Functional central limit theorems for occupancies and missing mass process in infinite urn models

Mikhail Chebunin*     Sergei Zuyev †

June 27, 2019

Abstract

We study the infinite urn scheme when the balls are sequentially distributed over an infinite number of urns labelled 1, 2, ... so that the urn \( j \) at every draw gets a ball with probability \( p_j \), \( \sum_j p_j = 1 \). We prove functional central limit theorems for discrete time and the poissonised version for the urn occupancies process, for the odd-occupancy and for the missing mass processes extending the known non-functional central limit theorems.

Keywords: infinite urn scheme, regular variation, functional CLT, occupancy process, missing mass process.

2010 Mathematics Subject Classification. Primary: 60F17, 60G22; Secondary: 60G15, 60G18

1 Introduction

In this paper we study the following classical urn model first considered by Karlin [12]: \( n \geq 1 \) balls are distributed one by one over an infinite number of urns enumerated from 1 to infinity. The ball distributed at step \( j = 1, 2 \ldots, \)

* Sobolev Institute of Mathematics SB RAS and Novosibirsk State University, Novosibirsk, Russia, E-mail: chebuninmikhail@gmail.com
† Chalmers University of Technology, Gothenburg, Sweden. E-mail: sergei.zuyev@chalmers.se
call it $j$th ball, gets into urn $i$ with probability $p_i$, $\sum_{i=1}^{\infty} p_i = 1$, independently of the other balls. Such multinomial occupancy schemes arise in many different applications, in Biology [11], Computer science [13], [14] and in many other areas, see, e.g., [10] and the references therein.

Let $X_j$ be the urn the $j$th ball gets into and let $J_i(n)$ be the number of balls the $i$th urn contains after $n$ balls are distributed:

$$J_i(n) = \sum_{j=1}^{n} \mathbb{I}_{X_j = i}.$$  

We are particularly interested in the asymptotic distribution of the number of urns containing at least $k \geq 1$ balls and containing exactly $k$ balls:

$$R_{n,k}^* = \sum_{i=1}^{\infty} \mathbb{I}_{J_i(n) \geq k}, \quad R_{n,k} = \sum_{i=1}^{\infty} \mathbb{I}_{J_i(n) = k} = R_{n,k}^* - R_{n,k+1}^*,$$  \hspace{1cm} (1)

of the number of urns with an odd number of balls and the asymptotic behaviour of the missing mass:

$$U_n = \sum_{i=1}^{\infty} \mathbb{I}_{J_i(n) \equiv 1 \mod 2}, \quad M_n = \sum_{i=1}^{\infty} np_i \mathbb{I}_{J_i(n) = 0},$$  \hspace{1cm} (2)

We also use notation $R_n \overset{\text{def}}{=} R_{n,1}^* = \sum_{k \geq 1} R_{n,k}$ for the number of non-empty urns. Renumbering the urns if necessary, we further assume that the sequence $(p_i)_{i \geq 1}$ is monotonely decaying and regularly varying, namely,

$$\alpha(x) = \max \{ i : p_i \geq 1/x \} = x^\theta L(x) \text{ with } \theta \in [0,1],$$  \hspace{1cm} (3)

where $L(x)$ is a slowly varying function as $x \to \infty$.

Alongside the the discrete time model, we will also consider its continuous time analogue when the balls are put into urns at the times of jumps of a homogeneous Poisson point processes $\Pi(t)$, $t \geq 0$ with intensity 1 on $\mathbb{R}_+$. According to the independent marking theorem for Poisson processes, \{ $J_i(\Pi(t)) \overset{\text{def}}{=} \Pi_i(t)$, $t \geq 0$ \} are independent homogeneous Poisson processes with intensities $p_i$. To ease the notation, we will write simply

$$R(t) \overset{\text{def}}{=} R_{\Pi(t),1}^*, \quad U(t) \overset{\text{def}}{=} U_{\Pi(t)} \text{ and } M(t) \overset{\text{def}}{=} M_{\Pi(t)} = \sum_{i=1}^{\infty} tp_i \mathbb{I}_{\Pi_i(t) = 0}.$$  

2
This paper extends the results of [7] and [6], where a functional central limit theorem (FCLT) was shown under condition (3) for the vector process $(R^*[nt], R^*[nt], \ldots, R^*[nt])_t \in [0,1]$ in the case $\theta \in (0,1]$.

Ordinary (not functional) central limit theorems for the above quantities were established under various conditions in [2], [3], [9], [10], [12], [13], [14]. In particular, under rather general conditions on the sequence $(p_i)$ involving an unbounded growth of the variances, the following results available: a strong law of large numbers and asymptotic normality of $R_n$, an asymptotic normality of the vector $(R_{n,1}, \ldots, R_{n,\nu})$, local limit theorems, etc. We acknowledge a novel method of a randomised decomposition for proving FCLTs for the processes of our kind developed in a recent paper [8], but we do not use it here.

We establish a FCLT for the odd-occupancy process and for the missing mass process when $\theta > 0$. Extending FCLT to the case $\theta = 0$ would require additional to (3) conditions. As it was mentioned in [12] and in [2], $\theta = 0$ does not imply that the variances grow to infinity and various asymptotic behaviour is possible for different statistics. We also argue that even an infinite growth of variances does not guarantee per se the required relative compactness.

When $\theta = 1$, we need a function

$$L^*(x) = \int_0^\infty L(xt)e^{-t} t^{-1} dt.$$ 

It is known (see [12]) that $L^*(x)$ is slowly varying when $x \to \infty$.

Finally, for $t \in [0,1]$ introduce the following notation:

$$\beta(n) = \begin{cases} \alpha(n), & \theta \in (0,1); \\ nL^*(n), & \theta = 1, \end{cases} \quad R_n(t) = \frac{R[nt] - \mathbf{E} R[nt]}{(\beta(n))^{1/2}},$$

$$U_n(t) = \frac{U[nt] - \mathbf{E} U[nt]}{(\beta(n))^{1/2}}, \quad M_n(t) = \frac{M[nt] - \mathbf{E} M[nt]}{(\alpha(n))^{1/2}}.$$ 

(4) (5)

We are now ready to formulate the main result of the paper.

