Asymptotic normality of urn models for clinical trials with delayed response

FEIFANG HU\textsuperscript{1} AND LI-XIN ZHANG\textsuperscript{2}

University of Virginia and Zhejiang University

\textsuperscript{1}Research supported by a grant R-155-000-015-112 from the National University of Singapore and a grant from University of Virginia.

\textsuperscript{2}Research supported by National Natural Science Foundation of China (No. 10071072) and National Natural Science Foundation of Zhejiang Province.


Key words. Asymptotic normality, generalized Friedman’s urn, staggered entry, treatment allocation, urn models.

Abbreviated Title: Asymptotic normality of urn models

Abstract

For response-adaptive clinical trials using a generalized Friedman’s urn design, we derive the asymptotic normality of sample fraction assigned to each treatment under staggered entry and delayed response. The rate of convergence and the law of the iterated logarithm are obtained for both the urn composition and the sample fraction. Some applications are also discussed.
1 Introduction

Response-adaptive designs involve the sequential selection of design points, chosen depending on the outcomes at previously selected design points, or treatments. Since the fundamental paper of Robbins (1952), adaptive designs have been extensively studied in the literature. Rosenberger (1996) provides a good review of adaptive designs.

An important family of adaptive design can be developed from the generalized Friedman’s urn (GFU) model [cf. Athreya and Karlin (1968) and Rosenberger (2002)]. It is also called the generalized Pólya urn (GPU) in the literature. The model is described as follows. Consider an urn containing particles of $K$ types, respectively representing $K$ ‘treatments’ in a clinical trial. Initially the urn contains $Y_0 = (Y_{01}, \ldots, Y_{0K})$ particles, where $Y_{0k}$ denotes the number of particles of type $k$, $k = 1, \ldots, K$. At stage $i$, $i = 1, \ldots, n$, a particle is drawn from the urn and replaced. If the particle is of type $k$, then the treatment $k$ is assigned to the $i$-th patient, $k = 1, \ldots, K$, $i = 1, \ldots, n$. We then wait for observing a random variable $\xi_i$, the response of the treatment at the patient $i$. After that, an additional $D_{k,q}(i)$ particles of type $q$, $q = 1, \ldots, K$ are added to the urn, where $D_{k,q}(i)$ is some function of $\xi_i$. This procedure is repeated throughout the $n$ stages. After $n$ stages, the urn composition is denoted by the row vector $Y_n = (Y_{n1}, \ldots, Y_{nK})$, where $Y_{nk}$ represents the number of particles of type $k$ in the urn after the $n$-th addition of particles. This relation can be written as the following recursive formula:

$$Y_n = Y_{n-1} + X_n D_n,$$  \hspace{1cm} (1.1)

where $D_n = D(\xi_n) = (D_{k,q}(n))_{k,q=1}^K$ is an $K \times K$ random matrix with $D_{k,q}(n)$ being its element in the $k$-th row and the $q$-th column, and $X_n$ is the result of the $n$-th draw, distributed according to the urn composition at the previous stages, i.e., if the $n$-th draw is type $k$ particle, then the $k$-th component of $X_n$ is 1 and other components are 0.

Furthermore, write $N_n = (N_{n1}, \ldots, N_{nK})$, where $N_{nk}$ is the number of times a type $k$ particle drawn in the first $n$ stages. In clinical trials, $N_{nk}$ represents the number of patients assigned to the treatment $k$ in the first $n$ trials. Obviously,

$$N_n = \sum_{k=1}^{n} X_n.$$  \hspace{1cm} (1.2)
Moreover, denote $H_i = \left( E(D_{k,q}(i)) \right)_{k,q=1}^K, \ i = 1, \ldots, n$. The matrices $D_i$’s are called the addition rules and $H_i$’s the generating matrices. A GFU model is said to be homogeneous if $H_i = H$ for all $i = 1, \ldots, n$.

Athreya and Karlin (1968) first considered the asymptotic properties of the GFU model with homogeneous generating matrix. Smythe (1996) defined the extended Pólya urn model (EPU) (a special class of GFU) and considered its asymptotic normality. For nonhomogeneous generating matrices, Bai and Hu (1999) establish strong consistency and asymptotic normality of $Y_n$ in GFU model.

Typically, clinical trials do not result in immediate outcomes, i.e., individual patient outcomes may not be immediately available prior to the randomization of the next patient. Consequently, the urn can not be updated immediately, but can be updated when the outcomes become available (See Wei (1988) and Bandyopadhyay and Biswas (1996) for further discussion). A motivating example was studied in Tamura, Faries, Andersen, and Heiligenstein (1994). Recently, Bai, Hu and Rosenberger (2002) established the asymptotic distribution of the urn composition $Y_n$ under very general delay mechanisms. But they did not obtain the asymptotic normality of the sample fractions $N_n$.

In a clinical trial, $N_n$ represents the number of patients assigned to each treatment. The asymptotic distribution of $N_n$ is essential to determine the required sample size of adaptive designs (Hu, 2002). In this paper, we establish the asymptotic normality of $N_n$ with delayed responses. We also obtain the law of the iterated logarithm of both $Y_n$ and $N_n$. The technique used in this paper is an approximation of Gaussian processes, which is different from the martingale approximation in Bai, Hu and Rosenberger (2002). Some applications of the theorems are discussed in the remarks.

