup D,$$

where  $P, CN, D$  are disjoint  $T$ -invariant measurable subsets of  $S$  and, restricted to each subset (if non-empty),  $T$  is positive, conservative null and dissipative, respectively. This decomposition generates a unique in law decomposition of the process  $X$  in (8) into a sum of 3 independent stationary S $\alpha$ S processes,  $X = X^P + X^{CN} + X^D$ , where  $X^P$  corresponds to a positive flow,  $X^{CN}$  corresponds to a conservative null flow, and  $X^D$  corresponds to a dissipative flow, with one or two of the components, possibly, vanishing. In fact, one can define all 3 processes as in (8), but integrating over  $P, CN, D$  correspondingly. It is known that  $X^P$  is non-ergodic,  $X^D$  is mixing, while  $X^{CN}$  is ergodic, and can be either mixing or non-mixing; see [21–23]. The processes generated by a dissipative flow have necessarily a mixed moving-average representation. The least understood family of processes are those generated by a conservative null flow.

We will show that the increment process of  $\mathbb{U}^{\alpha, \beta}$  is generated by a positive flow. In fact, a stationary S $\alpha$ S process  $X$  is generated by positive flow if and only if it can be represented as in (8), where now the control measure  $\mu$  is a probability measure invariant under action of the operators  $(T_t)_{t \in \mathbb{R}}$  (so that the factor  $(d\mu \circ T_t/d\mu)^{1/\alpha}$  disappears); see [22, Remark 2.6].

For this purpose, we first present a natural extension of  $\mathbb{U}^{\alpha, \beta}$  to a stochastic process indexed by  $t \in \mathbb{R}$ . We may and will assume that  $\Omega'$  is the space of Radon measures on  $\mathbb{R}$  equipped with the Borel  $\sigma$ -field corresponding to the topology of vague convergence,  $\mathbb{P}'$  is the law of the unit rate Poisson point process on  $\mathbb{R}$  and

$$N'(t)(\omega') := \begin{cases} \phi'([0, t]) & t \geq 0 \\ \phi'([t, 0)) & t < 0. \end{cases}$$

In this way, we now define

$$\mathbb{U}_t^{\alpha, \beta} := \iint_{\mathbb{R}_+ \times \Omega'} \mathbb{1}_{\{N'(tr)(\omega') \text{ odd}\}} \mathcal{M}_{\alpha, \beta}(dr, d\omega'), \quad t \in \mathbb{R}.$$

This definition extends (6).

**Proposition 2.** *The increment process of  $\mathbb{U}^{\alpha, \beta}$  defined as*

$$X_t := \mathbb{U}^{\alpha, \beta}(t+1) - \mathbb{U}^{\alpha, \beta}(t), \quad t \in \mathbb{R},$$

*is stationary and driven by a positive flow.*

*Proof.* Let  $m_\beta$  denote the measure  $\beta r^{-\beta-1} dr$  on  $\mathbb{R}_+$ . It follows from the stochastic integral representation of  $\mathbb{U}^{\alpha, \beta}$  that,

$$(9) \quad (X_t)_{t \in \mathbb{R}} \stackrel{f.d.d.}{=} \left( \iint_{\mathbb{R}_+ \times \Omega'} (\mathbb{1}_{\{N'((t+1)r)(\omega') \text{ odd}\}} - \mathbb{1}_{\{N'(tr)(\omega') \text{ odd}\}}) \mathcal{M}_{\alpha, \beta}(dr, d\omega') \right)_{t \in \mathbb{R}},$$

where  $\mathcal{M}_{\alpha, \beta}$  is the S $\alpha$ S random measure described above.

Let  $\theta_t$  be the standard left shift on the space of Radon measures on  $\mathbb{R}$ ,  $t \in \mathbb{R}$ , and define a group of measurable operators on  $\mathbb{R}_+ \times \Omega'$  by

$$T_t(r, \omega') := (r, \theta_{tr}(\omega')), \quad t \in \mathbb{R}.$$

If  $\nu$  is a probability measure on  $\mathbb{R}_+$  equivalent to  $m_\beta$  then the probability measure  $\mu = \nu \times \mathbb{P}'$  on  $\mathbb{R}_+ \times \Omega'$  is preserved by the operators  $(T_t)_{t \in \mathbb{R}}$ . Denote  $h = dm_\beta/d\nu$  and define

$$f(r, \omega') = h(r)^{1/\alpha} \mathbb{1}_{\{N'(r)(\omega') \text{ odd}\}},$$

and

$$c_t(r, \omega') = \mathbb{1}_{\{N'(tr)(\omega') \text{ even}\}} - \mathbb{1}_{\{N'(tr)(\omega') \text{ odd}\}}, \quad (r, \omega') \in \mathbb{R}_+ \times \Omega'.$$

If  $\mathcal{M}_\alpha$  is an S $\alpha$ S random measure on  $\mathbb{R}_+ \times \Omega'$  with control measure  $\mu$ , then, in law, (9) is the same as

$$(X_t)_{t \in \mathbb{R}} \stackrel{f.d.d.}{=} \int_{\mathbb{R}_+ \times \Omega'} c_t(r, \omega') f \circ T_t(r, \omega') \mathcal{M}_\alpha(dr, d\omega') \Bigg|_{t \in \mathbb{R}}.$$

Since  $(c_t)_{t \in \mathbb{R}}$  is, clearly, a  $\pm 1$ -valued cocycle, this will establish both stationarity of the increment process and the fact that it is driven by a positive flow once we check that the function  $f \in L^\alpha(\mu)$ . However,

$$\begin{aligned} \int |f|^\alpha d\mu &= \int \left( \mathbb{1}_{\{N'(r)(\omega') \text{ odd}\}} m_\beta(dr) \mathbb{P}'(d\omega') \right) \\ &= \int_0^\infty \beta r^{-\beta-1} \mathbb{P}'(N'(r) \text{ odd}) dr \\ &= \int_0^\infty \beta r^{-\beta-1} \frac{1}{2} (1 - e^{-2r}) dr = \Gamma(1 - \beta) 2^{\beta-1} < \infty. \end{aligned}$$

□

### 3. THE MULTIPARAMETER EVEN/ODD-OCCUPANCY PROCESS

Throughout, for  $d \in \mathbb{N}$ ,  $\mathbf{t} = (t_1, \dots, t_d) \in [0, 1]^d$ , we write

$$\lfloor n\mathbf{t} \rfloor = (\lfloor nt_1 \rfloor, \dots, \lfloor nt_d \rfloor) = (n_1, \dots, n_d)$$

and denote

$$\Lambda_d = \{0, 1\}^d \setminus \{(0, \dots, 0)\}.$$

Let  $\boldsymbol{\delta} \in \Lambda_d$  and consider the multiparameter even/odd-occupancy process

$$M_{\lfloor n\mathbf{t} \rfloor}^{\boldsymbol{\delta}} := \sum_{k=1}^\infty \prod_{j=1}^d \mathbb{1}_{\{Y_{n_j, k} = \delta_j \bmod 2\}}, \quad n \in \mathbb{N}, \mathbf{t} \in [0, 1]^d.$$

