A TWO-COLOR, RANDOMLY REINFORCED URN

PIETRO MULIERE, ANNA MARIA PAGANONI, AND PIERCESARE SECCHI

ABSTRACT. We study an urn containing balls of two or more colors. The urn is sequentially sampled. Each time a ball is extracted from the urn it is reintroduced in it together with a random number of balls of the same color: the distribution of the the number of added balls may depend on the color extracted. 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 adaptive design of experiments in a Bayesian framework.

1. INTRODUCTION: REINFORCED BERNOULLI PROCESSES

The law of an infinite sequence $X = (X_n, n = 1, 2, ...)$ of Bernoulli random variables is said to be *reinforced* (Walker and Muliere,2004) if

$$\begin{aligned} P(X_2 = 1 | X_1 = 1) &\geq P(X_1 = 1), \\ P(X_2 = 0 | X_1 = 0) &\geq P(X_1 = 0) \end{aligned} \tag{1.1}$$

and, for every $n \geq 1$ and $x_1, ..., x_n \in \{0, 1\}$,

$$\begin{aligned} P(X_{n+2} = 1 | X_1 = x_1, ..., X_n = x_n, X_{n+1} = 1) \\ \geq P(X_{n+1} = 1 | X_1 = x_1, ..., X_n = x_n). \end{aligned} \tag{1.2}$$

Initial conditions (1.1) together with condition (1.2) imply that

$$\begin{aligned} P(X_{n+2} = 0 | X_1 = x_1, ..., X_n = x_n, X_{n+1} = 0) \\ \geq P(X_{n+1} = 0 | X_1 = x_1, ..., X_n = x_n). \end{aligned}$$

for every $n \geq 1$ and $x_1, ..., x_n \in \{0, 1\}$.

The law of any infinite sequence of exchangeable Bernoulli random variables is reinforced; this follows from an application of de Finetti's Representation Theorem (Walker and Muliere, 2004). However, an infinite sequence of Bernoulli random variables with reinforced law need not be exchangeable.

Date: May 19, 2004.

1991 *Mathematics Subject Classification.* 60F15,62L05.

Key words and phrases. Reinforced processes, asymptotic exchangeability, urn schemes, gambling theory, two-arm bandit problems, sequential clinical trials.

Consider, for instance, a *generalized urn process* $X = (X_n, n = 1, 2, ...)$ of Hill, Lane and Sudderth (1980) with initial urn composition (b_0, w_0) and urn function f. The function f maps the unit interval to itself while b_0 and w_0 are two nonnegative real numbers with strictly positive sum. The law of X is defined by assuming that X_1 is a Bernoulli($f(z_0)$)) random variable with $z_0 = b_0/(b_0+w_0)$; for $n \geq 1$, the conditional distribution of X_{n+1} given $X_1, ..., X_n$ is a Bernoulli($f(Z_n)$) with

$$Z_n = \frac{b_0 + \sum_{i=1}^n X_i}{b_0 + w_0 + n}.$$

If f is nondecreasing, the law of $X = (X_n, n = 1, 2, ...)$ is reinforced. A celebrated special case is obtained when $f(x) = x$ for $x \in [0, 1]$: then X is a Pólya sequence and its law is exchangeable. However, X is not exchangeable if f is different from the identity, from a constant function or from a function of the type

$$f(x) = \begin{cases} 0, & x \in [0, \frac{b_0}{b_0+w_0}), \\ p, & x = \frac{b_0}{b_0+w_0}, \\ 1, & x \in (\frac{b_0}{b_0+w_0}, 1], \end{cases}$$

with $p \in (0, 1)$ (Hill, Lane and Sudderth, 1987).

An interesting urn scheme generating an infinite sequence of Bernoulli random variables whose law is reinforced and that is not exchangeable nor a generalized urn process, has been studied by Li, Durham and Flournoy (1996) and by Durham, Flournoy and Li (1998) under the name of *randomized Pólya urn*: it has applications in randomized sequential clinical trials. Let $\pi_0, \pi_1 \in [0, 1]$ be the probabilities of success for two treatments, say treatment 0 and treatment 1, respectively. An urn initially contains $b_0 > 0$ balls of color 1 and $w_0 > 0$ balls of color 0: when patient $n = 1, 2, ...$ needs to be assigned to a treatment, a ball is sampled from the urn and the patient is assigned to treatment 0 or treatment 1 according to the color of the sampled ball. Next, if the treatment is successful, the sampled ball is reintroduced in the urn with another of the same color, otherwise it is reintroduced in the urn alone; with this updated composition, the urn is ready for the assignment of patient $n+1$. Let $X = (X_n, n = 1, 2, ...)$ be the process of colors generated by this urn scheme: the law of X is reinforced. Durham, Flournoy and Li prove that, if $\pi_1 \neq \pi_0$, as n grows to infinity the composition of a randomized Pólya urn concentrates on the color corresponding to the treatment with highest probability of success. Hence, if $\pi_1 > \pi_0$, the conditional probability of assigning the $(n+1)$-th patient to treatment

1, given the treatment successes and failures observed on the previous n patients, converges almost surely to 1. This makes the model attractive for adaptive randomized sequential clinical trials.

The convergence result of Durham, Flournoy and Li is proved by embedding the sequence X in a continuous-time Yule process. We believe that, in the case of two treatments, a direct proof can be obtained by considering a generalized urn process Y embedded in X. In fact, let $T_1, T_2, ...$ indicate patients for which a successful treatment is observed: these are also the times when the urn composition changes. For every n, T_n is finite almost surely and we may define the process $Y = (Y_n, n = 1, 2, ...)$ such that $Y_n = X_{T_n}$ for $n = 1, 2, ...$ The sequence Y is a generalized urn process with initial urn composition (b_0, w_0) and urn function

$$f(x) = \frac{x\pi_1}{x\pi_1 + (1-x)\pi_0}, \quad x \in [0, 1].$$

Note that $f(0) = 0 = 1 - f(1)$ and f is continuous, strictly increasing and concave if $\pi_1 > \pi_0$; hence, in the terminology of Hill, Lane and Sudderth (1980), 1 is the unique downcrossing point for f and it follows from their Theorem 6.1 that the proportion of balls of color 1 in the urn generating the process Y converges to 1 almost surely. This must then also be true for the proportion of balls of color 1 in the urn generating the process X.