In the models which we consider in this paper, the urn will be updated by adding a fixed total number $\beta (\beta > 0)$ of balls into the urn whenever an outcome becomes available. The randomized Pólya urn (RPU) design of Durham, Flournoy and Li (1998) is not of such cases, where the number of balls added into the urn after an experiment can be zero and the urn composition only changes after each success. We are not sure whether or not our techniques can be used to study the RPU design with delayed responses.
2 Main Results

Typically, the urn models are simply not appropriate for today’s oft-performed long-term survival trials, where outcomes may not be ascertainable for many years. However, there are many trials where many or most outcomes are available during the recruitment period, even though individual patient outcomes may not be immediately available prior to the randomization of the next patient. Consequently, the urn can be updated when outcomes become available. We consider such trials, and assume that the delay times of the outcomes are not very long comparing with the time intervals for patients to entry to the trail. More clearly, if we denote an indicator function \( \delta_{jk}, j < k \), that takes the value 1 if the response of patient \( j \) occurs before patient \( k \) is randomized and 0 otherwise, then we will assume that \( P(\delta_{jk} = 0) = o((k-j)^{-\gamma}) \) as \( k-j \to \infty \) for some \( \gamma > 0 \) (See Assumption 2.1). Also, in clinical trials, it is more reasonable to think of censored observations when one considers delayed response, since we often do not wait a long time for a response to occur. However, in this paper we pay our main attention to the asymptotic properties, and so we assume each response will ultimately occur even though its occurrence may take a very long time. We would rather left the trials with censored observations to future studies, since censored observations will make the model more complex.

To describe the model clearly, we use the notation of Bai, Hu and Rosenberger (2002). Assume a multinomial response model with responses \( \xi_n = l \) if patient \( n \) had response \( l, l = 1, \ldots, L \), where \( \{\xi_n, n = 1, 2, \ldots\} \) is a sequence of independent random variables. Let \( J_n \) be the treatment indicator for the \( n \)-th patient, i.e., \( J_n = j \) if patient \( n \) was randomized to treatment \( j = 1, \ldots, K \). Then \( X_n = (X_{n1}, \ldots, X_{nk}) \) satisfies \( X_nJ_n = 1 \) and all other elements 0. We assume that the entry time of the \( n \)-th patient is \( t_n \), where \( \{t_n-t_{n-1}\} \) are independent for all \( n \). The response time of the \( n \)-th patient on treatment \( j \) with response \( l \) is denoted by \( \tau_n(j,l) \), whose distribution can depend on both the treatment assigned and the response observed. Let \( M_{jl}(n,m) \) be the indicator function that keeps track of when patients respond, and it takes the value 1 if \( t_n + \tau_n(j,l) \in (t_{n+m}, t_{n+m+1}] \) \( m \geq 0, \) and 0 otherwise. By definition, for every pair of \( n \) and \( j \), there is only one pair \( (l,m) \) such that \( M_{jl}(n,m) = 1 \) and \( M_{jl'}(n,m') = 0 \) for all \( (l,m) \neq (l',m') \). Also, for fixed \( n \) and \( j \), if the event \( \{\xi_n = l\} \) occurs, i.e., the response of the \( n \)-th patient is \( l \), then there is only one \( m \) such
that $M_{jl}(n, m) = 1$; while if the event $\{\xi_n = l\}$ does not occur, then $M_{jl}(n, m) = 0$ for all $m$. Consequently,

$$\sum_{m=0}^{\infty} M_{jl}(n, m) = I\{\xi_n = l\} \text{ for } j = 1, \ldots, K, l = 1, \ldots, L, n = 1, 2, \ldots.$$  \hspace{1cm} (2.3)

We can define $\mu_{jlm}(n) = E\{M_{jl}(n, m)\}$ as the probability that the $n$-th patient on treatment $j$ with response $l$ will respond after $m$ more patients are enrolled and before $m + 1$ more patients are enrolled. Then

$$\sum_{m=0}^{\infty} \mu_{jlm}(n) = P(\xi_n = l) \text{ for } j = 1, \ldots, K, l = 1, \ldots, L, n = 1, 2, \ldots.$$  \hspace{1cm} (2.3)

If we assume that $\{M_{jl}(n, m), n = 1, 2, \ldots\}$ are i.i.d. for fixed $j, l$, and $m$. Then $\mu_{jlm}(n) = \mu_{jlm}$ does not dependent on $n$.

For patient $n$, after observing $\xi_n = l$, $J_n = i$, we add $d_{ij}(\xi_n = l)$ balls of type $j$ to the urn, where the total number of balls added at each stage is constant; i.e., $\sum_{j=1}^{K} d_{ij}(l) = \beta$, where $\beta > 0$. Without loss of generality, we can assume $\beta = 1$. Let $D(\xi_n) = (d_{ij}(\xi_n), i, j = 1, \ldots, K)$. So, for given $n$ and $m$, if $M_{jl}(n, m) = 1$, then we add balls at the $n + m$-th stage (i.e., after the $n + m$-th patient is assigned and before the $n + m + 1$-th patient is assigned) according to $X_n D(l)$. Since $M_{jl'}(n, m) = 0$ for all $l' \neq l$, so $X_n D(l) = \sum_{l'=1}^{L} M_{jl'}(n, m)X_n D(l')$. Consequently, the numbers of balls of each type added to the urn after the $n$-th patient is assigned and before the $n + 1$-th patient is assigned are

$$W_n = \sum_{m=0}^{n-1} \sum_{l=1}^{L} M_{j_{n-m},l}(n - m, m)X_{n-m} D(l) = \sum_{m=1}^{n} \sum_{l=1}^{L} M_{j_{m},l}(m, n - m)X_{m} D(l).$$

Let $Y_n = (Y_{n1}, \ldots, Y_{nK})$ be the urn composition when the $n + 1$-th patient arrives to be randomized. Then

$$Y_n = Y_{n-1} + W_n.$$  \hspace{1cm} (2.4)
Note (2.3). By (2.4) we have

\[ Y_n - Y_0 = \sum_{k=1}^{n} W_k = \sum_{k=1}^{n} \sum_{m=1}^{k} \sum_{l=1}^{L} M_{J_m,l}(m, k - m)X_mD(l) \]

\[ = \sum_{m=1}^{n} \sum_{k=m}^{n} \sum_{l=1}^{L} M_{J_m,l}(m, k)X_mD(l) \]

\[ = \sum_{l=1}^{L} \sum_{m=1}^{n} \sum_{k=0}^{n-m} M_{J_m,l}(m, k)X_mD(l) \]

\[ = \sum_{l=1}^{L} \sum_{m=1}^{n} \sum_{k=0}^{n-m} M_{J_m,l}(m, k)X_mD(l) - \sum_{l=1}^{L} \sum_{m=1}^{n} \sum_{k=0}^{\infty} M_{J_m,l}(m, k)X_mD(l) \]

\[ = \sum_{m=1}^{n} X_mD(\xi_m) - \sum_{l=1}^{L} \sum_{m=1}^{\infty} \sum_{k=n-m+1}^{\infty} M_{J_m,l}(m, k)X_mD(l). \]