Let  $\mathbb{M}^{\boldsymbol{\delta}} = (\mathbb{M}_{\mathbf{t}}^{\boldsymbol{\delta}})_{\mathbf{t} \in [0, 1]^d}$  be a centered Gaussian random field with covariance function

$$\text{Cov}(\mathbb{M}_{\mathbf{t}}^{\boldsymbol{\delta}}, \mathbb{M}_{\mathbf{s}}^{\boldsymbol{\delta}}) = \int_0^\infty \text{Cov} \left( \mathbb{1}_{\{\bar{N}(r\mathbf{t}) = \boldsymbol{\delta} \bmod 2\}}, \mathbb{1}_{\{\bar{N}(r\mathbf{s}) = \boldsymbol{\delta} \bmod 2\}} \right) \beta r^{-\beta-1} dr,$$

where  $N$  is a standard Poisson process on  $\mathbb{R}_+$  and

$$(10) \quad \left\{ \bar{N}(n\mathbf{t}) = \boldsymbol{\delta} \bmod 2 \right\} \stackrel{\text{def}}{=} \{N(nt_j) = \delta_j \bmod 2 \text{ for all } j = 1, \dots, d\}.$$

The next result is a limit theorem for  $M_{\lfloor n\mathbf{t} \rfloor}^{\boldsymbol{\delta}}$ . It uses the normalization  $d_n = b_n^\alpha = n^\beta L(n)$ , where  $b_n$  is as in Theorem 1. We use the new notation to emphasize the fact that the normalization in Theorem 2 does not depend on  $\alpha$ .

**Theorem 2.** *Under the assumption (1), for all  $d \in \mathbb{N}$ ,  $\mathbf{t} \in [0, 1]^d$ ,  $\boldsymbol{\delta} \in \Lambda_d$ ,*

$$\lim_{n \rightarrow \infty} \frac{M_{[nt]}^{\boldsymbol{\delta}}}{d_n} = m_{\mathbf{t}}^{\boldsymbol{\delta}}$$

*in probability, where*

$$m_{\mathbf{t}}^{\boldsymbol{\delta}} := \lim_{n \rightarrow \infty} \frac{\mathbb{E} M_{[nt]}^{\boldsymbol{\delta}}}{d_n} = \int_0^\infty \mathbb{P} \left( \vec{N}(r\mathbf{t}) = \boldsymbol{\delta} \bmod 2 \right) \beta r^{-\beta-1} dr.$$

*Moreover,*

$$\frac{M_{[nt]}^{\boldsymbol{\delta}} - \mathbb{E} M_{[nt]}^{\boldsymbol{\delta}}}{\sqrt{d_n}} \Bigg|_{\mathbf{t} \in [0,1]^d} \Rightarrow (\mathbb{M}_{\mathbf{t}}^{\boldsymbol{\delta}})_{\mathbf{t} \in [0,1]^d}$$

*in  $D([0, 1]^d)$  with respect to the  $J_1$ -topology, and the limiting random field has a version with continuous sample paths.*

**Remark 1.** For the weak convergence in Theorem 2 (and in Proposition 3 below), we shall prove that it holds with respect to any topology generated by a complete separable metric on  $D([0, 1]^d)$  weaker than the uniform metric.

Only the first part of Theorem 2 is needed for Theorem 1. However, the Gaussian random field  $\mathbb{M}_{\mathbf{t}}^{\boldsymbol{\delta}}$  is of interest on its own. In fact, it was shown in [8, Theorem 2.3] that, when  $d = 1$ , the process  $M_n^1$  (which is simply the odd-occupancy process) satisfies the weak convergence in Theorem 2 with the limiting process  $\mathbb{M}^1$  being, up to a multiplicative constant, the bi-fractional Brownian motion [12, 15] with parameters  $H = 1/2, K = \beta$ . This is a centered Gaussian process with covariance function

$$\text{Cov}(\mathbb{M}_t^1, \mathbb{M}_s^1) = \Gamma(1 - \beta) 2^{\beta-2} \left( (s+t)^\beta - |s-t|^\beta \right) \quad \left( \begin{array}{l} s, t \geq 0. \end{array} \right)$$

Therefore, the limit obtained in Theorem 2 can be viewed as a random field generalization of the bi-fractional Brownian motion.

In order to analyze the multiparameter even/odd-occupancy process  $M_{[nt]}^{\boldsymbol{\delta}}$ , we introduce a Poissonization of the underlying urn sampling sequence  $(Y_n)_{n \in \mathbb{N}}$ . Let  $N$  be a standard Poisson process independent of  $(Y_n)_{n \in \mathbb{N}}$  and  $(\varepsilon_n)_{n \in \mathbb{N}}$ . Set

$$N_k(t) := \sum_{i=1}^{N(t)} \mathbb{1}_{\{Y_i=k\}}, \quad k \in \mathbb{N}, t \geq 0.$$

Clearly  $(N_k)_{k \in \mathbb{N}}$  are independent Poisson processes with respective parameters  $(p_k)_{k \in \mathbb{N}}$ . We use the notation  $\{\vec{N}_k(\mathbf{t}) = \boldsymbol{\delta} \bmod 2\}$  whose meaning is analogous to (10), and the Poissonized version of the multiparameter even/odd-occupancy process is

$$\widetilde{M}_{\mathbf{t}}^{\boldsymbol{\delta}} := \sum_{k=1}^{\infty} \mathbb{1}_{\{\vec{N}_k(\mathbf{t}) = \boldsymbol{\delta} \bmod 2\}}, \quad \mathbf{t} \in \mathbb{R}_+^d.$$

**Lemma 1.** *Under the assumption (1), for all  $d \in \mathbb{N}$ ,  $\mathbf{t} \in [0, 1]^d$ ,  $\boldsymbol{\delta} \in \Lambda_d$ ,*

$$\lim_{n \rightarrow \infty} \frac{\widetilde{M}_{nt}^{\boldsymbol{\delta}}}{d_n} = m_{\mathbf{t}}^{\boldsymbol{\delta}} \quad \text{in } L^2.$$

