MOX–Report No. 32/2011

A modified randomly reinforced urn design

Aletti, G; Ghiglietti, A; Paganoni, A.

MOX, Dipartimento di Matematica "F. Brioschi"
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox@mate.polimi.it http://mox.polimi.it

A modified randomly reinforced urn design

Giacomo Aletti[1], Andrea Ghiglietti[2], Anna Maria Paganoni[2]

July 19, 2011

[1] Dipartimento di Matematica "F. Enriques"
Università degli studi di Milano
via Saldini 50, 20133 Milano, Italy
giacomo.aletti@unimi.it

[2] MOX - Modellistica e Calcolo Scientifico
Dipartimento di Matematica "F. Brioschi"
Politecnico di Milano
via Bonardi 9, 20133 Milano, Italy
andrea.ghiglietti@mail.polimi.it
anna.paganoni@polimi.it

Abstract

We want to construct a response adaptive design, described in terms of two colors 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. Applications to sequential clinical trials are considered as well as connections with response-adaptive design of experiments.

Key words: Reinforced processes, urn schemes, sequential clinical trials, stochastic processes

1 Introduction

Consider a clinical trial with two competitive treatments, say R and W. We want to construct a response adaptive design, described in term of urn model, targeting any optimal fixed asymptotic allocation, in order to compare these models with the other ones studied in literature. A large class of response-adaptive randomized designs is based on urn models, a classical tool to guarantee a randomized device (Rosenberger 2002, Zhang, Hu and Cheung 2006), to construct designs targeting the best treatment (Muliere, Paganoni and Secchi 2006, May and Flournoy 2009) or to balance the allocations (Baldi Antognini and Giannerini 2007). The *two-color, Randomly Reinforced Urn* (RRU) introduced in Muliere, Paganoni and Secchi (2006) and studied in Aletti, May and Secchi (2009a, 2009b) is a randomized device able to target the optimal treatment, see Muliere, Paganoni and Secchi (2006). In this paper we modify the reinforcement scheme of the urn to asymptotically target an optimal allocation proportion. 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 variable on $(0, 1)$. We will interpret μ_R and μ_W as the laws of the responses to treatment R and W respectively. We assume that both the means $m_R = \int_\alpha^\beta x\mu_R(dx)$ and $m_W = \int_\alpha^\beta x\mu_W(dx)$ are strictly positive. Visualize an urn initially containing r_0 balls of color R and w_0 balls of color W. Set

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

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 distribution 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

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

Now iterate this sampling scheme forever. Thus, at time $n + 1$, given the sigma-field $\mathcal{F}_n$ generated by $X_1, ..., X_n, M_1, ..., M_n$ and $N_1, ..., N_n$, 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 distribution μ_R and μ_W respectively. Set

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

We thus generate an infinite sequence $X = (X_n, n = 1, 2, ..)$ of Bernoulli random variables, with X_n representing the color of the ball sampled from the urn at time n, and a process $(Z, D) = ((Z_n, D_n), n = 0, 1, 2...)$ 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 $\mathcal{F}_n$.

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 and finally Section 4 contains a simulation study on application of urn design to sequential clinical trials.

2 Upcrossing and reinforcements

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

$$\begin{aligned} \tau_j &= \begin{cases} \inf\{n > t_{j-1} : Z_n < d\} & \text{if } \{n > t_{j-1} : Z_n < d\} \neq \emptyset; \\ +\infty & \text{otherwise.} \end{cases} \\ t_j &= \begin{cases} \inf\{n > \tau_j : Z_n > u\} & \text{if } \{n > \tau_j : Z_n > u\} \neq \emptyset; \\ +\infty & \text{otherwise.} \end{cases} \end{aligned} \tag{2.1}$$

We call *j-th excursion* the random interval $(\tau_{j-1}, \tau_j]$, and we denote by

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

i.e., $\nu^Z_{[u,d]}$ counts the total number that the process Z crosses the strip $[u, d]$.