That is

\[ Y_n = Y_0 + \sum_{m=1}^{n} X_mD(\xi_m) + R_n, \quad (2.5) \]

where

\[ R_n = -\sum_{l=1}^{L} \sum_{m=1}^{n} \sum_{k=n-m+1}^{\infty} M_{J_m,l}(m, k)X_mD(l). \quad (2.6) \]

If there is no delay, i.e., \( M_{J_m,l}(m, k) = 0 \) for all \( k \geq 1 \) and all \( m \) and \( l \), then \( R_n = 0 \) and (2.5) reduces to

\[ Y_n = Y_0 + \sum_{m=1}^{n} X_mD(\xi_m), \quad (2.7) \]

which is just the model (1.1). We will show that \( R_n \) is small. So, it is natural that stochastic staggered entry and delay mechanisms do not affect the limiting distribution of the urn. But, it shall be mentioned that the distance between \( Y_n \) in (2.5) and that in (2.7) is not \( R_n \), since the distributions of \( X_n \) (with and without delayed responses) are different. So, the asymptotic properties of the model when delayed responses occur do not simply follow from those when delayed responses do not appear.

For simplicity, we assume that the responses \( \{\xi_n, n = 1, 2\ldots\} \) are i.i.d. random variables and let \( H = E[D(\xi_n)] \), i.e., we only consider the homogeneous case. For the nonhomogeneous case, if
there exist $V_{qij}, q, i, j = 1, \ldots, K$ and $H$ such that

$$\text{Cov}\{d_{qi}(\xi_n), d_{qj}(\xi_n)\} \to V_{qij}, \quad q, i, j = 1, \ldots, K$$

and

$$\sum_{n=1}^{n} \|E[D(\xi_n)] - H\| = o(n^{1/2}),$$

then we have the same asymptotic properties.

To give asymptotic properties of $N_n$, first we will require the following assumptions (same as Assumption 1 and 2 of Bai, Hu and Rosenberger (2002)):

**Assumption 2.1** For some $\gamma \in (0, 1)$,

$$\sum_{i=m}^{\infty} \mu_{ji}(n) = o(m^{-\gamma}) \text{ uniformly in } n.$$

The left of the above equality is just the probability of the event that the $n$-th patient on treatment $j$ responses $l$ after at least another $m$ patients are assigned.

Assumption 2.1 is for us to assume that the response times are not very long comparing with the entry time intervals. This assumption is satisfied if the response time $\tau_n(j, l)$ and the entry time $t_n$ satisfy (i) $t_2 - t_1, t_3 - t_2, \ldots$, are i.i.d. random variables with $E(t_2 - t_1)^2 < \infty$, and (ii) $\sup_n E|\tau_n(j, l)|^{\gamma'} < \infty$ for each $j, l$ and some $\gamma' > \gamma$ (See Bai, Hu and Rosenberger (2002)). So this assumption is widely satisfied.

**Assumption 2.2** Let $1 = (1, \ldots, 1)$. Note $H1' = 1'$. We assume that $H$ has the following Jordan decomposition:

$$T^{-1}HT = \text{diag}[1, J_2, \ldots, J_s],$$

where $J_s$ is a $\nu_t \times \nu_t$ matrix, given by

$$J_t = \begin{pmatrix}
\lambda_t & 1 & 0 & \ldots & 0 \\
0 & \lambda_t & 1 & \ldots & 0 \\
0 & 0 & \lambda_t & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & \lambda_t
\end{pmatrix}.$$
Let $v$ be the normalized left eigenvector of $H$ corresponding to its maximal eigenvalue 1. We may select the matrix $T$ so that its first column is $1'$ and the first row of $T^{-1}$ is $v$. Let $\lambda = \max\{Re(\lambda_2), \ldots, Re(\lambda_s)\}$ and $\nu = \max_j\{\nu_j : Re(\lambda_j) = \lambda\}$.

We also denote $\bar{H} = H - 1'v$, $\Sigma_1 = diag(v) - v'v$ and $\Sigma_2 = E[(D(\xi_n) - H)'diag(v)(D(\xi_n) - H)]$. We have the following asymptotic normality for $(Y_n, N_n)$.

**Theorem 2.1** Under Assumptions 2.1 and 2.2, if $\gamma > 1/2$ and $\lambda < 1/2$, then

$$n^{1/2}(Y_n/n - v, N_n/n - v) \xrightarrow{D} N(0, \Lambda),$$

where the $2K \times 2K$ matrix $\Lambda$ depends on $\Sigma_1$, $\Sigma_2$ and $H$, and is specified in (3.14).

**Remark 2.1** The covariance matrix $\Lambda$ has a very complicated form in Theorem 2.1 under the general conditions. If the matrix $H$ has a simple Jordan decomposition, the explicit form of $\Lambda$ can be obtained. Based on this asymptotic covariance matrix, we can determine the required sample size of a clinical trial as Hu (2002).

