Statistics and Probability Letters 119 (2016) 344–348

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Functional central limit theorems for certain statistics in an infinite urn scheme

Mikhail Chebunin [a], Artyom Kovalevskii [b,a,c,*]

[a] Novosibirsk State University, Novosibirsk, Russia
[b] Novosibirsk State Technical University, Novosibirsk, Russia
[c] Novosibirsk State University of Economics and Management, Novosibirsk, Russia

A R T I C L E   I N F O

Article history:
Received 11 February 2016
Received in revised form 25 June 2016
Accepted 31 August 2016
Available online 9 September 2016

Keywords:
Infinite urn scheme
Relative compactness
Slow variation

A B S T R A C T

We investigate a specific infinite urn scheme first considered by Karlin (1967). We prove functional central limit theorems for the total number of urns with at least k balls for any $k \geq 1$.

1. Introduction

Karlin (1967) studied an infinite urn scheme, that is, each of n balls goes to urn $i \geq 1$ with probability $p_i > 0$, $p_1 + p_2 + \cdots = 1$, independently of other balls. We assume $p_1 \geq p_2 \geq \cdots$. Let X_j be the box that the ball j is thrown into, and

$$R^*_{n,k} = \sum_{i=1}^{\infty} \mathbf{I}(\exists j_1 < \cdots < j_k \leq n: \, X_{j_1} = \cdots = X_{j_k} = i)$$

be the total number of urns with at least k balls. The number of nonempty urns is $R_n = R^*_{n,1}$. The total number of urns with k balls exactly is $R_{n,k} = R^*_{n,k} - R^*_{n,k+1}$. Let $J_i(n)$ be the number of n balls in urn i.

Let (see Karlin, 1967) $\Pi = \{\Pi(t), \, t \geq 0\}$ be a Poisson process with parameter 1. This process does not depend on $\{X_j\}_{j \geq 1}$. The Poissonized version of Karlin model assumes the total number of $\Pi(n)$ balls.

According to well-known thinning property of Poisson flows, stochastic processes $\{J_i(\Pi(t)) \stackrel{def}{=} \Pi_i(t), \, t \geq 0\}$ are Poisson with intensities p_i and are mutually independent for different i's. The definition implies that

$$R^*_{\Pi(n),k} = \sum_{i=1}^{\infty} \mathbf{I}(\Pi_i(n) \geq k), \qquad R_{\Pi(n),k} = \sum_{i=1}^{\infty} \mathbf{I}(\Pi_i(n) = k).$$

* Corresponding author at: Novosibirsk State Technical University, Novosibirsk, Russia.
E-mail addresses: chebuninmikhail@gmail.com (M. Chebunin), kovalevskiii@gmail.com (A. Kovalevskii).

http://dx.doi.org/10.1016/j.spl.2016.08.019
0167-7152/

Let $\alpha(x) = \max\{j | p_j \geq 1/x\}$. Following Karlin (1967), we assume that $\alpha(x) = x^\theta L(x)$, $0 \leq \theta \leq 1$. Here $L(x)$ is a slowly varying function as $x \to \infty$. Let for $t \in [0, 1]$, $k \geq 1$

$$Y^*_{n,k}(t) = \frac{R^*_{[nt],k} - \mathbf{E}R^*_{[nt],k}}{(\alpha(n))^{1/2}}, \qquad Z^*_{n,k}(t) = \frac{R^*_{\Pi(nt),k} - \mathbf{E}R^*_{\Pi(nt),k}}{(\alpha(n))^{1/2}},$$

$$Y_{n,k}(t) = \frac{R_{[nt],k} - \mathbf{E}R_{[nt],k}}{(\alpha(n))^{1/2}}, \qquad K_{k,\theta} = \begin{cases} -\Gamma(1-\theta), & k = 0; \\ \theta\Gamma(k-\theta), & k > 0. \end{cases}$$

The goal of our paper is to extend the following two theorems from Karlin (1967).

Theorem 1 (*Theorem 4 in Karlin (1967)*). *Let $\theta \in (0, 1]$. Then $(R_n - \mathbf{E}R_n)/B_n^{1/2}$ converges weakly to standard normal distribution, where*

$$B_n = \begin{cases} \Gamma(1-\theta)(2^\theta - 1)n^\theta L(n), & \theta \in (0, 1); \\ n\displaystyle\int_0^\infty \frac{e^{-1/y}}{y} L(ny)dy \stackrel{def}{=} nL^*(n), & \theta = 1. \end{cases}$$

Karlin (1967, Lemma 4) proved that the function $L^*(x)$ is slowly varying as $x \to \infty$.