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

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

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

Proof. We first note that

$$\begin{aligned} (Z_n)_n \text{ converges a.s.} &\overset{\forall 0<d<u<1}{\iff} P(\nu^Z_{[u,d]} = \infty) = 0 \\ &\overset{\forall 0<d<u<1}{\iff} 0 = \lim_{n\to\infty} P(\nu^Z_{[u,d]} \geq n) \\ &\qquad\qquad = \lim_{n\to\infty} P(\cap_{j=0}^n \{\tau_j < \infty\}) \end{aligned}$$

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

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

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

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

The fact that some $(p_n)_n$ might be zero is controlled by the assumption that $p_n = 0 \Rightarrow p_m = 0$, $\forall m > n$. $\square$

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

Definition 2.2 (General Urn Process)**.** Let $(\Omega, \mathcal{F}, (\mathcal{F}_n)_n, P)$ be a filtered space. A vector process $(R_n, W_n)_n$ on $(\Omega, \mathcal{F}, (\mathcal{F}_n)_n, P)$ is called *General Urn Process (GUP)* 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., $0 \le R_0 \le R_1 \le \dots R_n \le \dots$ and $0 \le W_0 \le W_1 \le \dots W_n \le \dots$) and $R_0 + W_0 > 0$. Let $D_n = R_n + W_n$, for $n \in \mathbb{N}$.

Lemma 2.3 (Reinforcements during excursions)**.** *For any GUP,*

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

Proof. For every $j \in \mathrm{N}_0$ we have that

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

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

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

From this we have immediately the following result

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

$\square$

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

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

A GUP $(R_n, W_n)_n$ 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

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

Finally, note that, given a GUP $(R_n, W_n)_n$, 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 2.4. *Given a Markov GUP, the process* $(Z_n)_n$ *converges a.s. if, for any* $0 < d < u < 1$*, there exists a function* $g : [0, \infty) \times [0, \infty) \to [0, 1]$*,* $(R_n, W_n)_n$ *is associated to the sequence* $(\tau_n)_n$ *defined in* (2.1)*, and*

$$\begin{aligned} & f(x, y, \cdot) \leq g(x', y') && \text{if } x + y \geq x' + y' \\ & g(c_1, c_2) < 1 && \text{for some } c_1, c_2 > 0 \end{aligned}$$

where f *is given in* (2.2)*.*

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

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

then, by Lemma 2.3,

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

Thesis follows by Theorem 2.1. □

3 Convergence Theorem

Let us consider the urn model described in Section 1. We have

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

$$lim_{n\to\infty} Z_n = \begin{cases} \eta & \text{if } \int_\alpha^\beta x\mu_R(dx) > \int_\alpha^\beta x\mu_W(dx), \\ \delta & \text{if } \int_\alpha^\beta x\mu_R(dx) < \int_\alpha^\beta x\mu_W(dx). \end{cases}$$

To get this task, 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 $m_R = \int_\alpha^\beta x\mu_R(dx)$ and $m_W = \int_\alpha^\beta x\mu_W(dx)$, the means of the patients responses to treatments.

Lemma 3.2 (Aletti, May and Secchi (2009a), Lemma A.2,A.3)**.** *Assume* $m_R = m_W = m$*. If* $D_0 \geq 2\beta$*, then*

$$\begin{aligned} & E(\sup_n |A_n|) \leq \frac{\beta}{D_0}; \\ & E(\langle M\rangle_\infty - \langle M\rangle_n | \mathcal{F}_n) \leq \frac{\beta}{D_0}, && \text{for any } n \geq 0. \end{aligned}$$

As a consequence, we get

Lemma 3.3. *Assume $m_R = m_W = m$. If $D_0 \geq 2\beta$, then*

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

for every $h > 0$.