**Remark 2.2** Assume $\lambda < 1/2$ and $\gamma > 1/2$. From the proof of Theorem 2.1, $\Lambda_{22}$ (the asymptotic covariance matrix of $n^{-1/2}N_n$) is a limit of

$$n^{-1} \sum_{m=1}^{n} B_{n,m} \Sigma_1 B_{n,m} + n^{-1} \sum_{m=1}^{n-1} \left\{ \left( \sum_{j=m}^{n-1} \frac{1}{j+1} B_{j,m} \right) \Sigma_2 \left( \sum_{j=m}^{n-1} \frac{1}{j+1} B_{j,m} \right) \right\}$$

$$= n^{-1}(I - v'1) \left[ \sum_{m=1}^{n} B_{n,m}^{*} \Sigma_1 B_{n,m} \right] (I - 1'v)$$

$$+ n^{-1}(I - v'1) \sum_{m=1}^{n-1} \left\{ \left( \sum_{j=m}^{n-1} \frac{1}{j+1} B_{j,m}^{*} \right) \Sigma_2 \left( \sum_{j=m}^{n-1} \frac{1}{j+1} B_{j,m} \right) \right\} (I - 1'v),$$

where $B_{n,i} = \prod_{j=i+1}^{n}(I + j^{-1}H)$ and $\bar{B}_{n,i} = \prod_{j=i+1}^{n}(I + j^{-1}H)$.

We may estimate $\Lambda_{22}$ based on the following procedure:

(i) Let $n_a = \sum_{i=2}^{n} \sum_{m=0}^{i-2} M_{j_i - m - 1, \xi_i - m - 1} (i - m - 1, m)$ represent the total number of responses before the $n$-th stage. Estimate $H$ by

$$\hat{H} = n^{-1}_a \sum_{i=2}^{n} \sum_{m=0}^{i-2} M_{j_i - m - 1, \xi_i - m - 1} (i - m - 1, m) diag(X_{i-m-1}) D(\xi_{i-m-1}),$$
where $M_{i-m-1, i-m-1}(i-m-1, m), X_{i-m-1},$ and $D_{i-m-1}$ are observed during the trial.

(ii) Let $W_i$ be the number of balls added to the urn after each response, which are observed during the trial. Estimate $\Sigma_1$ and $\Sigma_2$ by

$$\hat{\Sigma}_1 = \text{diag}(Y_n/|Y_n|) - Y_n'Y_n/|Y_n|^2$$

and

$$\hat{\Sigma}_2 = n^{-1} \sum_{i=1}^{n}(W_i - \bar{W})' \text{diag}(Y_n/|Y_n|)(W_i - \bar{W}),$$

respectively, where $\bar{W} = \frac{1}{n} \sum_{i=1}^{n} W_i$ and $|Y_n| = \sum_{j=1}^{K} Y_{nj}$ is the total number of particles in the urn after the $n$-th draw.

(iii) Define $\hat{B}_{n,i} = \prod_{j=i+1}^{n}(I + j^{-1}\bar{H})$ and estimate $\Lambda_{22}$ by

$$\hat{\Lambda}_{22} = n^{-1} \left\{ (I - (Y_n'/|Y_n|)I) \sum_{m=1}^{n-1} \hat{B}'_{n,m} \hat{\Sigma}_1 \hat{B}_{n,m}(I - Y_n'/|Y_n|) \right\}$$

$$+ (I - Y_n'/|Y_n|I) \sum_{m=1}^{n-1} \left\{ (\sum_{j=m}^{n-1} (j+1)^{-1}\hat{B}'_{j,m}) \hat{\Sigma}_2 (\sum_{j=m}^{n-1} (j+1)^{-1}\hat{B}_{j,m})(I - Y_n'/|Y_n|) \right\}.$$
3 Proofs

In the section, $C_0$, $C$ etc denote positive constants whose values may differ from line to line. For a vector $x$ in $\mathcal{R}^K$, we let $\|x\|$ be its Euclidean norm, and define the norm of an $K \times K$ matrix $M$ by $\|M\| = \sup\{\|xM\|/\|x\| : x \neq 0\}$. It is obvious that for any vector $x$ and matrices $M, M_1$,

$$\|xM\| \leq \|x\| \cdot \|M\|, \quad \|M_1M\| \leq \|M_1\| \cdot \|M\|.$$  

To prove the main theorems, we show the following two lemmas first.

**Lemma 3.1** If Assumption 2.1 is true, then

$$R_n = o(n^{1-\gamma}) \text{ in } L_1 \quad \text{and} \quad R_n = o(n^{1-\gamma'}) \text{ a.s. } \forall \gamma' < \gamma.$$  

**Proof** Obviously, for some constant $C_0$,

$$\max_{i \leq n} \|R_i\| \leq \max_l \|D(l)\| \sum_{j,l} \sum_{m=1}^{i} \sum_{k=i-m+1}^{\infty} M_{jl}(m,k) \leq C_0 \sum_{j,l} \sum_{m=1}^{n} \sum_{k=n-m+1}^{\infty} M_{jl}(m,k).$$

So,

$$E \max_{i \leq n} \|R_i\| \leq C \sum_{j,l} \sum_{m=1}^{n} \sum_{k=n-m+1}^{\infty} \mu_{jlk}(m) = \sum_{m=1}^{n} o((n-m)^{-\gamma}) = o(n^{1-\gamma}).$$

It also follows that

$$P\left\{ \max_{2^i \leq n \leq 2^{i+1}} \frac{\|R_n\|}{n^{1-\gamma}} \geq \epsilon \right\} \leq C \frac{E \max_{n \leq 2^{i+1}} \|R_n\|}{2^{i(1-\gamma)}} = o(2^{i(\gamma'-\gamma)}),$$

which is summable. This completes the proof of Lemma 3.1 by the Borel-Cantelli lemma.

Note that $|Y_n| = n + Y_0 1' + R_n 1'$. From Lemma 3.1, we get the following corollary, which is Lemma 1 of Bai, Hu and Rosenberger (2002).

