Bernoulli **10**(3), 2004, 447–463

Asymptotic normality of urn models for clinical trials with delayed response

FEIFANG HU[1] and LI-XIN ZHANG[2]

[1]*Department of Statistics, University of Virginia, Halsey Hall, Charlottesville VA 22904-4135, USA. E-mail: fh6e@virginia.edu*
[2]*Department of Mathematics, Zhejiang University, Xixi Campus, Zhejiang, Hangzhou 310028, P. R. China. E-mail: lxzhang@mail.hz.zj.cn*

For response-adaptive clinical trials using a generalized Friedman's urn design, we derive the asymptotic normality of the 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.

Keywords: Gaussian process; generalized Friedman's urn; staggered entry; treatment allocation; urn models

1. Introduction

Response-adaptive design involves the sequential selection of design points chosen depending on the outcomes at previously selected design points. The response-adaptive design has been extensively studied in the literature; see Rosenberger (1996), Flournoy and Rosenberger (1995) and Hu and Ivanova (2004) for details.

An important family of adaptive designs can be developed from the generalized Friedman's urn (GFU) model; see Athreya and Karlin (1968) and Rosenberger (2002). It is also called the generalized Pólya urn (GPU) model in the literature. The model is described as follows. Consider an urn containing balls of K types, representing K 'treatments' in a clinical trial. Initially the urn contains $\mathbf{Y}_0 = (Y_{01}, \ldots, Y_{0K})$ balls, where Y_{0k} denotes the number of balls of type k, $k = 1, \ldots, K$. At stage i, $i = 1, \ldots, n$, a ball is drawn from the urn and replaced. If the ball is of type k, then treatment k is assigned to the ith patient, $i = 1, \ldots, n$. We then wait to observe a random variable ξ_i, the response of the treatment by patient i. After that, an additional $D_{k,q}(i)$ balls of type q, $q = 1, \ldots, K$, are added to the urn, where $D_{k,q}(i)$ is some function of ξ_i. This procedure is repeated for n stages. After n stages, the urn composition is denoted by the row vector $\mathbf{Y}_n = (Y_{n1}, \ldots, Y_{nK})$, where Y_{nk} represents the number of balls of type k in the urn after the nth addition of balls. This relation can be written as the recursive formula

$$\mathbf{Y}_n = \mathbf{Y}_{n-1} + \mathbf{X}_n \mathbf{D}_n, \tag{1.1}$$

where $\mathbf{D}_n = \mathbf{D}(\xi_n) = (D_{k,q}(n))_{k,q=1}^{K}$ is a $K \times K$ random matrix with (k, q)th element $D_{k,q}(n)$, and $\mathbf{X}_n$ is the result of the nth draw, distributed according to the urn composition at the

1350–7265

previous stages – that is, if the nth draw is a ball of type k, then the kth component of $\mathbf{X}_n$ is 1 and the other components are 0.

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

$$\mathbf{N}_n = \sum_{k=1}^{n} \mathbf{X}_n. \tag{1.2}$$

Moreover, denote $\mathbf{H}_i = (\mathrm{E}(D_{k,q}(i)))_{k,q=1}^{K}$, $i = 1, \ldots, n$. The matrices $\mathbf{D}_i$ are called the *addition rules* and $\{\mathbf{H}_i\}$ are the *generating matrices*. A GFU model is said to be *homogeneous* if $\mathbf{H}_i = \mathbf{H}$ for all $i = 1, \ldots, n$.

Athreya and Karlin (1968) first considered the asymptotic properties of the GFU model with homogeneous generating matrices. Smythe (1996) defined the extended Pólya urn (EPU) model (a special class of the GFU model) and considered its asymptotic normality. For non-homogeneous generating matrices, Bai and Hu (1999) establish strong consistency and asymptotic normality of $\mathbf{Y}_n$ in the GFU model.

Typically, clinical trials do not result in immediate outcomes – that is, individual patient outcomes may not be immediately available prior to the randomization of the next patient. Consequently, the urn cannot 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 *et al.* (1994). Recently, Bai *et al.* (2002) established the asymptotic distribution of the urn composition $\mathbf{Y}_n$ under very general delay mechanisms. But they did not obtain the asymptotic normality of the sample fractions $\mathbf{N}_n$.

In a clinical trial, $\mathbf{N}_n$ represents the number of patients assigned to each treatment. The asymptotic distribution of $\mathbf{N}_n$ is essential to determine the required sample size of adaptive designs (Hu, 2003). In this paper, we establish the asymptotic normality of $\mathbf{N}_n$ with delayed responses. We also obtain the law of the iterated logarithm of both $\mathbf{Y}_n$ and $\mathbf{N}_n$. The technique used in this paper is an approximation of Gaussian processes, which is different from the martingale approximation in Bai *et al.* (2002). Some applications of the theorems are discussed in the remarks.

In this paper, we only consider urn models where the urn will be updated by adding a fixed total number β ($\beta > 0$) of balls when an outcome becomes available. The randomized Pólya urn (RPU) design of Durham *et al.* (1998) differs from such models in that the number of balls added into the urn after an outcome can be zero and the urn composition only changes after each success. Our techniques cannot be directly used to study the RPU design with delayed responses.

2. Main results

In many clinical trials, some outcomes may not be immediately available prior to the randomization of the next patient. Consequently, we can update the urn when outcomes become available (Wei, 1988). In this paper, we consider such trials, and assume that the delayed times of outcomes are not very long compared with the time interval between patients entering the trial. More precisely, we define an indicator function δ_{jk}, $j < k$, that takes the value 1 if the response of patient j occurs before patient k is assigned and 0 otherwise. We assume that $P(\delta_{jk} = 0) = o((k-j)^\gamma)$ as $k - j \to \infty$ for some $\gamma > 0$ (see Assumption 2.1). In clinical trials, it is sometimes important to consider censored observations. However, in this paper our main concern is with the asymptotic properties, and so we assume each response will ultimately occur even though its occurrence may take a very long time. We leave consideration of 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 *et al.* (2002). Assume a multinomial response model with responses $\xi_n = l$ if patient n gave 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 nth patient, that is, $J_n = j$ if patient n was randomized to treatment $j = 1, \ldots, K$. Then $\mathbf{X}_n = (X_{n1}, \ldots, X_{nk})$ has $X_{nJ_n} = 1$ and all other elements 0. We assume that the entry time of the nth patient is t_n, where $\{t_n - t_{n-1}\}$ are independent for all n. The response time of the nth 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 taking the value 1 if $t_n + \tau_n(j, l) \in (t_{n+m}, t_{n+m+1}]$ $m \geqslant 0$, 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, that is, the response of the nth 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,

