J. Appl. Prob. **50,** 486–498 (2013)
Printed in England

RANDOMLY REINFORCED URN DESIGNS WITH PRESPECIFIED ALLOCATIONS

GIACOMO ALETTI,* *Università degli Studi di Milano*
ANDREA GHIGLIETTI ** *** AND
ANNA MARIA PAGANONI,** **** *Politecnico di Milano*

Abstract

We construct a response adaptive design, described in terms of a two-color urn model, targeting fixed asymptotic allocations. We prove asymptotic results for the process of colors generated by the urn and for the process of its compositions. An application of the proposed urn model is presented in an estimation problem context.

Keywords: Reinforced process; urn scheme; sequential clinical trial; stochastic process

2010 Mathematics Subject Classification: Primary 60F15
Secondary 62L12

1. Introduction

Consider a clinical trial with two competitive treatments, say R and W. We want to construct a response-adaptive design, described in terms of an urn model, targeting any optimal, fixed asymptotic allocation, in order to compare these designs with others studied in the literature. A large class of response-adaptive randomized designs is based on urn models, a classical tool to guarantee a randomized device (see Rosenberger (2002) and Zhang *et al.* (2006)), to balance the allocations (see Baldi Antognini and Giannerini (2007)), or to construct designs which asymptotically assign all subjects to the best treatment (see Flournoy *et al.* (2012)). The *two-color, randomly reinforced urn* (RRU) introduced in Durham and Yu (1990), extended to the multi-color case in Durham *et al.* (1998), and studied in Muliere *et al.* (2006), Aletti *et al.* (2009), (2012), and May and Flournoy (2009), is a randomized device able to asymptotically allocate subjects to the optimal treatment; see Muliere *et al.* (2006). In this paper we modify the reinforcement scheme of the urn to asymptotically target an optimal allocation proportion. The term *target* refers to the limit of the urn proportion process. Let us consider two probability distributions, μ_R and μ_W, with support contained in $[\alpha, \beta]$, where $0 \leq \alpha \leq \beta < +\infty$ and a sequence $(U_n)_n$ of independent, uniform random variables on $(0, 1)$. We will interpret μ_R and μ_W as the laws of the responses to treatments R and W, respectively. We assume that both the means $m_R = \int_\alpha^\beta x\mu_R(\mathrm{d}x)$ and $m_W = \int_\alpha^\beta x\mu_W(\mathrm{d}x)$ are strictly positive. Visualize an urn initially containing r_0 balls of color R and w_0 balls of color W. Set

$$R_0 = r_0, \qquad W_0 = w_0, \qquad D_0 = R_0 + W_0, \qquad Z_0 = \frac{R_0}{D_0}.$$

Received 26 July 2011; revision received 23 July 2012.
* Postal address: Dipartimento di Matematica "F. Enriques", Università degli studi di Milano via Saldini 50, 20133 Milano, Italy. Email address: giacomo.aletti@unimi.it
** Postal address: Dipartimento di Matematica "F. Brioschi", Politecnico di Milano Piazza Leonardo da Vinci 32, 20133 Milano, Italy.
*** Email address: andrea.ghiglietti@mail.polimi.it
**** Email address: anna.paganoni@polimi.it

At time $n = 1$, a ball is sampled from the urn; its color is $X_1 = \mathbf{1}_{[0,Z_0]}(U_1)$, a random variable with Bernoulli(Z_0) distribution. Let M_1 and N_1 be two independent random variables with distributions μ_R and μ_W, respectively; assume that X_1, M_1, and N_1 are independent. Next, if the sampled ball is R, it is replaced in the urn together with $X_1 M_1$ balls of the same color if $Z_0 < \eta$, where $\eta \in (0, 1)$ is a suitable parameter; otherwise, the urn composition does not change. If the sampled ball is W, it is replaced in the urn together with $(1 - X_1)N_1$ balls of the same color if $Z_0 > \delta$, where $\delta < \eta \in (0, 1)$ is a suitable parameter; otherwise, the urn composition does not change. So we can update the urn composition in the following way:

$$\begin{gathered} R_1 = R_0 + X_1 M_1 \mathbf{1}_{[Z_0<\eta]}, \qquad W_1 = W_0 + (1 - X_1)N_1 \mathbf{1}_{[Z_0>\delta]}, \\ D_1 = R_1 + W_1, \qquad Z_1 = \frac{R_1}{D_1}. \end{gathered} \tag{1.1}$$

Now iterate this sampling scheme forever. Thus, at time $n + 1$, given the sigma-field $\mathcal{F}_n$ generated by $X_1, \dots, X_n$, $M_1, \dots, M_n$, and $N_1, \dots, N_n$, let $X_{n+1} = \mathbf{1}_{[0,Z_n]}(U_{n+1})$ be a Bernoulli(Z_n) random variable and, independently from $\mathcal{F}_n$ and X_{n+1}, assume that M_{n+1} and N_{n+1} are two independent random variables with distributions μ_R and μ_W, respectively. Set

$$\begin{gathered} R_{n+1} = R_n + X_{n+1} M_{n+1} \mathbf{1}_{[Z_n<\eta]}, \qquad W_{n+1} = W_n + (1 - X_{n+1})N_{n+1} \mathbf{1}_{[Z_n>\delta]}, \\ D_{n+1} = R_{n+1} + W_{n+1}, \qquad Z_{n+1} = \frac{R_{n+1}}{D_{n+1}}. \end{gathered} \tag{1.2}$$

We thus generate an infinite sequence $X = (X_n,\ n = 1, 2, \dots)$ of Bernoulli random variables, with X_n representing the color of the ball sampled from the urn at time n, and a process $(Z, D) = ((Z_n, D_n),\ n = 0, 1, 2 \dots)$ with values in $[0, 1] \times (0, \infty)$, where D_n represents the total number of balls in the urn before it is sampled for the $(n + 1)$th time, and Z_n is the proportion of balls of color R; we call X the process of colors generated by the urn, while (Z, D) is the process of its compositions. Let us observe that the process (Z, D) is a Markov sequence with respect to the filtration $\mathcal{F}_n$.