**Corollary 3.1** If Assumption 2.1 is true,

$$n^{-1}|Y_n| = 1 + o(n^{-\gamma}) \text{ in } L_1 \quad \text{and} \quad n^{-1}|Y_n| = 1 + o(n^{-\gamma'}) \text{ a.s. } \forall \gamma' < \gamma.$$
Lemma 3.2 Let $\mathcal{B}_{n,i} = \prod_{j=i+1}^{n}(I + j^{-1}\mathcal{H})$. Suppose matrices $Q_n$ and $P_n$ satisfy $Q_0 = P_0 = 0$ and

$$Q_n = P_n + \sum_{k=0}^{n-1} \frac{Q_k}{k+1} \mathcal{H}. \tag{3.1}$$

Then

$$Q_n = \sum_{m=1}^{n} \Delta P_m \mathcal{B}_{n,m} = P_n + \sum_{m=1}^{n-1} P_m \frac{\mathcal{H}}{m+1} \mathcal{B}_{n,m+1}, \tag{3.2}$$

where $\Delta P_m = P_m - P_{m-1}$. Also,

$$\|\mathcal{B}_{n,m}\| \leq C(n/m)^{\lambda} \log^{n-1}(n/m) \text{ for all } m = 1, \ldots, n, n \geq 1. \tag{3.3}$$

Proof: By (3.1),

$$Q_n = P_n - P_{n-1} + \frac{Q_n}{n} + P_{n-1} + \sum_{k=0}^{n-2} \frac{Q_k}{k+1} \mathcal{H}$$

$$= \Delta P_n + Q_{n-1}(I + n^{-1}\mathcal{H})$$

$$= \Delta P_n + \Delta P_{n-1}(I + n^{-1}\mathcal{H}) + Q_{n-2}(I + n^{-1}\mathcal{H})(I + (n-1)^{-1}\mathcal{H})$$

$$= \ldots = \sum_{m=1}^{n} \Delta P_m \mathcal{B}_{n,m}$$

$$= \sum_{m=1}^{n} P_m \mathcal{B}_{n,m} - \sum_{m=1}^{n-1} P_m \mathcal{B}_{n,m+1} = P_n + \sum_{m=1}^{n-1} P_m(\mathcal{B}_{n,m} - \mathcal{B}_{n,m+1})$$

$$= P_n + \sum_{m=1}^{n-1} P_m \frac{\mathcal{H}}{m+1} \mathcal{B}_{n,m+1}.$$

(3.2) is proved. For (3.3), first notice that

$$T^{-1} \mathcal{H} T = diag[0, J_2, \ldots, J_s],$$

$$T^{-1} \mathcal{B}_{n,m} T = \prod_{j=m+1}^{n} (I + j^{-1} diag[0, J_2, \ldots, J_s])$$

$$= diag[1, \prod_{j=m+1}^{n} (I + j^{-1} J_2), \ldots, \prod_{j=m+1}^{n} (I + j^{-1} J_s)]. \tag{3.4}$$

And also, by recalling (2.6) of Bai and Hu (1998), as $n > j \to \infty$, the $(h, h+i)$-th element of the block matrix $\prod_{j=m+1}^{n}(I + j^{-1} J_t)$ has the estimation

$$\frac{1}{i} \left( \frac{n}{m} \right)^{Re(\lambda_i)} \log^{i} \left( \frac{n}{m} \right) \left( 1 + o(1) \right).$$

11
It follows that
\[
\left\| \prod_{j=m+1}^{n} (I + j^{-1}J_{t}) \right\| \leq C(n/m)^{\Re(\lambda_{t})} \log^{\nu_{t}}(n/m).
\] (3.5)
Combining (3.4) and (3.5) yields (3.3).

**Proof of Theorem 2.1.** Let \( \mathcal{F}_{n} = \sigma(Y_{0}, \ldots, Y_{n}, \xi_{1}, \ldots, \xi_{n}) \) be the sigma algebra generated by the urn compositions \( \{Y_{0}, \ldots, Y_{n}\} \) and the responses \( \{\xi_{1}, \ldots, \xi_{n}\} \). Denote \( E_{n-1}\{\} = E\{\mathcal{F}_{n-1}\} \), \( q_{n} = X_{n} - E_{n-1}\{X_{n}\} \), \( Q_{n} = X_{n}D(\xi_{n}) - E_{n-1}\{X_{n}D(\xi_{n})\} \). Then \( \{Q_{n}, q_{n}, \mathcal{F}; n \geq 1\} \) is a sequence of \( \mathcal{R}^{2K} \)-valued martingale differences, and \( \xi_{n} \) is independent of \( \mathcal{F}_{n-1} \). Recall that \( |Y_{n}| = \sum_{j=1}^{K} Y_{nj} \) is the total number of particles in the urn after the \( n \)-th draw. Since the probability that a type \( k \) particle in the urn is drawn equals the number of the type \( k \) particles divided by the total number of particles, i.e., \( P(X_{nk} = 1) = \frac{Y_{n-1,k}}{|Y_{n-1}|}, k = 1, \ldots, K \), it is easily seen that \( E_{n-1}\{X_{n}\} = \frac{Y_{n-1}}{|Y_{n-1}|} \) and \( E_{n-1}\{X_{n}D(\xi_{n})\} = (\frac{Y_{n-1}}{|Y_{n-1}|})H \). Notice that \( vH = v \) and \( (\frac{Y_{n}}{|Y_{n}|} - v)1' = 1 - 1 = 0 \). From (2.5), it follows that
\[
Y_{n} - n\nu = \sum_{m=1}^{n} Q_{m} + \sum_{m=1}^{n-1} \frac{Y_{m-1}}{|Y_{m-1}|}H + Y_{0} + R_{n} - n\nu H
\]
\[
= \sum_{m=1}^{n} Q_{m} + \sum_{m=0}^{n-1} \left( \frac{Y_{m}}{|Y_{m}|} - \nu \right)H + Y_{0} + R_{n}
\]
\[
= \sum_{m=1}^{n} Q_{m} + \sum_{m=0}^{n-1} \left( \frac{Y_{m}}{|Y_{m}|} - \nu \right)H + Y_{0} + R_{n}
\]
\[
= \sum_{m=1}^{n} Q_{m} + \sum_{m=0}^{n-1} \frac{Y_{m} - m\nu}{m+1}H + R_{n}^{(1)},
\] (3.6)
where
\[
R_{n}^{(1)} = R_{n} + Y_{0} + \sum_{m=0}^{n-1} \left( \frac{Y_{m}}{|Y_{m}|} - \nu \right)(1 - \frac{|Y_{m}|}{m+1})H - \sum_{m=1}^{n} \frac{1}{m}\nu.
\] (3.7)
Also from (1.2), it follows that
\[
N_{n} = \sum_{m=1}^{n} [X_{m} - E_{m-1}(X_{m})] + \sum_{m=0}^{n-1} \frac{Y_{m}}{|Y_{m}|}
\]
\[
= \sum_{m=1}^{n} q_{m} + \sum_{m=0}^{n-1} \frac{Y_{m}}{m+1} + \sum_{m=0}^{n-1} \frac{Y_{m}}{|Y_{m}|}(1 - \frac{|Y_{m}|}{m+1}).
\] (3.8)
By (3.7), Lemma 3.1 and Corollary 3.1,
\[
R_{n}^{(1)} = o(n^{1-\gamma}) \text{ in } L_{1} \quad \text{and} \quad R_{n}^{(1)} = o(n^{1-\gamma'}) \text{ a.s. } \forall \gamma' < \gamma.
\] (3.9)
On the other hand, for the martingale difference sequence \( \{ Q_n \} \), we have