This argument convinced us to investigate whether the ideas driving Hill, Lane and Sudderth 1980's paper could be of use for studying the reinforced Bernoulli process generated by a *two-color, randomly reinforced urn*. Once again, visualize an urn initially containing b_0 balls of color 1 and w_0 balls of color 0. The urn is sequentially sampled. Let μ and ν be two probability distributions with bounded support contained in $[0, \infty)$; each time a ball of color 1 is sampled from the urn it is returned to the urn together with a random number of balls of color 1 having distribution μ; each time a ball of color 0 is sampled from the urn it is returned to the urn together with a random number of balls of color 0 having distribution ν. A more detailed definition is given in the next section before proving that, if μ is stochastically larger than ν, the sequence of proportions of balls of color 1 converges to a limit Z_∞. The process of colors generated by the urn is thus asymptotically exchangeable, i.e. given Z_∞ the colors generated by the urn are asymptotically conditionally i.i.d. with distribution Bernoulli(Z_∞). In section 4 we show that if, in addition, the supports of μ and ν are bounded away from 0 and the first moment of μ is greater than the first moment of ν, then Z_∞ is equal to 1 almost surely. The argument for this result follows closely that in Hill, Lane and Sudderth (1980) for studying convergence of generalized urn processes; like them, in section 3 we introduce a gambling problem connected with our urn, to be used as an auxiliary tool for proving properties of the distribution of Z_∞. A two-color, randomly reinforced urn implements a specific adaptive, randomized, sequential design for clinical trials on two treatments with a favorable response, like for instance survival time after treatment; a Bayesian perspective on the problem will be illustrated in Section 5 where connections with two-armed bandit problems will also be examined. Finally, a generalization to a $(k+1)$-color, randomly reinforced urn is considered in the last section of the paper.

2. Model specification and first asymptotic results

We are going to define a *two-color, randomly reinforced urn* with initial composition (b_0, w_0) and directed by two probability distributions μ and ν; b_0 and w_0 are two non negative real numbers with strictly positive sum, while μ and ν are two probability distributions with support contained in $[\alpha, \omega]$, where $0 \leq \alpha \leq \omega < \infty$. In fact, visualize an urn initially containing b_0 balls of color 1 and w_0 balls of color 0. Set

$$B_0 = b_0,\ W_0 = w_0,\ D_0 = B_0 + W_0,\ Z_0 = \frac{B_0}{D_0}. \tag{2.1}$$

At time $n = 1$, a ball is sampled from the urn; its color is X_1, a random variable with Bernoulli(Z_0) distribution. Let M_1 and N_1 be two independent random variables with distribution μ and ν, respectively; assume that X_1, M_1 and N_1 are independent. Next the sampled ball is replaced in the urn together with

$$X_1 M_1 + (1 - X_1) N_1$$

balls of the same color; set

$$B_1 = B_0 + X_1 M_1,\ W_1 = W_0 + (1 - X_1) N_1,\ D_1 = B_1 + W_1,\ Z_1 = \frac{B_1}{D_1}.$$

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} 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 and ν respectively. Set

$$B_{n+1} = B_n + X_{n+1} M_{n+1}, \qquad W_{n+1} = W_n + (1 - X_{n+1}) N_{n+1},$$
$$D_{n+1} = B_{n+1} + W_{n+1}, \qquad Z_{n+1} = \frac{B_{n+1}}{D_{n+1}}.$$

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 1; we call X the process of colors generated by the urn while (Z, D) is the process of its compositions. Note that, since (1.1) and (1.2) are satisfied, the law of X is reinforced. For $\mu = \nu$, a two-color, randomly reinforced urn becomes a *generalized Pólya urn* with parameters (b_0, w_0, ν), as defined and studied in May, Paganoni and Secchi (2002).

In the rest of the section we illustrate a few asymptotic results for the process of colors and the process of compositions generated by a two-color, randomly reinforced urn; we often assume that

$$\mu \geq_{\rm st} \nu. \tag{2.2}$$

This means that $\mathbf{E}[\phi(M)] \geq \mathbf{E}[\phi(N)]$ for all nondecreasing functions $\phi : \Re \to \Re$ for which expectations exist. In particular, when assumption (2.2) is satisfied, the random variable

$$\Delta_n = \mathbf{E}\left[\frac{M_{n+1}}{D_n + M_{n+1}} - \frac{N_{n+1}}{D_n + N_{n+1}} \middle| \mathcal{F}_n\right] \tag{2.3}$$

is almost surely non negative, for $n = 0, 1, 2, ...$

Theorem 2.1. *Assume that (2.2) holds. Then the sequence of proportions $Z = (Z_n, n = 1, 2, ...)$ is a bounded submartingale with respect to the filtration $\{\mathcal{F}_n\}$. Therefore, it converges almost surely to a random limit $Z_\infty \in [0, 1]$.*

Proof. Compute

$$\begin{aligned} &\mathbf{E}[Z_{n+1}|\mathcal{F}_n] \qquad (2.4)\\ =\ & \mathbf{E}\left[Z_n \frac{B_n + M_{n+1}}{B_n + W_n + M_{n+1}} + (1 - Z_n)\frac{B_n}{B_n + W_n + N_{n+1}} \middle| \mathcal{F}_n\right] \\ =\ & \mathbf{E}\left[Z_n \left(\frac{B_n + M_{n+1}}{B_n + W_n + M_{n+1}} + \frac{W_n}{B_n + W_n + N_{n+1}}\right) \middle| \mathcal{F}_n\right] \end{aligned}$$

and analogously

$$\begin{aligned} &\mathbf{E}[(1 - Z_{n+1})|\mathcal{F}_n] \qquad (2.5)\\ =\ & \mathbf{E}\left[(1 - Z_n)\left(\frac{W_n + N_{n+1}}{B_n + W_n + N_{n+1}} + \frac{B_n}{B_n + W_n + M_{n+1}}\right) \middle| \mathcal{F}_n\right]. \end{aligned}$$

Hence

$$\begin{aligned}
&\mathbf{E}[Z_{n+1} - Z_n|\mathcal{F}_n] \\
&\quad = \mathbf{E}[(1-Z_n)Z_{n+1} - Z_n(1-Z_{n+1})|\mathcal{F}_n] \\
&\quad = Z_n(1-Z_n)\,\mathbf{E}\left[\frac{B_n + M_{n+1}}{B_n + W_n + M_{n+1}} + \frac{W_n}{B_n + W_n + N_{n+1}}\right. \\
&\qquad\qquad \left. - \frac{W_n + N_{n+1}}{B_n + W_n + N_{n+1}} - \frac{B_n}{B_n + W_n + M_{n+1}} \right| \mathcal{F}_n\Bigg] \\
&\quad = Z_n(1-Z_n)\,\mathbf{E}\left[\frac{M_{n+1}}{B_n + W_n + M_{n+1}} - \frac{N_{n+1}}{B_n + W_n + N_{n+1}}\right| \mathcal{F}_n\Bigg] \\
&\quad = Z_n(1-Z_n)\Delta_n \\
&\quad \geq 0,
\end{aligned} \tag{2.6}$$

where Δ_n is defined in (2.3). ∎

Corollary 2.2. *Assume that (2.2) holds. Then the sequence of colors X is asymptotically exchangeable and*

$$\lim_{n\to\infty} P[X_{n+1} = x_1, ..., X_{n+k} = x_k] = \mathbf{E}[Z_\infty^{\sum_{i=1}^k x_i}(1-Z_\infty)^{k-\sum_{i=1}^k x_i}]$$

for every $k \geq 1$ and $x_1, ..., x_k \in \{0,1\}$.

