Proceedings of the Seventh Young Statisticians Meeting
Andrej Mrvar (Editor)
Metodološki zvezki, 21, Ljubljana: FDV, 2003

'Optimal' Randomized Designs for Sequential Experiments with Two Treatments

Alessandro Baldi Antognini[1]

Abstract

This paper starts with a discussion of the 'optimality' of sequential randomized designs for comparing two treatments and introduces the concept of "desirable" proportion of allocations to one of the treatments. The problem is finding a randomized design which converges to the desirable one almost certainly and also forces the procedure towards the desirable proportion even for small samples. When balance is optimal we show that Efron's Biased Coin Design (1971) and the class of Wei's designs (1978) are asymptotically desirable and propose extensions of the above mentioned algorithms that converge almost surely to any desired proportion, when the value is known. The Adjustable Biased Coin Design of Baldi Antognini and Giovagnoli (2003) also converges to balance and the convergence is faster than the other procedures.

1 Introduction

Assume we want to carry out a clinical trial to compare the efficacy of two drugs. Suppose also that subjects arrive sequentially at an experimental site and are assigned immediately to either treatment. In the great majority of experiments two requirements are:

1. the need for some form of possible restricted randomization to protect against several types of bias, including selection bias arising from being able to guess the next treatment allocation;

2. some type of optimality: for example, assuming a linear homoscedastic model a common optimality criterion is to seek to minimize the variances of the estimated treatment effects and this translates into a request for balance.

[1] Department of Statistical Sciences, University of Bologna, via Belle Arti 41, 40126, Bologna, Italy; baldi@stat.unibo.it

A classical statistical problem is finding a sequential randomized design which converge to the 'optimal' one almost certainly. However, since in many cases we do not know when the experiment is going to stop, there is a need for ad hoc designs that work sufficiently well for small samples too.

As regards balance, the problem was brought to the fore in an authoritative paper by Efron (1971) who proposed his - by now widely known - BCD (*Biased Coin Design*). Several authors have extended his suggestion to more complex algorithms, e.g. Wei (1978), Atkinson (1982), Smith (1984a, 1984b). However, for ethical reasons the experimental purpose may be optimizing the expected number of successfully treated patients, leading to the *Play-the-Winner* strategy (Zelen, 1969) or *randomized Play-the-Winner* (Wei and Durham, 1978). In the sequel we suppose that no major ethical issue is at stake, so that the main concern is maximum precision of our results. We shall not consider prognostic factors either. This paper starts with a discussion, in Section 2, of "goodness" of designs for comparing two treatments. Section 3 describes how a randomized sequential experiment forces to use a conditional inference setting and we introduce the concept of "desirable" procedure namely one that assigns a desired proportion of allocations of the two treatments. Section 4 deals with the problem of finding a randomized design which converges almost certainly to the desirable proportion when this proportion is known, and also forces the procedure towards such an allocation for small samples too. This allows us to use non-adaptive sequential procedures: in this case the randomized design is ancillary (Cox and Hinkley, 1975) which, by the ancillarity principle, provides a further justification for a conditional inference setting. Therefore in Section 4 we propose a generalization of a convergence theorem for randomized sequential designs and use it in Section 5 to show that when balance is desirable Wei's design (1978) and Efron's Biased Coin Design (1971) are asymptotically desirable. Section 6 discuss some extensions for a general desirable proportion. Finally, Section 7 shows how the Adjustable Biased Coin Design proposed by Baldi Antognini and Giovagnoli (2003) converges to the desirable proportion more rapidly than all the other coin designs considered so far.

2 Optimal designs for the allocation of two treatments

Let T_1 and T_2 be the treatments to compare. Suppose we allocate either of them sequentially to the experimental units, observing each time an outcome Z. The allocation at step k may depend, in a deterministic or radomized way, on the past allocations and/or on the past observations of the outcomes up to step $k-1$. Let $\delta_k = 1$ if the k-th unit is allocated to T_1, 0 otherwise. Also, $N_1(k)$ and $N_2(k)$ will denote the number of the first k observations that are allocated to T_1 and to T_2, respectively; hence $N_1(k) = \sum_{i=1}^{k} \delta_i$ and $N_2(k) = \sum_{i=1}^{k}(1-\delta_i) = k - N_1(k)$; $\pi_k = k^{-1}\sum_{i=1}^{k} \delta_i$ will be the proportion of allocations to T_1 (symmetrically, $1-\pi_k$ to T_2) after k observations. Let X_k, Y_k be the potential responses to the two treatments,