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

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

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

and

$$\sum_{l,m} \mu_{jlm}(n) = 1 \qquad \text{for } j = 1, \ldots, K.$$

If we assume that $\{M_{jl}(n, m), n = 1, 2, \ldots, \}$ are independent and identically distributed (i.i.d.) for fixed j, l, and m, then $\mu_{jlm}(n) = \mu_{jlm}$ does not depend 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; that is, $\sum_{j=1}^{K} d_{ij}(l) = \beta$, where $\beta > 0$. Without loss of generality, we can assume $\beta = 1$. Let $\mathbf{D}(\xi_n) = (d_{ij}(\xi_n), i, j = 1, \ldots, K)$. 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 $\mathbf{X}_n\mathbf{D}(l)$. Since $M_{jl'}(n, m)=0$ for all $l'\neq l$, $\mathbf{X}_n\mathbf{D}(l)=\sum_{l'=1}^{L} M_{jl'}(n, m)\mathbf{X}_n\mathbf{D}(l')$. Consequently, the numbers of balls of each type added to the urn after the nth patient is assigned and before the $(n+1)$th patient is assigned are

$$\mathbf{W}_n=\sum_{m=0}^{n-1}\sum_{l=1}^{L} M_{J_{n-m},l}(n-m, m)\mathbf{X}_{n-m}\mathbf{D}(l)=\sum_{m=1}^{n}\sum_{l=1}^{L} M_{J_m,l}(m, n-m)\mathbf{X}_m\mathbf{D}(l).$$

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

$$\mathbf{Y}_n=\mathbf{Y}_{n-1}+\mathbf{W}_n. \tag{2.4}$$

Note (2.3). By (2.4) we have

$$\begin{aligned}
\mathbf{Y}_n-\mathbf{Y}_0&=\sum_{k=1}^{n}\mathbf{W}_k=\sum_{k=1}^{n}\sum_{m=1}^{k}\sum_{l=1}^{L} M_{J_m,l}(m, k-m)\mathbf{X}_m\mathbf{D}(l)\\
&=\sum_{m=1}^{n}\sum_{k=m}^{n}\sum_{l=1}^{L} M_{J_m,l}(m, k-m)\mathbf{X}_m\mathbf{D}(l)\\
&=\sum_{m=1}^{n}\sum_{k=0}^{n-m}\sum_{l=1}^{L} M_{J_m,l}(m, k)\mathbf{X}_m\mathbf{D}(l)\\
&=\sum_{m=1}^{n}\sum_{l=1}^{L}\sum_{k=0}^{\infty} M_{J_m,l}(m, k)\mathbf{X}_m\mathbf{D}(l)-\sum_{l=1}^{L}\sum_{m=1}^{n}\sum_{k=n-m+1}^{\infty} M_{J_m,l}(m, k)\mathbf{X}_m\mathbf{D}(l)\\
&=\sum_{m=1}^{n}\sum_{l=1}^{L} I\{\xi_m=l\}\mathbf{X}_m\mathbf{D}(l)-\sum_{l=1}^{L}\sum_{m=1}^{n}\sum_{k=n-m+1}^{\infty} M_{J_m,l}(m, k)\mathbf{X}_m\mathbf{D}(l)\\
&=\sum_{m=1}^{n}\mathbf{X}_m\mathbf{D}(\xi_m)-\sum_{l=1}^{L}\sum_{m=1}^{n}\sum_{k=n-m+1}^{\infty} M_{J_m,l}(m, k)\mathbf{X}_m\mathbf{D}(l).
\end{aligned}$$

That is,

$$\mathbf{Y}_n=\mathbf{Y}_0+\sum_{m=1}^{n}\mathbf{X}_m\mathbf{D}(\xi_m)+\mathbf{R}_n, \tag{2.5}$$

where

$$\mathbf{R}_n=-\sum_{l=1}^{L}\sum_{m=1}^{n}\sum_{k=n-m+1}^{\infty} M_{J_m,l}(m, k)\mathbf{X}_m\mathbf{D}(l). \tag{2.6}$$

If there is no delay, that is, $M_{J_m,l}(m, k)=0$ for all $k\geqslant 1$ and all m and l, then $\mathbf{R}_n=\mathbf{0}$ and (2.5) reduces to

$$\mathbf{Y}_n = \mathbf{Y}_0 + \sum_{m=1}^{n} \mathbf{X}_m \mathbf{D}(\xi_m), \tag{2.7}$$

which is just model (1.1). We will show that $\mathbf{R}_n$ is small. So, it is natural that stochastic staggered entry and delay mechanisms do not affect the limiting distribution of the urn. However, we shall see that the distance between $\mathbf{Y}_n$ in (2.5) and that in (2.7) is not $\mathbf{R}_n$, since the distributions of $\mathbf{X}_n$ (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 $\mathbf{H} = \mathrm{E}[\mathbf{D}(\xi_n)]$; that is, we only consider the homogeneous case. For the non-homogeneous case, if there exist V_{qij}, $q, i, j = 1, \ldots, K$, and $\mathbf{H}$ such that

$$\mathrm{cov}\{d_{qi}(\xi_n), d_{qj}(\xi_n)\} \to \mathrm{V}_{qij}, \qquad q, i, j = 1, \ldots, K,$$

and

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

then we have the same asymptotic properties.

To give asymptotic properties of $\mathbf{N}_n$, first we will require the following assumptions (which are the same as Assumptions 1 and 2 of Bai *et al.* (2002):

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

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

The left-hand side of the above equality is just the probability of the event that the nth patient on treatment j has response l after at least another m patients are assigned.

Assumption 2.1 states that response time intervals are not very long compared with the entry time intervals. This assumption is satisfied if the entry times t_n are such that $t_2 - t_1$, $t_3 - t_2, \ldots,$ are i.i.d. random variables with $\mathrm{E}(t_2 - t_1)^2 < \infty$, and the response times $\tau_n(j, l)$ are such that $\sup_n \mathrm{E}|\tau_n(j, l)|^{\gamma'} < \infty$ for each j, l and some $\gamma' > \gamma$ (see Bai *et al.* 2002).

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

$$\mathbf{T}^1\mathbf{H}\mathbf{T} = \mathrm{diag}[1, \mathbf{J}_2, \ldots, \mathbf{J}_s],$$

where $\mathbf{J}_s$ *is a* $\nu_t \times \nu_t$ *matrix, given by*

$$\mathbf{J}_t = \begin{pmatrix} \lambda_t & 1 & 0 & \dots & 0 \\ 0 & \lambda_t & 1 & \dots & 0 \\ 0 & 0 & \lambda_t & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & \lambda_t \end{pmatrix}.$$

Let $\mathbf{v}$ *be the normalized left eigenvector of* $\mathbf{H}$ *corresponding to its maximal eigenvalue* 1. *We may select the matrix* $\mathbf{T}$ *so that its first column is* $\mathbf{1}'$ *and the first row of* $\mathbf{T}^{-1}$ *is* $\mathbf{v}$. *Let* $\lambda = \max\{\mathrm{Re}(\lambda_2), \dots, \mathrm{Re}(\lambda_s)\}$ *and* $\nu = \max_j\{\nu_j : \mathrm{Re}(\lambda_j) = \lambda\}$.

We also denote $\overline{\mathbf{H}} = \mathbf{H} - \mathbf{1}'\mathbf{v}$, $\mathbf{\Sigma}_1 = \mathrm{diag}(\mathbf{v}) - \mathbf{v}'\mathbf{v}$ and $\mathbf{\Sigma}_2 = \mathrm{E}[(\mathbf{D}(\xi_n) - \mathbf{H})'\mathrm{diag}(\mathbf{v}) (\mathbf{D}(\xi_n) - \mathbf{H})]$. We have the following asymptotic normality for $(\mathbf{Y}_n, \mathbf{N}_n)$.