*Proof.* Consider a Radon measure  $\nu$  on  $\mathbb{R}_+$  defined by  $\nu := \sum_{k=1}^{\infty} \delta_{1/p_k}$ . By the assumption (1), we have  $\nu(x) := \nu([0, x]) = x^\beta L(x)$ , for all  $x > 0$ . Observe that

$$\begin{aligned} \mathbb{E} \widetilde{M}_{n\mathbf{t}}^\delta &= \sum_{k=1}^{\infty} \mathbb{E} \mathbb{1}_{\{\vec{N}_k(n\mathbf{t}) = \delta \bmod 2\}} = \sum_{k=1}^{\infty} \mathbb{P} \left( \vec{N}(np_k \mathbf{t}) = \delta \bmod 2 \right) \\ &= \int_0^\infty \mathbb{P} \left( \vec{N}(n\mathbf{t}/x) = \delta \bmod 2 \right) \nu(dx) = \int_0^\infty \varphi_{\mathbf{t}} \left( \frac{n}{x} \right) \nu(dx) \end{aligned}$$

with  $\varphi_{\mathbf{t}}(s) := \mathbb{P}(\vec{N}(s\mathbf{t}) = \delta \bmod 2)$ . It is easy to see that  $\varphi_{\mathbf{t}}$  is differentiable and vanishes at zero. Integrating by parts gives us

$$\int_0^\infty \varphi_{\mathbf{t}} \left( \frac{n}{x} \right) \nu(dx) = \int_0^\infty \frac{1}{x^2} \varphi'_{\mathbf{t}} \left( \frac{1}{x} \right) \nu(nx) dx.$$

Assume, without loss of generality, that  $t_1 < \dots < t_d$  and  $\delta_1 = 1$ . We have the following explicit expression for  $\varphi_{\mathbf{t}}$ ,

$$\begin{aligned} \varphi_{\mathbf{t}}(s) &= \mathbb{P}(N(st_1) \text{ odd}) \prod_{k=2}^d \mathbb{P}(N(s(t_k - t_{k-1})) = |\delta_k - \delta_{k-1}| \bmod 2) \\ &= \frac{1 - e^{-2st_1}}{2^d} \prod_{k=2}^d \left[ 1 + (-1)^{|\delta_k - \delta_{k-1}|} e^{-2s(t_k - t_{k-1})} \right], \end{aligned}$$

from which we can easily deduce that for  $\mathbf{t}$  fixed,  $\varphi'_{\mathbf{t}}(s)$  is bounded and there exist  $T > 0$  and  $C > 0$  such that  $|\varphi'_{\mathbf{t}}(s)| \leq Ce^{-sT}$  for all  $s > 0$ . A standard argument using the Potter bounds ([23, Corollary 10.5.8]) tells us that

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{\nu(n)} \int_0^\infty \varphi_{\mathbf{t}} \left( \frac{n}{x} \right) \nu(dx) &= \int_0^\infty \frac{1}{x^2} \varphi'_{\mathbf{t}} \left( \frac{1}{x} \right) \left( \lim_{n \rightarrow \infty} \frac{\nu(nx)}{\nu(n)} dx \right) \\ &= \int_0^\infty \varphi'_{\mathbf{t}} \left( \frac{1}{x} \right) x^{\beta-2} dx = \int_0^\infty \varphi_{\mathbf{t}} \left( \frac{1}{x} \right) \beta x^{\beta-1} dx \\ &= \int_0^\infty \mathbb{P}(\vec{N}(tr) = \delta \bmod 2) \beta r^{-\beta-1} dr. \end{aligned}$$

Since

$$\text{Var}(\widetilde{M}_{n\mathbf{t}}^\delta) = \sum_{k=1}^{\infty} \text{Var} \left( \mathbb{1}_{\{\vec{N}_k(n\mathbf{t}) = \delta \bmod 2\}} \right) \leq \sum_{k=1}^{\infty} \mathbb{P}(\vec{N}_k(n\mathbf{t}) = \delta \bmod 2) = \mathbb{E} \widetilde{M}_{n\mathbf{t}}^\delta,$$

the  $L^2$  convergence follows.  $\square$

**Proposition 3.** *Under the assumption (1), for all  $d \in \mathbb{N}$ ,  $\delta \in \Lambda_d$ ,*

$$\left( \frac{\widetilde{M}_{n\mathbf{t}}^\delta - \mathbb{E} \widetilde{M}_{n\mathbf{t}}^\delta}{\sqrt{d_n}} \right)_{\mathbf{t} \in [0,1]^d} \Rightarrow (\mathbb{M}_{\mathbf{t}}^\delta)_{\mathbf{t} \in [0,1]^d}$$

*in  $D([0,1]^d)$  with respect to the  $J_1$ -topology, and the limiting random field has a version with continuous sample paths.*



*Proof.* We continue to use the notation in the proof of Lemma 1. For  $\mathbf{s}, \mathbf{t} \in [0, 1]^d$ ,

$$\begin{aligned} \text{Cov}(\widetilde{M}_{n\mathbf{t}}^\delta, \widetilde{M}_{n\mathbf{s}}^\delta) &= \sum_{k=1}^{\infty} \text{Cov} \left( \mathbb{1}_{\{\bar{N}_k(n\mathbf{t})=\delta \bmod 2\}}, \mathbb{1}_{\{\bar{N}_k(n\mathbf{s})=\delta \bmod 2\}} \right) \\ &= \int_0^\infty \text{Cov} \left( \mathbb{1}_{\{\bar{N}(nt/x)=\delta \bmod 2\}}, \mathbb{1}_{\{\bar{N}(ns/x)=\delta \bmod 2\}} \right) \nu(dx). \end{aligned}$$

Since  $d_n = \nu(n)$ , we can use the same argument as in the proof of Lemma 1 to show that

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{\text{Cov}(\widetilde{M}_{n\mathbf{t}}^\delta, \widetilde{M}_{n\mathbf{s}}^\delta)}{d_n} &= \int_0^\infty \text{Cov} \left( \mathbb{1}_{\{\bar{N}(rt)=\delta \bmod 2\}}, \mathbb{1}_{\{\bar{N}(rs)=\delta \bmod 2\}} \right) \beta r^{-\beta-1} dr \\ &= \text{Cov}(\mathbb{M}_{\mathbf{t}}^\delta, \mathbb{M}_{\mathbf{s}}^\delta). \end{aligned}$$

Write

$$\overline{M}_{n\mathbf{t}}^\delta = \widetilde{M}_{n\mathbf{t}}^\delta - \mathbb{E}\widetilde{M}_{n\mathbf{t}}^\delta.$$

For each  $n$ ,  $\overline{M}_{n\mathbf{t}}^\delta$  is the sum of independent random variables that are centered and uniformly bounded (by 2). Further,  $\sqrt{d_n} \rightarrow \infty$  as  $n \rightarrow \infty$ . Therefore, the Lindeberg–Feller condition holds. Together with the Cramér–Wold’s device, this shows the convergence of the finite-dimensional distributions in the statement of the proposition.