There are many experimental designs whose proportion of patients allocated to treatments converge to a fixed value, different from one or zero. Many of these procedures, like those targeting the optimal Neyman allocation, are no-adaptive designs. In this case, the limit proportion of assignment is independent of treatment responses, since we cannot use previous data to change and improve the strategy of the experiment. A very general adaptive design, targeting a fixed asymptotic allocation proportion, was introduced in Eisele and Woodroofe (1995). It was a biased coin procedure, in which the probability of assignment is modeled as a function of previous assignments and current estimates of the limiting proportion. The probability of allocation, under some very restrictive conditions, converges to the desired target allocation. After that, Melfi *et al.* (2001) and Hu and Zhang (2004) studied different versions of the doubly adaptive biased coin design by relaxing the conditions over the function described in Eisele's model. In Hu *et al.* (2009), the efficient randomized-adaptive design (ERADE) was presented, which consists of a family of response-adaptive randomization procedures that attain the Cramer Rao lower bounds. All these models are based on the function presented in Eisele and Woodroofe (1995). Our model is different because the probability of allocation cannot be expressed exclusively as a function of the previous assignments and the estimates of the limit proportion. The probability of allocation is determined by the composition of the urn, which is also influenced by the randomness of the reinforcements.

In this work we study the asymptotic behavior of the urn process. In particular, in Section 2 we prove some general results concerning urn processes. In Section 3 the convergence result on urn composition is proved. Finally, in Section 4 an application of the proposed urn model is presented in an estimation problem context.

2. Upcrossing/downcrossing and reinforcements

We are interested in studying the convergence of an adapted, bounded process $(Z_n)_n$. Without loss of generality, we will take $Z_n \in [0, 1]$ for all n. We study the crossing in both directions of a strip $[d, u]$, where $0 < d < u < 1$. More precisely, let $t_{-1} = -1$, and define, for every $j \in \mathbb{Z}_+$, the two stopping times

$$\tau_j = \begin{cases} \inf\{n > t_{j-1} \colon Z_n < d\} & \text{if } \{n > t_{j-1} \colon Z_n < d\} \neq \varnothing, \\ +\infty & \text{otherwise,} \end{cases} \tag{2.1a}$$

$$t_j = \begin{cases} \inf\{n > \tau_j \colon Z_n > u\} & \text{if } \{n > \tau_j \colon Z_n > u\} \neq \varnothing, \\ +\infty & \text{otherwise.} \end{cases} \tag{2.1b}$$

The random interval $(\tau_{j-1}, \tau_j]$ is called the *jth excursion*, and we denote it by

$$\nu^Z_{[d,u]} = \begin{cases} \sup\{j \colon \tau_j < \infty\} & \text{if } \tau_0 < +\infty, \\ 0 & \text{otherwise,} \end{cases}$$

that is, $\nu^Z_{[d,u]}$ counts the total number of times that the process Z crosses the strip $[d, u]$ in both directions, i.e. making both an upcrossing and a downcrossing.

Theorem 2.1. *The process $(Z_n)_n$ converges almost surely (a.s.) if and only if, for any $0 < d < u < 1$,*

$$\sum \mathbb{P}(\tau_{j+1} = \infty \mid \tau_j < \infty) = \infty,$$

with the convention that $\mathbb{P}(\tau_{j+1} = \infty \mid \tau_j < \infty) = 1$ if $\mathbb{P}(\tau_j = \infty) = 1$.

Proof. We first note that

$$\begin{aligned} & (Z_n)_n \text{ converges a.s.} \\ \iff\ & \mathbb{P}(\nu^Z_{[d,u]} = \infty) = 0 \quad \text{for all } 0 < d < u < 1 \\ \iff\ & 0 = \lim_{n\to\infty} \mathbb{P}(\nu^Z_{[u,d]} \geq n) = \lim_{n\to\infty} \mathbb{P}\Big(\bigcap_{j=0}^{n} \{\tau_j < \infty\}\Big) \quad \text{for all } 0 < d < u < 1 \end{aligned}$$

as a consequence of the countability of $\mathbb{Q}$ in $[0, 1]$. Now

$$\mathbb{P}(\{\tau_j < \infty,\ j = 0, \dots, n\}) = \mathbb{P}(\tau_0 < \infty) \prod_{j=1}^{n} \mathbb{P}(\tau_j < \infty \mid \tau_{j-1} < \infty),$$

and it is well known that if $(p_j)_j \subseteq (0, 1]$ then

$$\lim_{n\to\infty} \prod_{j=1}^{n} p_j = 0 \quad \iff \quad \sum_{j=1}^{\infty} (1 - p_j) = \infty.$$

The fact that some $(p_n)_n$ might be 0 is controlled by the assumption that $p_n = 0$ implies that $p_m = 0$ for all $m > n$.

https://doi.org/10.1239/jap/1371648956 Published online by Cambridge University Press

Now, we will prove the convergence of a general class of urn processes.

Definition 2.1. (*Birth urn process.*) Let $(\Omega, \mathcal{F}, (\mathcal{F}_n)_n, \mathbb{P})$ be a filtered space. A vector process $(R_n, W_n)_n$ on $(\Omega, \mathcal{F}, (\mathcal{F}_n)_n, \mathbb{P})$ is called a *birth urn process* (BUP) if $(R_n, W_n)_n$ is $(\mathcal{F}_n)_n$-adapted, the processes $(R_n)_n$ and $(W_n)_n$ are nonnegative and increasing (i.e. $0 \le R_0 \le R_1 \le \cdots \le R_n \le \cdots$ and $0 \le W_0 \le W_1 \le \cdots \le W_n \le \cdots$), and $R_0 + W_0 > 0$. Let $D_n = R_n + W_n$ for $n \in \mathbb{N}$.

Lemma 2.1. (Reinforcements during excursions.) *For any BUP,*

$$D_{\tau_j} \ge \frac{u(1-d)}{d(1-u)} D_{\tau_{j-1}} \ge \cdots \ge \left(\frac{u(1-d)}{d(1-u)}\right)^j D_{\tau_0}.$$

Proof. For every $j \in \mathbb{N}_0$, we have

- $R_{\tau_{j+1}} \ge R_{t_j} \Longrightarrow Z_{\tau_{j+1}} D_{\tau_{j+1}} \ge Z_{t_j} D_{t_j}$,
- $W_{t_j} \ge W_{\tau_j} \Longrightarrow (1 - Z_{t_j}) D_{t_j} \ge (1 - Z_{\tau_j}) D_{\tau_j}$.