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

$$n^{1/2}(\mathbf{Y}_n/n - \mathbf{v}, \mathbf{N}_n/n - \mathbf{v}) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathbf{\Lambda}),$$

where the $2K \times 2K$ *matrix* $\mathbf{\Lambda}$ *depends on* $\mathbf{\Sigma}_1$, $\mathbf{\Sigma}_2$ *and* $\mathbf{H}$, *and is specified in* (3.15).

Remark 2.1. The condition $\gamma < 1/2$ ensures the asymptotic normality of $n^{1/2}(\mathbf{N}_n/n - \mathbf{v})$. This condition implies that the probability that a patient will give a response (delayed) after at least m patients are assigned converges to 0 with speed $m^{-\gamma}$ ($\gamma < \frac{1}{2}$). This condition is not necessary for strong consistency of $\mathbf{N}_n/n$ (Theorem 2.2). $\lambda < \frac{1}{2}$ is a standard condition in urn models (see Rosenberger, 2002).

Remark 2.2. The covariance matrix $\mathbf{\Lambda}$ has a very complicated form in Theorem 2.1 under the general conditions. If the matrix $\mathbf{H}$ has a simple Jordan decomposition, the explicit form of $\mathbf{\Lambda}$ can be obtained. Based on this asymptotic covariance matrix, we can determine the requisite sample size of a clinical trial as in Hu (2003).

Remark 2.3. Recently, Hu and Rosenberger (2003) explicitly obtained the relationship between the power of a test and the variability of a randomization procedure (asymptotic covariance of $\mathbf{N}_n$) for binary responses, under the condition of asymptotic normality of $\mathbf{N}_n$. Theorem 2.1 ensures the results of Hu and Rosenberger (2003) with delayed responses. Therefore, it is important to calculate the covariance matrix. From (3.15), the asymptotic covariance matrix $\mathbf{\Lambda}_{22} = \mathbf{\Sigma}_1^{(1)} + \mathbf{\Sigma}_2^{(2)}$ does not depend on the delay mechanism. Also $\mathbf{\Lambda}_{22}$ is usually difficult to calculate. In practice, we need to estimate $\mathbf{\Lambda}_{22}$ based on the delay mechanism $M_{jl}(n, m)$. The procedure is given in Remark 2.4.

Remark 2.4. From the proof of Theorem 2.1, $\mathbf{\Lambda}_{22}$, the asymptotic covariance matrix of $n^{1/2}\mathbf{N}_n$, is a limit of

$$n^{-1}\sum_{m=1}^{n}\overline{\mathbf{B}}'_{n,m}\boldsymbol{\Sigma}_1\overline{\mathbf{B}}_{n,m}+n^{-1}\sum_{m=1}^{n-1}\left\{\left(\sum_{j=m}^{n-1}\frac{1}{j+1}\overline{\mathbf{B}}_{j,m}\right)'\boldsymbol{\Sigma}_2\left(\sum_{j=m}^{n-1}\frac{1}{j+1}\overline{\mathbf{B}}_{j,m}\right)\right\}$$

$$=n^1(\mathbf{I}-\mathbf{v}'\mathbf{1})\left[\sum_{m=1}^{n}\mathbf{B}'_{n,m}\boldsymbol{\Sigma}_1\mathbf{B}_{n,m}\right](\mathbf{I}-\mathbf{1}'\mathbf{v})$$

$$+n^1(\mathbf{I}-\mathbf{v}'\mathbf{1})\sum_{m=1}^{n-1}\left\{\left(\sum_{j=m}^{n-1}\frac{1}{j+1}\mathbf{B}'_{j,m}\right)\boldsymbol{\Sigma}_2\left(\sum_{j=m}^{n-1}\frac{1}{j+1}\mathbf{B}_{j,m}\right)\right\}(\mathbf{I}-\mathbf{1}'\mathbf{v}),$$

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

We may estimate $\boldsymbol{\Lambda}_{22}$ 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 nth stage. Estimate $\mathbf{H}$ by

$$\hat{\mathbf{H}}=n_{\mathrm{a}}^{-1}\sum_{i=2}^{n}\sum_{m=0}^{i-2}M_{J_{i-m-1},\xi_{i-m-1}}(i-m-1,\ m)\mathrm{diag}(\mathbf{X}_{i-m-1})\mathbf{D}(\xi_{i-m-1}),$$

where $M_{J_{i-m-1},\xi_{i-m-l}}(i-m-1,\ m)$, $\mathbf{X}_{i-m-1}$ and $\mathbf{D}(\xi_{i-m-1})$ are observed during the trial.

(ii) Let $\mathbf{W}_i$ be the number of balls added to the urn after each response, as observed during the trial. Estimate $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ by

$$\hat{\boldsymbol{\Sigma}}_1=\mathrm{diag}(\mathbf{Y}_n/|\mathbf{Y}_n|)-\mathbf{Y}'_n\mathbf{Y}_n/|\mathbf{Y}_n|^2$$

and

$$\hat{\boldsymbol{\Sigma}}_2=n_a^{-1}\sum_{i=1}^{n}(\mathbf{W}_i-\overline{\mathbf{W}})'\,\mathrm{diag}(\mathbf{Y}_n/|\mathbf{Y}_n|)(\mathbf{W}_i-\overline{\mathbf{W}}),$$

respectively, where $\overline{\mathbf{W}}=n_a^{-1}\sum_{i=1}^{n_a}\mathbf{W}_i$ and $|\mathbf{Y}_n|=\sum_{j=1}^{K}Y_{nj}$ is the total number of balls in the urn after the nth draw.

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

$$\hat{\boldsymbol{\Lambda}}_{22}=n^1\left\{(\mathbf{I}-(\mathbf{Y}'_n/|\mathbf{Y}_n|)\mathbf{1})\left[\sum_{m=1}^{n}\overline{\mathbf{B}}'_{n,m}\hat{\boldsymbol{\Sigma}}_1\overline{\mathbf{B}}_{n,m}\right](\mathbf{I}-\mathbf{1}'\mathbf{Y}_n/|\mathbf{Y}_n|)\right.$$

$$\left.+(\mathbf{I}-\mathbf{Y}'_n/|\mathbf{Y}_n|)\mathbf{1})\sum_{m=1}^{n-1}\left\{\left[\sum_{j=m}^{n-1}(j+1)^{-1}\overline{\mathbf{B}}'_{j,m}]\hat{\boldsymbol{\Sigma}}_2\left[\sum_{j=m}^{n-1}(j+1)^{-1}\hat{\mathbf{b}}_{j,m}\right](\mathbf{I}-\mathbf{1}'\mathbf{Y}_n/|\mathbf{Y}_n|)\right\}\right\}.$$

Based on $\hat{\boldsymbol{\Lambda}}_{22}$, we can assess the variation of designs. Bai *et al.* (2002) provide an estimator of $\boldsymbol{\Lambda}_{11}$ (the asymptotic covariance matrix of $n^{1/2}\mathbf{Y}_n/|\mathbf{Y}_n|$).

