isid/ms/2013/13
October 22, 2013
http://www.isid.ac.in/~statmath/eprints

Rate of Convergence and Large Deviation for the Infinite Color Pólya Urn Schemes

Antar Bandyopadhyay and Debleena Thacker

Indian Statistical Institute, Delhi Centre
7, SJSS Marg, New Delhi–110 016, India

Rate of Convergence and Large Deviation for the Infinite Color Pólya Urn Schemes

Antar Bandyopadhyay[*][†]
Debleena Thacker[‡]

Theoretical Statistics and Mathematics Unit
Indian Statistical Institute, Delhi Centre
7 S. J. S. Sansanwal Marg
New Delhi 110016
INDIA

October 23, 2013

Abstract

In this work we consider the *infinite color urn model* associated with a bounded increment random walk on $\mathbb{Z}^d$. This model was first introduced in [2]. We prove that the rate of convergence of the expected configuration of the urn at time n with appropriate centering and scaling is of the order $\mathcal{O}\left(\frac{1}{\sqrt{\log n}}\right)$. Moreover we derive bounds similar to the classical Berry-Essen bound. Further we show that for the expected configuration a *large deviation principle (LDP)* holds with a good rate function and speed $\log n$.

Keywords: *Berry-Essen bound, infinite color urn, large deviation principle, rate of convergence, urn models.*

AMS 2010 Subject Classification: *Primary: 60F05, 60F10; Secondary: 60G50.*

[*]E-Mail: antar@isid.ac.in

[†]Also affiliated with: Theoretical Statistics and Mathematics Unit, Indian Statistical Institute, Kolkata; 203 B. T. Road, Kolkata 700108, INDIA

[‡]E-Mail: thackerdebleena@gmail.com

1 Introduction

Pólya urn scheme is one of the most well studied stochastic process which has plenty of applications in various different fields. Since the time of its introduction by Pólya [17] there has been a vast number of different variants and generalizations [12, 11, 1, 15, 13, 14, 10, 16] studied in literature. In general one considers the model with finitely many colors and then it can be described simply by

> Start with an urn containing finitely many balls of different colors. At any time $n \geq 1$, a ball is selected uniformly at random from the urn, and its color is noted. The selected ball is then returned to the urn along with a set of balls of various colors which may depend on the color of the selected ball.

In [6] Blackwell and MacQueen introduced a version of the model with possibly infinitely many colors but with a very simple replacement mechanism. Recently the authors of this work has introduced [2] a new generalization of the classical model with infinite but countably many colors with replacement mechanism corresponding to random walks in d-dimension. This generalization is essentially different than that of the classical Pólya urn scheme, as well as the model introduced in [6], where the replacement mechanism is diagonal. The generalization by [2] considers replacement mechanism with non-zero off diagonal entries and provides a novel connection between the two classical models, namely, Pólya urn scheme and random walks on d-dimensional Euclidean space has been demonstrated. In the current work we exploit this connection to derive the *rate of convergence* and the *large deviation principle* for the $(n+1)^{\text{th}}$ selected color in the infinite color generalization of the Pólya urn scheme. In the following subsection we describe the specific model which we study.

1.1 Infinite Color Urn Model Associated with Random Walks

Let $(X_j)_{j\geq 1}$ be i.i.d. random vectors taking values in $\mathbb{Z}^d$ with probability mass function $p(\mathbf{u}) := \mathbf{P}(X_1 = \mathbf{u}), \mathbf{u} \in \mathbb{Z}^d$. We assume that the distribution of X_1 is bounded, that is there exists a non-empty finite subset $B \subseteq \mathbb{Z}^d$ such that $p(u) = 0$ for all $u \notin B$. Throughout this paper we take the convention of writing all vectors as row vectors. Thus for a vector $\mathbf{x} \in \mathbb{R}^d$ we will write $\mathbf{x}^T$ to denote it as a column vector. The notations $\langle \cdot, \cdot \rangle$ will denote the usual Euclidean inner product on $\mathbb{R}^d$ and $\| \cdot \|$ the the Euclidean norm. We will always write

$$\begin{aligned} \boldsymbol{\mu} &:= \mathbf{E}\left[X_1\right] \\ \Sigma &:= \mathbf{E}\left[X_1^T X_1\right] \\ e\left(\boldsymbol{\lambda}\right) &:= \mathbf{E}\left[e^{\langle \boldsymbol{\lambda}, X_1\rangle}\right],\ \boldsymbol{\lambda} \in \mathbb{Z}^d. \end{aligned} \tag{1}$$

When the dimension $d = 1$ we will denote the mean and variance simply by μ and σ^2 respectively.