Since $Z_{\tau_j} < d$ and $Z_{t_j} > u$ for every $j \in \mathbb{N}$, we find that

- $d D_{\tau_{j+1}} \ge u D_{t_j}$,
- $(1-u) D_{t_j} \ge (1-d) D_{\tau_j}$.

From this we immediately obtain

$$D_{\tau_j} \ge \frac{u(1-d)}{d(1-u)} D_{\tau_{j-1}} \ge \cdots \ge \left(\frac{u(1-d)}{d(1-u)}\right)^j D_{\tau_0},$$

completing the proof.

Given a sequence of stopping times $(\tau_n)_n$, it is always possible to define the counting process

$$C_n := \begin{cases} \displaystyle\sum_{j=1}^{\infty} \mathbf{1}_{\{\tau_j \le n\}} & \text{if } \tau_0 \le n, \\ -1 & \text{if } \tau_0 > n. \end{cases}$$

A BUP $(R_n, W_n)_n$ is associated to the sequence $(\tau_n)_n$ if $(R_n, W_n, C_n)_n$ is a time-homogeneous Markov process. In this case

$$\mathbb{P}(\tau_{i+1} < \infty \mid \tau_i < \infty) = f(R_{\tau_i}, W_{\tau_i}, i). \tag{2.2}$$

Finally, note that, given a generalized urn process $(R_n, W_n)_n$, it is always possible to define two adapted processes $\{D_n := R_n + W_n,\ n \in \mathbb{N}\}$ and $\{Z_n := R_n / D_n,\ n \in \mathbb{N}\}$.

Proposition 2.1. *Given a Markov BUP, the process $(Z_n)_n$ converges a.s. if, for any $0 < d < u < 1$, there exists a function $g\colon [0, \infty) \times [0, \infty) \to [0, 1]$, $(R_n, W_n)_n$ is associated to the sequence $(\tau_n)_n$ defined in (2.1a), and*

$$f(x, y, \cdot) \le g(x', y') \quad \textit{if } x + y \ge x' + y',$$
$$g(c_1, c_2) < 1 \quad \textit{for some } c_1, c_2 > 0,$$

where f is given in (2.2).

https://doi.org/10.1239/jap/1371648956 Published online by Cambridge University Press

Proof. On $\{\tau_0 = \infty\}$, we get $\nu^Z_{[u,d]} = 0$. On $\{\tau_0 < \infty\}$, if

$$j \geq \log_{u(1-d)/d(1-u)} \frac{c_1 + c_2}{D_{\tau_0}}$$

then, by Lemma 2.1,

$$\mathbb{P}(\tau_{j+1} = \infty \mid \tau_j < \infty) \geq 1 - g(c_1, c_2) = a > 0.$$

The proposition then follows from Theorem 2.1.

3. Convergence theorem

Let us consider the urn model described in Section 1.

Theorem 3.1. *The sequence of proportions $Z = (Z_n, n = 1, 2, \dots)$ of the urn process described in Section 1 converges a.s. to the following limit:*

$$\lim_{n\to\infty} Z_n = \begin{cases} \eta & \text{if } m_R > m_W, \\ \delta & \text{if } m_R < m_W. \end{cases}$$

To prove this theorem, we provide auxiliary results based on the Doob decomposition

$$Z_n = Z_0 + M_n + A_n,$$

where $(M_n)_n$ is a martingale and $(A_n)_n$ is a predictable process, both null at $n = 0$. Denote by $m_R = \int_\alpha^\beta x\mu_R(\mathrm{d}x)$ and $m_W = \int_\alpha^\beta x\mu_W(\mathrm{d}x)$ the means of the patients' responses to treatments.

Lemma 3.1. (Aletti *et al.* (2012, Lemmas A.2 and A.3).) *Assume that $m_R = m_W = m$. If $D_0 \geq 2\beta$ then*

$$\mathbb{E}\Big(\sup_n |A_n|\Big) \leq \frac{\beta}{D_0}, \qquad \mathbb{E}(\langle M\rangle_\infty - \langle M\rangle_n \mid \mathcal{F}_n) \leq \frac{\beta}{D_0} \quad \textit{for any } n \geq 0.$$

As a consequence, we obtain the following result.

Lemma 3.2. *Assume that $m_R = m_W = m$. If $D_0 \geq 2\beta$ then*

$$\mathbb{P}\Big(\sup_n |Z_n - Z_0| \geq h\Big) \leq \frac{\beta}{D_0}\bigg(\frac{4}{h^2} + \frac{2}{h}\bigg)$$

for every $h > 0$.

Proof. First note that, since $(M_n)_n$ is a martingale null at $n = 0$, we have, by Lemma 3.1 (choosing $n = 0$ in the second inequality),

$$\lim_{n\to\infty} \mathbb{E}(M_n^2) = \lim_{n\to\infty} \mathbb{E}(\langle M\rangle_n) \leq \frac{\beta}{D_0},$$

and, hence, by Doob's L^2-inequality,

$$\mathbb{P}\bigg(\bigg\{\sup_n |M_n| \geq \frac{h}{2}\bigg\}\bigg) \leq \lim_{n\to\infty} \frac{\mathbb{E}(M_n^2)}{(h/2)^2} \leq \frac{4\beta}{h^2 D_0} \quad \text{for any } h > 0.$$

We easily get

$$\begin{aligned}\mathbb{P}\Big(\sup_n |Z_n - Z_0| \ge h\Big) &\le \mathbb{P}\Big(\Big\{\sup_n |M_n| \ge \tfrac{1}{2}h\Big\} \cup \Big\{\sup_n |A_n| \ge \tfrac{1}{2}h\Big\}\Big)\\ &\le \mathbb{P}\Big(\Big\{\sup_n |M_n| \ge \tfrac{1}{2}h\Big\}\Big) + \mathbb{P}\Big(\Big\{\sup_n |A_n| \ge \tfrac{1}{2}h\Big\}\Big)\\ &\le \frac{\beta}{D_0}\Big(\frac{4}{h^2} + \frac{2}{h}\Big),\end{aligned}$$

completing the proof.