It remains to prove the tightness. Letting

$$A_\eta := \left\{ (\mathbf{t}, \mathbf{t}') \in [0, 1]^d \times [0, 1]^d : \max_{j=1, \dots, d} |t_j - t'_j| \leq \eta \right\}$$

it is enough to prove that for all  $\epsilon > 0$ ,

$$(11) \quad \lim_{\eta \downarrow 0} \limsup_{n \rightarrow \infty} \mathbb{P} \left( \sup_{(\mathbf{t}, \mathbf{t}') \in A_\eta} \frac{|\overline{M}_{n\mathbf{t}}^\delta - \overline{M}_{n\mathbf{t}'}^\delta|}{\sqrt{d_n}} > \epsilon \right) = 0.$$

It is easy to see that it suffices to show (11) with  $A_\eta$  replaced by

$$A_\eta^{(i)} := \left\{ (\mathbf{t}, \mathbf{t}') \in [0, 1]^d \times [0, 1]^d : 0 \leq t_i \leq t'_i \leq t_i + \eta, t_j = t'_j, j \neq i \right\},$$

for all  $i = 1, \dots, d$ , and we prove (11) for  $A_\eta^{(1)}$ . By [8, Lemma 3.7] which is due to a chaining argument, it suffices to establish the following: for all  $(\mathbf{t}, \mathbf{t}') \in A_\eta^{(1)}$ ,

$$(12) \quad \overline{M}_{n\mathbf{t}}^\delta - \overline{M}_{n\mathbf{t}'}^\delta \leq N(n(t_1 + \eta)) - N(nt_1) + n\eta \quad \text{almost surely,}$$

and for all  $p \in \mathbb{N}$  and  $\gamma \in (0, \beta)$ , there exists  $C_{p, \gamma} > 0$  such that for all  $(\mathbf{t}, \mathbf{t}') \in A_\eta^{(1)}$ ,

$$(13) \quad \mathbb{E} |\overline{M}_{n\mathbf{t}}^\delta - \overline{M}_{n\mathbf{t}'}^\delta|^{2p} \leq C_{p, \gamma} (|t_1 - t'_1|^{\gamma p} \nu(n)^p + |t_1 - t'_1|^\gamma \nu(n)).$$

Fix  $(\mathbf{t}, \mathbf{t}') \in A_\eta^{(1)}$  and to simplify the notation introduce  $B_k = \{N_k(n\mathbf{t}) = \delta \bmod 2\}$  and  $B'_k = \{N_k(n\mathbf{t}') = \delta \bmod 2\}$ . We have

$$B_k \Delta B'_k \subset \{N_k(nt'_1) - N_k(nt_1) \neq 0\}.$$

To show (12), it suffices to observe that

$$\begin{aligned} \overline{M}_{n\mathbf{t}}^\delta - \overline{M}_{n\mathbf{t}'}^\delta &= \sum_{k=1}^{\infty} (\mathbb{1}_{B_k} - \mathbb{1}_{B'_k} - \mathbb{P}(B_k) + \mathbb{P}(B'_k)) \left( \sum_{k=1}^{\infty} \mathbb{1}_{B_k \Delta B'_k} + \sum_{k=1}^{\infty} \mathbb{P}(B_k \Delta B'_k) \right) \\ &\leq N(nt'_1) - N(nt_1) + \mathbb{E}N(n(t'_1 - t_1)) \leq N(nt'_1) - N(nt_1) + n\eta. \end{aligned}$$

To show (13), by Rosenthal's inequality [20], for some constant  $C$  depending on  $p \in \mathbb{N}$  only,

$$\begin{aligned} & \mathbb{E} \left| \overline{M}_{nt}^\delta - \overline{M}_{nt}^\delta \right|^{2p} \\ & \leq C \left[ \sum_{k=1}^{\infty} \mathbb{E} \left( \mathbb{1}_{B_k} - \mathbb{1}_{B'_k} - \mathbb{P}(B_k) + \mathbb{P}(B'_k) \right)^{2p} + \sum_{k=1}^{\infty} \mathbb{E} \left( \mathbb{1}_{B_k} - \mathbb{1}_{B'_k} - \mathbb{P}(B_k) + \mathbb{P}(B'_k) \right)^2 \right]^p \end{aligned}$$

Since

$$\mathbb{E} |\mathbb{1}_A - \mathbb{1}_B - \mathbb{P}(A) + \mathbb{P}(B)|^{2p} \leq 2^{2p-1} \text{Var}(\mathbb{1}_A - \mathbb{1}_B) \leq 2^{2p-1} \mathbb{P}(A \Delta B),$$

the above expression does not exceed

$$C \left[ \sum_{k=1}^{\infty} \mathbb{P}(B_k \Delta B'_k) + \sum_{k=1}^{\infty} \mathbb{P}(B_k \Delta B'_k) \right]^p \leq C [V(n(t'_1 - t_1)) + V(n(t'_1 - t_1))]^p$$

with

$$V(t) = \sum_{k=1}^{\infty} \mathbb{P}(N_k(t) \neq 0) = \sum_{k=1}^{\infty} (1 - e^{-p_k t}).$$

By [8, Lemma 3.1], for each  $\gamma \in (0, \beta)$ , there exists  $C_\gamma$  such that

$$V(nt) \leq C_\gamma t^\gamma \nu(n) \text{ for all } t \in [0, 1], n \in \mathbb{N}.$$

Therefore, (13) follows. Therefore, we have proved (11), which also implies that the limit process has a version with continuous sample path [26, Theorem 5.6]. This completes the proof.  $\square$

*Proof of Theorem 2.* We will prove the second part of the theorem using a multivariate version of the change-of-time lemma from [6, p. 151] (the proof of which is the same as that of the univariate version). Since the limiting random field is continuous, the first part will then follow.