Proof. First note that, since $(M_n)_n$ is a martingale null at $n = 0$, we have, by Lemma 3.2 (choosing $n = 0$ in the second inequality) that

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

and hence, by Doob's L^2–inequality,

$$P(\{\sup_n |M_n| \geq h/2\}) \leq \lim_{n\to\infty} \frac{E(M_n^2)}{(h/2)^2} \leq \frac{4\beta}{h^2 D_0}$$

for any $h > 0$. We easily get

$$\begin{aligned} P(\sup_n |Z_n - Z_0| \geq h) &\leq P(\{\sup_n |M_n| \geq h/2\} \cup \{\sup_n |A_n| \geq h/2\}) \\ &\leq P(\{\sup_n |M_n| \geq h/2\}) + P(\{\sup_n |A_n| \geq h/2\}) \\ &\leq \frac{\beta}{D_0}\Big(\frac{4}{h^2} + \frac{2}{h}\Big) \end{aligned}$$

□

Proof of Theorem 3.1. We have an urn containing at the starting time R_0 red balls and W_0 white balls. Let us consider the case $m_R < m_W$; the opposite case ($m_R > m_W$) is completely analogous. In the case described in Muliere, Paganoni and Secchi (2006) the process $(Z_n)_{n\in\mathrm{N}}$ is a super-martingale converging to zero but, because of the barrier δ (see (1.2)), it's not like this anymore. Anyway, we want to prove that the process $(Z_n)_{n\in\mathrm{N}}$ still converges, but in this case the limit is equal to δ.

First of all, we will prove that

$$\liminf Z_n \leq \delta, \text{a.s}$$

By contradiction, there exists $l > \delta$ such that $P(\liminf Z_n \geq l) > 0$. Then, there exists n_0 such that $P(Z_n > \frac{l+\delta}{2}, \forall n \geq n_0) > 0$. This contradicts the fact that, by Markov property, $P(Z_n > \frac{l+\delta}{2}, \text{ev.}) = 0$, since it is a RRU with reinforcement with different means that goes to 0 (see Muliere, Paganoni and Secchi, 2006).

With the same argument, one may prove that $\limsup Z_n \geq \delta$, since the urn that eventually stays below δ is a RRU with reinforcement with different means that goes to 1 (again, see Muliere, Paganoni and Secchi, 2006).

In fact, one can prove more, with the arguments of Muliere, Paganoni and Secchi (2006): the barrier δ must be crossed infinitely times almost surely. With this result in mind, we will prove in a moment that $\liminf Z_n \geq \delta$. In fact, if there exists $l < \delta$ such that $P(\liminf Z_n \leq l) > 0$, then with positive probability the process must cross the strip $(\frac{l+\delta}{2}, \delta)$ infinite times. By Lemma 2.3, after a sufficiently large number of times, $D_n > \beta\frac{l+\delta}{\delta-l}$ and therefore, if k is any successive downcross of δ,

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

since each reinforced is bounded by β and $\frac{R_{k-1}}{D_{k-1}} = Z_{k-1} > \delta$.

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 (2.1), in order to apply Proposition 2.4. Let $i > \log_{\frac{u(1-d)}{d(1-u)}} \frac{\beta(1-d)}{D_{\tau_0}(d-\delta)}$ be fixed, so that, by Lemma 2.3, $D_{\tau_i} > \frac{\beta(1-d)}{d-\delta}$, and denote by $(\widehat{\cdot}_n)_{n\in\mathrm{N}}$ the renewed process on $\{\tau_i < \infty\}$: $(\widehat{R}_n, \widehat{W}_n) = (R_{\tau_i+n}, W_{\tau_i+n})$, $\widehat{D}_n = \widehat{R}_n + \widehat{W}_n = D_{\tau_i+n}$, $\widehat{Z}_n = \widehat{R}_n/\widehat{D}_n = Z_{\tau_i+n}$, $\widehat{U}_n = U_{\tau_i+n}$. 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}) and dynamic as in (1.1) and (1.2). We denote by $P_i(\cdot) = P(\cdot|\tau_i < \infty)$, and therefore, if