Proof. For $n \geq 1$, the conditional distribution of X_{n+1} given $\mathcal{F}_n$ is a Bernoulli(Z_n); on a set of probability one, it converges to a Bernoulli(Z_∞) as n grows to infinity. The result now follows from Lemma 8.2, part (b) in (Aldous, 1985). ∎

For $\epsilon > 0$, set

$$A_\epsilon = \{x \in [0,1] : x(1-x) > \epsilon\}.$$

Lemma 2.3. *Assume that $\alpha > 0$ and that $\int_\alpha^\omega x\mu(dx) > \int_\alpha^\omega x\nu(dx)$. If $Z_0 \in A_\epsilon$, then $P[Z$ exits from $A_\epsilon] = 1$.*

Proof. The argument is analogous to that used by Hill, Lane and Sudderth (1980) for proving their Proposition 3.1.

Fix $\epsilon > 0$ and let $T = \inf\{n \geq 0 : Z_n \notin A_\epsilon\}$. Since A_ϵ is open, T is a stopping time with respect to the filtration $\{\mathcal{F}_n\}$. For $n = 1, 2, 3, ...,$ let T_n be the minimum of T and n; for $k = 1, ..., n$, the indicator $\mathbf{1}_{[T_n \geq k]}$ is measurable with respect to $\mathcal{F}_{k-1}$. Thus

$$
\begin{aligned}
1 \geq \mathbf{E}[Z_{T_n}] \geq & & (2.7)\\
\geq \quad & \mathbf{E}\left[\sum_{k=1}^{n}(Z_k - Z_{k-1})\mathbf{1}_{[T_n \geq k]}\right] = \\
= \quad & \mathbf{E}\left[\sum_{k=1}^{n}\mathbf{E}[(Z_k - Z_{k-1})|\mathcal{F}_{k-1}]\mathbf{1}_{[T_n \geq k]}\right] \\
\geq \quad & \mathbf{E}\left[\sum_{k=1}^{n}\mathbf{E}[(Z_k - Z_{k-1})|\mathcal{F}_{k-1}]\mathbf{1}_{[T = \infty]}\right] \\
= \quad & \mathbf{E}\left[\sum_{k=1}^{n} Z_{k-1}(1 - Z_{k-1})\Delta_{k-1}\mathbf{1}_{[T = \infty]}\right] \\
\geq \quad & \varepsilon\, \mathbf{E}\left[\sum_{k=1}^{n}\Delta_{k-1}\mathbf{1}_{[T = \infty]}\right] \\
\geq \quad & \varepsilon \sum_{k=1}^{n}\frac{b_0 + w_0 + (k-1)\alpha}{(b_0 + w_0 + k\omega)^2}\,\mathbf{E}[M_k - N_k]P[T = \infty]
\end{aligned}
$$

on a set of probability one; the second equality follows from (2.6), the next to the last inequality holds because the proportions $Z_{k-1} \in A_\epsilon$ if $T = \infty$, while the last inequality is true because, with probability one,

$$b_0 + w_0 + (j-1)\alpha \leq D_j \leq b_0 + w_0 + (j-1)\omega \tag{2.8}$$

for every $j = 0, 1, ..., n$.

Thus, for $n \geq 1$,

$$1 \geq \varepsilon \sum_{k=1}^{n}\frac{b_0 + w_0 + (k-1)\alpha}{(b_0 + w_0 + k\omega)^2}\,\mathbf{E}[M_k - N_k]P[T = \infty]. \tag{2.9}$$

Since $\alpha > 0$,

$$\sum_{k=1}^{\infty}\frac{b_0 + w_0 + (k-1)\alpha}{(b_0 + w_0 + k\omega)^2} = \infty;$$

moreover, $\mathbf{E}[M_k - N_k] = \int_\alpha^\omega x\mu(dx) - \int_\alpha^\omega x\nu(dx) > 0$. Hence, $P[T = \infty]$ must be zero and $P[Z_n \textit{ exits from } A_\varepsilon] = 1$. ■

Theorem 2.1 and Lemma 2.3 admit a corollary that parallels Corollary 3.1 in Hill, Lane and Sudderth (1980). Before stating it, let us note that the process (Z, D) is a Markov sequence with respect to the filtration $\{\mathcal{F}_n\}$; hence the strong Markov property holds. Rephrased in our urn terminology it says that, if τ is an almost surely finite stopping time relative to the filtration $\{\mathcal{F}_n\}$ then, given the sigma-field $\mathcal{F}_\tau$, the conditional law of the random sequence $((Z_\tau, D_\tau), (Z_{\tau+1}, D_{\tau+1}), ...)$ is that of the process of compositions generated by a two-color, randomly reinforced urn with initial composition $(Z_\tau D_\tau, (1-Z_\tau)D_\tau)$ and directed by μ and ν.

Theorem 2.4. *Suppose that:*

(i) $\alpha > 0$;
(ii) *assumption (2.2) holds;*
(iii) $\int_\alpha^\omega x\mu(dx) > \int_\alpha^\omega x\nu(dx)$.

Then, $\lim_{n\to\infty} Z_n = Z_\infty \in \{0, 1\}$ *with probability one.*

Proof. Existence of the almost sure limit Z_∞ for the sequence of proportions $Z = (Z_n, n = 1, 2, ...)$ is guaranteed by (i) and Theorem 2.1.

Let $\epsilon > 0$; since A_ε is open,

$$P[Z_\infty \in A_\epsilon] = P[(Z_n, D_n) \text{ eventually in } A_\varepsilon \times (0, \infty)] = 0$$

where the last equality holds because of Lemma 2.3 and the strong Markov property. Since this is true for every $\epsilon > 0$, $P[Z_\infty \in \{0, 1\}] = 1$. ■

The last one is only an intermediate result: in fact, in Section 4 we are going to prove that

$$P(Z_\infty = 1) = 1 \tag{2.10}$$

when $b_0 > 0$ and assumptions (i)-(iii) of Theorem 2.4 are satistified.

3. An auxiliary gambling problem

We introduce a gambling problem for studying the distribution of the limit of the sequence of compositions of a two-color, randomly reinforced urn satisfying assumption (2.2).

Following Dubins and Savage (1965), a gambling problem is defined by three elements: the space S of the gambler's fortunes, a gambling house associating to each fortune $s \in S$ the gambles available to the player, and a utility function $u : S \to \Re$. The fortunes space of our gambling problem is $S = [0, 1] \times (0, \infty)$. The gambling house is described by the set $\mathcal{A}$ of all couples (μ, ν) of probability distributions with support contained in $[\alpha, \omega]$, with $0 < \alpha \leq \omega < \infty$, and such that $\mu \geq_{\text{st}} \nu$. When the gambler's fortune is $(z, d) \in S$ and the gambler selects $(\mu, \nu) \in \mathcal{A}$, his next fortune has distribution $\gamma((z, d), (\mu, \nu))$ and is equal to $((zd + M)/(d + M), d + M)$ with probability z and $((zd)/(d + N), d + N))$ with probability $1 - z$, where M and N are two independent random variables with distribution μ and ν respectively. Finally, for every $(z, d) \in S$, define

$$u(z, d) = \beta(1 - z)^{\beta - 1}$$

where β is an integer greater than or equal to 2.