Proof of Theorem 3.1. We have an urn initially containing R_0 red balls and W_0 white balls. Let us consider the case in which $m_R < m_W$; the opposite case ($m_R > m_W$) is completely analogous. In the case described in Muliere *et al.* (2006) the process $(Z_n)_{n\in\mathbb{N}}$ is a supermartingale converging to 0 but, because of the barrier δ (see (1.2)), it is not like this anymore. Anyway, we want to prove that the process $(Z_n)_{n\in\mathbb{N}}$ still converges, but in this case the limit is equal to δ.

First, we will prove that

$$\liminf Z_n \le \delta \quad \text{a.s.}$$

By contradiction, there exists $l > \delta$ such that $\mathbb{P}(\liminf Z_n \ge l) > 0$. Then, there exists n_0 such that $\mathbb{P}(Z_n > (l+\delta)/2 \text{ for all } n \ge n_0) > 0$. This contradicts the fact that, by Markov's property, $\mathbb{P}(Z_n > (l+\delta)/2 \text{ eventually}) = 0$, since it is an RRU with reinforcements with different means that goes to 0 (see Muliere *et al.* (2006)).

With the same argument, one may prove that $\limsup Z_n \ge \delta$, since the urn that eventually stays below δ is an RRU with reinforcements with different means that goes to 1 (again, see Muliere *et al.* (2006)).

In fact, one can prove more, with the arguments of Muliere *et al.* (2006): the barrier δ must be crossed infinitely times a.s. With this result in mind, we will prove in a moment that $\liminf Z_n \ge \delta$. In fact, if there exists $l < \delta$ such that $\mathbb{P}(\liminf Z_n \le l) > 0$, then, with positive probability, the process must cross the strip $((l+\delta)/2, \delta)$ an infinite number of times. By Lemma 2.1, after a sufficiently large number of times, $D_n > \beta(l+\delta)/(\delta-l)$ and, therefore, if k is any successive downcross of δ,

$$Z_k \ge \frac{R_{k-1}}{D_{k-1}+\beta} \ge \frac{\delta D_n}{D_n+\beta} > \frac{l+\delta}{2},$$

since each reinforced urn is bounded by β and $R_{k-1}/D_{k-1} = Z_{k-1} > \delta$. We have proved that $\liminf Z_n = \delta$ a.s.

Let d and u ($\delta < d < u$) be two arbitrary points, and let $(\tau_i)_i$ and $(t_i)_i$ be as in (2.1a) and (2.1b), in order to apply Proposition 2.1. Let

$$i > \log_{u(1-d)/d(1-u)} \frac{\beta(1-d)}{D_{\tau_0}(d-\delta)}$$

be fixed, so that, by Lemma 2.1, $D_{\tau_i} > \beta(1-d)/(d-\delta)$, and denote by $(\widehat{\cdot}_n)_{n\in\mathbb{N}}$ the renewed process on $\{\tau_i < \infty\}$: $(\hat{R}_n, \hat{W}_n) = (R_{\tau_i+n}, W_{\tau_i+n})$, $\hat{D}_n = \hat{R}_n + \hat{W}_n = D_{\tau_i+n}$, $\hat{Z}_n = \hat{R}_n/\hat{D}_n = Z_{\tau_i+n}$, and $\hat{U}_n = U_{\tau_i+n}$. The Markov property of the original urn ensures that, on $\{\tau_i < \infty\}$, the process $(\widehat{\cdot}_n)_n$ started afresh a new urn with initial composition (R_{τ_i}, W_{τ_i}) with dynamics (1.1) and (1.2). We denote by $P_i(\cdot) = \mathbb{P}(\cdot \mid \tau_i < \infty)$, and, therefore, if

$$t = \begin{cases} \inf\{n\colon \hat{Z}_n > u\} & \text{if } \{n\colon \hat{Z}_n > u\} \neq \varnothing,\\ +\infty & \text{otherwise,}\end{cases}$$

then we have

$$P_i(t < \infty) = P_i(t_i < \infty) \geq \mathbb{P}(\tau_{i+1} < \infty \mid \tau_i < \infty). \tag{3.1}$$

Define the sequences $(t_n^*, \tau_n^*)_n$ of stopping times which indicate the $(\hat{Z}_n)_n$-crosses of the border δ: let $t_{-1}^* = -1$, and define, for every $j \in \mathbb{Z}_+$, the two stopping times

$$\tau_j^* = \begin{cases} \inf\{n > t_{j-1}^* \colon \hat{Z}_n \leq \delta\} & \text{if } \{n > t_{j-1}^* \colon \hat{Z}_n \leq \delta\} \neq \varnothing, \\ +\infty & \text{otherwise,} \end{cases}$$

$$t_j^* = \begin{cases} \inf\{n > \tau_j^* \colon \hat{Z}_n > \delta\} & \text{if } \{n > \tau_j^* \colon \hat{Z}_n > \delta\} \neq \varnothing, \\ +\infty & \text{otherwise.} \end{cases}$$

Note that

$$\frac{R}{R+W} \leq \delta, \qquad R + W > \frac{\beta(1-d)}{d-\delta} \quad \Longrightarrow \quad \frac{R+x}{R+W+x} < d \quad \text{for all } x \leq \beta,$$

and, hence, since the reinforcements are bounded by β, we have

$$\hat{Z}_{t_j^*-1} \leq \delta, \qquad \hat{D}_{t_j^*-1} > \frac{\beta(1-d)}{d-\delta} \quad \Longrightarrow \quad \hat{Z}_{t_j^*} < d \quad \Longrightarrow \quad \hat{R}_{t_j^*} < \hat{W}_{t_j^*-1}\frac{d}{1-d}. \tag{3.2}$$