$$t = \begin{cases} \inf\{n : \widehat{Z}_n > u\} & \text{if } \{n : \widehat{Z}_n > u\} \neq \emptyset; \\ +\infty & \text{otherwise} \end{cases}$$

then we have

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

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

$$\begin{aligned} \tau_j^* &= \begin{cases} \inf\{n > t_{j-1}^* : \widehat{Z}_n \leq \delta\} & \text{if } \{n > t_{j-1}^* : \widehat{Z}_n \leq \delta\} \neq \emptyset; \\ +\infty & \text{otherwise.} \end{cases} \\ t_j^* &= \begin{cases} \inf\{n > \tau_j^* : \widehat{Z}_n > \delta\} & \text{if } \{n > \tau_j^* : \widehat{Z}_n > \delta\} \neq \emptyset; \\ +\infty & \text{otherwise.} \end{cases} \end{aligned} \tag{3.2}$$

Note that,

$$\begin{matrix} \dfrac{R}{R+W} \leq \delta, \\ (R+W) > \dfrac{\beta(1-d)}{d-\delta} \end{matrix} \qquad \Longrightarrow \qquad \frac{R+x}{R+W+x} < d, \forall x \leq \beta,$$

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

$$\left.\begin{aligned} &\widehat{Z}_{t_j^*-1} \le \delta, \\ &\widehat{D}_{t_j^*-1} > \frac{\beta(1-d)}{d-\delta} \end{aligned}\right\} \implies \widehat{Z}_{t_j^*} < d \implies \widehat{R}_{t_j^*} < \widehat{W}_{t_j^*-1}\frac{d}{1-d}. \tag{3.3}$$

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

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

The new urn process is a urn process that starts with $\widetilde{Z}_0 = \frac{u+d}{2}$, it is reinforced at time $n+1$ only when $\widehat{Z}_n > \delta$ with nonnegative reinforcements that have the same mean m_R and it is rebuilt at time $n+1$ only when $\widehat{Z}_n \le \delta$.

We will prove by induction that, for any n,

$$\widetilde{Z}_n > \widehat{Z}_n, \qquad \widetilde{W}_n \le \widehat{W}_n, \qquad \widetilde{R}_n > \widehat{R}_n \tag{3.4}$$

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

In fact, by construction we have that

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

which immediately implies $\widetilde{R}_0 > \widehat{R}_0$. Assume (3.4) by induction hypothesis. . We divide the two cases:

$[\widehat{Z}_n \le \delta]$: $\widetilde{W}_{n+1} = \widehat{W}_{n+1}$ by construction. By (3.3), $\widehat{Z}_{n+1} < d < \widetilde{Z}_n = \widetilde{Z}_{n+1}$ and hence $\widetilde{R}_{n+1} > \widehat{R}_{n+1}$;

$[\widehat{Z}_n > \delta]$: Since $\widetilde{X}_{n+1} = \mathbf{1}_{[0,\widetilde{Z}_n]} \ge \mathbf{1}_{[0,\widehat{Z}_n]} = \widehat{X}_{n+1}$ by construction, we get

$$\begin{aligned} \widehat{R}_{n+1} - \widehat{R}_n = \widehat{X}_{n+1}\widehat{M}_{n+1} \le \widetilde{X}_{n+1}\widetilde{M}_{n+1} = \widetilde{R}_{n+1} - \widetilde{R}_n, \\ \widehat{W}_{n+1} - \widehat{W}_n = (1-\widehat{X}_{n+1})\widehat{N}_{n+1} \ge (1-\widetilde{X}_{n+1})\widetilde{N}_{n+1} = \widetilde{W}_{n+1} - \widetilde{W}_n. \end{aligned}$$