Let  $(\tau_n)_{n \in \mathbb{N}_0}$  be the arrival times of the Poisson process  $N$  (with  $\tau_0 = 0$ ). For  $\mathbf{t} \in \mathbb{R}_+^d$ , set  $\boldsymbol{\tau}_{\lfloor n\mathbf{t} \rfloor} = (\tau_{\lfloor nt_1 \rfloor}, \dots, \tau_{\lfloor nt_d \rfloor}) \in \mathbb{R}_+^d$ . By the strong law of large numbers and monotonicity,

$$\lambda_n \equiv \left( \frac{\boldsymbol{\tau}_{\lfloor n\mathbf{t} \rfloor}}{n} \wedge \mathbf{2} \right) \Big|_{\mathbf{t} \in [0, 2]^d} \rightarrow \mathbf{id} \quad \text{almost surely,}$$

as  $n \rightarrow \infty$ , where  $\mathbf{id}$  is the identity function from  $[0, 2]^d$  to  $[0, 2]^d$ . By the multivariate change-of-time lemma,

$$\left( \frac{\overline{M}_{n((\boldsymbol{\tau}_{\lfloor n\mathbf{t} \rfloor}/n) \wedge \mathbf{2})}^\delta}{\sqrt{d_n}} \right) \Big|_{\mathbf{t} \in [0, 2]^d} \Rightarrow (\mathbb{M}_{\mathbf{t}}^\delta)_{\mathbf{t} \in [0, 2]^d}.$$

In particular, the convergence also holds if the random fields are restricted to  $\mathbf{t} \in [0, 1]^d$ . Since

$$\left( \frac{M_{\lfloor n\mathbf{t} \rfloor}^\delta - \mathbb{E} M_{\lfloor n\mathbf{t} \rfloor}^\delta}{\sqrt{d_n}} \right) \Big|_{\mathbf{t} \in [0, 1]^d} = \left( \frac{\overline{M}_{n(\boldsymbol{\tau}_{\lfloor n\mathbf{t} \rfloor}/n)}^\delta}{\sqrt{d_n}} \right)_{\mathbf{t} \in [0, 1]^d},$$

and

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( \left( \frac{\boldsymbol{\tau}_{\lfloor n\mathbf{t} \rfloor}}{n} \wedge \mathbf{2} \right) \Big|_{\mathbf{t} \in [0, 1]^d} = \left( \frac{\boldsymbol{\tau}_{\lfloor n\mathbf{t} \rfloor}}{n} \right) \Big|_{\mathbf{t} \in [0, 1]^d} \right) = 1,$$

the desired result follows.  $\square$

## 4. PROOF OF THEOREM 1

*Proof of convergence of finite-dimensional distributions.* Let  $d \geq 1$  and fix  $\mathbf{t} = (t_1, \dots, t_d) \in [0, 1]^d$  and  $\mathbf{a} = (a_1, \dots, a_d) \in \mathbb{R}^d$ . Note that

$$\begin{aligned}
 (14) \quad \mathbb{E} \exp \left( i \sum_{j=1}^d a_j \mathbb{U}_{t_j}^{\alpha, \beta} \right) &= \exp \left( \int_{\mathbb{R}_+ \times \Omega'} \sum_{j=1}^d a_j \mathbb{1}_{\{N'(rt_j) \text{ odd}\}} m_\beta(dr) \mathbb{P}'(d\omega') \right) \\
 &= \exp \left( - \int_{\mathbb{R}_+ \times \Omega'} \sum_{\delta \in \Lambda_d} |\langle \mathbf{a}, \delta \rangle|^\alpha \mathbb{1}_{\{\bar{N}'(r\mathbf{t}) = \delta \bmod 2\}} \beta r^{-\beta-1} dr d\mathbb{P}' \right) \\
 &= \exp \left( - \sum_{\delta \in \Lambda_d} |\langle \mathbf{a}, \delta \rangle|^\alpha m_\delta^\delta \right).
 \end{aligned}$$

Similarly, with the notation  $n_j = \lfloor nt_j \rfloor$ ,

$$\begin{aligned}
 \mathbb{E} \exp \left( i \sum_{j=1}^d a_j \frac{U_{n_j}}{b_n} \right) &= \mathbb{E} \exp \left( i \sum_{j=1}^d \frac{a_j}{b_n} \sum_{k=1}^\infty \varepsilon_k \mathbb{1}_{\{Y_{n_j, k} \text{ odd}\}} \right) \\
 &= \mathbb{E} \exp \left( i \sum_{\delta \in \Lambda_d} \frac{\langle \mathbf{a}, \delta \rangle}{b_n} \sum_{k=1}^\infty \varepsilon_k \prod_{j=1}^d \mathbb{1}_{\{Y_{n_j, k} = \delta_j \bmod 2\}} \right).
 \end{aligned}$$

Let  $\mathcal{Y}$  denote the  $\sigma$ -algebra generated by  $(Y_i)_{i \in \mathbb{N}}$ . Then, the above expression becomes

$$\mathbb{E} \left\{ \prod_{\delta \in \Lambda_d} \mathbb{E} \left[ \exp \left( i \frac{\langle \mathbf{a}, \delta \rangle}{b_n} \sum_{\ell=1}^{M_{[nt]}^\delta} \varepsilon_\ell \right) \middle| \mathcal{Y} \right] \right\} = \mathbb{E} \prod_{\delta \in \Lambda_d} \left( \phi \left( \frac{\langle \mathbf{a}, \delta \rangle}{b_n} \right)^{M_{[nt]}^\delta} \right),$$

with  $\phi(\theta) = \mathbb{E} \exp(i\theta \varepsilon_1)$  being the characteristic function of  $\varepsilon_1$ . Therefore,

$$\begin{aligned}
 (15) \quad \mathbb{E} \exp \left( i \sum_{j=1}^d a_j \frac{U_{n_j}}{b_n} \right) &= \mathbb{E} \prod_{\delta \in \Lambda_d} \left( \phi \left( \frac{\langle \mathbf{a}, \delta \rangle}{b_n} \right)^{M_{[nt]}^\delta} \right) = \mathbb{E} \exp \left( \sum_{\delta \in \Lambda_d} M_{[nt]}^\delta \log \phi \left( \frac{\langle \mathbf{a}, \delta \rangle}{b_n} \right) \right) \\
 &= \mathbb{E} \exp \left( - \sum_{\delta \in \Lambda_d} \left( \frac{M_{[nt]}^\delta}{b_n^\alpha} \sigma_\varepsilon^\alpha |\langle \mathbf{a}, \delta \rangle|^\alpha \frac{\log \phi(c_n)}{-\sigma_\varepsilon^\alpha |c_n|^\alpha} \right) \right),
 \end{aligned}$$

with  $c_n := \langle \mathbf{a}, \delta \rangle / b_n$ . Recall that assumption (5) implies that (see e.g. [7, Theorem 8.1.10])

$$\log \phi(\theta) \sim \phi(\theta) - 1 \sim -\sigma_\varepsilon^\alpha |\theta|^\alpha \text{ as } \theta \rightarrow 0.$$

By Theorem 2 and dominated convergence theorem the expression in (15) converges to the characteristic function in (14).  $\square$