Let $S_n := X_0 + X_1 + \cdots + X_n, n \geq 0$ be the random walk on $\mathbb{Z}^d$ starting at X_0 and with increments $\left(X_j\right)_{j\geq 1}$ which are independent. Needless to say that $\left(S_n\right)_{n\geq 0}$ is Markov chain with state-space $\mathbb{Z}^d$, initial distribution given by the distribution of X_0 and the transition matrix $R := \left(\left(p\left(\mathbf{u} - \mathbf{v}\right)\right)\right)_{u,v\in\mathbb{Z}^d}$.

In [2] the following infinite color generalization of Pólya urn scheme was introduced where the colors were indexed by $\mathbb{Z}^d$. Let $U_n := \left(U_{n,\mathbf{v}}\right)_{\mathbf{v}\in\mathbb{Z}^d} \in [0,\infty)^{\mathbb{Z}^d}$ denote the configuration of the urn at time n, that is,

$$\mathbf{P}\left((n+1)^{\text{th}} \text{ selected ball has color } \mathbf{v} \,\Big|\, U_n, U_{n-1}, \cdots, U_0\right) \propto U_{n,\mathbf{v}},\ \mathbf{v} \in \mathbb{Z}^d.$$

Starting with U_0 which is a probability distribution we define $\left(U_n\right)_{n\geq 0}$ recursively as follows

$$U_{n+1} = U_n + C_{n+1}R \tag{2}$$

where $C_{n+1} = \left(C_{n+1,\mathbf{v}}\right)_{\mathbf{v}\in\mathbb{Z}^d}$ is such that $C_{n+1,V} = 1$ and $C_{n+1,\mathbf{u}} = 0$ if $\mathbf{u} \neq V$ where V is a random color chosen from the configuration U_n. In other words

$$U_{n+1} = U_n + R_V$$

where R_V is the V^{th} row of the replacement matrix R. Following [2] we define the process $\left(U_n\right)_{n\geq 0}$ as the *infinite color urn model* with initial configuration U_0 and replacement matrix R. We will also refer it as the *infinite color urn model associated with the random walk* $\left(S_n\right)_{n\geq 0}$ *on* $\mathbb{Z}^d$. Throughout this paper we will assume that $U_0 = \left(U_{0,\mathbf{v}}\right)_{\mathbf{v}\in\mathbb{Z}^d}$ is such that $U_{0,\mathbf{v}} = 0$ for all but finitely many $\mathbf{v} \in \mathbb{Z}^d$.

It is worth noting that $\sum_{\mathbf{u}\in\mathbb{Z}^d} U_{n,\mathbf{u}} = n + 1$ for all $n \geq 0$. So if Z_n denotes the $(n+1)^{\text{th}}$ selected color then

$$\mathbf{P}\left(Z_n = \mathbf{v} \,\Big|\, U_n, U_{n-1}, \cdots, U_0\right) = \frac{U_{n,\mathbf{v}}}{n+1} \Rightarrow \mathbf{P}\left(Z_n = \mathbf{v}\right) = \frac{\mathbf{E}\left[U_{n,\mathbf{v}}\right]}{n+1}. \tag{3}$$

In other words the expected configuration of the urn at time n is given by the distribution of Z_n.

1.2 Outline of the Main Contribution of the Paper

In [2] the authors studied the asymptotic distribution of Z_n, in particular, it has been proved (see Theorem 2.1 of [2]) that as $n \to \infty$,

$$\frac{Z_n - \boldsymbol{\mu} \log n}{\sqrt{\log n}} \xrightarrow{d} N_d\left(\mathbf{0}, \Sigma\right). \tag{4}$$

In Section 2 we find the rate of convergence for the above asymptotic and show that classical Berry-Essen type bound hold at any dimension $d \geq 1$, which is of the order $\mathcal{O}\left(\frac{1}{\sqrt{\log n}}\right)$.