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

$$T_m = \begin{cases} \inf\{n \leq \tau^*_m - t^*_{m-1} : \widetilde{Z}_{t^*_{m-1}+n} \leq d \text{ or } \widetilde{Z}_{t^*_{m-1}+n} \geq u\} \\ \qquad \text{if } \{n \leq \tau^*_m - t^*_{m-1} : \widetilde{Z}_{t^*_{m-1}+n} \leq d \text{ or } \widetilde{Z}_{t^*_{m-1}+n} \geq u\} \neq \emptyset; \\ +\infty \qquad \text{otherwise,} \end{cases}$$

then we have stated that

$$P_i(T_m < \infty) \geq P_i(t < \infty | \{t^*_{m-1} < t < \tau^*_m\}). \tag{3.5}$$

Now, as a consequence of Lemma 3.3 and the fact that $\widetilde{D}_{t^*_{m-1}} \geq \widetilde{D}_0 \geq D_{\tau_i}$, if we set $h = \frac{u-d}{2}$, we get

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

Thus define the function $g : [0, \infty) \times [0, \infty) \to [0, 1]$ in the following way

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

and note that $g(8\beta/h^2, 4\beta/h) = 1/2$ and g is monotone in $x + y$. We can apply Proposition 2.4 to get the thesis, since, by (3.1) and (3.5)

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

□

Remark 3.4. Note that in the proof it was never necessary to know which is the type of distribution that generates the reinforcements. Indeed, we don't need all information about the probability laws, but 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 fixed the sign of the difference of the means.

Remark 3.5. Consider the particular case in which the reinforcements are independent Bernoulli variables, with parameters π_R for the red balls and π_W for the white balls. In this situation, our model is equivalent to the one studied by Hill, Lane and Sudderth (1980), in which the urn function f can be defined like follows:

$$f(x) = \frac{x\pi_R 1_{\{x<\eta\}}}{x\pi_R 1_{\{x<\eta\}} + (1-x)\pi_W 1_{\{x>\delta\}}}$$

Looking at the expression above and applying the Theorem 4.1. of Hill, Lane and Sudderth (1980), we can reach to the same result of convergence proved in this paper.

Now let us consider the target allocation $\rho(m_R, m_W) = \eta\mathbf{1}_{[m_R>m_W]} + \delta\mathbf{1}_{[m_R<m_W]}$, we have shown that for the reinforcement scheme introduced here Z_n converges almost surely to ρ. By using the same martingale argument of Melfi, Page and Geraldes (2001) we can prove that $N_R(n)/n \to \rho$ almost surely. This results allows us to force the design to be asymptotic 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 infinity almost surely as long as n increases to infinity. This allows us to define the following adaptive consistence estimators based on the observed responses until time n, with random sample sizes $N_R(n)$ and $N_W(n)$ respectively:

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

We can apply the results proved in Melfi, Page and Geraldes (2001) to state the following

Proposition 3.6. *The estimators $\overline{M}(n)$ and $\overline{N}(n)$ are consistent estimators of m_R and m_W, respectively. Moreover as $n \to \infty$,*

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

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

4 A simulation study

In this section we present a possible application which concerns the convergence theorem proved in this paper. Let us consider to have an unknown treatment W, whose we want to know the mean effect on patients. In statistical terms, this means we are interested in finding out which is the mean of the patients responses distribution to this treatment. Then, we introduce a well-known treatment R, having the propriety that its mean effect on patients can be chosen arbitrarily. The aim of the experiment is to infer the mean effect of the treatment W by modifying suitably the mean effect of treatment R.
Let us consider K urns with the same initial composition (r_0, w_0). Red balls are associated with treatment R, while white balls with treatment W. The model applied for each urn process is the one described in

Section 1. We will denote with $Z^j = (Z^j_n)_{n\in\mathbb{N}}$ the process of the urn proportion in the j^{th} urn, for $j \in \{1, 2, .., K\}$. For every urn the convergence Theorem 3.1 tells us that