assume that $\{(X_k, Y_k)\}_{k\in\mathbb{N}}$ are *i.i.d.* and the distributions of X and Y depend on the vector parameters θ_1, θ_2 respectively

$$X_i \sim \mathcal{L}_{\theta_1}(X) \text{ and } Y_i \sim \mathcal{L}_{\theta_2}(Y) \quad \forall i \geq 1.$$

At stage k exactly one of X_k, Y_k is observed and the observed response Z_k can be represented by $Z_k = \delta_k X_k + (1-\delta_k)Y_k$, with a probability model

$$\mathcal{L}_\theta(Z_k|\delta_k) = \begin{cases} \mathcal{L}_{\theta_1}(X) & \text{if } \delta_k = 1 \\ \mathcal{L}_{\theta_2}(Y) & \text{if } \delta_k = 0 \end{cases}, \quad \forall k \geq 1 \tag{2.1}$$

where $\theta = (\theta_1, \theta_2) \in \Theta \subseteq \mathbb{R}^t$. In many instances there is a given proportion π^* (to T_1) which can be regarded as optimal by the experimenter. For example, assuming a linear homoscedastic model for the responses X and Y, if we are interested in estimating the difference between the mean responses with maximum precision, a balanced design (i.e. $\pi^* = 1/2$) minimizes the variance of the OLS estimated difference; in case of heteroschedasticity of the observations, let σ_1 and σ_2 be the standard deviations corresponding to the two treatments, the proportion which minimizes the estimated difference variance is $\pi^* = \sigma_1/(\sigma_1+\sigma_2)$ (the so-called Neyman allocation), which may be unknown or known on the basis of some previous experiment. For independent binary responses let p_1 and p_2 denote the success probabilities corresponding to the two treatments, if we take the p_j $(j = 1, 2)$ to be the unknown parameters and we are interested in estimating the difference between the two success probabilities, then the proportion which minimizes the estimated difference variance is unknown, $\pi^* = \dfrac{\sqrt{p_1(1-p_1)}}{\sqrt{p_1(1-p_1)} + \sqrt{p_2(1-p_2)}}$, because of its dependence on the parameters.

In this paper the word *optimal* is taken in Kiefer's sense (see Silvey, 1980) to refer just to some inferential aspect. This is not the only way to decide that a given proportion is optimal from the experimenter's point of view: there may be ethical reasons for assining one treatment more often than the other, i.e. urn designs (Bandyopadhyay and Biswas, 2000), two-armed bandit problems (Berry and Fristedt, 1985) etc... It is not however the viewpoint of the present paper, where we said that the aim of the experiment is inference without ethical issues regarding the subjects' health being involved.

According to Kiefer's theory, to find the optimal design we may look at the average Fisher information matrix. Let $L(\theta|\cdot)$ denote the likelihood function, under the hypothesis of independence of the observations conditionally on the design we obtain

$$M(\theta) = \frac{1}{n}\sum_{i=1}^{n} E\left[I(\theta|\delta_i)\right] = \frac{1}{n}\sum_{i=1}^{n} E\left[E\left(\frac{-\partial^2 \log L(\theta|Z_i, \delta_i)}{\partial\theta_i \partial\theta_j}|\delta_i\right)\right]$$

where $I(\theta|\delta_i)$ is the conditional Fisher information. The expected value inside the ( ) brackets is taken with respect to the conditional models, namely (2.1), while that in [ ] refers to the distribution of the design $\delta^{(n)}$. So, to calculate $M(\theta)$ we must know all the process $\left\{Z^{(n)}, \delta^{(n)}\right\}$: even in the simplest cases this turns out to be prohibitive.