In the remainder of this section we consider the case  $\alpha \in (0, 1)$  and prove the tightness of the sequence of processes  $(U_{[nt]}/b_n)_{t \in [0, 1]}$ ,  $n \geq 1$  in the  $J_1$ -topology on the space  $D([0, 1])$ . For  $\kappa > 0$  we decompose  $U_n = U_n^{+, \kappa} + U_n^{-, \kappa}$  with

$$\begin{aligned}
 U_n^{+, \kappa} &= \sum_{k \geq 1} \varepsilon_k \mathbb{1}_{\{|\varepsilon_k| > \kappa b_n\}} \mathbb{1}_{\{Y_{n, k} \text{ odd}\}} \\
 U_n^{-, \kappa} &= \sum_{k \geq 1} \varepsilon_k \mathbb{1}_{\{|\varepsilon_k| \leq \kappa b_n\}} \mathbb{1}_{\{Y_{n, k} \text{ odd}\}}.
 \end{aligned}$$

We start by showing that for any  $\alpha \in (0, 2)$  and any  $\kappa > 0$ , the sequence of the laws of the processes  $(U_{[nt]}^{+, \kappa}/b_n)_{t \in [0, 1]}$ ,  $n \geq 1$ , is tight in the  $J_1$ -topology. To this end define

$$\mathcal{T}^{(\kappa)}(n) = \{i \in \{1, \dots, n\} : Y_i = k \text{ for } k \text{ such that } |\varepsilon_k| > \kappa b_n\}.$$

**Lemma 2.** *For any  $\kappa > 0$ ,*

$$\lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \mathbb{P} \left( \exists i_1, i_2 \in \mathcal{T}^{(\kappa)}(n) \text{ such that } |i_1 - i_2| < n\delta \right) \neq 0.$$

*Proof.* Fix  $\kappa > 0$  and let  $\epsilon > 0$ . Let  $\Theta_n = \sum_{k \geq 1} \mathbb{1}_{\{|\varepsilon_k| > \kappa b_n\}} p_k$ ,  $n \geq 1$ . We first prove that

$$(16) \quad \exists C = C(\epsilon, \kappa) < +\infty \text{ such that } \limsup_{n \rightarrow \infty} \mathbb{P}(\Theta_n > C/n) \leq \epsilon.$$

To see this, choose  $c > 0$ . Recalling the notation  $\nu(x) = x^\beta L(x)$  we consider  $\Theta_n^{(c)} = \sum_{k > \nu(cn)} \mathbb{1}_{\{|\varepsilon_k| > \kappa b_n\}} p_k$ . Note that by (5),

$$\begin{aligned} \mathbb{P}(\Theta_n^{(c)} \neq \Theta_n) &\leq \mathbb{P}(\exists k \leq \nu(cn) \text{ such that } |\varepsilon_k| > \kappa b_n) \\ &= 1 - (1 - \mathbb{P}(|\varepsilon_1| > \kappa b_n))^{\lfloor \nu(cn) \rfloor} \rightarrow 1 - e^{-c^\beta C_\epsilon \kappa^{-\alpha}}, \text{ as } n \rightarrow \infty, \end{aligned}$$

and so we can choose  $c = c(\epsilon, \kappa)$  such that

$$(17) \quad \limsup_{n \rightarrow \infty} \mathbb{P}(\Theta_n^{(c)} \neq \Theta_n) \leq \epsilon/2.$$

Further,

$$\sum_{k > \nu(cn)} p_k = \sum_{k \geq 1} p_k \mathbb{1}_{\{p_k < 1/cn\}} = \int_{(cn, \infty)} \frac{1}{x} \nu(dx) \sim \frac{\beta}{1 - \beta} (cn)^{\beta-1} L(n), \text{ as } n \rightarrow \infty,$$

where the equivalence is due to integration by parts and an application of a Karamata Theorem (see [10, Theorem 1 p. 281]). Thus,

$$\mathbb{E}\Theta_n^{(c)} \sim \frac{\beta}{1 - \beta} (cn)^{\beta-1} L(n) \mathbb{P}(|\varepsilon_1| > \kappa b_n) \sim \frac{\beta}{1 - \beta} c^{\beta-1} C_\epsilon \kappa^{-\alpha} n^{-1}, \text{ as } n \rightarrow \infty.$$

Now (16) follows from the Markov inequality and (17). By (16),

$$\begin{aligned} \limsup_{n \rightarrow \infty} \mathbb{P} \left( \exists i_1, i_2 \in \mathcal{T}^{(\kappa)}(n) \text{ such that } |i_1 - i_2| < n\delta \right) &\leq \limsup_{n \rightarrow \infty} \mathbb{P} \left( \exists i_1, i_2 \in \mathcal{T}^{(\kappa)}(n) \text{ such that } |i_1 - i_2| < n\delta \mid \Theta_n \leq C/n \right) \leq \epsilon. \end{aligned}$$

Letting  $(B_i^{(n)})_{i \in \mathbb{N}}$  be i.i.d. Bernoulli random variables with parameter  $C/n$ , we can bound the first term above by

$$\limsup_{n \rightarrow \infty} \mathbb{P} \left( \exists i_1, i_2 \in \{1, \dots, n\} \text{ such that } |i_1 - i_2| < n\delta \text{ and } B_{i_1}^{(n)} = B_{i_2}^{(n)} = 1 \right)$$

The latter probability has a limit, equal to the probability that a Poisson point process with intensity  $C$  over  $[0, 1]$  has two points less than  $\delta$  apart. This probability goes to zero as  $\delta \downarrow 0$ . This completes the proof.  $\square$

**Proposition 4.** *For all  $\alpha \in (0, 2)$ ,  $\kappa > 0$ , the sequence of processes  $(U_{[nt]}^{+, \kappa}/b_n)_{t \in [0, 1]}$ ,  $n \geq 1$ , is tight in the  $J_1$ -topology on  $D([0, 1])$ .*

*Proof.* Since  $\sup_{s \in [r, t]} |U_{[nr]}^{+, \kappa} - U_{[ns]}^{+, \kappa}| = 0$  as soon as  $\mathcal{T}^{(\kappa)}([nt]) \setminus \mathcal{T}^{(\kappa)}([nr]) = \emptyset$ , from the preceding lemma we infer that for all  $\eta > 0$ ,

$$\lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \mathbb{P} \left( \sup_{\substack{r \leq s \leq t \leq 1 \\ |r-t| < \delta}} U_{[nr]}^{+, \kappa} - U_{[ns]}^{+, \kappa} \wedge U_{[ns]}^{+, \kappa} - U_{[nt]}^{+, \kappa} > \eta \right) = 0,$$

which yields the tightness of  $(U_{[n \cdot]}^{+, \kappa})_{n \geq 1}$  (see [6]).  $\square$