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

When $m_R = m_W$ we don't have the explicit form of the asymptotic distribution of the urn proportion Z_n. Nevertheless, we know it converges to a random variable Z_∞ whose distribution has no atoms and its support $S_\infty = [\delta, \eta]$.
The idea is to use this theorem in order to find out which is the unknown mean effect of treatment W. At the beginning, we set the mean effect of treatment R to a standard value $m_{R,0}$. As say, in each urn the reinforcements of red and white balls will follow distributions with different means, respectively $m_{R,0}$ and m_W. Then, we let start the K urn processes simultaneously. At each step we will have an array composed by K urn proportions which we can use to compute the empirical cumulative distribution function $\widehat{F}_n$ for the random variable Z_n. Thanks to the Theorem 3.1, for every $x \in [0, 1]$, $\widehat{F}_n(x)$ must converge to

$$\begin{cases} F_\eta(x) = 1_{\{x\geq\eta\}} & \text{if } m_{R,0} > m_W, \\ \\ F_\delta(x) = 1_{\{x\geq\delta\}} & \text{if } m_{R,0} < m_W. \end{cases}$$

If $m_{R,0} = m_W$ we assume that both the reinforcements have the same distributions. Since we know the distribution of the reinforcement of red balls and the exact value $m_{R,0}$, we have all the information to compute offline the asymptotic cumulative distribution $\widehat{F}_e$,

$$\frac{1}{K}\sum_{j=1}^{K} 1_{\{Z^j_n < x\}} \simeq \widehat{F}_e(x), \qquad \text{for large n.}$$

We can say this approach works because we expect that the asymptotic distribution of urn process do not depend so much on the particular kind of reinforcement, whenever they present the same mean effect.

At each step, once every urn has been reinforced, we use the normalized Wasserstein distance (d_W) in order to compute the distance between Z_n with the three asymptotic possible distributions. Then, we take the minimum among these distances and if it is lower than a suitable quantity α we can assume the proportion Z_n has reached its limit. Otherwise, the urn processes go on with the next draws. We stop the algorithm at step $\widetilde{n}$ if

$$\min\{d_W(Z_{\widetilde{n}}, \delta_\eta)\ , d_W(Z_{\widetilde{n}}, Z_\infty)\ , d_W(Z_{\widetilde{n}}, \delta_\delta)\} =$$

$$\min\left\{\int_0^1 |F_{\widetilde{n}}(x) - F_\eta(x)|dx\ ,\ \int_0^1 |F_{\widetilde{n}}(x) - \widehat{F}_e(x)|dx\ ,\ \int_0^1 |F_{\widetilde{n}}(x) - F_\delta(x)|dx\right\} < \alpha$$

where α is an arbitrarily parameter, fixed in advance.

When the stopping rule is verified, different scenarios are possible. If the minimum distance is $d(Z_{\widetilde{n}}, \delta_\eta)$ we can assume $m_{R,0}$ was greater than the unknown mean m_W. For this reason, we change the composition of treatment R to increase the mean effect at another value $m_{R,1} > m_{R,0}$. Alternatively, if the lowest distance was $d(Z_{\widetilde{n}}, \delta_\delta)$ we decrease the mean effect of treatment R in order to have a new mean $m_{R,1} < m_{R,0}$. In any case, we can suppose the difference between the two means is decreased. At this point we can start over with K urns with the same initial composition (r_0, w_0). The model is the same as before with the only difference that the mean of the reinforcement of red balls has been updated.

The whole experiment go on until the stopping rule is verified by $d(Z_{\widetilde{n}}, Z_\infty) < \alpha$. If we define i_0 as the number of times we have update the mean m_R, we can suppose that $m_{R,i_0} = m_W$ and this is a good estimate of the unknown mean m_W. We made some simulation studies and we report here some graphics that illustrate the estimation procedure.