Remark 3.1. If A is a countable subset of $[\alpha, \omega]$ and, for all $(\mu, \nu) \in \mathcal{A}$, both probability distributions μ and ν have support contained in A, we may describe our gambling problem by means of a countable space of fortunes S and refer to the gambling theory developed in Maitra and Sudderth (1996) which avoids technical issues of measurability.

A gambler's strategy σ is an infinite sequence $\sigma_0, \sigma_1, ...$ such that $\sigma_0 = a_0 \in \mathcal{A}$ and $\sigma_n(s_1, ..., s_n) \in \mathcal{A}$, for every $n \geq 1$ and $s_1, ..., s_n \in S$. When his initial fortune is $(z, d) \in S$ and the gambler chooses a strategy σ, this determines the law of a stochastic process $(Z, D) = ((z, d), (Z_1, D_1), (Z_2, D_2), ...)$ with values in S; (Z_1, D_1) has distribution $\gamma((z, d), \sigma_0)$ and, given $(Z_1, D_1) = (z_1, d_1), ..., (Z_n, D_n) = (z_n, d_n)$, the conditional distribution of (Z_{n+1}, D_{n+1}) is equal to

$$\gamma((z_n, d_n), \sigma_n((z_1, d_1), ...(z_n, d_n))),$$

for $n = 1, 2, 3, ...$ The strategy σ generates a payoff to the gambler equal to

$$u(\sigma) = \mathbf{E}[\limsup_{n \to \infty} u(Z_n, D_n)]. \tag{3.1}$$

This is consistent with the payoff treated in Dubins and Savage (1965) because the utility function u is bounded and thus the Fatou equation in Sudderth (1971) is in force (see also Theorem 2.2 in Maitra and Sudderth (1996)). In fact, with an argument analogous to that illustrated by equations (2.4)-(2.6) in the proof of Theorem 2.1, one can show that, for every gambler's strategy σ, the sequence $Z = (Z_n, n = 1, 2, ...)$ of the first components of the process (Z, D) is a bounded submartingale and converges almost surely to a random variable Z_∞; hence

$$u(\sigma) = \mathbf{E}[\beta(1 - Z_\infty)^{\beta - 1}].$$

The process $D = (D_n, n = 1, 2, ...)$ diverges to infinity with probability one for any strategy σ available to the gambler; this happens because, for any $(\mu, \nu) \in \mathcal{A}$, both probabilities distributions have support contained in $[\alpha, \omega]$ and $\alpha > 0$.

A *constant* strategy consists in fixing $(\mu_0, \nu_0) \in \mathcal{A}$ and letting

$$\sigma_0 = \sigma_n(s_1, ..., s_n) = (\mu_0, \nu_0),$$

for every $n \geq 1$ and $s_1, ..., s_n \in S$; that is the gambler plays constantly the gamble identified by (μ_0, ν_0). We indicate such a constant strategy as $(\mu_0, \nu_0)^\infty$. If $(z, d) \in S$ is the initial fortune of a gambler who chooses a constant strategy $(\mu_0, \nu_0)^\infty$, the process of states (Z, D) has the same law as the process with the same name generated by a two-color, randomly reinforced urn with initial composition equal to $(zd, (1-z)d)$ and directed by μ_0, and ν_0.

The value of the gambling problem at the initial fortune $(z, d) \in S$ is defined as

$$V(z, d) = \sup_\sigma u(\sigma);$$

a strategy σ is *optimal* at (z, d) if $u(\sigma) = V(z, d)$. With the next theorem we find an expression for the value of our gambling problem and an optimal strategy for the gambler playing it. Before stating the result, let us define for $a, b > 0$

$$B(a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$$

where $\Gamma(a)$ is the gamma function evaluated at a. For $a \in [\alpha, \omega]$, we use δ_a for the point mass distribution at a.

Theorem 3.2. *For every initial fortune* $(z, d) \in S$,

$$V(z, d) = \beta \frac{\Gamma(\frac{d}{\omega})\Gamma(\frac{(1-z)d}{\omega} + \beta - 1)}{\Gamma(\frac{(1-z)d}{\omega})\Gamma(\frac{d}{\omega} + \beta - 1)} \tag{3.2}$$

and the constant strategy $(\delta_\omega, \delta_\omega)^\infty$ *is optimal at* (z, d).

Proof. For every initial fortune $(z, d) \in S$, let $Q(z, d) = u((\delta_\omega, \delta_\omega)^\infty)$. When the gambler chooses the constant strategy $(\delta_\omega, \delta_\omega)^\infty$, the law of the sequence $Z = (Z_n, n = 1, 2, ...)$ is that of the infinite sequence of random proportions of balls of color 1 generated by a Pólya urn initially containing d balls, of which a number zd are of color 1 and a number $(1-z)d$ of color 0. At time $n = 1, 2, ...$, a ball is sampled from the urn and it is reintroduced in it together with other ω balls of the same color; the number of balls contained in the urn becomes $d + (n+1)\omega$, of which a fraction Z_{n+1} are of color 1. It is well known that the sequence of random proportions generated by this Pólya urn converges to a random variable Z_∞ with Beta distribution with parameters $(zd/\omega, (1-z)d/\omega)$.

Thus

$$
\begin{aligned}
&Q(z,d) \\
&\quad = \int_0^1 \beta(1-t)^{\beta-1} B\left(\frac{zd}{\omega}, \frac{(1-z)d}{\omega}\right) t^{\frac{zd}{\omega}-1}(1-t)^{\frac{(1-z)d}{\omega}-1} dt \\
&\quad = \beta \frac{\Gamma\left(\frac{d}{\omega}\right)\Gamma\left(\frac{(1-z)d}{\omega}+\beta-1\right)}{\Gamma\left(\frac{(1-z)d}{\omega}\right)\Gamma\left(\frac{d}{\omega}+\beta-1\right)}.
\end{aligned} \tag{3.3}
$$

In order to prove that $(\delta_\omega, \delta_\omega)^\infty$ is optimal, we show that Q is excessive, i.e. that for all $(z,d) \in S$ and $(\mu, \nu) \in \mathcal{A}$

$$
Q(z,d) \geq \int Q(u,v)\gamma((z,d),(\mu,\nu))(du,dv), \tag{3.4}
$$

and that

$$
Q(\sigma) \geq u(\sigma) \tag{3.5}
$$

for every gambler's strategy σ; the real number $Q(\sigma)$ is defined in (3.1) after substitution of u with Q. The result then follows from Theorem 3.3.1 in Dubins and Savage (1965) or Theorem 5.1 in Maitra and Sudderth (1996).