*Proof of tightness of  $((U_{[nt]}/b_n)_{t \in [0,1]})$  (when  $\alpha \in (0, 1)$ ).* Let  $\alpha \in (0, 1)$ . In view of Proposition 4, it is sufficient to show that for any  $\eta > 0$ ,

$$(18) \quad \lim_{\kappa \rightarrow 0} \limsup_{n \rightarrow \infty} \mathbb{P} \left( \sup_{t \in [0,1]} \frac{U_{[nt]}^{-, \kappa}}{b_n} > \eta \right) = 0.$$

Note that

$$\begin{aligned} \mathbb{P} \left( \sup_{t \in [0,1]} \frac{U_{[nt]}^{-, \kappa}}{b_n} > \eta \right) &\leq \mathbb{P} \left( \sum_{k \geq 1} \frac{\varepsilon_k}{b_n} \mathbf{1}_{\{|\varepsilon_k| \leq \kappa b_n\}} \mathbf{1}_{\{Y_{n,k} > 0\}} > \eta \right) \\ &\leq \mathbb{P} \left( \sum_{k=1}^{K_n} \left( \frac{\varepsilon_k}{b_n} \mathbf{1}_{\{|\varepsilon_k| \leq \kappa b_n\}} \right) > \eta \right) \end{aligned}$$

where  $K_n$  is the number of nonempty boxes at time  $n$  in the infinite urn scheme. Since for large  $n$ ,

$$\begin{aligned} \mathbb{E} \left( \sum_{k=1}^{K_n} \left( \frac{\varepsilon_k}{b_n} \mathbf{1}_{\{|\varepsilon_k| \leq \kappa b_n\}} \right) \right) &= \mathbb{E}(K_n) \mathbb{E} \left( \frac{\varepsilon_1}{b_n} \mathbf{1}_{\{|\varepsilon_1| \leq \kappa b_n\}} \right) \\ &\leq C b_n^\alpha b_n^{-1} \kappa b_n \mathbb{P}(|\varepsilon_1| > \kappa b_n) \rightarrow C C_\varepsilon \kappa^{1-\alpha} \end{aligned}$$

(for a finite constant  $C$ ) by [11, Proposition 2] and Karamata's theorem, (18) follows by Markov's inequality.  $\square$

**Remark 2.** Whether or not the full weak convergence in Theorem 1 holds when  $\alpha \in [1, 2)$  remains an open question. In this case, it is not even clear to us whether  $\mathbb{U}^{\alpha, \beta}$  has a càdlàg modification: sufficient conditions are given, for example, in [4, Theorem 4.3], but they are not satisfied here.

## 5. DISCUSSIONS

There are a few limit theorems for other statistics in [8] that we have not addressed yet. We provide a brief discussions here focusing on other processes that appear in the limit. As for the proofs, they do not require new ideas (if one ignores the tightness issues).

For the odd-occupancy process  $U_n$  in (2), one can write, for  $\beta < \alpha$ ,

$$\begin{aligned} U_n &= \sum_{k=1}^{\infty} \varepsilon_k \mathbf{1}_{\{Y_{n,k} \text{ odd}\}} = \sum_{k=1}^{\infty} \left( \varepsilon_k (\mathbf{1}_{\{Y_{n,k} \text{ odd}\}} - \mathbb{P}(Y_{n,k} \text{ odd})) \right) + \sum_{k=1}^{\infty} \varepsilon_k \mathbb{P}(Y_{n,k} \text{ odd}) \\ &=: U_n^{(1)} + U_n^{(2)}, \end{aligned}$$

and one could eventually prove that

$$(19) \quad \frac{1}{b_n} \left( U_{[nt]}, U_{[nt]}^{(1)}, U_{[nt]}^{(2)} \right)_{t \in [0,1]} \xrightarrow{f.d.d.} \sigma_\varepsilon \left( \mathbb{U}_t^{\alpha, \beta}, \mathbb{U}_t^{\alpha, \beta, (1)}, \mathbb{U}_t^{\alpha, \beta, (2)} \right)$$

as  $n \rightarrow \infty$ , with

$$\begin{aligned}\mathbb{U}_t^{\alpha,\beta,(1)} &= \iint_{\mathbb{R}_+ \times \Omega'} [\mathbb{1}_{\{N'(tr)(\omega') \text{ odd}\}} - \mathbb{P}'(N'(tr) \text{ odd})] \mathcal{M}_{\alpha,\beta}(dr, d\omega') \\ \mathbb{U}_t^{\alpha,\beta,(2)} &= \iint_{\mathbb{R}_+ \times \Omega'} \mathbb{P}'(N'(tr) \text{ odd}) \mathcal{M}_{\alpha,\beta}(dr, d\omega'),\end{aligned}$$

where here and below  $\mathcal{M}_{\alpha,\beta}$  and  $N'$  are as before. We need the constraint  $\beta < \alpha$  so that  $U_n^{(1)}$ ,  $U_n^{(2)}$ ,  $\mathbb{U}_t^{\alpha,\beta,(1)}$  and  $\mathbb{U}_t^{\alpha,\beta,(2)}$  are well defined. Such a weak convergence, for the original randomized Karlin model ( $\varepsilon_k \in \{\pm 1\}$  and  $\alpha = 2$ ), has been proved in [8]. An appealing feature is that the corresponding decomposition of

$$\mathbb{U}_t^{2,\beta} = \mathbb{U}_t^{2,\beta,(1)} + \mathbb{U}_t^{2,\beta,(2)}$$

recovers a decomposition of fractional Brownian motion by a bi-fractional Brownian motion and another smooth self-similar Gaussian process discovered in [15], and in particular, in this case the two processes are independent. For  $\alpha \in (0, 2)$ , the convergence of finite-dimensional distributions to the decomposition still holds, although  $\mathbb{U}_t^{\alpha,\beta,(1)}$  and  $\mathbb{U}_t^{\alpha,\beta,(2)}$  are no longer independent. The convergence in (19) could be established by computing characteristic functions and applying the same conditioning trick.

Another statistics considered in [8, 13] is the *occupancy process*

$$Z_n := \sum_{k=1}^{\infty} \varepsilon_k \mathbb{1}_{\{Y_{n,k} > 0\}}.$$

Correspondingly, the limit process is

$$\mathbb{Z}_t^{\alpha,\beta} = \iint_{\mathbb{R}_+ \times \Omega'} \mathbb{1}_{\{N'(tr) > 0\}} \mathcal{M}_{\alpha,\beta}(dr, d\omega'), \quad t \geq 0.$$

At the same time, this is nothing but a time-changed SaS Lévy process, as one can verify by computing the characteristic functions that

$$\left(\mathbb{Z}_t^{\alpha,\beta}\right)_{t \geq 0} \stackrel{f.d.d.}{=} \left(\mathbb{Z}^\alpha(t^\beta)\right)_{t \geq 0},$$

where  $(\mathbb{Z}^\alpha(t))_{t \geq 0}$  is an SaS Lévy process ( $\mathbb{E} e^{i\theta \mathbb{Z}^\alpha(1)} = e^{-|\theta|^\alpha}$ ,  $\theta \in \mathbb{R}$ ). A similar decomposition for  $\mathbb{Z}^{\alpha,\beta}$ , and the corresponding limit theorem as in (19) can also be established, again by computing characteristic functions. The corresponding results for the Gaussian case ( $\alpha = 2$ ) have already been investigated in [8, Theorem 2.1].