We now define a process $(\tilde{\cdot}_n)_{n\in\mathbb{N}}$ to set a new urn, coupled with $(\hat{\cdot}_n)_{n\in\mathbb{N}}$, with the following features:

$$\begin{aligned}
\tilde{W}_0 &= \hat{W}_0, \\
\tilde{R}_0 &= \tilde{W}_0 \frac{u+d}{2-u-d}, \\
\tilde{X}_{n+1} &= \mathbf{1}_{[0,\tilde{Z}_n]}(\hat{U}_{n+1}), \\
\tilde{M}_{n+1} &= \hat{M}_{n+1} + (m_W - m_R), \\
\tilde{N}_{n+1} &= \hat{N}_{n+1}, \\
\tilde{R}_{n+1} &= (\tilde{R}_n + \tilde{X}_{n+1}\tilde{M}_{n+1})\,\mathbf{1}_{[\hat{Z}_n > \delta]} + \tilde{W}_n \frac{u+d}{2-u-d}\,\mathbf{1}_{[\hat{Z}_n \leq \delta]}, \\
\tilde{W}_{n+1} &= (\tilde{W}_n + (1-\tilde{X}_{n+1})\tilde{N}_{n+1})\,\mathbf{1}_{[\hat{Z}_n > \delta]} + \tilde{W}_n\,\mathbf{1}_{[\hat{Z}_n \leq \delta]}, \\
\tilde{D}_{n+1} &= \tilde{R}_{n+1} + \tilde{W}_{n+1}, \\
\tilde{Z}_{n+1} &= \frac{\tilde{R}_{n+1}}{\tilde{D}_{n+1}}.
\end{aligned}$$

The new urn process starts with $\tilde{Z}_0 = (u+d)/2$, it is reinforced at time $n+1$ only when $\hat{Z}_n > \delta$, and has nonnegative reinforcements with the same means m_R; besides, the process is set equal to $(u+d)/2$ at time $n+1$ whenever $\hat{Z}_n \leq \delta$.

We will prove by induction that, for any n,

$$\tilde{Z}_n > \hat{Z}_n, \qquad \tilde{W}_n \leq \hat{W}_n, \qquad \tilde{R}_n > \hat{R}_n. \tag{3.3}$$

In other words, we will show that $(\tilde{Z}_n)_{n\in\mathbb{N}}$ is always above the original process $(\hat{Z}_n)_{n\in\mathbb{N}}$.

In fact, by construction we have

$$\tilde{Z}_0 = \frac{d+u}{2} > d > \hat{Z}_0, \qquad \tilde{W}_0 = \hat{W}_0,$$

which immediately implies that $\tilde{R}_0 > \hat{R}_0$. Assume that (3.3) holds by the induction hypothesis. We consider two cases.

Case 1: $\hat{Z}_n \le \delta$. $\tilde{W}_{n+1} = \hat{W}_{n+1}$ by construction. By (3.2), $\hat{Z}_{n+1} < d < \tilde{Z}_n = \tilde{Z}_{n+1}$ and, hence, $\tilde{R}_{n+1} > \hat{R}_{n+1}$.

Case 2: $\hat{Z}_n > \delta$. Since $\tilde{X}_{n+1} = \mathbf{1}_{[0,\tilde{Z}_n]} \ge \mathbf{1}_{[0,\hat{Z}_n]} = \hat{X}_{n+1}$ by construction, we obtain

$$\hat{R}_{n+1} - \hat{R}_n = \hat{X}_{n+1}\hat{M}_{n+1} \le \tilde{X}_{n+1}\tilde{M}_{n+1} = \tilde{R}_{n+1} - \tilde{R}_n,$$
$$\hat{W}_{n+1} - \hat{W}_n = (1 - \hat{X}_{n+1})\hat{N}_{n+1} \ge (1 - \tilde{X}_{n+1})\tilde{N}_{n+1} = \tilde{W}_{n+1} - \tilde{W}_n.$$

Note that, for any $m \ge 1$, the process $(\tilde{Z}_{t^*_{m-1}+n})_{n=0}^{\tau^*_m - t^*_{m-1}}$ is an urn process reinforced by distributions with the same means and initial composition $(\tilde{R}_{t^*_{m-1}}, \tilde{W}_{t^*_{m-1}})$. Therefore, if T_m is the stopping time for $(\tilde{Z}_{t^*_{m-1}+n})_n$ to exit from (d, u) before τ^*_m, i.e.

$$T_m = \begin{cases} \inf\{n \le \tau^*_m - t^*_{m-1} : \tilde{Z}_{t^*_{m-1}+n} \le d \text{ or } \tilde{Z}_{t^*_{m-1}+n} \ge u\} & \text{if } \{n \le \tau^*_m - t^*_{m-1} : \tilde{Z}_{t^*_{m-1}+n} \le d \\ & \quad \text{or } \tilde{Z}_{t^*_{m-1}+n} \ge u\} \ne \varnothing, \\ +\infty & \text{otherwise,} \end{cases}$$

then we have stated that

$$P_i(T_m < \infty) \ge P_i(t < \infty \mid \{t^*_{m-1} < t < \tau^*_m\}). \tag{3.4}$$

Now, as a consequence of Lemma 3.2 and the fact that $\tilde{D}_{t^*_{m-1}} \ge \tilde{D}_0 \ge D_{\tau_i}$, if we set $h = (u-d)/2$, we get

$$P_i(T_m < \infty) \le \mathbb{P}\Big(\sup_n |\tilde{Z}_{t^*_{m-1}+n} - \tilde{Z}_{t^*_{m-1}}| \ge h\Big) \le \min\bigg(\frac{\beta}{D_{\tau_i}}\bigg(\frac{4}{h^2} + \frac{2}{h}\bigg), 1\bigg).$$

Thus, define the function $g\colon [0,\infty) \times [0,\infty) \to [0,1]$ as

$$g(x,y) := \min\bigg(\frac{\beta}{x+y}\bigg(\frac{4}{h^2} + \frac{2}{h}\bigg), 1\bigg),$$

and note that $g(8\beta/h^2, 4\beta/h) = \frac{1}{2}$ and g is monotone in $x + y$. We can apply Proposition 2.1 to get the thesis, since, by (3.1) and (3.4),

$$\begin{aligned} \mathbb{P}(\tau_{i+1} < \infty \mid \tau_i < \infty) &\le \sum_m P_i(t < \infty \mid \{t^*_{m-1} < t < \tau^*_m\}) P_i(\{t^*_{m-1} < t < \tau^*_m\}) \\ &\le \sup_m P_i(t < \infty \mid \{t^*_{m-1} < t < \tau^*_m\}) \\ &\le g(R_{\tau_i}, W_{\tau_i}). \end{aligned}$$

This completes the proof.