Theorem 2.2. *Under Assumptions 2.1 and 2.2, if $\lambda < 1$, then for any $\kappa > (\frac{1}{2}) \vee \lambda \vee (1-\gamma)$, almost surely,*

$$n^{-\kappa}(\mathbf{Y}_n - n\mathbf{v}) \to \mathbf{0} \quad and \quad n^{-\kappa}(\mathbf{N}_n - n\mathbf{v}) \to \mathbf{0}.$$

Further, if $\gamma < \frac{1}{2}$ and $\lambda < \frac{1}{2}$, then, almost surely,

$$\mathbf{Y}_n - n\mathbf{v} = O(\sqrt{n \log\log n}) \quad and \quad \mathbf{N}_n - n\mathbf{v} = O(\sqrt{n \log\log n}).$$

Remark 2.5. Bai *et al.* (2002) establish the strong consistency of both $\mathbf{Y}_n$ and $\mathbf{N}_n$. Here we obtain the rate of strong consistency as well as the law of the iterated logarithm for both $\mathbf{Y}_n$ and $\mathbf{N}_n$. The result is new and can also be applied in cases without delayed responses.

3. Proofs

In this section, C_0, C, etc. denote positive constants whose values may differ from line to line. For a vector $\mathbf{x}$ in $\mathbb{R}^K$, we let $\|\mathbf{x}\|$ be its Euclidean norm, and define the norm of a $K \times K$ matrix $\mathbf{M}$ by $\|\mathbf{M}\| = \sup\{\|\mathbf{xM}\|/\|\mathbf{x}\| : \mathbf{x} \neq \mathbf{0}\}$. It is obvious that, for any vector $\mathbf{x}$ and matrices $\mathbf{M}$, $\mathbf{M}_1$,

$$\|\mathbf{xM}\| \leqslant \|\mathbf{x}\| \cdot \|\mathbf{M}\|, \qquad \|\mathbf{M}_1\mathbf{M}\| \leqslant \|\mathbf{M}_1\| \cdot \|\mathbf{M}\|.$$

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

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

$$\mathbf{R}_n = o(n^{1-\gamma}) \quad in \quad L_1 \quad and \quad \mathbf{R}_n = o(n^{1-\gamma'}) \ a.s. \ \forall \gamma' < \gamma. \tag{3.1}$$

Proof. For some constant C,

$$\max_{i \leqslant n} \|\mathbf{R}_i\| \leqslant \max_l \|\mathbf{D}(l)\| \{\max_{i \leqslant n}$$

So,

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

It also follows that

$$P\left\{ \max_{2^i \leqslant n \leqslant 2^{i+1}} \frac{\|\mathbf{R}_n\|}{n^{1-\gamma'}} \geqslant \epsilon \right\} \leqslant C \frac{\mathrm{E} \max_{n \leqslant 2^{i+l}} \|\mathbf{R}_n\|}{2^{i(1-\gamma')}} = o(2^{i(\gamma'-\gamma)}),$$

which is summable, whence (3.1) follows by the Borel–Cantelli lemma. □

Note that $|\mathbf{Y}_n| = n + \mathbf{Y}_0\mathbf{1}' + \mathbf{R}_n\mathbf{1}'$. From Lemma 3.1, we obtain the following corollary, which is Lemma 1 of Bai *et al.* (2002).

Corollary 3.1. *If Assumption 2.1 is true,*

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

Lemma 3.2. *Let* $\overline{\mathbf{B}}_{n,i} = \prod_{j=i+1}^{n}(\mathbf{I} + j^{-1}\overline{\mathbf{H}})$. *Suppose matrices* $\mathbf{Q}_n$ and $\mathbf{P}_n$ *satisfy* $\mathbf{Q}_0 = \mathbf{P}_0 = \mathbf{0}$ *and*

$$\mathbf{Q}_n = \mathbf{P}_n + \sum_{k=1}^{n-1}\frac{\mathbf{Q}_k}{k+1}\overline{\mathbf{H}}. \tag{3.2}$$

Then

$$\mathbf{Q}_n = \sum_{m=1}^{n}\Delta\mathbf{P}_m\overline{\mathbf{B}}_{n,m} = \mathbf{P}_n + \sum_{m=1}^{n-1}\mathbf{P}_m\frac{\overline{\mathbf{H}}}{m+1}\overline{\mathbf{B}}_{n,m+1}, \tag{3.3}$$

where $\Delta\mathbf{P}_m = \mathbf{P}_m - \mathbf{P}_{m-1}$. *Also, for some constant* C,

$$\|\overline{\mathbf{B}}_{n,m}\| \leqslant C(n/m)^{\lambda}\log^{\nu-1}(n/m) \qquad \text{for all } m = 1, \ldots, n,\ n \geqslant 1. \tag{3.4}$$

Proof. By (3.2),

$$\begin{aligned}
\mathbf{Q}_n &= \mathbf{P}_n - \mathbf{P}_{n-1} + \frac{\mathbf{Q}_{n-1}}{n} + \mathbf{P}_{n-1} + \sum_{k=0}^{n-2}\frac{\mathbf{Q}_k}{k+1}\overline{\mathbf{H}} \\
&= \Delta\mathbf{P}_n + \mathbf{Q}_{n-1}(\mathbf{I} + n^{1}\overline{\mathbf{H}}) \\
&= \Delta\mathbf{P}_n + \Delta\mathbf{P}_{n-1}(\mathbf{I} + n^{-1}\overline{\mathbf{H}}) + \mathbf{Q}_{n-2}(\mathbf{I} + n^{-1}\overline{\mathbf{H}})(\mathbf{I} + (n-1)^{-1}\overline{\mathbf{H}}) \\
&= \ldots = \sum_{m=1}^{n}\Delta\mathbf{P}_m\overline{\mathbf{B}}_{n,m} \\
&= \sum_{m=1}^{n}\mathbf{P}_m\overline{\mathbf{B}}_{n,m} - \sum_{m=1}^{n-1}\mathbf{P}_m\overline{\mathbf{B}}_{n,m+1} = \mathbf{P}_n + \sum_{m=1}^{n-1}\mathbf{P}_m(\overline{\mathbf{B}}_{n,m} - \overline{\mathbf{B}}_{n,m+1}) \\
&= \mathbf{P}_n + \sum_{m=1}^{n-1}\mathbf{P}_m\frac{\overline{\mathbf{H}}}{m+1}\overline{\mathbf{B}}_{n,m+1},
\end{aligned}$$

and (3.3) is proved. For (3.4), first notice that

$$\mathbf{T}^{-1}\overline{\mathbf{H}}\mathbf{T} = \text{diag}[0, \mathbf{J}_2, \ldots, \mathbf{J}_s],$$

$$\mathbf{T}^{-1}\overline{\mathbf{B}}_{n,m}\mathbf{T} = \prod_{j=m+1}^{n} (\mathbf{I} + j^{-1}\,\mathrm{diag}[0, \mathbf{J}_2, \ldots, \mathbf{J}_s])$$

$$= \mathrm{diag}\left[1, \prod_{j=m+1}^{n} (\mathbf{I} + j^{-1}\mathbf{J}_2), \ldots, \prod_{j=m+1}^{n} (\mathbf{I} + j^{-1}\mathbf{J}_s)\right]. \tag{3.5}$$

Also, recalling (2.6) of Bai and Hu (1999), as $n > j \to \infty$, the $(h, h+i)$th element of the block matrix $\prod_{j=m+1}^{n}(\mathbf{I} + j^{-1}\mathbf{J}_t)$ is

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

It follows that

$$\left\| \prod_{j=m+1}^{n} (\mathbf{I} + j^{-1}\mathbf{J}_t) \right\| \leqslant C\left(\frac{n}{m}\right)^{\mathrm{Re}(\lambda_T)} \log^{\nu_t - 1}\left(\frac{n}{m}\right). \tag{3.6}$$

Combining (3.5) and (3.6) yields (3.4).