**Acknowledgments.** The first author would like to thank the hospitality and financial support from Taft Research Center and Department of Mathematical Sciences at University of Cincinnati, for his visits in 2016 and 2017. The second author's research was partially supported by NSF grant DMS-1506783 and the ARO grant W911NF-12-10385 at Cornell University. The third author's research was partially supported by the NSA grants H98230-14-1-0318 and H98230-16-1-0322, the ARO grant W911NF-17-1-0006, and Charles Phelps Taft Research Center at University of Cincinnati.

## REFERENCES

- [1] Aaronson, J. (1997). *An introduction to infinite ergodic theory*, volume 50 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI.
- [2] Alsmeyer, G., Iksanov, A., and Marynych, A. (2017). Functional limit theorems for the number of occupied boxes in the Bernoulli sieve. *Stochastic Process. Appl.*, 127(3):995–1017.

- [3] Bahadur, R. R. (1960). On the number of distinct values in a large sample from an infinite discrete distribution. *Proc. Nat. Inst. Sci. India Part A*, 26(supplement II):67–75.
- [4] Basse-O'Connor, A. and Rosiński, J. (2013). On the uniform convergence of random series in Skorohod space and representations of càdlàg infinitely divisible processes. *Ann. Probab.*, 41(6):4317–4341.
- [5] Beran, J., Feng, Y., Ghosh, S., and Kulik, R. (2013). *Long-memory processes*. Springer, Heidelberg. Probabilistic properties and statistical methods.
- [6] Billingsley, P. (1999). *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition. A Wiley-Interscience Publication.
- [7] Bingham, N. H., Goldie, C. M., and Teugels, J. L. (1987). *Regular variation*, volume 27 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge.
- [8] Durieu, O. and Wang, Y. (2016). From infinite urn schemes to decompositions of self-similar Gaussian processes. *Electron. J. Probab.*, 21:Paper No. 43, 23.
- [9] Durieu, O. and Wang, Y. (2017). From random partitions to fractional Brownian sheets. *Bernoulli*, to appear. Available at <https://arxiv.org/abs/1709.00934>.
- [10] Feller, W. (1971). *An introduction to probability theory and its applications. Vol. II*. Second edition. John Wiley & Sons Inc., New York.
- [11] Gnedin, A., Hansen, B., and Pitman, J. (2007). Notes on the occupancy problem with infinitely many boxes: general asymptotics and power laws. *Probab. Surv.*, 4:146–171.
- [12] Houdré, C. and Villa, J. (2003). An example of infinite dimensional quasi-helix. In *Stochastic models (Mexico City, 2002)*, volume 336 of *Contemp. Math.*, pages 195–201. Amer. Math. Soc., Providence, RI.
- [13] Karlin, S. (1967). Central limit theorems for certain infinite urn schemes. *J. Math. Mech.*, 17:373–401.
- [14] Krengel, U. (1985). *Ergodic theorems*, volume 6 of *de Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin. With a supplement by Antoine Brunel.
- [15] Lei, P. and Nualart, D. (2009). A decomposition of the bifractional Brownian motion and some applications. *Statist. Probab. Lett.*, 79(5):619–624.
- [16] Owada, T. and Samorodnitsky, G. (2015). Functional central limit theorem for heavy tailed stationary infinitely divisible processes generated by conservative flows. *Ann. Probab.*, 43(1):240–285.
- [17] Pipiras, V. and Taqqu, M. S. (2017a). *Long-range dependence and self-similarity*, volume 45 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press.
- [18] Pipiras, V. and Taqqu, M. S. (2017b). *Stable non-Gaussian self-similar processes with stationary increments*. Springer.
- [19] Pitman, J. (2006). *Combinatorial stochastic processes*, volume 1875 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin. Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002, With a foreword by Jean Picard.
- [20] Rosenthal, H. P. (1970). On the subspaces of  $L^p$  ( $p > 2$ ) spanned by sequences of independent random variables. *Israel J. Math.*, 8:273–303.
- [21] Rosiński, J. (1995). On the structure of stationary stable processes. *Ann. Probab.*, 23(3):1163–1187.
- [22] Samorodnitsky, G. (2005). Null flows, positive flows and the structure of stationary symmetric stable processes. *Ann. Probab.*, 33(5):1782–1803.
- [23] Samorodnitsky, G. (2016). *Stochastic processes and long range dependence*. Springer, Cham, Switzerland.

- [24] Samorodnitsky, G. and Taqqu, M. S. (1994). *Stable non-Gaussian random processes*. Stochastic Modeling. Chapman & Hall, New York. Stochastic models with infinite variance.
- [25] Spitzer, F. (1964). *Principles of random walk*. The University Series in Higher Mathematics. D. Van Nostrand Co., Inc., Princeton, N.J.-Toronto-London.
- [26] Straf, M. L. (1972). Weak convergence of stochastic processes with several parameters. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory*, pages 187–221. Univ. California Press, Berkeley, Calif.

OLIVIER DURIEU, LABORATOIRE DE MATHÉMATIQUES ET PHYSIQUE THÉORIQUE, UMR-CNRS 7350, FÉDÉRATION DENIS POISSON, FR-CNRS 2964, UNIVERSITÉ FRANÇOIS-RABELAIS DE TOURS, PARC DE GRANDMONT, 37200 TOURS, FRANCE.

*E-mail address:* `olivier.durieu@univ-tours.fr`

GENNADY SAMORODNITSKY, SCHOOL OF OPERATIONS RESEARCH AND INFORMATION ENGINEERING AND DEPARTMENT OF STATISTICAL SCIENCE, CORNELL UNIVERSITY, 220 RHODES HALL, ITHACA, NY 14853, USA.

*E-mail address:* `gs18@cornell.edu`

YIZAO WANG, DEPARTMENT OF MATHEMATICAL SCIENCES, UNIVERSITY OF CINCINNATI, 2815 COMMONS WAY, CINCINNATI, OH, 45221-0025, USA.

*E-mail address:* `yizao.wang@uc.edu`