To prove (3.4), let $m \in [\alpha, \omega]$, $(z,d) \in S$ and observe that, because of (3.3),

$$
\begin{aligned}
&Q(\frac{zd+m}{d+m}, d+m) \\
&\quad = \beta \frac{\Gamma\left(\frac{d+m}{\omega}\right)\Gamma\left(\frac{(1-z)d}{\omega}+\beta-1\right)}{\Gamma\left(\frac{(1-z)d}{\omega}\right)\Gamma\left(\frac{d+m}{\omega}+\beta-1\right)} \\
&\quad = \beta\left[\frac{((1-z)d+\omega(\beta-2))(((1-z)d+\omega(\beta-3))\cdots((1-z)d)}{(d+m+\omega(\beta-2))((d+m+\omega(\beta-3))\cdots(d+m)}\right];
\end{aligned}
$$

hence $Q((zd+m)/(d+m), d+m)$ is a decreasing function of m. This implies that, for all independent random variables M and N with distributions μ and ν respectively, $(\mu,\nu) \in \mathcal{A}$,

$$
\mathbf{E}[Q(\frac{zd+M}{d+M}, d+M)] \leq \mathbf{E}[Q(\frac{zd+N}{d+N}, d+N)]
$$

since $\mu \geq_{\rm st} \nu$; therefore,

$$\begin{aligned} &\int Q(u,v)\gamma((z,d),(\mu,\nu))(du,dv) \qquad (3.6)\\ &\quad = z\,\mathbf{E}[Q(\frac{zd+M}{d+M},d+M)] + (1-z)\,\mathbf{E}[Q(\frac{zd}{d+N},d+N)]\\ &\quad \leq z\,\mathbf{E}[Q(\frac{zd+N}{d+N},d+N)] + (1-z)\,\mathbf{E}[Q(\frac{zd}{d+N},d+N)]\\ &\quad = \mathbf{E}[zQ(\frac{zd+N}{d+N},d+N) + (1-z)Q(\frac{zd}{d+N},d+N)]. \end{aligned}$$

Now define, for all couples of non negative real numbers (b,w), all $n \in [\alpha,\omega]$ and all integers $r \geq 0$, the quantity

$$\begin{aligned} &g(b,w,n,r)\\ &\quad = \frac{b}{w+b}\left[\frac{(w+b)(w+b+\omega)\cdots(w+b+r\omega)}{(w+b+n)(w+b+n+\omega)\cdots(w+b+n+r\omega)}\right] +\\ &\quad\quad + \frac{w}{w+b}\left[\frac{(w+b)(w+b+\omega)\cdots(w+b+r\omega)}{w(w+\omega)\cdots(w+r\omega)}\cdot\right.\\ &\quad\quad\quad \left.\cdot\frac{(w+n)(w+n+\omega)\cdots(w+n+r\omega)}{(w+b+n)(w+b+n+\omega)\cdots(w+b+n+r\omega)}\right]. \end{aligned}$$

Notice that $g(b,w,n,0) = 1$; with a few elementary, but tedious, computations one can show that

$$\begin{aligned} &g(b,w,n,r) - g(b,w,n,r+1)\\ &\quad = \frac{(bn)(b+w+\omega)\cdots(b+w+r\omega)}{(b+w+n)(b+w+n+\omega)\cdots(b+w+n+(r+1)\omega)}\cdot\\ &\quad\quad \cdot\left[1 - \frac{(w+n)\cdot(w+n+r\omega)}{(w+\omega)\cdots(w+(r+1)\omega)}\right]\\ &\quad \geq 0; \end{aligned}$$

hence $g(b,w,n,r) \leq 1$ for all $r \geq 0$. Fix $n \in [\alpha,\omega]$ and check that

$$\begin{aligned} zQ(\frac{zd+n}{d+n},d+n) &+ (1-z)Q(\frac{zd}{d+n},d+n) \qquad (3.7)\\ &= Q(z,d)g(zd,(1-z)d,n,\beta-2)\\ &\leq Q(z,d). \end{aligned}$$

Equation (3.4) now follows from (3.6) and (3.7).

To prove (3.5), let σ be a strategy available to the gambler; under σ the sequence $Z = (Z_n, n = 1,2,...)$ converges almost surely to a random variable Z_∞ while $\lim_{n\to\infty} D_n = \infty$ with probability one. By Lemma 5.2 in Hill, Lane and Sudderth (1980), on a set of probability one the sequence of Beta distributions with parameters $(Z_nD_n/\omega,(1-Z_n)D_n/\omega)$ converges in distribution to δ_{Z_∞}. Thus

$$\begin{aligned}
Q(\sigma) &= \mathbf{E}[\lim_{n\to\infty} Q(Z_n,D_n)] \\
&= \mathbf{E}[\lim_{n\to\infty}\int_0^1 \beta(1-t)^{\beta-1}B\left(\frac{Z_nD_n}{\omega},\frac{(1-Z_n)D_n}{\omega}\right)t^{\frac{Z_nD_n}{\omega}-1}(1-t)^{\frac{(1-Z_n)D_n}{\omega}-1}dt] \\
&= \mathbf{E}[\beta(1-Z_\infty)^{\beta-1}] \\
&= u(\sigma).
\end{aligned}$$

■

4. On the distribution of Z_∞

We are finally ready to prove (2.10). Consider a two-color, randomly reinforced urn with initial composition (b_0,w_0) and directed by the probability distributions μ and ν with support in $[\alpha,\omega]$, with $0\le\alpha\le\omega<\infty$.

Theorem 4.1. *Suppose that* $b_0>0$ *and*

(i) $\alpha>0$;
(ii) *assumption (2.2) is satisfied by the probability distributions* μ *and* ν;
(iii) $\int_\alpha^\omega x\mu(dx)>\int_\alpha^\omega x\nu(dx)$.

Then $Z_\infty=\lim_{n\to\infty}Z_n=1$ *almost surely.*

Proof. It follows from Theorem 2.4 that $P[Z_\infty\in\{0,1\}]=1$. Suppose that

$$P[Z_\infty=0]=\epsilon>0. \tag{4.1}$$

Let $(z,d)\in(0,1)\times(0,\infty)$ be such that $(b_0,w_0)=(zd,(1-z)d)$.