It is easy to see that (4) implies

$$\frac{Z_n}{\log n} \xrightarrow{d} \boldsymbol{\mu} \text{ as } n \to \infty \Rightarrow \frac{Z_n}{\log n} \xrightarrow{p} \boldsymbol{\mu} \text{ as } n \to \infty. \tag{5}$$

So it is then natural to ask whether the sequence of measures $\left(\mathbf{P}\left(\frac{Z_n}{\log n} \in \cdot\right)\right)_{n\geq 2}$ satisfy a *large deviation principle (LDP)*. In Section 3 we show that the above sequence of measures satisfy a LDP with a good rate function and speed $\log n$. We also give an explicit representation of the rate function in terms of rate function of a marked Poisson process with intensity one and the markings given by the i.i.d. increments $\left(X_j\right)_{j\geq 1}$.

1.3 Fundamental Representation

We end the introduction with the following very important observation made in [2] (see Theorem 3.1 in [2])

$$Z_n \stackrel{d}{=} Z_0 + \sum_{j=1}^{n} I_j X_j \tag{6}$$

where $\left(X_j\right)_{j\geq 1}$ are as above and $\left(I_j\right)_{j\geq 1}$ are independent Bernoulli variables such that $I_j \sim \text{Bernoulli}\left(\frac{1}{j+1}\right)$ and are independent of $\left(X_j\right)_{j\geq 1}$. $Z_0 \sim U_0$ and is independent of $\left(\left(X_j\right)_{j\geq 1}, \left(I_j\right)_{j\geq 1}\right)$.

Note that using this representation the asymptotic normality (4) follows immediately as an application of the Lindeberg Central Limit Theorem [5]. We use this representation to derive the Berry-Essen type bounds and also the LDP.

2 Berry-Essen Bounds for the Expected Configuration

In this section we show that the rate of convergence of (4) is of the order $\mathcal{O}\left(\frac{1}{\sqrt{\log n}}\right)$. In fact we show that the Berry-Essen type bound holds for the color of the $(n+1)^{\text{th}}$-selected ball.

2.1 Berry-Essen Bound for $d = 1$

We first consider the case when the associated random walk is a one dimensional walk and the set of colors are indexed by the set of integers $\mathbb{Z}$.

Theorem 1. *Suppose $U_0 = \delta_0$ then*

$$\sup_{x\in\mathbb{R}} \left| \mathbf{P}\left(\frac{Z_n - \mu h_n}{\sqrt{n\rho_2}} \leq x\right) - \Phi\left(x\right) \right| \leq 2.75 \times \frac{\sqrt{n}\rho_3}{\rho_2^{3/2}} = \mathcal{O}\left(\frac{1}{\sqrt{\log n}}\right), \tag{7}$$

where $h_n := \sum_{j=1}^{n} \frac{1}{j+1}$, Φ is the standard normal distribution function and

$$\rho_2 := \frac{1}{n}\left(\sigma^2 h_n - \mu^2 \sum_{j=1}^{n} \frac{1}{(j+1)^2}\right) \tag{8}$$

and

$$\rho_3 := \frac{1}{n}\left(\sum_{j=1}^{n} \frac{1}{j+1}\mathbf{E}\left[\left|X_1 - \frac{\mu}{j+1}\right|^3\right] + |\mu|^3 \sum_{j=1}^{n} \frac{j}{(j+1)^4}\right). \tag{9}$$

Proof. We first note that when $U_0 = \delta_0$ then (6) can be written as

$$Z_n \stackrel{d}{=} \sum_{j=1}^{n} I_j X_j \tag{10}$$

where $(X_j)_{j\geq 1}$ are i.i.d. increments of the random walk $(S_n)_{n\geq 0}$, $(I_j)_{j\geq 1}$ are independent Bernoulli variables such that $I_j \sim \text{Bernoulli}\left(\frac{1}{j+1}\right)$ and are independent of $(X_j)_{j\geq 1}$.