\[
E\{ \| Q_n \|^4 \} \leq 4E\| D(\xi_n) \|^4 \leq C_0
\]  

(3.10)

and

\[
E_{n-1}\{ Q_n Q_n' \} = E_{n-1}\{ \| Q_n \|^2 \} \leq 2E\| D(\xi_n) \|^2 \leq C_1.
\]

So,

\[
\sum_{m=1}^{n} E_{n-1}\{ Q_m Q_m' \} \leq C_1 n.
\]  

(3.11)

It follows that

\[
E\| \sum_{m=1}^{n} Q_m \|^2 = E[\sum_{m=1}^{n} E_{n-1}\{ Q_m Q_m' \}] \leq C_1 n.
\]

Hence

\[
\sum_{m=1}^{n} Q_m = O(n^{1/2}) \text{ in } L_2.
\]

Also, by (3.11) and the law of the iterated logarithm for martingales, we have

\[
\sum_{m=1}^{n} Q_m = O((n \log \log n)^{1/2}) \text{ a.s.}
\]

With the above two estimations for \( \sum_{m=1}^{n} Q_m \), by (3.6) and Lemma 3.2 we conclude that

\[
|Y_n - n\nu| = \left\| \sum_{i=1}^{n} Q_i + R_n^{(1)} \right\| \right. + \left. \sum_{i=1}^{n-1} \sum_{m=1}^{m} \left( \| Q_i + R_m^{(1)} \| \right) \frac{\| H \|}{m+1} \right. \left. \| B_{n,m+1} \| \right.
\leq C \sum_{m=1}^{n} \left( \| \sum_{i=1}^{m} Q_i + R_m^{(1)} \| \right) \frac{1}{m+1} \left( \frac{n}{m} \right)^{\lambda \log \nu - 1} \left( \frac{n}{m} \right)
\leq O_L(1) \sum_{m=1}^{n} (m^{1/2} + o(m^{1-\gamma})) \frac{1}{m+1} \left( \frac{n}{m} \right)^{\lambda \log \nu - 1} \left( \frac{n}{m} \right)
= O(n^{1/2})(1 + n^{\lambda - 1/2} \log \nu n + o(n^{1/2-\gamma})) = O(n^{1/2})
\]

and

\[
|Y_n - n\nu| = O((n \log \log n)^{1/2})(1 + n^{\lambda - 1/2} \log \nu n + o(n^{1/2-\gamma})),
\]

(3.12)

\[
= O((n \log \log n)^{1/2}) \text{ a.s.,}
\]
where $1/2 < \gamma' < \gamma$. It follows that
\[
\frac{Y_n}{|Y_n|} - v = O(n^{-1/2}) \text{ in } L_1 \quad \text{and} \quad \frac{Y_n}{|Y_n|} - v = O\left(\frac{\log \log n}{n}\right)^{1/2} \text{ a.s.}
\]

Now, recalling $\Sigma_1 = \text{diag}(v) - v'v$ and $\Sigma_2 = E[(D(\xi_n) - H)\text{diag}(v)(D(\xi_n) - H)]$, we have
\[
E_{n-1}[Q'_n Q_n] = \sum_{l=1}^{L} D(l)'\text{diag}\left(\frac{Y_{n-1}'}{|Y_{n-1}|}\right)D(l)P(\xi_n = l) - H'\left[\frac{Y_{n-1}'}{|Y_{n-1}|}\right]\frac{Y_{n-1}'}{|Y_{n-1}|}H
= E[(D(\xi_n)'\text{diag}(v)D(\xi_n)) - H'v'H + O(n^{-1/2})] \text{ in } L_1
= \Sigma_2 + H'\Sigma_1 H + O(n^{-1/2}),
\]
\[
E_{n-1}[Q'_n q_n] = \text{diag}\left(\frac{Y_{n-1}}{|Y_{n-1}|}\right) - \left[\frac{Y_{n-1}'}{|Y_{n-1}|}\right]\frac{Y_{n-1}}{|Y_{n-1}|} = \Sigma_1 + O(n^{-1/2}) \text{ in } L_1
\]
and
\[
E_{n-1}[q'_n Q_n] = \text{diag}\left(\frac{Y_{n-1}}{|Y_{n-1}|}\right)H - \frac{Y_{n-1}'}{|Y_{n-1}|}\frac{Y_{n-1}}{|Y_{n-1}|}H = \Sigma_1 H + O(n^{-1/2}) \text{ in } L_1.
\]
Notice (3.10) and $\|q_n\| \leq 2$. By using Theorem 8 of Monrad and Philipp (1991), one can find two independent sequences $\{Z^{(1)}_n\}$ and $\{Z^{(2)}_n\}$ of i.i.d $d$-dimensional standard normal random variables such that
\begin{equation}
\left\{
\begin{aligned}
\sum_{m=1}^{n} Q_m &= \sum_{m=1}^{n} (Z^{(2)}_m \Sigma_2^{1/2} + Z^{(1)}_m \Sigma_1^{1/2} H) + o(n^{1/2-\tau}) \quad \text{a.s.} \\
\sum_{m=1}^{n} q_m &= \sum_{m=1}^{n} Z^{(1)}_m \Sigma_1^{1/2} + o(n^{1/2-\tau}) \quad \text{a.s.}
\end{aligned}
\right.
\end{equation}

Here $\tau > 0$ depends only on $d$. Without loss of generality, we assume $\tau < (\gamma - 1/2) \wedge (1/2 - \lambda)$.