Case 1: $(zd,(1-z)d)\in L=\{(b,w)\in(0,\infty)\times[0,\infty):\min(b,w)>\omega\}$ Consider the gambling problem described in the previous section and a gambler with initial fortune equal to (z,d) who decides to use the strategy $(\mu,\nu)^\infty$ for playing it. Then, (4.1) implies that

$$\epsilon\beta\le u((\mu,\nu)^\infty)\le V(z,d). \tag{4.2}$$

However, because of Theorem 3.2,

$$\begin{aligned}
&V(z,d)\\
&=\beta\frac{\Gamma\left(\frac{d}{\omega}\right)\Gamma\left(\frac{(1-z)d}{\omega}+\beta-1\right)}{\Gamma\left(\frac{(1-z)d}{\omega}\right)\Gamma\left(\frac{d}{\omega}+\beta-1\right)}\\
&=\int_0^1 B\left(\frac{zd}{\omega},\frac{(1-z)d}{\omega}\right)t^{\frac{zd}{\omega}-1}(1-t)^{\frac{(1-z)d}{\omega}-1}\beta(1-t)^{\beta-1}dt.
\end{aligned}$$

For β going to infinity, a Beta distribution with parameters $(1, \beta)$ converges in distribution to δ_0; this fact and (4.2) imply that

$$0 = \lim_{\beta \to \infty} V(z, d) \geq \lim_{\beta \to \infty} \beta \epsilon = \infty; \tag{4.3}$$

hence (4.1) is false. Note that the first equality in (4.3) holds because the integrand

$$B\left(\frac{zd}{\omega}, \frac{(1-z)d}{\omega}\right) t^{\frac{zd}{\omega}-1}(1-t)^{\frac{(1-z)d}{\omega}-1}$$

is a bounded and continuous function defined on $[0, 1]$ that is equal to 0 for $t = 0$, since $zd > \omega$.

Case 2: $(zd, (1-z)d) \in L^c \bigcap \{(b, w) \in (0, \infty) \times [0, \infty) : w > 0\}$. Let $\tau = \inf\{n : (Z_n D_n, (1 - Z_n)D_n) \in L\}$; τ is a stopping time with respect to the filtration $\{\mathcal{F}_n\}$. Moreover, since $\min(b_0, w_0) > 0$ and $\alpha > 0$, $P[\tau < \infty] = 1$. The strong Markov property implies that, given the sigma-field generated by τ, the conditional law of the process $((Z_\tau, D_\tau), (Z_{\tau+1}, D_{\tau+1}), ...)$ is the same has that of the process of compositions generated by a two-color, randomly reinforced urn with initial composition $(Z_\tau D_\tau, (1-Z_\tau)D_\tau)$ and directed by the probability distributions μ and ν; the sequence of proportions of balls of color 1 generated by this urn converges to 1 almost surely because of Case 1. This must then also be true for the two-color, randomly reinforced urn with initial composition (b_0, w_0) and directed by μ and ν.

Case 3: $(1 - z)d = 0$. This is the simplest case, since $Z_n = 1$ for all $n = 1, 2, ...$ if $w_0 = 0$. ■

When $\mu = \nu$, a two-color, randomly reinforced urn with initial composition (b_0, w_0) and directed by μ and ν is a generalized Pólya urn with parameters (b_0, w_0, ν) as in May, Paganoni and Secchi (2002); in this case the distribution of Z_∞ has no atoms. We don't know the exact distribution of Z_∞, but Theorem 3.2 implies the following inequalities for its moments:

$$\mathbf{E}[(1 - Z_\infty)^r] \leq \frac{\Gamma(\frac{b_0+w_0}{\omega})\Gamma(\frac{w_0}{\omega} + r)}{\Gamma(\frac{w_0}{\omega})\Gamma(\frac{b_0+w_0}{\omega} + r)} \tag{4.4}$$

and

$$\mathbf{E}[Z_\infty^r] \leq \frac{\Gamma(\frac{b_0+w_0}{\omega})\Gamma(\frac{b_0}{\omega} + r)}{\Gamma(\frac{b_0}{\omega})\Gamma(\frac{b_0+w_0}{\omega} + r)} \tag{4.5}$$

for $r = 1, 2, ...$ To obtain (4.4), consider a gambler with initial fortune $(z, d) = (b_0/(b_0 + w_0), b_0 + w_0)$ who plays the constant strategy $(\nu, \nu)^\infty$.

Then, for $\beta = r + 1$,

$$\frac{1}{\beta} u((\nu, \nu)^{\infty}) \leq \frac{1}{\beta} V(z, d)$$

because of (3.2). This is in fact inequality (4.4) written with a different notation. Inequality (4.5) now follows by simmetry from (4.4). For instance, for $r = 1$, (4.4) and (4.5) imply that $\mathbf{E}(Z_\infty) = b_0/(b_0 + w_0)$, something we can get directly from Theorem 2.1 since the sequence of proportions $Z = (Z_n, n = 0, 1, 2, ...)$ is a bounded martingale for $\mu = \nu$.

5. A Bayesian perspective

Suppose that the probability distributions μ and ν directing a two-color, randomly reinforced urn represent the laws of a favorable covariate with positive values observed for a patient after the patient has experienced 'treatment 1' or 'treatment 0', respectively; for instance, the covariate may result from a blood count or be a function of patient's survival time after treatment. If μ and ν are unknown, a researcher designing a sequential clinical trial may be tempted to use the urn for allocating patients to treatments, knowing that, in the long run, the urn will select the better treatment with higher and higher probability, if such a treatment exists; better here refers to the treatment with associated a stochastically larger covariate with greatest expectation.

A Bayesian researcher can take μ and ν to be random, with joint law incorporating prior information. For instance, following Ferguson (1973), he could assume that μ and ν are independent Dirichlet processes, with parameters ϕ_1 and ϕ_0 respectively; ϕ_1 and ϕ_0 are finite measures with support contained in $[\alpha, \omega]$, with $0 < \alpha \leq \omega < \infty$. In this case, let X be the process of colors and (Z, D) the process of compositions generated by a two-color, randomly reinforced urn with initial composition

$$b_0 = \int_\alpha^\omega x \phi_1(dx), \ w_0 = \int_\alpha^\omega x \phi_0(dx)$$

and directed by μ and ν. The color of the ball extracted from the urn when it is sampled for the n-th time indicates the treatment that patient n will experience; M_n indicates the the value of the covariate for patient n if he has been subject to 'treatment 1' while N_n represents the value of the same covariate when he is subject to 'treatment 0'. Then, at stage $n = 1, 2, 3, ...$ of the trial, after having observed the covariate's values for the first n patients, the conditional expected value for the covariate on the next patient experiencing 'treatment 1' is

$$\mathbf{E}[\int_\alpha^\omega x\mu(dx)|\mathcal{F}_n^*] = \frac{b_0 + \sum_{i=1}^n X_i M_i}{\phi_1([\alpha,\omega]) + \sum_{i=1}^n X_i} = \frac{D_n}{\phi_1([\alpha,\omega]) + \sum_{i=1}^n X_i} Z_n \tag{5.1}$$

while the conditional expected value for the covariate on the next patient experiencing 'treatment 0' is:

$$\mathbf{E}[\int_\alpha^\omega x\nu(dx)|\mathcal{F}_n^*] = \frac{w_0 + \sum_{i=1}^n (1-X_i) M_i}{\phi_0([\alpha,\omega]) + \sum_{i=1}^n (1-X_i)} = \frac{D_n}{\phi_0([\alpha,\omega]) + \sum_{i=1}^n (1-X_i)} (1-Z_n); \tag{5.2}$$