Now observe that

$$n\rho_2 = \sum_{j=1}^{n} \mathbf{E}\left[\left(I_j X_j - \mathbf{E}\left[I_j X_j\right]\right)^2\right] \text{ and } n\rho_3 = \sum_{j=1}^{n} \mathbf{E}\left[\left|I_j X_j - \mathbf{E}\left[I_j X_j\right]\right|^3\right].$$

Thus from the *Berry-Essen Theorem* for the independent but non-identical increments (see Theorem 12.4 of [4]) we get

$$\sup_{x\in\mathbb{R}} \left| \mathbf{P}\left(\frac{\sum_{j=1}^{n} I_j X_j - \mu h_n}{\sqrt{n\rho_2}} \leq x \right) - \Phi\left(x\right) \right| \leq 2.75 \times \frac{\sqrt{n}\rho_3}{\rho_2^{3/2}}. \tag{11}$$

The equations (10) and (11) implies the inequality in (7).

Finally to prove the last part of the equation (7) we note that from definition $n\rho_2 \sim C_1 \log n$ and $n\rho_3 \sim C_2 \log n$ where $0 < C_1, C_2 < \infty$ are some constants. Thus

$$\frac{\sqrt{n}\rho_3}{\rho_2^{3/2}} = \mathcal{O}\left(\frac{1}{\sqrt{\log n}}\right).$$

This completes the proof of the theorem. □

Following result follows easily from the above theorem by observing the facts $h_n \sim \log n$ and $n\rho_2 \sim C_1 \log n$.

Theorem 2. *Suppose $U_{0,k} = 0$ for all but finitely many $k \in \mathbb{Z}$ then there exists a constant $C > 0$ such that*

$$\sup_{x\in\mathbb{R}} \left| \mathbf{P}\left(\frac{Z_n - \mu \log n}{\sigma\sqrt{\log n}} \leq x \right) - \Phi\left(x\right) \right| \leq C \times \frac{\sqrt{n}\rho_3}{\rho_2^{3/2}} = \mathcal{O}\left(\frac{1}{\sqrt{\log n}}\right), \tag{12}$$

Φ is the standard normal distribution function and ρ_2 and ρ_3 are as defined in (8) *and* (9) *respectively.*

It is worth noting that unlike in Theorem 1 the constant C which appears in (12) above, is not a universal constant, it may depend on the increment distribution, as well as on U_0.

2.2 Berry-Essen bound for $d \geq 2$

We now consider the case when the associated random walk is $d \geq 2$ dimensional and the colors are indexed by $\mathbb{Z}^d$. Before we present our main result we introduce few notations.

Notations: For a vector $\mathbf{x} \in \mathbb{R}^d$ we will write the coordinates as $\left(x^{(1)}, x^{(2)}, \cdots, x^{(d)}\right)$. For example the coordinates of $\boldsymbol{\mu}$ will be written as $\left(\mu^{(1)}, \mu^{(2)}, \cdots, \mu^{(d)}\right)$. For a matrix $A = ((a_{ij}))_{1\leq i,j\leq d}$ we denote by $A\left(i,j\right)$

the $(d-1)\times(d-1)$ sub-matrix of A, obtained by deleting the i^{th} row and j^{th} column. Let

$$\rho_2^{(d)} := \frac{1}{n}\sum_{j=1}^{n}\frac{1}{(j+1)}\frac{\det\left(\Sigma-\frac{1}{j+1}M\right)}{\det\left(\Sigma(1,1)-\frac{1}{j+1}M(1,1)\right)}, \tag{13}$$

where $M := \left(\left(\mu^{(i)}\mu^{(j)}\right)\right)_{1\le i,j\le d}$ and

$$\rho_3^{(d)} := \frac{1}{nd}\sum_{j=1}^{n}\sum_{i=1}^{d}\gamma_n^3\left(i\right)\beta_j\left(i\right), \tag{14}$$

where

$$\gamma_n^2(i) := \max_{1\le j\le n}\frac{\det\left(\Sigma(i,i)-\frac{1}{(j+1)}M(i,i)\right)}{\det\left(\Sigma(1,1)-\frac{1}{j+1}M(1,1)\right)}$$

and

$$\beta_j(i) = \frac{1}{j+1}\mathbf{E}\left[\left|X_1^{(i)}-\frac{\mu^{(i)}}{j+1}\right|^3\right]+\frac{j}{(j+1)^4}\left|\mu^{(i)}\right|^3.$$

For any two vectors $\mathbf{x}$ and $\mathbf{y}\in\mathbb{R}^d$ we will write $\mathbf{x}\le\mathbf{y}$, if the inequality holds coordinate wise. Finally for a positive definite matrix B, we write $B^{1/2}$ for the unique positive definite square root of it.