Theorem 2 (*Theorem 5 in Karlin (1967)*). *Let $\theta \in (0, 1)$, $r_1 < \cdots < r_\nu$ be ν positive integers. Then random vector $(Y_{n,r_1}(1), \ldots, Y_{n,r_\nu}(1))$ converges weakly to the multivariate normal distribution with zero expectation and covariances*

$$c_{r_i,r_j} = \begin{cases} -\dfrac{\theta\Gamma(r_i + r_j - \theta)}{r_i! r_j!} 2^{\theta - r_i - r_j}, & i \neq j; \\ \dfrac{\theta}{\Gamma(r_i + 1)} \left(\Gamma(r_i - \theta) - 2^{-2r_i + \theta} \dfrac{\Gamma(2r_i - \theta)}{\Gamma(r_i + 1)} \right), & i = j. \end{cases}$$

Here we briefly mention some related results on this model. Dutko (1989) generalized Theorem 1 by proving asymptotic normality of R_n if $\mathbf{Var}\, R_n \to \infty$ as $n \to \infty$. This condition always holds if $\theta \in (0, 1]$ but can hold too for $\theta = 0$. Gnedin et al. (2007) focused on study of conditions for convergence $\mathbf{Var}\, R_n \to \infty$. Barbour and Gnedin (2009) extended Theorem 2 on the case of $\theta = 0$ if variances go to infinity. They found conditions for convergence of covariances to a limit and identified four types of limiting behavior of variances. Barbour (2009) proved theorems on approximation of the number of cells with k balls by translated Poisson distribution. Key (1992, 1996) studied the limit behavior of statistics $R_{n,1}$. Hwang and Janson (2008) proved local limit theorems for finite and infinite number of cells. Zakrevskaya and Kovalevskii (2001) proved consistency for one parametric family of an estimator of $\theta \in (0, 1)$ which is an implicit function of R_n. Chebunin (2014) constructed an R_n-based explicit parameter estimator for $\theta \in (0, 1)$ and proved its consistency. Durieu and Wang (2015) established a functional central limit theorem for a randomization of process R_n: each indicator is multiplied independently by a random variable taking values in ± 1 with equal probabilities. The limiting Gaussian process is a sum of independent self-similar processes in this case.

Now we formulate the main result of the paper.

Theorem 3. (i) *Let $\theta \in (0, 1)$ and $\nu \geq 1$ be an integer. Then process $(Y^*_{n,1}(t), \ldots, Y^*_{n,\nu}(t),\ 0 \leq t \leq 1)$ converges weakly in the uniform metric in $D([0, 1]^\nu)$ to ν-dimensional Gaussian process with zero expectation and covariance function $(c^*_{ij}(\tau, t))_{i,j=1}^\nu$: for $\tau \leq t$, $i, j \in \{1, \ldots, \nu\}$ (taking $0^0 = 1$)*

$$c^*_{ij}(\tau, t) = \begin{cases} \displaystyle\sum_{s=0}^{i-1} \sum_{m=0}^{j-s-1} \frac{\tau^s (t-\tau)^m K_{m+s,\theta}}{t^{m+s-\theta} s! m!} - \sum_{s=0}^{i-1} \sum_{m=0}^{j-1} \frac{\tau^s t^m K_{m+s,\theta}}{(t+\tau)^{m+s-\theta} s! m!}, & i < j; \\ \displaystyle t^\theta \sum_{m=0}^{j-1} \frac{K_{m,\theta}}{m!} - \sum_{s=0}^{i-1} \sum_{m=0}^{j-1} \frac{\tau^s t^m K_{m+s,\theta}}{(t+\tau)^{m+s-\theta} s! m!}, & i \geq j; \end{cases}$$

$c^*_{ij}(\tau, t) = c^*_{ji}(t, \tau)$.

(ii) *Let $\theta = 1$. Then process $\left(\frac{R_{[nt]} - \mathbf{E}R_{[nt]}}{(nL^*(n))^{1/2}},\ 0 \leq t \leq 1\right)$ converges weakly in the uniform metrics in $D(0, 1)$ to a standard Wiener process.*

The limiting ν-dimensional Gaussian process is self-similar with Hurst parameter $H = \theta/2 < 1/2$. Its first component coincides in distribution with the first component of the limiting process in Theorem 1 in Durieu and Wang (2015). The above Karlin's theorems are partial cases of Theorem 3 due to $c_{ij} = c^*_{ij}(1, 1) - c^*_{i+1,j}(1, 1) - c^*_{i,j+1}(1, 1) + c^*_{i+1,j+1}(1, 1)$. Section 2 contains the proof of Theorem 3.