Proof of Theorem 2.1. Let $\mathcal{F}_n = \sigma(\mathbf{Y}_0, \ldots, \mathbf{Y}_n, \xi_1, \ldots, \xi_n)$ be the sigma algebra generated by the urn compositions $\{\mathbf{Y}_0, \ldots, \mathbf{Y}_n\}$ and the responses $\{\xi_1, \ldots, \xi_n\}$. Denote $\mathrm{E}_{n-1}\{\cdot\} = \mathrm{E}\{\cdot|\mathcal{F}_{n-1}\}$, $\mathbf{q}_n = \mathbf{X}_n - \mathrm{E}_{n-1}\{\mathbf{X}_n\}$, $\mathbf{Q}_n = \mathbf{X}_n\mathbf{D}(\xi_n) - \mathrm{E}_{n-1}\{\mathbf{X}_n\mathbf{D}(\xi_n)\}$. Then $\{(\mathbf{Q}_n, \mathbf{q}_n), \mathcal{F};\ n \geqslant 1\}$ is a sequence of $\mathbb{R}^{2K}$-valued martingale differences, and ξ_n is independent of $\mathcal{F}_{n-1}$. Recall that $|\mathbf{Y}_n| = \sum_{j=1}^{K} Y_{nj}$ is the total number of balls in the urn after the nth draw. Since the probability that a ball of type k is drawn from the urn equals the number of balls of type k divided by the total number of balls, that is, $P(X_{nk} = 1) = Y_{n-1,k}/|\mathbf{Y}_{n-1}|$, $k = 1, \ldots, K$, it is easily seen that $\mathrm{E}_{n-1}\{\mathbf{X}_n\} = \mathbf{Y}_{n-1}/|\mathbf{Y}_{n-1}|$ and $\mathrm{E}_{n-1}\{\mathbf{X}_n\mathbf{D}(\xi_n)\} = (\mathbf{Y}_{n-1}/|\mathbf{Y}_{n-1}|)\mathbf{H}$. Notice that $\mathbf{vH} = \mathbf{v}$ and $(\mathbf{Y}_n/|\mathbf{Y}_n| - \mathbf{v})\mathbf{1}' = 1 - 1 = 0$. From (2.5), it follows that

$$\mathbf{Y}_n - n\mathbf{v} = \sum_{m=1}^{n} \mathbf{Q}_m + \sum_{m=1}^{n} \frac{\mathbf{Y}_{m-1}}{|\mathbf{Y}_{m-1}|}\mathbf{H} + \mathbf{Y}_0 + \mathbf{R}_n - n\mathbf{vH}$$

$$= \sum_{m=1}^{n} \mathbf{Q}_m + \sum_{m=0}^{n-1} \left(\frac{\mathbf{Y}_m}{|\mathbf{Y}_m|} - \mathbf{v}\right)\mathbf{H} + \mathbf{Y}_0 + \mathbf{R}_n$$

$$= \sum_{m=1}^{n} \mathbf{Q}_m + \sum_{m=0}^{n-1} \left(\frac{\mathbf{Y}_m}{|\mathbf{Y}_m|} - \mathbf{v}\right)\overline{\mathbf{H}} + \mathbf{Y}_0 + \mathbf{R}_n$$

$$= \sum_{m=1}^{n} \mathbf{Q}_m + \sum_{m=0}^{n-1} \frac{\mathbf{Y}_m - m\mathbf{v}}{m+1}\overline{\mathbf{H}} + \mathbf{R}_n^{(1)}, \tag{3.7}$$

where

$$\mathbf{R}_n^{(1)} = \mathbf{R}_n + \mathbf{Y}_0 + \sum_{m=0}^{n-1}\left(\frac{\mathbf{Y}_m}{|\mathbf{Y}_m|} - \mathbf{v}\right)\left(1 - \frac{|Y_m|}{m+1}\right)\overline{\mathbf{H}} - \sum_{m=1}^{n}\frac{1}{m}\mathbf{v}. \tag{3.8}$$

Also from (1.2), it follows that

$$\begin{aligned}\mathbf{N}_n &= \sum_{m=1}^{n}[\mathbf{X}_m - \mathrm{E}_{m-1}(\mathbf{X}_m)] + \sum_{m=0}^{n-1}\frac{\mathbf{Y}_m}{|\mathbf{Y}_m|} \\ &= \sum_{m=1}^{n}\mathbf{q}_m + \sum_{m=0}^{n-1}\frac{\mathbf{Y}_m}{m+1} + \sum_{m=0}^{n-1}\frac{\mathbf{Y}_m}{|\mathbf{Y}_m|}\left(1 - \frac{|\mathbf{Y}_m|}{m+1}\right).\end{aligned} \tag{3.9}$$

By (3.8), Lemma 3.1 and Corollary 3.1,

$$\mathbf{R}_n^{(1)} = o(n^{1-\gamma}) \text{ in } L_1 \quad \text{and} \quad \mathbf{R}_n^{(1)} = o(n^{1-\gamma'}) \text{ a.s. } \forall \gamma' < \gamma. \tag{3.10}$$

On the other hand, for the martingale difference sequence $\{\mathbf{Q}_n\}$, we have, for some constants C_0 and C_1,

$$\mathrm{E}\{\|\mathbf{Q}_n\|^4\|\} \leqslant 4\mathrm{E}\|\mathbf{D}(\xi_n)\|^4 \leqslant C_0 \tag{3.11}$$

and

$$\mathrm{E}_{n-1}\{\mathbf{Q}_n\mathbf{Q}_n'\} = \mathrm{E}_{n-1}\{\|\mathbf{Q}_n\|^2\} \leqslant 2\mathrm{E}\|\mathbf{D}(\xi_n)\|^2 \leqslant C_1.$$

So

$$\sum_{m=1}^{n}\mathrm{E}_{n-1}\{\mathbf{Q}_m\mathbf{Q}_m'\} \leqslant C_1 n. \tag{3.12}$$

It follows that

$$\mathrm{E}\left\|\sum_{m=1}^{n}\mathbf{Q}_m\right\|^2 = \mathrm{E}\left[\sum_{m=1}^{n}\mathrm{E}_{n-1}\{\mathbf{Q}_m\mathbf{Q}_m'\}\right] \leqslant C_1 n.$$