$\mathcal{F}_n^*$ is the sigma-field generated by the random variables $X_1, ..., X_n$ and $X_1M_1 + (1-X_1)N_1, ..., X_nM_n + (1-X_n)N_n$. Equations (5.1) and (5.2) follow from independence of μ and ν and from Ferguson (1973) where it is shown that if μ is a Dirichlet process with parameter ϕ_1, then the expectation of $\mu((-\infty,x])$ is $\phi_1^{-1}([\alpha,\omega])\phi_1((-\infty,x])$ for every $x \in \Re$, and the conditional distribution of μ, given a sample $M_1, ..., M_k$ of size $k \geq 1$ from μ, is a Dirichlet process with parameter $\phi_1 + \sum_{i=1}^k \delta_{M_i}$. The researcher may thus consider the urn as a randomizing device that assigns patient $n+1$ to a treatment with a probability proportional to the product of the total number of patients who experienced in the past such treatment and the best Bayesian prediction under square loss for the value of the covariate for patients experiencing such treatment: we are here interpreting the prior information contained in the quantities $\phi_1([\alpha,\omega])$ and $\phi_0([\alpha,\omega])$ as if these were sample sizes of patients who experienced the two treatments prior to the actual sequential trial.

In fact, the urn implements a strategy for a two-armed bandit problem, a gambling problem where a player has to make sequential selections from two independent stochastic processes (arms); at each stage of the problem, the gambler chooses to observe an arm based on past selections and observations; see Berry and Fristed (1985) for a general theory for bandit problems. The case where observations on arm $i = 1, 0$ are conditionally i.i.d. given that their marginal distribution μ and ν respectively are Dirichlet processes with parameters ϕ_1 and ϕ_0, has been treated by Chattopadhyay (1994) who generalized results of Clayton and Berry (1985). For a Bayesian, the theory of bandit problems is the right mathematical setting for sequential clinical trials: of course, for each sequence of selections and observations a payoff need to be specified and the worth of a strategy is then judged with respect to the expected payoff it generates. For instance, Clayton and Berry (1985) and Chattopadhyay (1994) consider as payoff the sum of the first $N \geq 1$ covariate's values that, in our notation, is

$$\sum_{i=1}^{N} X_i M_i + \sum_{i=1}^{N} (1 - X_i) N_i. \tag{5.3}$$

The existence of an optimal strategy is then guaranteed, but it seems next to impossible to specify it explicitly unless N is very small. However, Clayton and Berry note that, with a payoff as in (5.3), optimal strategies possess a monotonicity property:

> roughly, the larger the observation from arm 1, the greater the inclination to continue selecting that arm.

This is yet another instance of reinforcement; the strategy implemented by our two-color, randomly reinforced urn has the same property. Note that this strategy is randomized in the sense that, at each stage of the problem, the strategy selects an arm by means of a randomizing device. In general, the use of a randomized strategy does not give an advantage to the gambler: there is always a non-randomized strategy generating a greater or equal expected payoff. Nevertheless, in sequential clinical trials non-randomized strategies are not appealing since they are open to experimental bias: for instance, Rosenberger and Lachin (2002) and Hu and Rosenberger (2003) claim that:

> randomization should be preserved in clinical trials at all costs, because it mitigates certain biases and provides a basis for inference.

6. Concluding remark: a $(k+1)$-color, randomly reinforced urn

The arguments of the previous sections may be easily extended to cover the case where the urn contains more than two colors and the reinforcement's distribution associated with one particular color dominates the reinforcement's distributions associated with the other colors.

Let $k \geq 1$, $B_0 = (b_0(1), ..., b_0(k+1))$ a vector of $k+1$ positive real numbers and $\mu, \nu_1, ..., \nu_k$ be $k+1$ probability distributions with support contained in $[\alpha, \omega]$, where $0 \leq \alpha \leq \omega < \infty$. The vector B_0 represents the initial urn composition with $b_0(i) > 0$ indicating the number of balls of color $i = 1, ..., k+1$ contained in the urn before it is sampled for the first time. Set

$$D_0 = \sum_{i=1}^{k+1} B_0(i) \text{ and } Z_0 = \frac{1}{D_0} B_0.$$

Let X_1 be a random vector with values in $\{0,1\}^{k+1}$ and Multinomial$(1, Z_0)$ distribution, and $M_1 \in [\alpha, \omega]^{k+1}$ a random vector with probability distribution equal to the product $\mu \times \nu_1 \cdots \times \nu_k$; assume that X_1 and M_1 are independent. Next, set

$$B_1 = B_0 + X_1 * M_1,\ D_1 = \sum_{i=1}^{k+1} B_1(i) \text{ and } Z_1 = \frac{1}{D_1} B_1,$$

where the operator $*$ represents the element by element product between two vectors with same length. At time $n+1$, given the sigma-field $\mathcal{F}_n$ generated by $X_1, ..., X_n$ and $M_1, ..., M_n$, let $X_{n+1} \in \{0,1\}^{k+1}$ be a random vector with Multinomial$(1, Z_n)$ distribution and, independently of $\mathcal{F}_n$ and X_{n+1}, assume that M_{n+1} is a random vector with values in $[\alpha, \omega]^{k+1}$ and with distribution equal to the product $\mu \times \nu_1 \cdots \times \nu_k$. Set

$$B_{n+1} = B_n + X_{n+1} * M_{n+1},\ D_{n+1} = \sum_{i=1}^{k+1} B_{n+1}(i),\ Z_{n+1} = \frac{1}{D_{n+1}} B_{n+1}.$$

And so on forever, thus generating the process of the urn compositions $(Z, D) = ((Z_n, D_n), n = 0, 1, ...)$ with values in $[0,1]^{k+1} \times (0, \infty)$ and the process of colors $X = (X_n, n = 1, 2, ...)$ with values in $\{0,1\}^{k+1}$.

Theorems 2.1 and 4.1 are easily generalized, if we assume that the probability distribution μ dominates every probability distribution ν_i, for $i = 1, ..., k$.

Theorem 6.1. *Assume that, for all* $i = 1, ..., k$,

$$\mu \geq_{st} \nu_i.$$

Then the sequence $Z(1) = (Z_n(1), n = 0, 1, 2, ...)$, *of the proportions of balls of color 1 in the urn, is a bounded submartingale with respect to the filtration* $\{\mathcal{F}_n\}$ *and it converges to a random limit* $Z_\infty(1) \in [0,1]$.