**Theorem 1.** When $\theta \in (0,1]$, the vector process

$$(R_n(t), U_n(t), M_n(t)), \quad t \in [0,1],$$

converges weakly in the uniform metric on $D((0,1)^3)$ to a 3-dimensional Gaussian process with zero mean and the covariance function $C(\tau,t)$ with
the following components: when $\theta \in (0, 1)$, $\tau \leq t$,
\[
c_{R,R}(\tau, t) = \Gamma(1 - \theta)((\tau + t)^\theta - t^\theta),
\]
\[
c_{U,U}(\tau, t) = \Gamma(1 - \theta)2^{\theta-2}((t + \tau)^\theta - (t - \tau)^\theta),
\]
\[
c_{M,M}(\tau, t) = \theta \Gamma(2 - \theta) \left( \frac{\tau}{t^{1-\theta}} - \frac{t\tau}{(t + \tau)^{2-\theta}} \right),
\]
\[
c_{R,U}(\tau, t) = \frac{(t + \tau)/2}{2^{\theta-2}((t + \tau)^{1-\theta} - (t - \tau)^{1-\theta})},
\]
\[
c_{R,M}(\tau, t) = \frac{t}{2(2t + \tau)^{1-\theta} - \frac{\tau}{2(2t - \tau)^{1-\theta}}},
\]
\[
c_{M,U}(\tau, t) = \frac{t}{2(2t + \tau)^{1-\theta} - \frac{\tau}{2(2t - \tau)^{1-\theta}}},
\]
When $\theta = 1$, $\tau \leq t$, $C(\tau, t)$ is given by
\[
c_{R,R}(\tau, t) = \tau, \quad c_{U,U}(\tau, t) = 2\tau, \quad c_{M,M}(\tau, t) = \tau^2,
\]
\[
c_{R,U}(\tau, t) = \tau, \quad c_{R,U}(t, \tau) = (t + \tau)/2,
\]
\[
c_{R,M}(\tau, t) = c_{R,M}(t, \tau) = c_{U,M}(\tau, t) = c_{U,M}(t, \tau) = 0.
\]

2 Proof of Theorem 1

We start with formulating a couple of lemmas proved in [7]. We will generally use the letter $C$ and its variants to denote a constant whose value is of no importance for us and note in parentheses the parameters it depends upon. This should not lead to a confusion when the same notation is used for, actually, different constants in different contexts, the same way $O(1)$ notation is used.

Lemma 2. When $\theta > 0$, there exist $n_0 \geq 1$ and $C(\theta) < \infty$ such that
\[
\frac{\mathbb{E} R(n\delta)}{\beta(n)} \leq C(\theta)\delta^{\theta/2}
\]
holds for any \( \delta \in [0, 1] \) and \( n \geq n_0 \).

**Lemma 3.** For any \( \varepsilon, \delta \in (0, 1) \) there exists an \( N = N(\varepsilon, \delta) \) such that for any \( n \geq N \),

\[
P(\forall t \in [0, 1] \; \exists \tau : |\tau - t| \leq \delta, \; \Pi(n\tau) = [nt]) \geq 1 - \varepsilon.
\]

In preparation of the proof, let us introduce some further notation and establish a few inequalities we will be using.

In view of (5), let

\[
U_n^*(t) = \frac{U(nt) - EU(nt)}{(\beta(n))^{1/2}}, \quad \quad U_n^{**}(t) = \frac{U([nt]) - EU([nt])}{(\beta(n))^{1/2}} \quad (6)
\]

\[
M_n^*(t) = \frac{M(nt) - EM(nt)}{(\alpha(n))^{1/2}}, \quad \quad M_n^{**}(t) = \frac{M([nt]) - EM([nt])}{(\alpha(n))^{1/2}}. \quad (7)
\]

For any two positive \( \tau_1 \leq \tau_2 \), define

\[
U(\tau_2) - U(\tau_1) = \sum_{i=1}^{\infty} 1 \mathbb{I}\{\Pi_i(\tau_2) \text{ is odd}\} - 1 \mathbb{I}\{\Pi_i(\tau_1) \text{ is odd}\}
\]

\[
= \sum_{i=1}^{\infty} 1 \mathbb{I}\{\Pi_i(\tau_2) \text{ is odd, } \Pi_i(\tau_1) \text{ is even}\}
\]

\[
- 1 \mathbb{I}\{\Pi_i(\tau_2) \text{ is even, } \Pi_i(\tau_1) \text{ is odd}\}
\]

\[
= \sum_{i=1}^{\infty} u_i(\tau_1, \tau_2) = \sum_{i=1}^{\infty} u_i = \sum_{i=1}^{\infty} u'_i - u''_i,
\]

their expectations are denoted by

\[
\overline{u}_i = \overline{u}'_i - \overline{u}''_i = \overline{u}_i(\tau_1, \tau_2) \overset{\text{def}}{=} EU'_i - EU''_i.
\]

Similarly for \( M(t) \), write

\[
M(\tau_2) - M(\tau_1) = \sum_{i=1}^{\infty} (\tau_2 - \tau_1) p_i \mathbb{I}\{\Pi_i(\tau_2) = 0\} - \tau_1 p_i \mathbb{I}\{\Pi_i(\tau_1) = 0, \Pi_i(\tau_2) > 0\}
\]

\[
= \sum_{i=1}^{\infty} m_i(\tau_1, \tau_2) = \sum_{i=1}^{\infty} m_i = \sum_{i=1}^{\infty} m'_i - m''_i,
\]

\[
\overline{m}_i = \overline{m}'_i - \overline{m}''_i = \overline{m}_i(\tau_1, \tau_2) \overset{\text{def}}{=} EM'_i - EM''_i.
\]
Clearly, for all natural $k$,
\[
E |u_i - \bar{u}_i|^k = |1 + \pi_i|^k \bar{u}_i'' + |\bar{u}_i|^k (1 - \bar{u}_i'' - \bar{u}_i') + |1 - \bar{u}_i|^k \bar{u}_i'
\leq 2^k (\bar{u}_i' + \bar{u}_i'') + |\bar{u}_i|^k \leq (2^k + 1) (\bar{u}_i' + \bar{u}_i'')
= (2^k + 1) \left[ \sum_{j=0}^{\infty} P\{\Pi_j(\tau_1) = 2j, \ \Pi_j(\tau_2) - \Pi_j(\tau_1) \text{ is odd} \} \right.
+ \sum_{j=0}^{\infty} P\{\Pi_j(\tau_1) = 2j + 1, \ \Pi_j(\tau_2) - \Pi_j(\tau_1) \text{ is odd} \} \right]
= (2^k + 1) P\{\Pi_j(\tau_2 - \tau_1) \text{ is odd} \}
< (2^k + 1) P\{\Pi_j(\tau_2 - \tau_1) > 0 \}.
\]