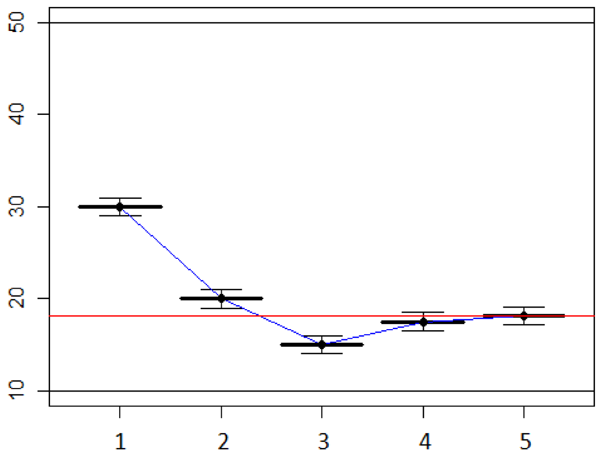

Figure 1: Graphic shows the updating process of m_R :(30,20,15,17.5,18.125). Five iterations were necessary. The red line represents the unknown mean $m_W = 18.195$. The width of vertical intervals indicates the standard deviation of reinforcement distribution ($\sigma = 1$).

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

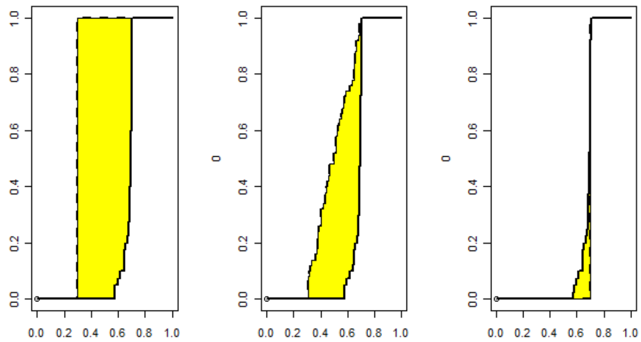

Figure 2: Normalized Wasserstein distances (yellow zone) for $d(Z_{\widetilde{n}}, \delta_\delta)$ (left panel), $d(Z_{\widetilde{n}}, Z_\infty)$ (central panel) and $d(Z_{\widetilde{n}}, \delta_\eta)$ (right panel) in the case of $m_R = 30$ and $m_W = 18.195$ (first iteration). The limit of the process seems to be $\eta = 0.7$.

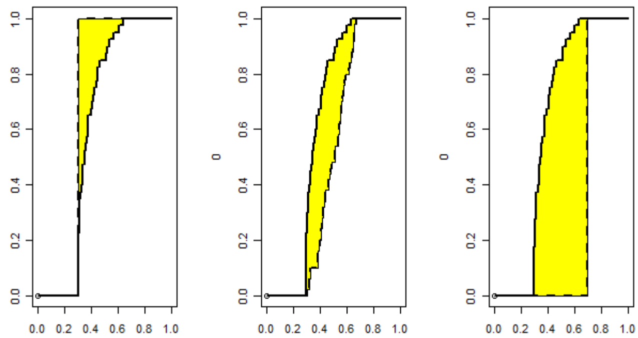

Figure 3: Normalized Wasserstein distances (yellow zone) for $d(Z_{\widetilde{n}}, \delta_\delta)$ (left panel), $d(Z_{\widetilde{n}}, Z_\infty)$ (central panel) and $d(Z_{\widetilde{n}}, \delta_\eta)$ (right panel) in the case of $m_R = 15$ and $m_W = 18.195$ (third iteration). The limit of the process seems to be $\delta = 0.3$.

References

[1] Aletti, G., May, C. and Secchi, P. (2009a). A functional equation whose unknown is $P([0;1])$ valued. *Mox Report n. 16/2009*, Dipartimento di Matematica, Politecnico di Milano. [Online] http://mox.polimi.it/it/progetti/pubblicazioni/quaderni/16-2009.pdf

[2] Aletti, G., May, C. and Secchi, P. (2009b). A Central Limit Theorem, and related results, for two-color randomly reinforced urn. *Advances in Applied Probability*, 41, 829–844.

[3] Baldi Antognini, A. and Giannerini, S. (2007). Generalized Polya urn designs with null balance. *Journal of Applied Probability*, 44, 661–669.