Proof. . In the proof of Theorem 2.1 substitute Z_n with $Z_n(1)$, B_n with $B_n(1)$, M_n with $M_n(1)$ and set $W_n = \sum_{i=2}^{k+1} B_n(i)$. When computing the analogous of equations (2.4)-(2.5), for $n = 1, 2, ...$, let

$$N_n = \sum_{i=2}^{k+1} \widetilde{X}_n(i-1) M_n(i),$$

where $(\widetilde{X}_n(1), ..., \widetilde{X}_n(k))$ has values in $\{0,1\}^k$ and, given $\mathcal{F}_n$, is independent of X_{n+1} and of M_{n+1} with conditional distribution equal to a Multinomial$(1, (1 - Z_n(1))^{-1}(Z_n(2), ..., Z_n(k+1)))$. Note that, given $\mathcal{F}_n$, the conditional distribution of N_{n+1} is a mixture of the probability distributions $\nu_1, ..., \nu_k$; μ is stochastically larger of such mixture because $\mu \geq_{st} \nu_i$ for all $i = 1, ..., k$. ■

Theorem 6.2. *Suppose that:*

(i) $\alpha > 0$

and, for $i = 1, ..., k$,

(ii) $\mu \geq_{st} \nu_i$;

(iii) $\int_\alpha^\omega x\mu(dx) > \int_\alpha^\omega x\nu_i(dx)$.

Then $\lim_{n\to\infty} Z_n(1) = 1$ *almost surely.*

Proof. As for the last theorem, Lemma 2.3 and Theorem 2.4 are easily reformulated and proved in the new $(k+1)$-color setting after observing that μ is stochastically larger than any mixture of the probability distributions ν_i's because of (ii), and the first moment of μ is strictly greater than the first moment of any mixture of the probability distributions ν_i's if (iii) is true. Hence $\lim_{n\to\infty} Z_n(1) \in \{0, 1\}$ almost surely.

To complete the proof, introduce a gambling problem, with space of fortunes $S = [0,1]^{k+1} \times (0, \infty)$ and gambling house described by the set $\mathcal{A}$ of all $(k+1)$-tuples of probability distributions $(\mu, \nu_1, ..., \nu_k)$ with support contained in $[\alpha, \omega]$ and such that $\mu \geq_{st} \nu_i$ for $i = 1, ..., k$; when the gambler fortune is $(z, d) \in S$ and the gambler selects $(\mu, \nu_1, ..., \nu_k) \in \mathcal{A}$, his next fortune has distribution $\gamma((z, d), (\mu, \nu_1, ..., \nu_k))$ and is equal to

$$\left(\frac{1}{d + \sum_{i=1}^{k+1} X(i)M(i)} (dz + X * M), d + \sum_{i=1}^{k+1} X(i)M(i) \right)$$

where X is a random vector with values in $\{0,1\}^{k+1}$ with Multinomial$(1, z)$ distribution and M is random vector, independent of X, with values in $[\alpha, \omega]^{k+1}$ and probability distribution $\mu \times \nu_1 \times \cdots \times \nu_k$. Finally, for β an integer greater than or equal to 2, let $u(z, d) = \beta(1 - z(1))^{\beta-1}$ be the utility function at the fortune $(z, d) \in S$.

There are no difficulties in adapting the proof of Theorem 3.2 for showing that, for every initial fortune $(z, d) \in S$, this new gambling problem has value

$$V(z, d) = \beta \frac{\Gamma(\frac{d}{\omega})\Gamma(\frac{(1-z(1))d}{\omega} + \beta - 1)}{\Gamma(\frac{(1-z(1))d}{\omega})\Gamma(\frac{d}{\omega} + \beta - 1)}$$

and the constant strategy $(\delta_\omega, \delta_\omega, ..., \delta_\omega)^\infty$ is optimal.

Now, go through the argument of Theorem 4.1 for proving that $Z_\infty(1) = 1$ almost surely. ■

7. Acknowledgement

We thank Bill Sudderth for an interesting discussion that gave us the impetus to the generalization to the $(k+1)$-color urn.

References

1. Aldous, D.J. (1985). Exchangeability and related topics. *École d'été de probabilités de Saint-Flour*, XIII-1983,1–198, Lecture Notes in Math. 1117, Springer, Berlin.
2. Berry, D.A. and Fristedt, B. (1985). *Bandit problems: sequential allocation of experiments.* Chapman and Hall, New York.
3. Chattopadhyay, M.K. (1994). Two-armed Dirichlet bandits with discounting. *The Annals of Statistics*, 22(3), 1212–1221.
4. Clayton, M. K. and Berry, D.A. (1985). Bayesian nonparametric bandits. *The Annals of Statistics*, 13, 1523–1534.
5. Dubins, L. and Savage, L. J. (1965). *How to gamble if you must: inequalities for stochastic processes*, MacGraw-Hill, New York. 2nd edition (1976) Dover, New York.
6. Durham, S.C., Flournoy, N. and Li, W. (1998). A sequential design for maximizing the probability of a response. *The Canadian Journal of Statistics*, 26(3), 479–495.
7. Ferguson, T.S (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1, 209–230.
8. Hill, B.M., Lane, D. and Sudderth, W. (1980). A strong law for some generalized urn processes. *The Annals of Probability*, 8, 214–226.
9. Hill, B.M., Lane, D. and Sudderth, W. (1987). Exchangeable urn processes. *The Annals of Probability*, 15, 1586–1592.
10. Hu, F. and Rosenberger, W.F. (2003). Optimality, variability, power: evaluating response-adaptive randomization procedures for treatment comparison. *JASA*, 98 (463), 671–678.
11. Li, W., Durham, S.D. and Flournoy, N. (1997). Randomized Pólya urn design. *1996 Proceedings of the Biometrics Section of the American Statistical Association*, 166–170.
12. May, C., Paganoni, A.M. and Secchi, P. (2002). On a two-color generalized Pólya urn. *Quaderno di Dipartimento* n. 505/P, Aprile 2002, Dipartimento di Matematica, Politecnico di Milano.
13. Maitra, A. and Sudderth, D.W. (1996). *Discrete gambling and stochastic games.* Springer-Verlag, New York.
14. Rosenberger, W. F. and Lachin, J.M. (2002). *Randomization in clinical trials: theory and practice*, Wiley, New York.
15. Walker, S. G. and Muliere, P. (2004). Reinforcement and finite exchangeability. *Submitted.*

Pietro Muliere, Istituto di Metodi Quantitativi, Università "L. Bocconi", Milano -Italy
E-mail address: pietro.muliere@uni-bocconi.it

Anna Maria Paganoni, Dipartimento di Matematica "F. Brioschi",Politecnico di Milano,Piazza Leonardo da Vinci32, 20123 Milano - Italy
E-mail address: annap@mate.polimi.it

Piercesare Secchi, Dipartimento di Matematica "F. Brioschi", Politecnico di Milano, Piazza Leonardo da Vinci 32, 20123 Milano - Italy
E-mail address: secchi@mate.polimi.it