Hence

$$\sum_{m=1}^{n}\mathbf{Q}_m = O(n^{1/2}) \text{ in } L_2.$$

Also, by (3.12) and the law of the iterated logarithm for martingales, we have

$$\sum_{m=1}^{n}\mathbf{Q}_m = O((n\log\log n)^{1/2}) \text{ a.s.}$$

With the above two equations for $\sum_{m=1}^{n}\mathbf{Q}_m$, by (3.7) and Lemma 3.2 we conclude that

$$
\begin{aligned}
\|\mathbf{Y}_n - n\mathbf{v}\| &= \left\| \left(\sum_{i=1}^{n} \mathbf{Q}_i + \mathbf{R}_n^{(1)} \right) + \sum_{m=1}^{n-1} \left(\sum_{i=1}^{m} \mathbf{Q}_i + \mathbf{R}_m^{(1)} \right) \frac{\overline{\mathbf{H}}}{m+1} \overline{\mathbf{B}}_{n,m+1} \right\| \\
&\leqslant \left\| \sum_{i=1}^{n} \mathbf{Q}_i + \mathbf{R}_n^{(1)} \right\| + \sum_{m=1}^{n-1} \left\| \sum_{i=1}^{m} \mathbf{Q}_i + \mathbf{R}_m^{(1)} \right\| \frac{\|\overline{\mathbf{H}}\|}{m+1} \|\overline{\mathbf{B}}_{n,m+1}\| \\
&\leqslant C \sum_{m=1}^{n} \left(\left\| \sum_{i=1}^{m} \mathbf{Q}_i + \mathbf{R}_m^{(1)} \right\| \right) \frac{1}{m+1} \left(\frac{n}{m} \right)^{\lambda} \log^{\nu-1} \left(\frac{n}{m} \right) \\
&= \mathrm{O}_{L_1}(1) \sum_{m=1}^{n} (m^{1/2} + o(m^{1-\gamma})) \frac{1}{m+1} \left(\frac{n}{m} \right)^{\lambda} \log^{\nu-1} \left(\frac{n}{m} \right) \\
&= O(n^{1/2})(1 + n^{\lambda-1/2} \log^{\nu} n + o(n^{1/2-\gamma})) = O(n^{1/2})
\end{aligned}
$$

and

$$
\begin{aligned}
\|\mathbf{Y}_n - n\mathbf{v}\| &= O((n \log\log n)^{1/2})(1 + n^{\lambda-1/2} \log^{\nu} n + o(n^{1/2-\gamma'})) \\
&= O((n \log\log n)^{1/2}) \text{ a.s.,}
\end{aligned}
\tag{3.13}
$$

where $\frac{1}{2} < \gamma' < \gamma$. It follows that

$$
\frac{\mathbf{Y}_n}{|\mathbf{Y}_n|} - \mathbf{v} = O(n^{1/2}) \text{ in } L_1 \quad \text{and} \quad \frac{\mathbf{Y}_n}{|\mathbf{Y}_n|} - \mathbf{v} = O\left(\left(\frac{\log\log n}{n} \right)^{1/2} \right) \text{ a.s.}
$$

Now, recalling $\boldsymbol{\Sigma}_1 = \operatorname{diag}(\mathbf{v}) - \mathbf{v}'\mathbf{v}$ and $\boldsymbol{\Sigma}_2 = \mathrm{E}[(\mathbf{D}(\xi_n) - \mathbf{H})' \operatorname{diag}(\mathbf{v})(\mathbf{D}(\xi_n) - \mathbf{H})]$, we have

$$
\begin{aligned}
\mathrm{E}_{n-1}[\mathbf{Q}_n'\mathbf{Q}_n] &= \sum_{l=1}^{L} \mathbf{D}(l)' \operatorname{diag}\left(\frac{\mathbf{Y}_{n-1}}{|\mathbf{Y}_{n-1}|} \right) \mathbf{D}(l) P(\xi_n = l) - \mathbf{H}' \frac{\mathbf{Y}_{n-1}'}{|\mathbf{Y}_{n-1}|} \frac{\mathbf{Y}_{n-1}}{|\mathbf{Y}_{n-1}|} \mathbf{H} \\
&= \mathrm{E}[(\mathbf{D}(\xi_n)' \operatorname{diag}(\mathbf{v}) \mathbf{D}(\xi_n)] - \mathbf{H}'\mathbf{v}'\mathbf{v}\mathbf{H} + O(n^{1/2}) \qquad \text{in } L_1 \\
&= \boldsymbol{\Sigma}_2 + \mathbf{H}'\boldsymbol{\Sigma}_1\mathbf{H} + O(n^{1/2}), \\
\mathrm{E}_{n-1}[\mathbf{q}_n'\mathbf{q}_n] &= \operatorname{diag}\left(\frac{\mathbf{Y}_{n-1}}{|\mathbf{Y}_{n-1}|} \right) - \frac{\mathbf{Y}_{n-1}'}{|\mathbf{Y}_{n-1}|} \frac{\mathbf{Y}_{n-1}}{|\mathbf{Y}_{n-1}|} = \boldsymbol{\Sigma}_1 + O(n^{1/2}) \qquad \text{in } L_1
\end{aligned}
$$

and

$$
\mathrm{E}_{n-1}[\mathbf{q}_n'\mathbf{Q}_n] = \operatorname{diag}\left(\frac{\mathbf{Y}_{n-1}}{|\mathbf{Y}_{n-1}|} \right) \mathbf{H} - \frac{\mathbf{Y}_{n-1}'}{|\mathbf{Y}_{n-1}|} \frac{\mathbf{Y}_{n-1}}{|\mathbf{Y}_{n-1}|} \mathbf{H} = \boldsymbol{\Sigma}_1 \mathbf{H} + O(n^{1/2}) \qquad \text{in } L_1.
$$

Recall (3.11) and notice that $\|\mathbf{q}_n\| \leqslant 2$. By using Theorem 8 of Monrad and Philipp (1991), one can find two independent sequences $\{\mathbf{Z}_n^{(1)}\}$ and $\{\mathbf{Z}_n^{(2)}\}$ of i.i.d. d-dimensional standard normal random variables such that

$$
\begin{aligned}
\sum_{m=1}^{n} \mathbf{Q}_m &= \sum_{m=1}^{n} (\mathbf{Z}_m^{(2)} \mathbf{\Sigma}_2^{1/2} + \mathbf{Z}_m^{(1)} \mathbf{\Sigma}_1^{1/2} \mathbf{H}) + o(n^{1/2-\tau}) \qquad \text{a.s.,} \\
\sum_{m=1}^{n} \mathbf{q}_m &= \sum_{m=1}^{n} \mathbf{Z}_m^{(1)} \mathbf{\Sigma}_1^{1/2} + o(n^{1/2-\tau}) \qquad \text{a.s.}
\end{aligned} \tag{3.14}
$$

Here $\tau > 0$ depends only on d. Without loss of generality, we assume $\tau < (\gamma - \frac{1}{2}) \wedge (\frac{1}{2} - \lambda)$. If we define $\mathbf{G}_n^{(i)}$ by $\mathbf{G}_0^{(i)} = 0$ and

$$
\mathbf{G}_n^{(i)} = \sum_{m=1}^{n} \mathbf{Z}_m^{(i)} \mathbf{\Sigma}_i^{1/2} + \sum_{m=1}^{n-1} \frac{\mathbf{G}_m^{(i)}}{m+1} \overline{\mathbf{H}},
$$

$i = 1, 2$, then by (3.7), (3.8), (3.10) and (3.14),

$$
\mathbf{Y}_n - n\mathbf{v} - (\mathbf{G}_n^{(2)} + \mathbf{G}_n^{(1)} \mathbf{H}) = \sum_{m=1}^{n-1} \frac{\mathbf{Y}_m - m\mathbf{v} - (\mathbf{G}_m^{(2)} + \mathbf{G}_m^{(1)} \mathbf{H})}{m+1} \overline{\mathbf{H}} + o(n^{1/2-\tau}) \qquad \text{a.s.}
$$