In the contest of the sequential allocation of two treaments several authors (see for instance Robbins, Simons and Starr (1967), Eisele (1994), Melfi, Page and Geraldes (2001)) adopt a conditional inference approach, assuming the sequence of allocations as predetermined. The (conditional) average Fisher information

$$M\left(\theta|\delta^{(n)}\right) = \frac{1}{n}\sum_{i=1}^{n} I\left(\theta|\delta_i\right) = \frac{1}{n}\sum_{i=1}^{n} E\left[\frac{-\partial^2 \log L\left(\theta|Z_i,\delta_i\right)}{\partial\theta_i\partial\theta_j}\,|\delta_i\right] \tag{2.2}$$

is well known to be asymptotically equal to the inverse of the asymptotic variance-covariance of MLE's conditional on the design. Thus, (2.2) is a measure of the precision of our ML estimates.

3 Sequential randomized designs and desirable proportion

Suppose now that subjects arrive sequentially at an experimental site and are assigned immediately to either treatment. We may want to decide each assignment on the basis of the outcome of the previous ones. We may also want introduce a component of randomization in the assignments to protect against several types of bias, including selection bias. Let $\{\Im_n, n \geq 1\}$ be an increasing sequence of σ-algebras such that (X_k, Y_k) is $\Im_k$-measurable for every k and let:

$$Z^{(n)} = (Z_1, Z_2, \ldots, Z_n) \quad \text{and} \quad \delta^{(n)} = (\delta_1, \delta_2, \ldots, \delta_n) \quad \forall n \geq 1.$$

A sequential randomized design consists of a sequence of random variables $\delta_1, ..., \delta_n, ...$ where δ_n depends on the history of the process $\left\{Z^{(n)}, \delta^{(n)}\right\}$. The design can be represented by δ_1 and the sequence of conditional probabilities:

$$\Pr\left(\delta_{k+1} = 1|\Im_k\right) = \Pr\left(\delta_{k+1} = 1|\delta^{(k)}, Z^{(k)}\right) \quad \forall k \geq 1.$$

For any fixed sample size n, by (2.1) it follows that

$$M\left(\theta|\delta^{(n)}\right) = diag\left(\pi_n i\left(\theta_1\right), (1-\pi_n) i\left(\theta_2\right)\right), \tag{3.1}$$

where

$$i\left(\theta_1\right) = E\left[\frac{-\partial^2 \log L\left(\theta_1|X\right)}{\partial\theta_{1h}\partial\theta_{1k}}\right] \quad \text{and} \quad i\left(\theta_2\right) = E\left[\frac{-\partial^2 \log L\left(\theta_2|Y\right)}{\partial\theta_{2h}\partial\theta_{2k}}\right]$$

are the Fisher information matrices generated by an observation on T_1 and T_2, respectively. Thus the (conditional) average Fisher information at step n depends on the design only through the proportion π_n and we can write $M\left(\theta|\pi_n\right)$. In the theory of optimal designs, an optimality criterion is a real function ϕ which measures "lack of precision" of the experiment. Accordingly, it is taken to be a continuous, bounded and strictly convex function defined on the variance-covariance matrices of the parameters estimates; extending the definition to the present setting we take ϕ to be defined on the set of all $t \times t$ non-singular square matrices.

Definition 1 *The proportion* ${}_{\phi}\pi^*(\theta) \in (0;1)$ *of allocations to* T_1 *is said to be desirable, if it is* $\phi-$*optimal in the Kiefer sense, for conditional inference, namely if and only if:*

$$ {}_{\phi}\pi^*(\theta) = \arg \max_{\pi \in (0;1)} \phi \left[M\left(\theta|\pi\right) \right]. \tag{3.2}$$

We recall that:

1. $M\left(\theta|\pi\right)$ is obtained replacing π_n in (3.1) with a generic $\pi \in (0;1)$.
2. Since a sequential randomized design is a stochastic process, for any given sample size n, $N_1\left(n\right)$ and $N_2\left(n\right)$ are random variables.