Similarly,
\[
E |m_i' - \bar{m}_i'|^k \leq 2^{k-1} (E |m_i'|^k + |\bar{m}_i'|^k) = 2^{k-1} (\tau_2 - \tau_1)^k \rho_i^k (e^{-\tau_2 \rho_i} + e^{-\tau_2 \rho_i})
< 2^k k! (1 - e^{-(\tau_2 - \tau_1) \rho_i}) = 2^k k! P\{\Pi_i(\tau_2 - \tau_1) > 0 \},
E |m_i'' - \bar{m}_i''|^k \leq 2^{k-1} (E |m_i''|^k + |\bar{m}_i''|^k) < 2^{k-1} \tau_1^k \rho_i^k e^{-\tau_1 \rho_i} (1 - e^{-(\tau_2 - \tau_1) \rho_i})
< 2^k k! (1 - e^{-(\tau_2 - \tau_1) \rho_i}) = 2^k k! P\{\Pi_i(\tau_2 - \tau_1) > 0 \}.
\]

As a result,
\[
E |m_i - \bar{m}_i|^k < 4^k k! P\{\Pi_i(\tau_2 - \tau_1) > 0 \}.
\]

We are using the same notation $u_i, \ m_i$ and $\bar{u}_i, \ \bar{m}_i$ without explicitly specifying the corresponding values of $\tau_1 < \tau_2$, this should not create a confusion. The following lemma will be used in the proof of a relative compactness of the process $M_n(t)$.

**Lemma 4.** Let $\theta \in (0, 1]$ and $\delta \in [0, 1]$. Then there exist $n_0 \geq 1$ and $C(\theta) < \infty$ such that
\[
\frac{\text{var}(M(nt_2) - M(nt_1))}{\alpha(n)} \leq C(\theta) \delta^{\theta/2}
\]
for all $t_2 - t_1 = \delta \geq 0$ and $n \geq n_0$.

**Proof.** Put $\tau_2 = nt_2$ and $\tau_1 = nt_1$. Since the variance of an indicator does
not exceed its expectation, we have that

\[
\text{var}(M(\tau_2) - M(\tau_1)) = \sum_{i=1}^{\infty} \mathbb{E}(m_i - \bar{m})^2 = \sum_{i=1}^{\infty} \mathbb{E}(m_i')^2 - (\bar{m}')^2 + \mathbb{E}(m_i'')^2
\]

\[
\leq \sum_{i=1}^{\infty} (\tau_2 - \tau_1)^2 p_i^2 e^{-\tau_2 p_i} + \tau_1^2 p_i^2 e^{-\tau_1 p_i} + (\tau_2 - \tau_1) p_i e^{-(\tau_2 - \tau_1) p_i}
\]

\[
\leq 2 \frac{(\tau_2 - \tau_1)^2}{\tau_2} \mathbb{E} R_{\Pi(\tau_2),2} + \mathbb{E} R^{*}_{\Pi(\tau_2-\tau_1),2} + 6 \frac{\tau_1^2}(\tau_2 - \tau_1) \mathbb{E} R_{\Pi(\tau_2),3}.
\]

By \cite{12} Th. 2.1 and (23),

\[
\lim_{x \to \infty} \frac{\mathbb{E} R^{*}_{\Pi(x),2}}{\alpha(x)} = \Gamma(2 - \theta) < 2,
\]

therefore there exists an \( x_1 > 1 \) such that for all \( x \geq x_1 \),

\[
\mathbb{E} R_{\Pi(x),2} + \mathbb{E} R_{\Pi(x),3} < \mathbb{E} R^{*}_{\Pi(x),2} < 2 \alpha(x).
\]

According to Karamata (see, e.g. \cite{5} Th. 2.1, Eq. A6.2.10), there exists an \( x_2 > 0 \) such that for all \( x \) and \( \delta \in (0, 1] \) satisfying \( x \delta \geq x_2 \), one has

\[
\frac{L(x \delta)}{L(x)} \leq 2 \delta^{-1/2}.
\]

Let \( n\delta > \max\{x_1, x_2\} = x_0 \), then

\[
\frac{\mathbb{E} R^{*}_{\Pi(n\delta),2}}{\alpha(n)} \leq 2 \frac{(n\delta)^2 \theta L(n\delta)}{n^{\theta} L(n)} \leq 4 \delta^{\theta/2}, \quad \frac{\max(\mathbb{E} R_{\Pi(nt_2),2}, \mathbb{E} R_{\Pi(nt_2),3})}{\alpha(n)} \leq 4 t_2^{\theta/2}.
\]

Choose \( n_0 \) such that for all \( n \geq n_0 \) we have \( n^{\theta} L(n) \geq n^{\theta/2} \). Then, provided \( nt_2 \leq x_0 \),

\[
\frac{\mathbb{E} R_{\Pi(n\delta),2}}{\alpha(n)} \leq \frac{\mathbb{E} R_{\Pi(n\delta)}}{\alpha(n)} \leq \frac{n\delta}{n^{\theta/2}} = \frac{(n\delta)^{\theta/2} \delta^{\theta/2}}{x_0^{1/2} \delta^{\theta/2}} \leq x_0^{1/2} t_2^{\theta/2},
\]

\[
\frac{\max(\mathbb{E} R_{\Pi(nt_2),2}, \mathbb{E} R_{\Pi(nt_2),3})}{\alpha(n)} \leq x_0^{1/2} t_2^{\theta/2}.
\]

Now take \( c = \max\{4, x_0^{1/2}\} \). Since \( t_2 - t_1 = \delta \geq 0 \), then for all \( n \geq n_0 \) we obtain

\[
\frac{\text{var}(M(nt_2) - M(nt_1))}{\alpha(n)} \leq 2c \frac{\delta^2}{t_2^{-2/\theta/2}} + \delta^{\theta/2} + 6c \frac{t_2^2 \delta}{t_2^{-3-\theta/2}} \leq 9c \cdot \delta^{\theta/2}.
\]

\qed
We are ready to prove Theorem 1. The proof is broken into four steps.

**Step 1: Covariance.** The first rather technical step consists in establishing a formulae for the covariances which is put in Appendix.