By Lemma 3.2 again,

$$
\mathbf{Y}_n - n\mathbf{v} - (\mathbf{G}_n^{(2)} + \mathbf{G}_n^{(1)} \mathbf{H}) = \sum_{m=1}^{n} o(m^{1/2-\tau}) \frac{1}{m+1} \left(\frac{n}{m}\right)^{\lambda} \log^{\nu-1}\left(\frac{n}{m}\right) = o(n^{1/2-\tau}) \qquad \text{a.s.}
$$

And then by (3.9) and Corollary 3.1,

$$
\begin{aligned}
\mathbf{N}_n - n\mathbf{v} &= \sum_{m=1}^{n} \mathbf{q}_m + \sum_{m=1}^{n-1} \frac{\mathbf{Y}_m - m\mathbf{v}}{m+1} + o(n^{1-\gamma'}) \qquad \text{a.s.} \\
&= \sum_{m=1}^{n} \mathbf{Z}_m^{(1)} \mathbf{\Sigma}_1^{1/2} + \sum_{m=1}^{n-1} \frac{\mathbf{G}_m^{(1)}}{m+1} \mathbf{H} + \sum_{m=1}^{n-1} \frac{\mathbf{G}_m^{(2)}}{m+1} + o(n^{1-\tau}) \qquad \text{a.s.} \\
&= \sum_{m=1}^{n} \mathbf{Z}_m^{(1)} \mathbf{\Sigma}_1^{1/2} + \sum_{m=1}^{n-1} \frac{\mathbf{G}_m^{(1)}}{m+1} \overline{\mathbf{H}} + \sum_{m=1}^{n-1} \frac{\mathbf{G}_m^{(2)}}{m+1} + o(n^{1-\tau}) \qquad \text{a.s.} \\
&= \mathbf{G}_n^{(1)} + \sum_{m=1}^{n-1} \frac{\mathbf{G}_m^{(2)}}{m+1} + o(n^{1-\tau}) \qquad \text{a.s.,}
\end{aligned}
$$

where we use the fact $\mathbf{G}_n^{(1)} \mathbf{1}' = \mathbf{0}$, which is implied by $\mathbf{\Sigma}_1 \mathbf{1}' = \mathbf{0}$. Note that $\mathbf{G}_n^{(i)}$ and $\sum_{m=1}^{n-1} (m+1)^1 \mathbf{G}_m^{(i)}$, $i = 1, 2$, are normal random variables. To finish the proof, it suffices to calculate their covariances. Note that $\overline{\mathbf{B}}_{n,m} = (n/m)^{\overline{\mathbf{H}}}(1 + o(1))$, where $a^{\overline{\mathbf{H}}}$ is defined to be $\sum_{i=0}^{\infty} (1/i!) \overline{\mathbf{H}}^i \log^i a$. Notice that $\|(n/m)^{\overline{\mathbf{H}}}\| \leqslant (n/m)^{\lambda} \log^{\nu-1}(n/m)$. By (3.3),

$$\operatorname{var}(\mathbf{G}_n^{(i)}) = \sum_{m=1}^{n} \operatorname{var}(\mathbf{Z}_m^{(i)} \boldsymbol{\Sigma}_i^{1/2} \overline{\mathbf{B}}_{n,m}) = \sum_{m=1}^{n} \overline{\mathbf{B}}'_{n,m} \boldsymbol{\Sigma}_i \overline{\mathbf{B}}_{n,m}$$
$$= \sum_{m=1}^{n} \left(\frac{n}{m}\right)^{\overline{\mathbf{H}}'} \boldsymbol{\Sigma}_i \left(\frac{n}{m}\right)^{\overline{\mathbf{H}}} + o(1) \sum_{m=1}^{n} \left(\frac{n}{m}\right)^{2\lambda} \log^{2(\nu-1)}\left(\frac{m}{m}\right)$$
$$= \sum_{m=1}^{n} \left(\frac{n}{m}\right)^{\overline{\mathbf{H}}'} \boldsymbol{\Sigma}_i \left(\frac{n}{m}\right)^{\overline{\mathbf{H}}} + o(n)$$
$$= n\int_0^1 \left(\frac{1}{y}\right)^{\overline{\mathbf{H}}'} \boldsymbol{\Sigma}_i \left(\frac{1}{y}\right)^{\overline{\mathbf{H}}} + o(n) =: n\boldsymbol{\Sigma}_1^{(i)} + o(n).$$

Also,

$$\operatorname{var}\left(\sum_{m=1}^{n-1} \frac{\mathbf{G}_m^{(i)}}{m+1}\right) = \sum_{m=1}^{n-1} \operatorname{var}\left(\mathbf{Z}_m^{(i)} \boldsymbol{\Sigma}_i^{1/2} \sum_{j=m}^{n-1} \frac{1}{j+1} \overline{\mathbf{B}}_{j,m}\right)$$
$$= \sum_{m=1}^{n-1} \left(\sum_{j=m}^{n-1} \frac{1}{j+1} \overline{\mathbf{B}}_{j,m}\right)' \boldsymbol{\Sigma}_i \left(\sum_{j=m}^{n-1} \frac{1}{j+1} \overline{\mathbf{B}}_{j,m}\right)$$
$$= n\int_0^1 \mathrm{d}x \int_x^1 \frac{1}{v}\left(\frac{1}{v}\right)^{\overline{\mathbf{H}}'} \mathrm{d}v \boldsymbol{\Sigma}_i \int_x^1 \frac{1}{u}\left(\frac{1}{u}\right)^{\overline{\mathbf{H}}} \mathrm{d}u + o(n) =: n\boldsymbol{\Sigma}_2^{(i)} + o(n)$$

and

$$\operatorname{cov}\left(\mathbf{G}_n^{(i)}, \sum_{m=1}^{n-1} \frac{\mathbf{G}_m^{(i)}}{m+1}\right) = \sum_{m=1}^{n-1} \overline{\mathbf{B}}'_{n,m} \boldsymbol{\Sigma}_i \left(\sum_{j=m}^{n-1} \frac{1}{j+1} \overline{\mathbf{B}}_{j,m}\right)$$
$$= n\int_0^1 \left(\frac{1}{x}\right)^{\overline{\mathbf{H}}'} \mathrm{d}x \boldsymbol{\Sigma}_i \int_x^1 \frac{1}{u}\left(\frac{1}{u}\right)^{\overline{\mathbf{H}}} \mathrm{d}u + o(n) =: n\boldsymbol{\Sigma}_3^{(i)} + o(n).$$

Note that $\|(1/y)^{\overline{\mathbf{H}}}\| \leqslant c(1/y)^{\lambda} \log^{\nu-1}(1/y)$ for $0 < y \leqslant 1$, and $\lambda < \frac{1}{2}$. The above integrals are well defined. So

$$n^{1/2}(\mathbf{Y}_n/n - \mathbf{v}, \mathbf{N}_n/n - \mathbf{v}) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \boldsymbol{\Lambda}), \tag{3.15}$$

where

$$\boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{\Lambda}_{11}, & \boldsymbol{\Lambda}_{12} \\ \boldsymbol{\Lambda}_{21}, & \boldsymbol{\Lambda}_{22} \end{pmatrix}, \boldsymbol{\Lambda}_{11} = \boldsymbol{\Sigma}_1^{(2)} + \mathbf{H}'\boldsymbol{\Sigma}_1^{(1)}\mathbf{H}, \boldsymbol{\Lambda}_{22} = \boldsymbol{\Sigma}_1^{(1)} + \boldsymbol{\Sigma}_2^{(2)}$$

and

$$\boldsymbol{\Lambda}_{12} = \boldsymbol{\Sigma}_1^{(1)}\mathbf{H} + \boldsymbol{\Sigma}_3^{(2)}. \qquad \square$$