If we define $G^{(i)}_n$ by $G^{(i)}_0 = 0$ and
\[
G^{(i)}_n = \sum_{m=1}^{n} Z^{(i)}_m \Sigma_i^{1/2} + \sum_{m=1}^{n-1} \frac{G^{(i)}_m}{m+1} H,
\]
i = 1, 2, then by (3.6), (3.7), (3.9) and (3.13),
\[
Y_n - n v - (G^{(2)}_n + G^{(1)}_n H) = \sum_{m=1}^{n-1} \frac{Y_m - m v - (G^{(2)}_m + G^{(1)}_m H)}{m+1} H + o(n^{1/2-\tau}) \quad \text{a.s.}
\]
By Lemma 3.2 again,
\[
Y_n - n v - (G^{(2)}_n + G^{(1)}_n H) = \sum_{m=1}^{n} o(m^{1/2-\tau}) \frac{1}{m+1} \left(\frac{n}{m}\right)^{\lambda} \log^{\nu-1}\left(\frac{n}{m}\right) = o(n^{1/2-\tau}) \text{ a.s.}
\]
And then by (3.8) and Corollary 3.1,

\[ N_n - n \nu = \sum_{m=1}^{n} q_m + \sum_{m=1}^{n-1} \frac{Y_m - m \nu}{m + 1} + o(n^{1-\gamma}) \quad a.s. \]

\[ = \sum_{m=1}^{n} Z_{1}^{(1)} \Sigma_{1}^{1/2} + \sum_{m=1}^{n-1} \frac{G_{m}^{(1)}}{m + 1} H + \sum_{m=1}^{n-1} \frac{G_{m}^{(2)}}{m + 1} + o(n^{1-\gamma}) \quad a.s. \]

\[ = \sum_{m=1}^{n} Z_{1}^{(1)} \Sigma_{1}^{1/2} + \sum_{m=1}^{n-1} \frac{G_{m}^{(1)}}{m + 1} H + \sum_{m=1}^{n-1} \frac{G_{m}^{(2)}}{m + 1} + o(n^{1-\gamma}) \quad a.s. \]

\[ = G_{n}^{(1)} + \sum_{m=1}^{n-1} \frac{G_{m}^{(2)}}{m + 1} + o(n^{1-\gamma}) \quad a.s., \]

where we use the fact \( G_{n}^{(1)} 1' = 0 \), which is implied by \( \Sigma_{1} 1' = 0 \). Note that \( G_{n}^{(i)} \) and \( \sum_{m=1}^{n-1} (m + 1)^{-1} G_{m}^{(i)} \), \( i = 1, 2 \), are normal random variables. To finish the proof, it suffices to calculate their co-variances. Note that \( B_{n,m} = (n/m)^{\nu}(1 + o(1)) \) and \( \|(n/m)^{\nu}\| \leq (n/m)^{\lambda} \log^{\nu-1}(n/m) \), where \( a^{\nu} \) is defined to be \( \sum_{i=0}^{\nu} \frac{1}{i!} \nu^{i} \). By (3.2),

\[ \text{Var}(G_{n}^{(i)}) = \sum_{m=1}^{n} \text{Var}(Z_{1}^{(i)} \Sigma_{1}^{1/2} B_{n,m}) = \sum_{m=1}^{n} B_{n,m} \Sigma_{i} \Sigma_{m} \]

\[ = \sum_{m=1}^{n} (n/m)^{\nu} \Sigma_{1}(n/m)^{\nu} + o(1) \sum_{m=1}^{n} (n/m)^{2\lambda} \log^{2(\nu-1)}(m/m) \]

\[ = \sum_{m=1}^{n} (n/m)^{\nu} \Sigma_{1}(n/m)^{\nu} + o(n) \]

\[ = n \int_{0}^{1} \left( \frac{1}{y} \right)^{\nu} \Sigma_{i} \left( \frac{1}{y} \right)^{\nu} + o(n) =: n \Sigma_{1}^{(i)} + o(n). \]

Also,

\[ \text{Var} \left( \sum_{m=1}^{n-1} \frac{G_{m}^{(i)}}{m + 1} \right) = \sum_{m=1}^{n-1} \text{Var} \left( Z_{1}^{(i)} \Sigma_{1}^{1/2} \sum_{j=m}^{n-1} \frac{1}{j + 1} B_{j,m} \right) \]

\[ = \sum_{m=1}^{n-1} \left( \sum_{j=m}^{n-1} \frac{1}{j + 1} B_{j,m} \right) \Sigma_{i} \left( \sum_{j=m}^{n-1} \frac{1}{j + 1} B_{j,m} \right) \]

\[ = n \int_{0}^{1} dx \int_{x}^{1} \frac{1}{v} \left( \frac{1}{v} \right)^{\nu} dv \Sigma_{i} \int_{x}^{1} \frac{1}{u} \left( \frac{1}{u} \right)^{\nu} du + o(n) =: n \Sigma_{2}^{(i)} + o(n) \]

and

\[ \text{Cov}(G_{n}^{(i)}, \sum_{m=1}^{n-1} \frac{G_{m}^{(i)}}{m + 1}) = \sum_{m=1}^{n-1} B_{n,m} \Sigma_{i} \left( \sum_{j=m}^{n-1} \frac{1}{j + 1} B_{j,m} \right) \]