Theorem 3. *Suppose $U_0=\delta_{\mathbf{0}}$ then there exists an universal constant $C\left(d\right)>0$ which may depend on the dimension d such that*

$$\sup_{\mathbf{x}\in\mathbb{R}^d}\left|\mathbf{P}\left(\left(Z_n-\boldsymbol{\mu}h_n\right)\Sigma_n^{-1/2}\le\mathbf{x}\right)-\Phi_d\left(\mathbf{x}\right)\right|\le C\left(d\right)\frac{\sqrt{n}\rho_3^{(d)}}{\left(\rho_2^{(d)}\right)^{3/2}}=\mathcal{O}\left(\frac{1}{\sqrt{\log n}}\right), \tag{15}$$

where $\Sigma_n := \sum_{j=1}^{n}\frac{1}{j+1}\left(\Sigma-\frac{1}{j+1}M\right)$ and Φ_d is the distribution function of a standard d-dimensional normal random vector.

Proof. Like in the one dimensional case, we start by observing that when $U_0=\delta_0$ then (6) can be written as

$$Z_n \stackrel{d}{=} \sum_{j=1}^{n}I_jX_j \tag{16}$$

where $(X_j)_{j\geq 1}$ are i.i.d. increments of the random walk $(S_n)_{n\geq 0}$, $(I_j)_{j\geq 1}$ are independent Bernoulli variables such that $I_j \sim \text{Bernoulli}\left(\frac{1}{j+1}\right)$ and are independent of $(X_j)_{j\geq 1}$.

Now the proof of the inequality in (15) follows from equation (D) of [3] which deals with d-dimensional version of the classical Berry-Essen inequality for independent but non-identical summands, which in our case are the random variables $(I_jX_j)_{j\geq 1}$. It is enough to notice that

$$\beta_j(i) = \mathbf{E}\left[\left|I_jX_1^{(i)} - \mathbf{E}\left[I_jX_j^{(i)}\right]\right|^3\right],$$

and

$$\Sigma_n = \sum_{j=1}^{n} \mathbf{E}\left[\left(I_jX_j - \mathbf{E}\left[I_jX_j\right]\right)^T\left(I_jX_j - \mathbf{E}\left[I_jX_j\right]\right)\right].$$

Finally to prove the last part of the equation (15) just like in the one dimensional case we note that from definition $n\rho_2^{(d)} \sim C_1' \log n$ and $n\rho_3^{(d)} \sim C_2' \log n$ where $0 < C_1', C_2' < \infty$ are some constants. Thus

$$\frac{\sqrt{n}\rho_3}{\rho_2^{3/2}} = \mathcal{O}\left(\frac{1}{\sqrt{\log n}}\right).$$

This completes the proof of the theorem. □

Remark 1. If we define that $\Sigma(1,1) = 1$ and $M(1,1) = 0$ when $d = 1$ then Theorem 1 follows from the above theorem except in Theorem 1 the constant is more explicit.

Just like in the one dimensional case the following result follows easily from the above theorem by observing $h_n \sim \log n$.

Theorem 4. *Suppose $U_0 = (U_{0,\mathbf{v}})_{\mathbf{v}\in\mathbb{Z}^d}$ is such that $U_{0,\mathbf{v}} = 0$ for all but finitely many $\mathbf{v} \in \mathbb{Z}^d$ then there exists a constant $C > 0$ which may depend on the increment distribution, such that*

$$\sup_{\mathbf{x}\in\mathbb{R}^d}\left|\mathbf{P}\left(\left(\frac{Z_n - \boldsymbol{\mu}\log n}{\sqrt{\log n}}\right)\Sigma^{-1/2} \leq \mathbf{x}\right) - \Phi_d(\mathbf{x})\right| \leq C \times \frac{\sqrt{n}\rho_3^{(d)}}{\left(\rho_2^{(d)}\right)^{3/2}} = \mathcal{O}\left(\frac{1}{\sqrt{\log n}}\right), \tag{17}$$

where Φ_d is the distribution function of a standard d-dimensional normal random vector.