**Step 2: Convergence of finite-dimensional distributions.** Along the lines of the proof of [9, Th. 12], one can show

\[ m \geq 1, \quad 0 < t_1 < t_2 < \ldots < t_m \leq 1 \]

the triangular array of \( m \)-dimensional vectors (independent in \( k \) for every \( n \))

\[ \left\{ \frac{\mathbb{I}(\Pi_k(nt_j) \text{ is odd}) - \mathbb{P}(\Pi_k(nt_j) \text{ is odd})}{\sqrt{\beta(n)}}, \ j \leq m, \ k \leq n \right\}_{n \geq 1} \]

satisfies the Lindeberg condition (see, e.g., [5, Th. 6.2]). Similarly, the convergence of the finite-dimensional distributions is shown for the process \( M_n^*(t) \).

**Step 3: Relative compactness.** We shall follow the following plan:

(a) prove the continuity of the limiting process;

(b) prove that \( U_n^* (M_n^*) \) and \( U_n^{**} (M_n^{**}) \) are sufficiently close;

(c) prove the relative compactness of \( U_n^{**} (M_n^{**}) \).

\textbf{a(U)} Take \( \tau_1 = nt_1, \ \tau_2 = nt_2 \) for \( 0 < t_1 < t_2 < 0 \). Then

\[ \mathbb{E}(U_n^*(t_2) - U_n^*(t_1))^2 = \mathbb{E} \left( \sum_{i=1}^{\infty} (u_i - \overline{u}_i) \right)^2 / \beta(n) = \sum_{i=1}^{\infty} \mathbb{E}(u_i - \overline{u}_i)^2 / \beta(n) \]

\[ \leq 5 \sum_{i=1}^{\infty} \mathbb{P}(\Pi_i(\tau_2 - \tau_1) > 0) / \beta(n) = 5 \mathbb{E} R_{\Pi(\tau_2 - \tau_1)} / \beta(n) \leq 5C(\theta)(t_2 - t_1)^{\theta/2}. \]

We have used above the independence of the summands, inequality (8) and Lemma 2.

Since the covariance function has a limit, [1, Th. 1.4] will imply that the limiting Gaussian process a.s. has a continuous modification on \([0, 1]\).
Since the trajectories of the limiting Gaussian process belong a.s. to the class $C(0, 1)$, then the weak convergence in the Skorohod topology implies the weak convergence in the uniform metric, see, e.g., [4]. Therefore, it is sufficient to prove the relative compactness of $\{U^*_n\}_{n \geq n_0}$ (with $n_0$ as in Lemma 2) in the Skorohod topology.

\textbf{b(U)} Since with probability one we have

$$|U(nt) - U([nt])| \leq \Pi(nt) - \Pi([nt]) \leq \Pi([nt] + 1) - \Pi([nt]),$$

then

$$\mathbb{E}|U(nt) - U([nt])| \leq 1.$$

Hence, for all $\eta > 0$,

$$\mathbb{P}(\sup_{0 \leq t \leq 1}|U^*_n(t) - U^*_{n*}(t)| > \eta)$$

$$\leq \mathbb{P}(\sup_{0 \leq t \leq 1}(|U(nt) - U([nt])| + \mathbb{E}|U(nt) - U([nt])|) > \eta \sqrt{\beta(n)})$$

$$\leq \mathbb{P}(\sup_{0 \leq t \leq 1} (\Pi([nt] + 1) - \Pi([nt]) + 1) > \eta \sqrt{\beta(n)})$$

$$= \mathbb{P}(\sup_{0 \leq m \leq n} (\Pi(m + 1) - \Pi(m) + 1) > \eta \sqrt{\beta(n)})$$

$$\leq \sum_{m=0}^{n} \mathbb{P}(\Pi(m + 1) - \Pi(m) + 1 > \eta \sqrt{\beta(n)})$$

$$\leq \sum_{m=0}^{n} \frac{\mathbb{E}e^{\Pi(m+1)-\Pi(m)+1}}{e^{\eta \sqrt{\beta(n)}}} = (n + 1) \frac{\mathbb{E}e^{\Pi(1)}}{e^{\eta \sqrt{\beta(n)-1}}} = (n + 1)e^{-\eta \sqrt{\beta(n)}} \to 0$$

when $n \to \infty$. Therefore, it is sufficient to show the relative compactness of $\{U^*_{n*}\}_{n \geq n_0}$ (with $n_0$ as in Lemma 2) in the Skorokhod topology.

\textbf{c(U)} For any $t_1, t_2 \in [0, 1]$ satisfying $\frac{1}{2n} \leq t_2 - t_1$ we have that

$$[nt_2] - [nt_1] \leq n(t_2 - t_1) + 1 \leq n(t_2 - t_1) + 2n(t_2 - t_1) = 3n(t_2 - t_1)$$

$$\leq 3n(t_2 - t_1) \cdot (2n(t_2 - t_1))^3 = 24n^4(t_2 - t_1)^4. \quad (10)$$

Put $k = \lceil 16/\theta \rceil + 1$, $\tau_1 = [nt_1]$, $\tau_2 = [nt_2]$. 

\begin{flushright}
9
\end{flushright}
Recall the Rosenthal inequality \[15\]: if \( \varphi_i \) are independent random variables with \( E \varphi_i = 0 \), then for all \( k \geq 2 \) there exists a constant \( c(k) \) such that

\[
E \left| \sum_i \varphi_i \right|^k \leq c(k) \max \left\{ \sum_i E|\varphi_i|^k, \left( \sum_i E\varphi_i^2 \right)^{k/2} \right\}. \tag{11}
\]

For all \( n \geq n_0 \) (with \( n_0 \) as in Lemma 2) we then have

\[
E \left| U^{**}_n(t_2) - U^{**}_n(t_1) \right|^k = \frac{E \left| \sum_{i=1}^\infty (u_i - \bar{u}_i) \right|^k}{(\beta(n))^{k/2}} \leq \frac{c(k)}{(\beta(n))^{k/2}} \left( \sum_{i=1}^\infty E|u_i - \bar{u}_i|^k + \left( \sum_{i=1}^\infty E(u_i - \bar{u}_i)^2 \right)^{k/2} \right)
\]

\[
\leq \frac{C(k)}{(\beta(n))^{k/2}} \left( \sum_{i=1}^\infty P(\Pi_i(\tau_2 - \tau_1) > 0) + \left( \sum_{i=1}^\infty P(\Pi_i(\tau_2 - \tau_1) > 0) \right)^{k/2} \right)
\]

\[
= \frac{C(k)}{(\beta(n))^{k/2}} \left( E R(\tau_2 - \tau_1) + (E R(\tau_2 - \tau_1))^{k/2} \right)
\]

\[
\leq \frac{C(k)}{(\beta(n))^{k/2}} \left( 24n^4(t_2 - t_1)^4 + (E R(3n(t_2 - t_1)))^{k/2} \right) \leq \tilde{C}(\theta)(t_2 - t_1)^4,
\]

where \( c(k), C(k) \) and \( \tilde{C}(\theta) \) depend only on their arguments.