Proof of Theorem 2.2. By (3.12),

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

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

Then, by (3.9) and Corollary 3.1, it follows that

$$\mathbf{N}_n - n\mathbf{v} = \sum_{m=1}^{n} \mathbf{q}_m + \sum_{m=0}^{n-1} \frac{\mathbf{Y}_m - m\mathbf{v}}{m+1} + \sum_{m=1}^{n} o(m^{-\gamma'}) \qquad \forall \gamma' < \gamma$$

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

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

whenever $\kappa > \frac{1}{2} \vee \lambda \vee (1-\gamma)$, and

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

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

whenever $\lambda < \frac{1}{2}$ and $\gamma < \frac{1}{2}$. □

Remark 3.1. Write $\mathbf{T}^*\mathbf{\Sigma}_1^{(i)}\mathbf{T} = (\mathbf{\Sigma}_{ghi},\ g,\ h = 1, \ldots, s)$ and $\mathbf{T} = (\mathbf{1}', \mathbf{T}_2, \ldots, \mathbf{T}_s)$, where $\mathbf{T}$ is defined in Assumption 2.2. Note that

$$\mathbf{T}^*\mathbf{\Sigma}_1^{(i)}\mathbf{T} = \int_0^1 \left(\frac{1}{y}\right)^{\mathrm{diag}[0,\mathbf{J}_2^*,\ldots,\mathbf{J}_s^*]} \mathbf{T}^*\mathbf{\Sigma}_i\mathbf{T} \left(\frac{1}{y}\right)^{\mathrm{diag}[0,\mathbf{J}_2,\ldots,\mathbf{J}_s]} \mathrm{d}y$$

$$= \int_0^1 \mathrm{diag}\left[1, \left(\frac{1}{y}\right)^{\mathbf{J}_2^*}, \ldots, \left(\frac{1}{y}\right)^{\mathbf{J}_s^*}\right] \mathbf{T}^*\mathbf{\Sigma}_i\mathbf{T}\,\mathrm{diag}\left[1, \left(\frac{1}{y}\right)^{\mathbf{J}_2}, \ldots, \left(\frac{1}{y}\right)^{\mathbf{J}_s}\right] \mathrm{d}y.$$

Also

$$\left(\frac{1}{y}\right)^{\mathbf{J}_t} = \left(\frac{1}{y}\right)^{\lambda_t} \sum_{a=0}^{\nu_{t-1}} \frac{\hat{\mathbf{J}}_t^a}{a!} \log^a\left(\frac{1}{y}\right),$$

where

$$\hat{\mathbf{J}}_t = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & 0 & \ldots & 0 \end{pmatrix}, \qquad \hat{\mathbf{J}}_t^2 = \begin{pmatrix} 0 & 0 & 1 & \ldots & 0 \\ 0 & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & 0 & \ldots & 0 \end{pmatrix}, \ldots.$$

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

$$\sum_{a'=0}^{a-1}\sum_{b'=0}^{b-1}\frac{1}{a'!b'!}\int_0^1\left(\frac{1}{y}\right)^{\bar{\lambda}_g+\lambda_h}\log^{a'+b'}\left(\frac{1}{y}\right)[\mathbf{T}_g^*\boldsymbol{\Sigma}_i\mathbf{T}_h]_{a-a',b-b'}$$

$$=\sum_{a'=0}^{a-1}\sum_{b'=0}^{b-1}\frac{(a'+b')!}{a'!b'!(1-\bar{\lambda}_g-\lambda_h)^{a'+b'+1}}[\mathbf{T}_g^*\boldsymbol{\Sigma}_i\mathbf{T}_h]_{a-a',b-b'}.$$

This agrees with the results of Bai and Hu (1999) and Bai *et al.* (2002). One can calculate $\boldsymbol{\Sigma}_2^{(i)}$ and $\boldsymbol{\Sigma}_3^{(i)}$ similarly.

Acknowledgements

Professor Hu was supported by grant DMS-0204232 from the National Science Foundation (USA). Professor Zhang was supported by grants from the National Natural Science Foundation of China and the National Natural Science Foundation of Zhejiang Province. Professor Hu is also affiliated with the Division of Biostatistics and Epidemiology, Department of Health Evaluation Sciences, School of Medicine, University of Virginia. Part of his research was done while visiting the Department of Biometrics, Cornell University. He thanks the Department for its hospitality and support. Special thanks go to anonymous referees for their constructive comments, which led to a much improved version of the paper.

References

Athreya, K.B. and Karlin, S. (1968) Embedding of urn schemes into continuous time branching processes and related limit theorems. *Ann. Math. Statist.*, **39**, 1801–1817.

Bai, Z.D. and Hu, F. (1999) Asymptotic theorem for urn models with nonhomogeneous generating matrices. *Stochastic Process. Appl.*, **80**, 87–101.

Bai, Z.D., Hu, F. and Rosenberger, W.F. (2002) Asymptotic properties of adaptive designs for clinical trials with delayed response. *Ann. Statist.*, **30**, 122–139.

Bandyopadhyay, U. and Biswas, A. (1996) Delayed response in randomized play-the-winner rule: a decision theoretic approach. *Calcutta Statist. Assoc. Bull.*, **46**, 69–88.

Durham, S.D., Flournoy, N. and Li, W. (1998) A sequential design for maximizing the probability of a favourable response. *Canad. J. Statist.*, **26**, 479–495.

Flournoy, N. and Rosenberger, W.F. (eds.) (1995) *Adaptive Designs*. Hayward, CA: Institute of Mathematical Statistics.

Hu, F. (2003) Sample size and power of randomized design. Submitted for publication.

Hu, F. and Ivanova, A. (2004) Adaptive designs. In S.-C. Chow (ed.), *Encyclopedia of Biopharmaceutical Statistics*. New York: Marcel Dekker.

Hu, F. and Rosenberger, W.F. (2003) Optimality, variability, power: evaluating response-adaptive randomization procedures for treatment comparisons. *J. Amer. Statist. Assoc.*, **98**, 671–678.

Monrad, D. and Philipp, W. (1991) Nearby variables with nearby conditional laws and a strong approximation theorem for Hilbert space valued martingales. *Probab. Theory Related Fields*, **88**, 381–404.

Rosenberger, W.F. (1996) New directions in adaptive designs. *Statist. Sci.*, **11**, 137–149.

Rosenberger, W.F. (2002) Randomized urn models and sequential design (with discussions). *Sequential Anal.*, **21**, 1–28.

Smythe, R.T. (1996) Central limit theorems for urn models. *Stochastic Process. Appl.*, **65**, 115–137.

Tamura, R.N., Faries, D.E., Andersen, J.S. and Heiligenstein, J.H. (1994) A case study of an adaptive clinical trial in the treatment of out-patients with depressive disorder. *J. Amer. Statist. Assoc.*, **89**, 768–776.

Wei, L.J. (1988) Exact two-sample permutation tests based on the randomized play-the-winner rule. *Biometrika*, **75**, 603–606.

Received August 2002 and revised January 2004