3 Large Deviations for the Expected Configuration

In this section we discuss the asymptotic behavior of the tail probabilities of $\frac{Z_n}{\log n}$. Following standard notations are used in rest of the paper. For any subset $A \subseteq \mathbb{R}^d$ we write A° to denote the *interior* of A and $\bar{A}$ to denote the *closer* of A under the usual Euclidean metric.

Theorem 5. *The sequence of measures* $\mathbf{P}\left(\frac{Z_n}{\log n} \in \cdot\right)_{n\geq 2}$ *satisfy a LDP with rate function* $I(\cdot)$ *and speed* $\log n$*, that is,*

$$-\inf_{\mathbf{x}\in A^\circ} I(\mathbf{x}) \leq \varliminf_{n\to\infty} \frac{\log \mathbf{P}\left(\frac{Z_n}{\log n} \in A\right)}{\log n} \leq \varlimsup_{n\to\infty} \frac{\log \mathbf{P}\left(\frac{Z_n}{\log n} \in A\right)}{\log n} \leq -\inf_{\mathbf{x}\in \bar{A}} I(\mathbf{x}) \tag{18}$$

where $I(\cdot)$ *is the Fenchel-Legendre dual of* $e(\cdot) - 1$*, that is for* $x \in \mathbb{R}^d$*,*

$$I(x) = \sup_{\boldsymbol{\lambda}\in\mathbb{R}^d} \{\langle \mathbf{x}, \boldsymbol{\lambda}\rangle - e(\boldsymbol{\lambda}) + 1\}. \tag{19}$$

Moreover $I(\cdot)$ *is convex and a* good rate function*.*

Proof. We start with the representation (6)

$$Z_n \stackrel{d}{=} Z_0 + \sum_{j=1}^{n} I_j X_j$$

where as earlier $(X_j)_{j\geq 1}$ are i.i.d. increments of the random walk $(S_n)_{n\geq 0}$ on $\mathbb{Z}^d$ and $(I_j)_{j\geq 1}$ are independent Bernoulli variables such that $I_j \sim \text{Bernoulli}\left(\frac{1}{j+1}\right)$ and are independent of $(X_j)_{j\geq 1}$. $Z_0 \sim U_0$ and is independent of $\left((X_j)_{j\geq 1}, (I_j)_{j\geq 1}\right)$. Now without loss of any generality we may assume that $Z_0 = \mathbf{0}$ with probability one, that is, $U_0 = \delta_{\mathbf{0}}$.

Consider the following scaled *logarithmic moment generating function* of Z_n,

$$\Lambda_n(\boldsymbol{\lambda}) := \frac{1}{\log n} \log \mathbb{E}\left[e^{\langle \boldsymbol{\lambda}, Z_n\rangle}\right]. \tag{20}$$

From (6) it follows that

$$\mathbb{E}\left[e^{\langle \boldsymbol{\lambda}, Z_n\rangle}\right] = \frac{1}{n+1}\Pi_n\left(e\left(\boldsymbol{\lambda}\right)\right)$$

where $\Pi_n(z) = \prod_{j=1}^{n}\left(1+\frac{z}{j}\right)$, $z \in \mathbb{C}$. Using Gauss's formula (see page 178 of [8]) we have

$$\lim_{n\to\infty} \frac{\Pi_n(z)}{n^z}\Gamma(z+1) = 1 \tag{21}$$

and the convergence happens uniformly on compact subsets of $\mathbb{C}\backslash\{-1,-2,\ldots\}$. Therefore we get

$$\Lambda_n(\boldsymbol{\lambda}) \longrightarrow e(\boldsymbol{\lambda}) - 1 < \infty \;\; \forall\, \boldsymbol{\lambda} \in \mathbb{R}^d. \tag{22}$$

Thus the LDP as stated in (18) follows from the Gärtner-Ellis Theorem (see Remark (a) on page 45 of [9] or page 66 of [7]).