Above, we have used \((11)\) in the first inequality, \((8)\) in the second and finally, \((10)\) and Lemma 2 alongside with the bound

\[
E R(\tau_2 - \tau_1) \leq E(\Pi([nt_2]) - \Pi([nt_1])) = [nt_2] - [nt_1]. \tag{12}
\]

If \( 0 \leq t_2 - t_1 < \frac{1}{n} \), then \( [nt_1] = [nt] \) or \( [nt_2] = [nt] \) for all \( t \in [t_1, t_2] \), therefore

\[
Q \overset{\text{def}}{=} E(|U^{**}_n(t) - U^{**}_n(t_1)|^{k/2}|U^{**}_n(t_2) - U^{**}_n(t)|^{k/2}) = 0 \leq (t_2 - t_1)^2.
\]

If \( t_2 - t_1 \geq 1/n \), then there are the following three cases:

1. if \( t_2 - t \geq \frac{1}{2n} \) and \( t - t_1 \geq \frac{1}{2n} \), then the Cauchy–Schwarz inequality implies

\[
Q \leq \tilde{C}(\theta)(t_2 - t)^2 \cdot (t - t_1)^2 \leq \tilde{C}(\theta)(t_2 - t_1)^2.
\]
2. If \( t_2 - t \geq \frac{1}{2n}, t - t_1 < \frac{1}{2n} \), then since
\[
|U([nt]) - U([nt_1])| \leq_{a.s.} \Pi([nt]) - \Pi([nt_1]) \leq_{st} \Pi(1),
\]
the same inequality yields
\[
Q \leq \left( \tilde{C}(\theta)(t_2 - t)^4 \cdot \mathbb{E} \left( \frac{\Pi(1) + 1}{\sqrt{\beta(n)}} \right)^k \right)^{1/2} \leq \tilde{C}(\theta)(t_2 - t_1)^2.
\]

3. If \( t_2 - t < \frac{1}{2n}, t - t_1 \geq \frac{1}{2n} \), then since
\[
|U([nt_2]) - U([nt])| \leq_{a.s.} \Pi([nt_2]) - \Pi([nt]) \leq_{st} \Pi(1),
\]
we have that
\[
Q \leq \left( \mathbb{E} \left( \frac{\Pi(1) + 1}{\sqrt{\beta(n)}} \right)^k \cdot \tilde{C}(\theta)(t - t_1)^4 \right)^{1/2} \leq \tilde{C}(\theta)(t_2 - t_1)^2.
\]

Now the relative compactness follows from, e.g., [4, Th. 13.5].

**a(M)** Because the covariance function has a limit, it is sufficient to appeal to Lemma 3 and [1, Th. 1.4] to establish existence of an almost sure continuous on \([0, 1]\) modification of the limiting Gaussian process. Since the trajectories of this process are a.s. in \( C(0, 1) \), then the weak convergence in the Skorohod topology implies the uniform convergence, see [4]. Thus it is sufficient to prove a relative compactness of the family \( \{M^*_n\} \) in the Skorohod topology (here \( n_0 \) is the same as in Lemma 2).

**b(M)** Set \( \tau_2 = nt \) and \( \tau_1 = [nt] \). Since \( \tau_2 - \tau_1 \leq 1 \), then
\[
\mathbb{E} |M(\tau_2) - M(\tau_1)| \leq \sum_{i=1}^{\infty} (\tau_2 - \tau_1)p_i e^{-p_i \tau_2} + \tau_1 p_i e^{-p_i \tau_1} (1 - e^{-p_i (\tau_2 - \tau_1)})
\]
\[
\leq \sum_{i=1}^{\infty} p_i e^{-p_i \tau_2} + e^{-p_i} p_i (\tau_2 - \tau_1) < \sum_{i=1}^{\infty} 2p_i = 2.
\]
Let $m_i''' = m_i'''(\tau_1, \tau_1 + 1)$ and $m_i''' = E m_i'''$. Then we have almost surely,

$$|M(\tau_2) - M(\tau_1)| \leq \sum_{i=1}^{\infty} (m_i' + m_i'') \leq \sum_{i=1}^{\infty} (p_i + m_i''')$$

$$= 1 + \sum_{i=1}^{\infty} (m_i''' + m_i''' - \bar{m}_i''') < 2 + \sum_{i=1}^{\infty} (m_i''' - \bar{m}_i''')$$

We know that for any integer $k \geq 2$

$$E |m_i''' - E m_i'''|^k < 2^k k! P(\Pi_i(\tau_1 + 1 - \tau_1) > 0) = 2^k k!(1 - e^{-p_i}) < 2^k k! p_i.$$ 

Using the independence of the terms and Rosenthal inequality, for any $k \geq 2$,

$$E \left( \sum_{i=1}^{\infty} (m_i''' - \bar{m}_i''') \right)^k \leq c(k) \left( \sum_{i=1}^{\infty} E |m_i''' - \bar{m}_i'''|^k + \left( \sum_{i=1}^{\infty} E(m_i''' - \bar{m}_i''')^2 \right)^{k/2} \right)$$

$$< c(k)(2^k k! + 4^k) = C(k).$$

Hence, for $k \geq \lceil 2/\theta \rceil + 1$ and all $\eta > 0$

$$P( \sup_{0 \leq t \leq 1} |M_n^*(t) - M_n^{**}(t)| > \eta)$$

$$\leq P( \sup_{0 \leq t \leq 1} (|M(nt) - M([nt])| + E |M(nt) - M([nt])|) > \eta \sqrt{\alpha(n)})$$

$$\leq P \left( \max_{0 \leq [nt] \leq n} \left( \sum_{i=1}^{\infty} m_i''' - E m_i''' \right) + 4 \right) > \eta \sqrt{\alpha(n)}$$

$$\leq \sum_{[nt]=m \in \{0,1,\ldots,n\}} P \left( \sum_{i=1}^{\infty} m_i''' - E m_i''' \right) + 4 > \eta \sqrt{\alpha(n)}$$

$$\leq \sum_{m=0}^{n} \frac{C(k)}{(\eta \sqrt{\alpha(n)} - 4)^k} = \frac{C(k)(n+1)}{(\eta \sqrt{\alpha(n)} - 4)^k} \to 0 \text{ when } n \to \infty.$$ 