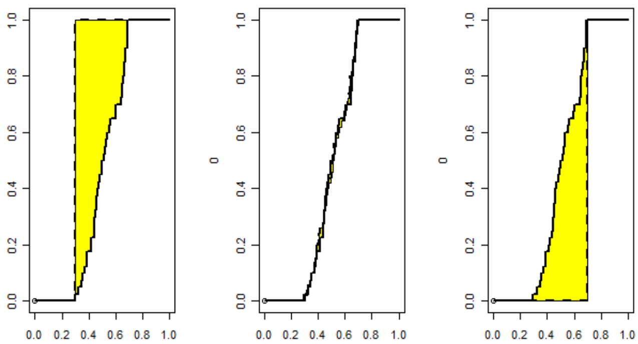

Figure 4: Normalized Wasserstein distances (yellow zone) for $d(Z_{\widetilde{n}}, \delta_\delta)$ (left panel), $d(Z_{\widetilde{n}}, Z_\infty)$ (central panel) and $d(Z_{\widetilde{n}}, \delta_\eta)$ (right panel) in the case of $m_R = 18.125$ and $m_W = 18.195$ (fifth iteration). The limit of the process seems to be Z_∞, a random variable with no atoms.

[4] Hill, B.M., Lane, D. and Sudderth, W. (1980). A Strong Law for Some Generalized Urn Processes. *The Annals of probability*, 8, 214-226.

[5] May, C. and Flournoy, N.(2009). Asymptotics in response-adaptive designs generated by a two-color, randomly reinforced urn. *The Annals of Statistics*, 37, 1058–1078.

[6] Melfi, F., Page, C. and Geraldes, M. (2001). An adaptive randomized design with application to estimation. *The Canadian Journal of Statistics*, 29, 107–116.

[7] Muliere, P., Paganoni, A.M. and Secchi, P. (2006). A randomly reinforced urn. *Journal of Statistical Planning and Inference*, 136, 1853–1874.

[8] Rosenberger, W. F. (2002). Randomized urn models and sequential design, *Sequential Analysis*, 21(1&2), 1–28.

[9] Zhang, Li-X., Hu, F. and Cheung, S.H. (2006) Asymptotic theorems of sequential estimation-adjusted urn models. *The Annals of Applied Probability*, 16, 340–369.

MOX Technical Reports, last issues

Dipartimento di Matematica "F. Brioschi",
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

33/2011 Antonietti, P.F.; Beirao da Veiga, L.; Lovadina, C.; Verani, M.
Hierarchical a posteriori error estimators for the mimetic discretization of elliptic problems

32/2011 Aletti, G; Ghiglietti, A; Paganoni, A.
A modified randomly reinforced urn design

34/2011 Benacchio, T.; Bonaventura, L.
A spectral collocation method for the one dimensional shallow water equations on semi-infinite domains

31/2011 Astorino, M.; Becerra Sagredo, J.; Quarteroni, A.
A modular lattice Boltzmann solver for GPU computing processors

30/2011 Nobile, F.; Pozzoli, M.; Vergara, C.
Time accurate partitioned algorithms for the solution of fluid-structure interaction problems in haemodynamics

29/2011 Morin, P.; Nochetto, R.H.; Pauletti, S.; Verani, M.
AFEM for Shape Optimization

28/2011 Pischiutta, M.; Formaggia, L.; Nobile, F.
Mathematical modelling for the evolution of aeolian dunes formed by a mixture of sands: entrainment-deposition formulation

27/2011 Antonietti, P.F.; Bigoni, N.; Verani, M.
A Mimetic Discretization of Elliptic Control Problems

26/2011 Secchi, P.; Vantini, S.; Vitelli, V.
Bagging Voronoi classifiers for clustering spatial functional data

25/2011 de Luca, M.; Ambrosi, D.; Robertson, A.M.; Veneziani, A.; Quarteroni, A.
Finite element analysis for a multi-mechanism damage model of cerebral arterial tissue