We next note that $I(\cdot)$ is a convex function because it is the Fenchel-Legendre dual of $e(\boldsymbol{\lambda}) - 1$ which is finite for all $\boldsymbol{\lambda} \in \mathbb{R}^d$.

Finally, we will show that $I(\cdot)$ is good rate function, that is, the level sets $A(\alpha) = \{\mathbf{x}\colon I(\mathbf{x}) \le \alpha\}$ are compact for all $\alpha > 0$. Since I is a rate function so by definition it is lower semicontinuous. So it is enough to prove that $A(\alpha)$ is bounded for all $\alpha \in \mathbb{R}$.

Observe that for all $\mathbf{x} \in \mathbb{R}^d$,

$$I(\mathbf{x}) \ge \sup_{\|\boldsymbol{\lambda}\|=1} \{\langle \mathbf{x}, \boldsymbol{\lambda}\rangle - e(\boldsymbol{\lambda}) + 1\}.$$

Now the function $\boldsymbol{\lambda} \mapsto e(\boldsymbol{\lambda})$ is continuous and $\{\boldsymbol{\lambda}\colon \|\boldsymbol{\lambda}\| = 1\}$ is a compact set. So $\exists\, \boldsymbol{\lambda}_0 \in \{\boldsymbol{\lambda}\colon \|\boldsymbol{\lambda}\| = 1\}$ such that $\sup_{|\boldsymbol{\lambda}|=1} e(\boldsymbol{\lambda}) = e(\boldsymbol{\lambda}_0)$. Therefore for $\|\mathbf{x}\| \ne 0$ choosing $\boldsymbol{\lambda} = \frac{\mathbf{x}}{\|\mathbf{x}\|}$, we have $I(\mathbf{x}) \ge \|\mathbf{x}\| - e(\boldsymbol{\lambda}_0) + 1$. So if $\mathbf{x} \in A(\alpha)$ then

$$\|\mathbf{x}\| \le (\alpha + e(\boldsymbol{\lambda}_0) - 1).$$

This proves that the level sets are bounded, which completes the proof. □

Our next result is an easy consequence of (19) which can be used to compute explicit formula for the rate function I in many examples in one or higher dimensions.

Theorem 6. *The rate function I is same as the rate function for the large deviation of the empirical means of i.i.d. random vectors with distribution corresponding to the distribution of the following random vector*

$$W = \sum_{i=1}^{N} X_i, \tag{23}$$

where $N \sim$ Poisson(1) and is independent of $(X_j)_{j\ge1}$ which are the i.i.d. increments of the associated random walk.

Proof. We first observe that $\log \mathbf{E}\left[e^{\langle \boldsymbol{\lambda}, W\rangle}\right] = e\left(\boldsymbol{\lambda}\right) - 1$. The rest then follows from (19) and Cramér's Theorem (see Theorem 2.2.30 of [9]). $\square$

Remark 2. Using Theorem 6 we can conclude that the tail of the asymptotic distribution of Z_n can be approximated by the tail of a marked Poisson process with intensity one where the markings are given by the i.i.d. increments of the associated random walk.

For $d = 1$, one can get more information about the rate function I, in particular following result it follows from Theorem 6 and Lemma 2.2.5 of [9].

Proposition 7. *Suppose $d = 1$ then $I(x)$ is non-decreasing when $x \geq \mu$ and non-increasing when $x \leq \mu$. Moreover*

$$I(x) = \begin{cases} \sup\limits_{\lambda \geq 0}\{x\lambda - e(\lambda) + 1\} & \text{if } x \geq \mu \\ \sup\limits_{\lambda \leq 0}\{x\lambda - e(\lambda) + 1\} & \text{if } x \leq \mu. \end{cases} \tag{24}$$

In particular, $I(\mu) = \inf_{x \in \mathbb{R}} I(x)$.

Following is an immediate corollary of the above result and Theorem 5.

Corollary 8. *Let $d = 1$ then for any $\epsilon > 0$*

$$\lim_{n \to \infty} \frac{1}{\log n} \log \mathbf{P}\left(\frac{Z_n}{\log n} \geq \mu + \epsilon\right) = -I\left(\mu + \epsilon\right) \tag{25}$$

and

$$\lim_{n \to \infty} \frac{1}{\log n} \log \mathbf{P}\left(\frac{Z_n}{\log n} \leq \mu - \epsilon\right) = -I\left(\mu - \epsilon\right). \tag{26}$$

We end the section with explicit computations of the rate functions for two examples of infinite color urn models associated with random walks on one dimensional integer lattice.