Therefore, it is sufficient to show the local compactness of $\{M_n^{**}\}_{n \geq n_0}$ in the Skorohod topology.
\textbf{c(M)} Let \( t_1, t_2 \in [0, 1] \) and \( \frac{1}{2n} \leq t_2 - t_1 \), then (10) holds. Set \( k = [16/\theta] + 1, \) \\
\( \tau_1 = [nt_1], \) \( \tau_2 = [nt_2]. \)

Again, by independence and the Rosenthal inequality,

\[
\mathbb{E} |M_{n}^{**}(t_2) - M_{n}^{**}(t_1)|^k = \frac{\mathbb{E} \left( \sum_{i=1}^{\infty} (m_i - \bar{m}_i)^k \right)}{(\alpha(n))^{k/2}}
\leq \frac{c(k)}{(\alpha(n))^{k/2}} \left( \sum_{i=1}^{\infty} \mathbb{E} |m_i - \bar{m}_i|^k + \left( \sum_{i=1}^{\infty} \mathbb{E} (m_i - \bar{m}_i)^2 \right)^{k/2} \right)
\leq \frac{C(\beta)}{(\alpha(n))^{k/2}} \left( \sum_{i=1}^{\infty} \mathbb{P}(\Pi_i(\tau_2 - \tau_1) > 0) + \left( \text{var}(M(\tau_2) - M(\tau_1)) \right)^{k/2} \right)
= \frac{C(k)}{(\alpha(n))^{k/2}} \left( \mathbb{E} R(\tau_2 - \tau_1) + \left( \text{var}(M(\tau_2) - M(\tau_1)) \right)^{k/2} \right)
\leq \frac{C(k)}{(\alpha(n))^{k/2}} \left( 24n^4(t_2 - t_1)^4 + (C(\theta)\alpha(n)(\tau_2 - \tau_1)/n)^{k/2} \right) \leq \tilde{C}(\theta)(t_2 - t_1)^4,
\]

where \( c(k), C(k) \) and \( \tilde{C}(\theta) \) depend only on their arguments.

Above, we have used inequalities (9), (10) and Lemmas 3, 2 alongside with the bound

\[
\mathbb{E} R(\tau_2 - \tau_1) \leq \mathbb{E}(\Pi([nt_2] - [nt_1])) = [nt_2] - [nt_1].
\]

When \( 0 \leq t_2 - t_1 < \frac{1}{n} \), then \([nt_1] = [nt]\) or \([nt_2] = [nt]\) for any \( t \in [t_1, t_2] \).
Thus

\[
Q \overset{\text{def}}{=} \mathbb{E}([M_{n}^{**}(t) - M_{n}^{**}(t_1)]^{k/2}[M_{n}^{**}(t_2) - M_{n}^{**}(t)]^{k/2}) = 0 \leq (t_2 - t_1)^2.
\]

When \( t_2 - t_1 \geq 1/n \), we have the following three cases:

1. if \( t_2 - t \geq \frac{1}{2n}, \ t - t_1 \geq \frac{1}{2n}, \) then the Cauchy–Schwarz inequality gives

\[
Q \leq \tilde{C}(\theta)(t_2 - t)^2 \cdot (t - t_1)^2 \leq \tilde{C}(\theta)(t_2 - t_1)^2;
\]

2. if \( t_2 - t \geq \frac{1}{2n}, \ t - t_1 < \frac{1}{2n}, \) then since for any \( l \geq 2, \)

\[
\mathbb{E} |M([nt]) - M([nt_1]) - \mathbb{E}(M([nt]) - M([nt_1])|^l
\]

13
\[ \leq \mathbb{E} \left( 4 + \left| \sum_{i=1}^{\infty} m''_n([nt_1] + 1, [nt_1]) - \mathbb{E} m''_n([nt_1] + 1, [nt_1]) \right| \right)^t < C(l), \]

the Cauchy–Schwarz inequality yields the bound

\[ Q \leq \left( \tilde{C}(\theta) (t_2 - t)^4 \cdot \frac{C(k)}{\alpha(n)^{k/2}} \right)^{1/2} \leq \tilde{C}(\theta) (t_2 - t_1)^2; \]

3. finally, \( t_2 - t < \frac{1}{2n}, t - t_1 \geq \frac{1}{2n} \), is similar to the previous case.

Thus the required compactness follows from [4, Th. 13.5].

**Step 4: Approximation of the initial process.** Since \( \Pi(t) \) is monotone, the Strong Law of Large Numbers implies that for any \( \varepsilon, \delta \in (0, 1) \) there is an integer \( N = N(\varepsilon, \delta) \) such that for all \( n \geq N \) one has

\[ P(\forall \tau \in [0, 1] \exists \tau : |\tau - t| \leq \delta, \Pi(n\tau) = [nt]) \overset{def}{=} P(A(n)) \geq 1 - \varepsilon, \]

see Lemma [3]. Here and below, \( F \) stands for \( R, U \) or \( M \). The relative compactness of the distributions \( \{F^*_n\}_{n \geq n_0} \) implies that for any \( \varepsilon \in (0, 1) \) and \( \eta > 0 \) there exist \( \delta \in (0, 1) \) and an integer \( N_1 = N_1(\varepsilon, \eta) \) such that for all \( n \geq N_1 \),

\[ P(\sup_{|t-\tau| \leq \delta} |F^*_n(\tau) - F^*_n(t)| \geq \eta) \leq \varepsilon. \]

Hence, since

\[ P(F_n(t) = F^*_n(\tau) | \Pi(n\tau) = [nt]) = 1, \]

then for all \( n \geq \max(N, N_1) \),

\[ P\left( \sup_{0 \leq t \leq 1} |F_n(t) - F^*_n(t)| \geq \eta \right) \leq P\left( \sup_{0 \leq t \leq 1} |F_n(t) - F^*_n(t)| \geq \eta, A(n) \right) + \varepsilon \leq P\left( \sup_{|t-\tau| \leq \delta} |F^*_n(\tau) - F^*_n(t)| \geq \eta \right) + \varepsilon \leq 2\varepsilon. \]

which proves Theorem [1].