Remark 3.1. Note that in the proof of Theorem 3.1 it was never necessary to specify the type of distribution generating the reinforcements. Indeed, we do not need all the information about the probability laws, we deal only with the means of those distributions. In particular, in the proof we only needed to know which of the two reinforcements has the greatest mean. For this reason, all the results still hold if we change the dynamic of the process, maintaining a fixed sign for the difference in the means.

Remark 3.2. Consider a Pólya urn initially containing r_0 red balls and w_0 white balls. Let $X = (X_n)_{n\in\mathbb{N}}$ be a generalized urn process of the sampled balls, and let f be the corresponding urn function, i.e. the function f that maps the interval (0,1) to itself and such that the law of X is defined by assuming that X_1 is a Bernoulli($f(z_0)$), where $z_0 = r_0/(r_0 + w_0)$ and, for $n \geq 1$, the conditional distribution of X_{n+1} given $X_1, \dots, X_n$ is a Bernoulli($f(Z_n)$), where

$$Z_n = \frac{r_0 + \sum_{i=1}^n X_i}{r_0 + w_0 + n}.$$

If $f(x) = x$ for every $x \in [0, 1]$, we obtain the Pólya sequence. Now, consider the urn model described in the introduction, in the particular case in which reinforcements are independent Bernoulli variables, with parameters π_R for the red balls and π_W for the white balls. In this situation, this model is equivalent to a generalized Pólya urn in which the urn function f can be defined as follows:

$$f(x) = \frac{x\pi_R\, \mathbf{1}_{[x<\eta]}}{x\pi_R\, \mathbf{1}_{[x<\eta]} + (1-x)\pi_W\, \mathbf{1}_{[x>\delta]}} = \begin{cases} 1 & \text{if } x < \delta, \\ \dfrac{x\pi_R}{x\pi_R + (1-x)\pi_W} & \text{if } \delta < x < \eta, \\ 0 & \text{if } x > \eta. \end{cases}$$

Looking at the expression above, we can reach to the same convergence result proved in this paper, by applying the Theorem 4.1 of Hill *et al.* (1980). The convergence theorem proved in this paper is more general, because it also holds when reinforcements do not follow Bernoulli distributions.

Now, let us consider $\rho(m_R, m_W) = \eta\, \mathbf{1}_{[m_R > m_W]} + \delta\, \mathbf{1}_{[m_R < m_W]}$. We have shown that, for the reinforcement scheme introduced here, Z_n converges a.s. to ρ, so we denote ρ as the target allocation. By using the same martingale argument as Melfi *et al.* (2001), we can prove that $N_R(n)/n \to \rho$ a.s. This results allows us to force the design to be asymptotically balanced or unbalanced for a fixed suitable quantity: in fact, $(N_R(n) - N_W(n))/n \to 2\rho - 1$. Moreover, consider an estimation problem of the means m_R and m_W of the responses to treatments. The limit of the process ρ is within the open interval (0, 1) and so both the sequences $N_R(n) = \sum_{i=1}^n X_i$ and $N_W(n) = \sum_{i=1}^n (1 - X_i)$ diverge to ∞ a.s. as long as n increases to ∞. This allows us to define the following adaptive consistent estimators based on the observed responses until time n, with random sample sizes $N_R(n)$ and $N_W(n)$, respectively:

$$\bar{M}(n) = \frac{\sum_{i=1}^n X_i M_i}{N_R(n)} \quad \text{and} \quad \bar{N}(n) = \frac{\sum_{i=1}^n (1 - X_i) N_i}{N_W(n)}.$$

We can apply the results proved in Melfi *et al.* (2001) to state the following proposition.

Proposition 3.1. *The estimators $\bar{M}(n)$ and $\bar{N}(n)$ are consistent estimators of m_R and m_W, respectively. Moreover, as $n \to \infty$,*

$$\left(\sqrt{N_R(n)}\frac{\bar{M}(n) - m_R}{\sigma_R}, \sqrt{N_W(n)}\frac{\bar{N}(n) - m_W}{\sigma_W}\right) \to (Z_1, Z_2)$$

in distribution, where (Z_1, Z_2) are independent standard normal random variables.

4. A simulation study

In this section we present a simulation study that takes advantage of the convergence theorem proved in Section 3. We provide a method to estimate an unknown parameter by using the proposed response-adaptive design. Let us consider a treatment W, whose mean effect on subjects is unknown. Let us model the patient's response to the treatment W with a random variable with distribution μ_W. The goal of the study is to estimate its mean effect $m_W = \int x\mu_W(\mathrm{d}x)$. Consider another treatment, denoted as R, and suppose that its random effect on patients follows a known distribution μ_R; let us assume that its mean m_R depends on the given dose, which can be suitably modified by the experimenter. We consider a response-adaptive design based on the urn model introduced in Section 1, with μ_R and μ_W modeling the patients' responses to treatment R and W, respectively. The inference on m_W is performed by monitoring over time the urn composition Z_n.

In this simulation study we consider K urns with the same initial composition (r_0, w_0). Red balls are associated with treatment R, while white balls are associated with treatment W. We denote by $Z^j = (Z_n^j)_{n\in\mathbb{N}}$ the process of the urn proportion in the jth urn for $j \in \{1, 2, \dots, K\}$. The reinforced scheme applied to each urn is that described in Section 1. Hence, for each urn, convergence Theorem 3.1 holds, and

$$\lim_{n\to\infty} Z_n^j = \begin{cases} \eta & \text{if } m_R > m_W, \\ \delta & \text{if } m_R < m_W. \end{cases}$$

When $m_R = m_W$, we do not have an explicit form for the limit distribution of the urn proportion Z_n. Nevertheless, we know that it converges to a random variable Z_e, whose distribution has no atoms and support $S_e = [\delta, \eta]$.