Example 1. Our first example is the case when the random walk is trivial, which moves deterministically one step at a time. In other words $X_1 = 1$ with probability one. In this case $\mu = 1$ and $\sigma^2 = 1$. Also the moment generating function of X_1 is given by $e\left(\lambda\right) := e^{\lambda}$, $\lambda \in \mathbb{R}$. By Theorem 6 the rate function for the associated infinite color urn model is same as the rate function for a Poisson random variable with mean 1, that is

$$I(x) = \begin{cases} +\infty & \text{if } x < 0 \\ 1 & \text{if } x = 0 \\ x \log x - x + 1 & \text{if } x > 0 \end{cases} \tag{27}$$

Thus for this example one can prove a *Poisson approximation* for Z_n.

Example 2. Our next example is the case when the random walk is the *simple symmetric random walk* on the one dimensional integer lattice. For this case we note that $\mu = 0$, $\sigma^2 = 1$ and the moment generating function X_1 is $e(\lambda) = \cosh \lambda$, $\lambda \in \mathbb{R}$. The rate function for the associated infinite color urn model turns out to be

$$I(x) = x \sinh^{-1} x - \sqrt{1 + x^2} + 1. \tag{28}$$

References

[1] A. Bagchi and A. K. Pal. Asymptotic normality in the generalized Pólya-Eggenberger urn model, with an application to computer data structures. *SIAM J. Algebraic Discrete Methods*, 6(3):394–405, 1985.

[2] Antar Bandyopadhyay and Debleena Thacker. Pólya urn schemes with infinitely many colors. (http://arxiv.org/pdf/1303.7374v2.pdf), 2013.

[3] Harald Bergström. On the central limit theorem in the case of not equally distributed random variables. *Skand. Aktuarietidskr.*, 32:37–62, 1949.

[4] R. N. Bhattacharya and R. Ranga Rao. *Normal approximation and asymptotic expansions*. John Wiley & Sons, New York-London-Sydney, 1976. Wiley Series in Probability and Mathematical Statistics.

[5] Patrick Billingsley. *Probability and measure*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, second edition, 1986.

[6] David Blackwell and James B. MacQueen. Ferguson distributions via Pólya urn schemes. *Ann. Statist.*, 1:353–355, 1973.

[7] Arijit Chakrabarty. *When is a Truncated Heavy Tail Heavy?* PhD thesis, Cornell University, 2010.

[8] John B. Conway. *Functions of one complex variable*, volume 11 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1978.

[9] Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications*. Jones and Bartlett Publishers, Boston, MA, 1993.

[10] Philippe Flajolet, Philippe Dumas, and Vincent Puyhaubert. Some exactly solvable models of urn process theory. In *Fourth Colloquium on Mathematics and Computer Science Algorithms, Trees, Combinatorics and Probabilities*, Discrete Math. Theor. Comput. Sci. Proc., AG, pages 59–118. Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2006.

[11] David A. Freedman. Bernard Friedman's urn. *Ann. Math. Statist*, 36:956–970, 1965.

[12] Bernard Friedman. A simple urn model. *Comm. Pure Appl. Math.*, 2:59–70, 1949.

[13] Raúl Gouet. Strong convergence of proportions in a multicolor Pólya urn. *J. Appl. Probab.*, 34(2):426–435, 1997.

[14] Svante Janson. Functional limit theorems for multitype branching processes and generalized Pólya urns. *Stochastic Process. Appl.*, 110(2):177–245, 2004.

[15] Robin Pemantle. A time-dependent version of Pólya's urn. *J. Theoret. Probab.*, 3(4):627–637, 1990.

[16] Robin Pemantle. A survey of random processes with reinforcement. *Probab. Surv.*, 4:1–79, 2007.

[17] G. Pólya. Sur quelques points de la théorie des probabilités. *Ann. Inst. H. Poincaré*, 1(2):117–161, 1930.