**Acknowledgements.** MC’s research is supported by RSF Grant 17-11-01173. He also acknowledges hospitality of Chalmers university where a part of this work has been done. The authors are thankful to Sergey Foss for his interest in this research and valuable comments.
Appendix

An explicit expression for the covariance between $R(\tau)$ and $R(t)$ can be found in [7]. Take $\tau \leq t$. The

\[
c_{U,U}(\tau, t) = \text{var}(U(\tau), U(t))
\]

\[
= \sum_{k=1}^{\infty} \mathbb{P}(\Pi_k(\tau), \Pi_k(t) \text{ is odd}) - \mathbb{P}(\Pi_k(\tau) \text{ is odd})\mathbb{P}(\Pi_k(t) \text{ is odd})
\]

\[
= \frac{1}{4} \sum_{k=1}^{\infty} \left( (1 - e^{-2p_k\tau})(1 + e^{-2p_k(t-\tau)}) - (1 - e^{-2p_k\tau})(1 - e^{-2p_k(t+\tau)}) \right)
\]

\[
= \frac{1}{4} \sum_{k=1}^{\infty} e^{-2p_k(t-\tau)} - e^{-2p_k(t+\tau)} = \frac{1}{2} \mathbb{E}(U(t + \tau) - U(t - \tau)).
\]

Hence (since $\beta(nt) / \beta(n) \to t^\theta$ as $n \to \infty$)

\[
c_{U,U}(\tau, t) = \lim_{n \to \infty} c_{U,U}(n\tau, nt) / \alpha(n) = \Gamma(1 - \theta)2^{\theta-2}((t + \tau)^\theta - (t - \tau)^\theta), \theta \in (0, 1),
\]

\[
c_{U,U}(\tau, t) = \lim_{n \to \infty} c_{U,U}(n\tau, nt) / nL^*(n) = 2/\tau, \theta = 1.
\]

cf. [12, Eq. (21)].

Next,

\[
c_{M,M}(\tau, t) = \text{var}(M(\tau), M(t))
\]

\[
= \sum_{k=1}^{\infty} \mathbb{E}(tp_i \mathbb{I}(\Pi_i(t) = 0) - tp_i e^{-tp_i})(\tau p_i \mathbb{I}(\Pi_i(\tau) = 0) - \tau p_i e^{-\tau p_i})
\]

\[
= \sum_{k=1}^{\infty} t\tau p_i^2 e^{-tp_i}(1 - e^{-\tau p_i}) = \frac{2\tau}{t} \mathbb{E} R_{\Pi(t),2} - \frac{2t}{(t+\tau)^2} \mathbb{E} R_{\Pi(t+\tau),2}.
\]

Since $\alpha(nt) / \alpha(n) \to t^\theta$ when $n \to \infty$),

\[
c_{M,M}(\tau, t) = \lim_{n \to \infty} c_{M,M}(n\tau, nt) / \alpha(n) = \theta \Gamma(2 - \theta) \left( \frac{\tau}{t^{1-\theta}} - \frac{t\tau}{(t+\tau)^{2-\theta}} \right),
\]

cf. [12, Eq. (23)].
Continuing,

\[ c_{RU}^*(\tau, t) = \text{var}(R(\tau), U(t)) = \sum_{k=1}^{\infty} \text{var}(1 - \mathbb{I}(\Pi_k(\tau) = 0), \mathbb{I}(\Pi_k(t) \text{ is odd})) \]

\[ = - \sum_{k=1}^{\infty} \text{var}(\mathbb{I}(\Pi_k(\tau) = 0), \mathbb{I}(\Pi_k(t) \text{ is odd})) \]

\[ = - \sum_{k=1}^{\infty} P(\Pi_k(\tau) = 0, \Pi_k(t) \text{ is odd}) - P(\Pi_k(\tau) = 0) P(\Pi_k(t) \text{ is odd}) \]

\[ = - \frac{1}{2} \sum_{k=1}^{\infty} \left( e^{-p_k \tau} (1 - e^{-2p_k(t-\tau)}) - e^{-p_k \tau} (1 - e^{-2p_k t}) \right) = \frac{1}{2} \sum_{k=1}^{\infty} \left( e^{-p_k (2t-\tau)} - e^{-p_k (2t+\tau)} \pm 1 \right) \]

\[ = \frac{1}{2} E(R(2t + \tau) - R(2t - \tau)). \]

Similarly,

\[ c_{RU}^*(t, \tau) = \text{var}(R(t), U(\tau)) = - \sum_{k=1}^{\infty} \text{var}(\mathbb{I}(\Pi_k(t) = 0), \mathbb{I}(\Pi_k(\tau) \text{ is odd})) \]

\[ = \frac{1}{2} \sum_{k=1}^{\infty} e^{-p_k t} (1 - e^{-2p_k \tau}) = \frac{1}{2} \sum_{k=1}^{\infty} \left( e^{-p_k t} - e^{-p_k (2\tau + t)} \pm 1 \right) \]

\[ = \frac{1}{2} E(R(2t + \tau) - R(t)). \]

Because \( \frac{\beta(nt)}{\beta(n)} \to t^\theta \) when \( n \to \infty \), for \( \theta \in (0, 1) \) we have that

\[ c_{RU}(\tau, t) = \lim_{n \to \infty} \frac{c_{RU}^*(n\tau, nt)}{\alpha(n)} = (1 - \theta)((2t + \tau)^\theta - (2t - \tau)^\theta)/2, \]

\[ c_{RU}(t, \tau) = \lim_{n \to \infty} \frac{c_{RU}^*(nt, n\tau)}{\alpha(n)} = (1 - \theta)((2t + \tau)^\theta - t^\theta)/2. \]