At the beginning of the experiment, we choose an initial dose for treatment R. Let us call $m_{R,1}$ the patient response mean corresponding to that dose. Then, the reinforcements of red and white balls follow distributions with means $m_{R,1}$ and m_W, respectively. We start K mutually independent urn processes simultaneously. At each step, we draw a ball from each urn and we update the composition of each urn independently, following the model described in Section 1. After n draws and reinforcements, we have K urn proportions Z_n^j, $j \in \{1, 2, \dots, K\}$, which can be used to compute the empirical cumulative distribution function $\hat{F}_n$ for the random variable Z_n. Thanks to Theorem 3.1, for every $x \in [0, 1]$, $\hat{F}_n(x)$ must converge to

$$F_\eta(x) = \mathbf{1}_{[x\geq\eta]} \quad \text{if } m_W < m_{R,1}, \qquad F_\delta(x) = \mathbf{1}_{[x\geq\delta]} \quad \text{if } m_W > m_{R,1}.$$

If $m_W = m_{R,1}$, we can compute offline $\hat{F}_e(x)$, the asymptotic cumulative distribution of Z_e. This calculation requires the simulation of M urn processes with m draws for each one; the number of urns M and the number of draws m can be arbitrarily large. So we have

$$\hat{F}_e(x) \simeq \frac{1}{M}\sum_{i=1}^{M} \mathbf{1}_{[Z_m^i < x]} \quad \text{for large } m \text{ and } M.$$

At each step, once each urn has been reinforced, we use the Wasserstein distance (d_W) to compute the distances between the empirical cumulative distribution function $\hat{F}_n$ and the three asymptotic possible distributions F_η, $\hat{F}_e$, and F_δ. When one of these three distances is small enough, we have a good estimate of the distribution of the limit proportion Z_n, and so we can state if m_W is less than, equal to, or greater than $m_{R,1}$. Let

$$\begin{aligned}\xi &= \min\{d_W(Z_n, \delta_\eta), d_W(Z_n, Z_e), d_W(Z_n, \delta_\delta)\} \\ &= \min\left\{\int_0^1 |F_n(x) - F_\eta(x)|\,\mathrm{d}x, \int_0^1 |F_n(x) - \hat{F}_e(x)|\,\mathrm{d}x, \int_0^1 |F_n(x) - F_\delta(x)|\,\mathrm{d}x\right\}.\end{aligned}$$

When ξ is less than a suitable small parameter α, fixed in advance, the drawing process ends and different scenarios are possible. If $\xi = d_W(Z_n, Z_e)$, we conclude that $m_{R,1} = m_W$. Otherwise, if $\xi = d_W(Z_n, \delta_\delta)$, we conclude that m_W is greater than $m_{R,1}$. Hence, we change the given dose for treatment R to increase the mean effect at a new suitable value $m_{R,2} > m_{R,1}$. If $\xi = d_W(Z_n, \delta_\eta)$, we conclude that m_W is less than $m_{R,1}$, so the dose is changed in order to decrease the mean effect $m_{R,2} < m_{R,1}$. In any case, we can suppose that the difference between the two means is decreased ($|m_{R,2} - m_W| < |m_{R,1} - m_W|$). At this point, we start over with K urn processes, with the same initial composition (r_0, w_0). Although the reinforcement scheme applied is the same as before, the probability law of the reinforcements of red balls is not, because the mean is changed.

The whole study goes on until both the conditions $\xi = d_W(Z_n, Z_e)$ and $\xi < \alpha$ are satisfied. Call i_0 the number of times the random response mean to treatment R has been changed. Then m_{R,i_0} is an estimate of the unknown mean m_W. We made some simulation studies and we present here some graphics that illustrate this estimation procedure.

The simulation study was carried out with $K = 40$ urns. Parameters were fixed at $\delta = 0.3$, $\eta = 0.7$, and $\alpha = 0.05$. Responses to treatment W are assumed to be normal random variables with mean m_W and standard deviation $\sigma = 1$. Responses to treatment R are assumed to be normal random variables with mean $m_{R,i}$ and standard deviation $\sigma = 1$. As explained before, the mean is changed every time ξ is less than α. The parameter m_W was sampled by a uniform $(10, 50)$. At the beginning, the response mean to treatment R was set equal to 30 ($m_{R,1} = 30$). After changing m_R four times ($i_0 = 5$), the conditions $\xi = d(Z_n, Z_e)$ and $\xi < \alpha$ have been satisfied; this allows us to conclude that $m_W = m_{R,5}$ (see Figures 1–4). The cumulative

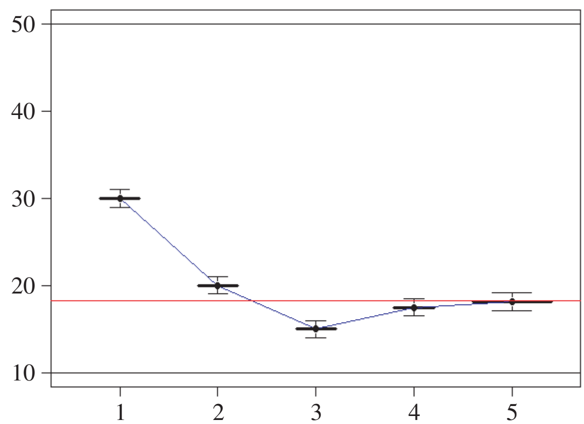

FIGURE 1: Graphic shows the different values assumed by m_R during the experiment: $(m_{R,1}, m_{R,2}, m_{R,3}, m_{R,4}, m_{R,5}) = (30, 20, 15, 17.5, 18.125)$. Five changes were necessary to reach a satisfactory estimate of the mean m_W. The x-axis represents the number of times m_R was changed, while the y-axis indicates the response means to treatments. The middle (red) line represents the unknown mean $m_W = 18.195$. The width of vertical intervals indicates the standard deviation of reinforcement distribution ($\sigma = 1$).

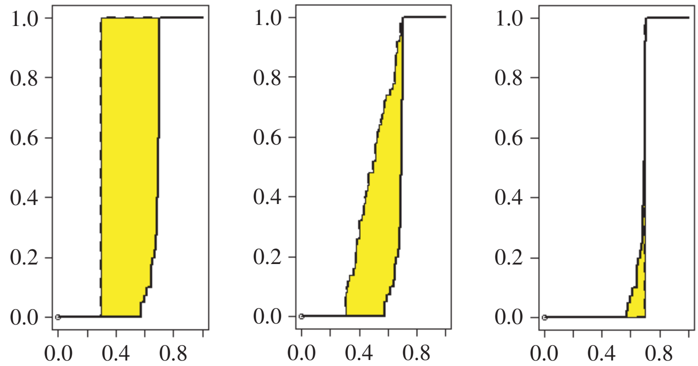

FIGURE 2: Wasserstein distances (area of shaded zones) for $d_W(Z_n, \delta_\delta)$ (*left*), $d_W(Z_n, Z_e)$ (*middle*), and $d_W(Z_n, \delta_\eta)$ (*right*) in the case where $m_{R,1} = 30$ and $m_W = 18.195$ (first iteration). Since $d_W(Z_n, \delta_\eta) < \alpha$, the limit of the process seems to be $\eta = 0.7$.