\[ = n \int_{0}^{1} \left( \frac{1}{x} \right)^{\nu} dx \Sigma_{i} \int_{x}^{1} \frac{1}{u} \left( \frac{1}{u} \right)^{\nu} du + o(n) =: n \Sigma_{3}^{(i)} + o(n). \]
Note that $\| (1/y) \overline{H} \| \leq c(1/y)^{\lambda\log^{\nu-1}(1/y)}$ for $0 < y \leq 1$, and $\lambda < 1/2$. The above integrals are well defined. So,

$$n^{1/2}(Y_n/n - v, N_n/n - v) \xrightarrow{D} N(0, \Lambda),$$

(3.14)

where $\Lambda = \begin{pmatrix} \Lambda_{11}, \Lambda_{12} \\ \Lambda_{21}, \Lambda_{22} \end{pmatrix}$, $\Lambda_{11} = \Sigma_1^{(2)} + H' \Sigma_1 \Sigma_1 + \Sigma_2^{(2)}$ and $\Lambda_{12} = \Sigma_1 \Sigma_2 + \Sigma_3^{(2)}$.

**Proof of Theorem 2.2.** By (3.12),

$$Y_n - n v = O((n \log \log n)^{1/2}) (1 + n^{\lambda-1/2} \log^{\nu} n + o(n^{1/2-\gamma'})) \quad \forall \gamma' < \gamma$$

= \begin{cases} 
   o(n^{\kappa}) \text{ a.s.}, & \text{if } \kappa > (1/2) \lor \lambda \lor (1 - \gamma), \\
   O((n \log \log n)^{1/2}) \text{ a.s.}, & \text{if } \lambda < 1/2 \land \gamma > 1/2. 
\end{cases}

Then, by (3.8) and Corollary 3.1, it follows that

$$N_n - n v = \sum_{m=1}^n q_m + \sum_{m=0}^{n-1} \frac{Y_m - m v}{m + 1} + \sum_{m=1}^n o(m^{1-\gamma'}) \quad \forall \gamma' < \gamma$$

= $O(\sqrt{n \log \log n}) + \sum_{m=0}^{n-1} \frac{o(m^{\kappa})}{m + 1} + o(n^{1-\gamma'}) \quad \forall \gamma' < \gamma$

= $o(n^{\kappa})$ a.s.

whenever $\kappa > (1/2) \lor \lambda \lor (1 - \gamma)$, and

$$N_n - n v = O(\sqrt{n \log \log n}) + \sum_{m=0}^{n-1} \frac{o(\sqrt{m \log \log m})}{m + 1} + o(n^{1-\gamma'}) \quad \forall \gamma' < \gamma$$

= $O(\sqrt{n \log \log n})$ a.s.

whenever $\lambda < 1/2$ and $\gamma > 1/2$. Theorem 2.2 is now proved.

**Remark 3.1** Write $T^*\Sigma_1 T = (\Sigma_{ghi}, g, h = 1, \ldots, s)$ and $T = (T_1, T_2, \ldots, T_s)$, where $T$ is defined in Assumption 2.2. Note that

$$T^*\Sigma_1 T = \int_0^1 \left( \frac{1}{y} \right)^{\operatorname{diag}[0, J_2^t, \ldots, J_s^t]} T^* \Sigma_1 T \left( \frac{1}{y} \right)^{\operatorname{diag}[0, J_2^t, \ldots, J_s^t]} dy$$

$$= \int_0^1 \operatorname{diag}[1, (\frac{1}{y}) J_2^t, \ldots, (\frac{1}{y}) J_s^t] T^* \Sigma_1 T \operatorname{diag}[1, (\frac{1}{y}) J_2^t, \ldots, (\frac{1}{y}) J_s^t] dy.$$
where

\[
\hat{J}_t = \begin{pmatrix}
0 & 1 & 0 & \ldots & 0 \\
0 & 0 & 1 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & 0
\end{pmatrix},
\hat{J}_t^2 = \begin{pmatrix}
0 & 0 & 1 & \ldots & 0 \\
0 & 0 & 0 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & 0
\end{pmatrix},
\ldots
\]

So for \(g, h = 2, \ldots, s\), \(\Sigma_{11i} = \mathbf{1}_i'\mathbf{1} = 0\), \(\Sigma_{1gi} = \Sigma_{g1i} = \int_0^1 (1/y)\mathbf{J}_g^* \mathbf{T}_g^* \mathbf{1}' dy = 0\), and, the \((a, b)\) element of matrix \(\Sigma_{ghi} = \int_0^1 (1/y)\mathbf{J}_g^* \mathbf{T}_g^* \mathbf{T}_h (1/y)\mathbf{J}_h dy\) is

\[
\sum_{a'=0}^{a-1} \sum_{b'=0}^{b-1} \frac{1}{a'!b'!} \int_0^1 \left( \frac{1}{y} \right) \lambda_{g}^{a'} + \lambda_{h}^{b'} \log^{a'+b'} \left( \frac{1}{y} \right) [\mathbf{T}_g^* \mathbf{1}]_a-a',b-b'.
\]

This agrees with the results of Bai and Hu (1999) and Bai, Hu and Rosenberger (2002). Similarly, one can calculate \(\Sigma_{2}^{(i)}\) and \(\Sigma_{3}^{(i)}\).

**Acknowledgments.** Professor Hu is also affiliated with Department of Statistics and Applied Probability, National University of Singapore. Part of his research was done while visiting the Department of Biometrics, Cornell University. He thanks the Department for its hospitality and support.
References


Feifang Hu
Department of Statistics
University of Virginia
Halsey Hall
Charlottesville, Virginia 22904-4135
Email: fh6e@virginia.edu

Lixin Zhang
Department of Mathematics
Zhejiang University, Xixi Campus
Zhejiang, Hangzhou 310028
P.R. China
Email: lxzhang@mail.hz.zj.cn