For \( \theta = 1 \) this reduces to

\[ c_{RU}(\tau, t) = \lim_{n \to \infty} \frac{c_{RU}^*(n\tau, nt)}{nL^*(n)} = \tau, \]

\[ c_{RU}(t, \tau) = \lim_{n \to \infty} \frac{c_{RU}^*(nt, n\tau)}{nL^*(n)} = (t + \tau)/2. \]
Next, 

\[
c^*_\text{MU}(\tau, t) = \text{var}(M(\tau), U(t)) = \sum_{k=1}^{\infty} \tau p_k \text{ var}(\mathbb{I}(\Pi_k(\tau) = 0), \mathbb{I}(\Pi_k(t) \text{ is odd}))
\]

\[
= \frac{1}{2} \sum_{k=1}^{\infty} \tau p_k \left( e^{-p_k(2\tau + t)} - e^{-p_k(2t - \tau)} \right)
\]

\[
= \frac{\tau}{2(2t + \tau)} \mathbb{E} M(2t + \tau) - \frac{\tau}{2(2t - \tau)} \mathbb{E} M(2t - \tau).
\]

and

\[
c^*_\text{MU}(t, \tau) = \text{var}(M(t), U(\tau)) = \frac{1}{2} \sum_{k=1}^{\infty} t p_k \left( e^{-p_k(2\tau + t)} - e^{-p_k t} \right)
\]

\[
= \frac{t}{2(2\tau + t)} \mathbb{E} M(2\tau + t) - \frac{1}{2} \mathbb{E} M(t).
\]

Finally,

\[
c^*_\text{RM}(\tau, t) = \text{var}(R(\tau), M(t)) = \sum_{k=1}^{\infty} \text{var}(1 - \mathbb{I}(\Pi_k(\tau) = 0), t p_k \mathbb{I}(\Pi_k(t) = 0))
\]

\[
= - \sum_{k=1}^{\infty} t p_k \text{ var}(\mathbb{I}\{\Pi_k(\tau) = 0\}, \mathbb{I}\{\Pi_k(t) = 0\})
\]

\[
= - \sum_{k=1}^{\infty} t p_k \left( e^{-p_k t} - e^{-p_k(\tau + t)} \right) = \frac{t}{\tau + t} \mathbb{E} M(\tau + t) - \mathbb{E} M(t).
\]

and

\[
c^*_\text{RM}(t, \tau) = \text{var}(R(t), M(\tau)) = \frac{\tau}{\tau + t} \mathbb{E} M(\tau + t) - \frac{\tau}{t} \mathbb{E} M(t).
\]
Because $\frac{\alpha(nt)}{\alpha(n)} \to t^\theta$ when $n \to \infty$, for $\theta \in (0,1)$ we obtain

$$c_{RM}(\tau, t) = \lim_{n \to \infty} \frac{c^*_{RM}(n\tau, nt)}{\alpha(n)} = \theta \Gamma(1 - \theta) \left( \frac{t}{(t + \tau)^{1-\theta}} - t^{\theta} \right),$$

$$c_{RM}(t, \tau) = \lim_{n \to \infty} \frac{c^*_{RM}(nt, n\tau)}{\alpha(n)} = \theta \Gamma(1 - \theta) \left( \frac{\tau}{(t + \tau)^{1-\theta}} - \frac{\tau}{t^{1-\theta}} \right),$$

$$c_{MU}(\tau, t) = \lim_{n \to \infty} \frac{c^*_{MU}(n\tau, nt)}{\alpha(n)} = \theta \Gamma(1 - \theta) \left( \frac{\tau}{2(2t + \tau)^{1-\theta}} - \frac{\tau}{2(2t - \tau)^{1-\theta}} \right),$$

$$c_{MU}(t, \tau) = \lim_{n \to \infty} \frac{c^*_{MU}(nt, n\tau)}{\alpha(n)} = \theta \Gamma(1 - \theta) \left( \frac{t}{2(2\tau + t)^{1-\theta}} - \frac{t^{\theta}}{2} \right),$$

cf. [12, Eq. (23)].

Clearly, $L(n) \to 0$ as $n \to \infty$. According to [12, Lem. 4], in the case $\theta = 1$ the function $L^*(n) \to 0$ when $n \to \infty$ is slowly varying and

$$\lim_{n \to \infty} \frac{L(n)}{L^*(n)} \overset{\text{def}}{=} \lim_{n \to \infty} \delta_n = 0. \quad (13)$$

Therefore, in the case $\theta = 1$,

$$c_{RM}(\tau, t) = \lim_{n \to \infty} \frac{c^*_{RM}(n\tau, nt)}{\alpha(n)} \sqrt{\delta_n} = 0, \quad c_{RM}(t, \tau) = \lim_{n \to \infty} \frac{c^*_{RM}(nt, n\tau)}{\alpha(n)} \sqrt{\delta_n} = 0,$$

$$c_{MU}(\tau, t) = \lim_{n \to \infty} \frac{c^*_{MU}(n\tau, nt)}{\alpha(n)} \sqrt{\delta_n} = 0, \quad c_{MU}(t, \tau) = \lim_{n \to \infty} \frac{c^*_{MU}(nt, n\tau)}{\alpha(n)} \sqrt{\delta_n} = 0.$$

References

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

[2] A.D. Barbour and A.V. Gnedin. Small counts in the infinite occupancy scheme. *Electronic J. Probab.*, 14(13):365–384, 2009.

[3] 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(1):249–287, 2017.

[4] P. Billingsley. *Convergence of Probability Measures*. Wiley, 2nd edition, 1999.
[5] A.A. Borovkov. *Probability Theory*. Universitext, 2013.

[6] M.G. Chebunin. Functional central limit theorem in an infinite urn scheme for distributions with superheavy tails. *Siberian Electronic Mathematical Reports*, 14:1289–1298, 2017.

[7] M.G. Chebunin and A. Kovalevskii. Functional central limit theorems for certain statistics in an infinite urn scheme. *Stats. Prob. Letters*, 119:344–348, 2016.

[8] O. Durieu and Y. Wang. From infinite urn schemes to decompositions of self-similar Gaussian processes. *Electronic J. of Prob.*, 21(43):1–23, 2016.

[9] M. Dutko. Central limit theorems for infinite urn models. *Ann. Probab.*, 17:1255–1263, 1989.

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

[11] I.J. Good and G.H. Toulmin. The number of new species, and the increase in population coverage, when a sample is increased. *Biometrika*, 43(1/2):45–63, 1956.

[12] S. Karlin. Central limit theorems for certain infinite urn schemes. *J. of Mathematics and Mechanics*, 17(4):373–401, 1967.

[13] A. Muratov and S. Zuyev. Bit flipping and time to recover. *J. Appl. Prob.*, 53(3):1–17, 2016.

[14] A. Orlitsky, N. Santhanam, and J. Zhang. Universal compression of memoryless sources over unknown alphabets. *IEEE Trans. Inform. Theory*, 50(7):1469–1481, 2004.

[15] H.P. Rosenthal. On the subspaces of $l_p$ ($p > 2$) spanned by sequences of independent random variables. *Israel J. Math.*, 8(3):273–303, 1970.