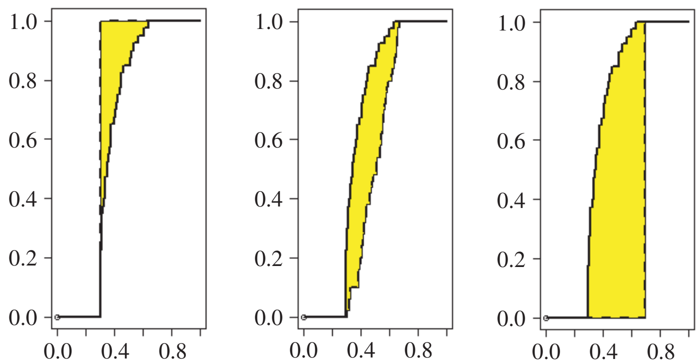

FIGURE 3: Wasserstein distances (area of shaded zones) for $d_W(Z_n, \delta_\delta)$ (*left*), $d_W(Z_n, Z_e)$ (*middle*), and $d_W(Z_n, \delta_\eta)$ (*right*) in the case where $m_{R,3} = 15$ and $m_W = 18.195$ (third iteration). Since $d_W(Z_n, \delta_\delta) < \alpha$, the limit of the process seems to be $\delta = 0.3$.

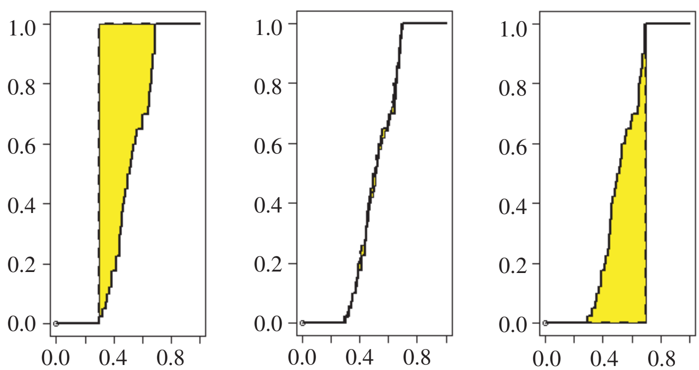

FIGURE 4: Wasserstein distances (area of shaded zones) for $d_W(Z_n, \delta_\delta)$ (*left*), $d_W(Z_n, Z_e)$ (*middle*), and $d_W(Z_n, \delta_\eta)$ (*right*) in the case where $m_{R,5} = 18.125$ and $m_W = 18.195$ (fifth iteration). Since $d_W(Z_n, Z_e) < \alpha$, the limit of the process seems to be Z_e, a random variable with no atoms.

https://doi.org/10.1239/jap/1371648956 Published online by Cambridge University Press

distribution $\hat{F}_e$ was computed with $M = 200$ urns and $m = 10^3$ draws for each one. This procedure provided an estimate of $m_W = m_{R,5} = 18.125$. In fact, the result of the started random extraction for m_W was equal to 18.195.

Acknowledgement

We thank Piercesare Secchi for stimulating discussions that gave us the impetus to prove the convergence result.

References

Aletti, G., May, C. and Secchi, P. (2009). A central limit theorem, and related results, for two-color randomly reinforced urn. *Adv. Appl. Prob.* **41,** 829–844.

Aletti, G., May, C. and Secchi, P. (2012). A functional equation whose unknown is $P([0, 1])$ valued. *J. Theoret. Prob.* **25,** 1207–1232.

Baldi Antognini, A. and Giannerini, S. (2007). Generalized Pólya urn designs with null balance. *J. Appl. Prob.* **44,** 661–669.

Durham, S. D. and Yu, K. F. (1990). Randomized play-the-leader rules for sequential sampling from two populations. *Prob. Eng. Inf. Sci.* **4,** 355–367.

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

Eisele, J. R. and Woodroofe, M. B. (1995). Central limit theorems for doubly adaptive biased coin designs. *Ann. Statist.* **23,** 234–254.

Flournoy, N., May, C. and Secchi, P. (2012). Asymptotically optimal response-adaptive designs for allocating the best treatment: an overview. *Internat. Statist. Rev.* **80,** 293–305.

Hill, B. M., Lane, D. and Sudderth, W. (1980). A strong law for some generalized urn processes. *Ann. Prob.* **8,** 214–226.

Hu, F. and Zhang, L.-X. (2004). Asymptotic proprieties of doubly adaptive biased coin designs for multitreatment clinical trials. *Ann. Statist.* **32,** 268–301.

Hu, F., Zhang, L.-X. and He, X. (2009). Efficient randomized-adaptive designs. *Ann. Statist.* **37,** 2543–2560.

May, C. and Flournoy, N. (2009). Asymptotics in response-adaptive designs generated by a two-color, randomly reinforced urn. *Ann. Statist.* **37,** 1058–1078.

Melfi, V. F., Page, C. and Geraldes, M. (2001). An adaptive randomized design with application to estimation. *Canad. J. Statist.* **29,** 107–116.

Muliere, P., Paganoni, A. M. and Secchi, P. (2006). A randomly reinforced urn. *J. Statist. Planning Infer.* **136,** 1853–1874.

Rosenberger, W. F. (2002). Randomized urn models and sequential design. *Sequent. Anal.* **21,** 1–41.

Zhang, Li-X., Hu, F. and Cheung, S. H. (2006). Asymptotic theorems of sequential estimation-adjusted urn models. *Ann. Appl. Prob.* **16,** 340–369.

Note added in proof

A simplified proof of Theorem 3.1 has recently been formulated and is available from the authors upon request.