Definition 2 *A sequential randomized design is said to be asymptotically desirable if and only if:*

$$\lim_{n\to\infty} \frac{N_1\left(n\right)}{n} = {}_{\phi}\pi^*(\theta) \quad \textit{almost certainly} \quad \forall \theta \in \Theta.$$

4 Asymptotically desirable designs for a known proportion

From now on we will drop the reference to ϕ. In general $\pi^*(\cdot)$ is a known function of the unknown population parameters. However in some cases the conditional information matrix may not depend on the parameters of interest, or the adopted criterion may be such that this dependence cancels in (3.2). An example is D-optimality where, for the previous assumptions:

$$\pi^*(\theta) = \arg \max_{\pi \in (0;1)} \det M\left(\theta|\pi\right) = \frac{1}{2} \quad \forall \theta \in \Theta.$$

Thus the desirable proportion may be known *a priori*

$$\pi^*(\theta) = \pi_0 \quad \forall \theta \in \Theta.$$

When $\pi^*(\theta)$ is unknown because of its dependence on θ, a natural approach is the use of adaptive algorithms, namely to let δ_{k+1} depend on the previous allocations and observations $\left\{Z^{(k)}, \delta^{(k)}\right\}$. Otherwise it does not seem necessary to use adaptive procedures.

Suppose now that the desirable proportion is a known constant π_0 and let

$$P\left(\delta_{k+1} = 1 | \delta^{(k)}, Z^{(k)}\right) = \Phi_{\pi_0}\left(\delta^{(k)}\right), \quad \forall k \geq 1$$

be our non-adaptive sequential algorithm. The most natural procedure - in the absence of further experimental aims - consists in assigning at each step treatment T_1 with probability π_0, independently from the previous allocations:

$$\Phi_{\pi_0}\left(\delta^{(k)}\right) = \pi_0 \quad \forall k \geq 1, \tag{4.1}$$

This algorithm represents a generalization of the completely randomized design when $\pi_0 = 1/2$. In this case $\delta_1, \delta_2, \ldots, \delta_n, \ldots$ is a sequence of indipendent and identically distributed random variables and by the strong law of large numbers the design generated by (4.1) is asymptotically desirable:

$$\lim_{n\to\infty} \frac{N_1(n)}{n} = \pi_0 \quad a.s.$$

However, since in many cases we do not know when the experiment is going to stop, there is a need for *ad hoc* designs that work sufficiently well for small samples too. Thus it may seem natural to impose that the desing at each step be as close as possible to the desirable proportion π_0, however this would imply a deterministic decision through out, which is adverse to the randomization demands. In a sequential setting, the compromise consists in considering a randomized design which at each step is, with a high probability, near the desirable proportion: this allows us to stop the experiment at any time and find ourselves, with very high probability, in an excellent setting for inferential purposes.

4.1 A general asymptotic result

Definition 3 *Let $g : [0;1] \to [0;1]$. A point $x_0 \in [0;1]$ is a downcrossing of $g(\cdot)$ if:*

$$\forall x < x_0 \quad g(x) \geq x_0 \quad \text{and} \quad \forall x > x_0 \quad g(x) \leq x_0$$

Theorem 4 *Suppose the set of discontinuities of function $g : [0;1] \to [0;1]$ is nowhere dense and define a sequential design by $\Phi_{\pi_0}\left(\delta^{(n)}\right) = g\left(\frac{1}{n}\sum_{i=1}^{n}\delta_i\right)$. If the point $\pi_0 \in (0;1)$ is the only downcrossing of $g(\cdot)$, then*

$$\lim_{n\to\infty} \frac{N_1(n)}{n} = \pi_0 \quad a.s.$$

Proof. *By Corollary 2.1 and Theorem 4.1 of Hill, Suddert and Lane (1980)* ■

Corollary 5 *Let $D \subseteq \mathbb{R}$ and $h : D \to [0;1]$ be a non-increasing function with a set of discontinuities nowhere dense and let $k : [0;1] \to D$ be a continuous and strictly increasing function. Define $\Phi_{\pi_0}(\delta^{(n)}) = h \circ k(\frac{1}{n}\sum_{i=1}^{n}\delta_i)$. If $\exists d^* \in D$ such that $h(d^*) = k^{-1}(d^*) \in (0;1)$, then:*

$$\lim_{n\to\infty} \frac{N_1(n)}{n} = k^{-1}(d^*) \quad a.s.$$

Proof. Let $g = h \circ k : [0;1] \to [0;1]$. Then g is a non-increasing function with a nowhere dense set of discontinuities. Put $x^* = k^{-1}(d^*)$, we obtain $x^* = h(d^*) = h \circ k(x^*) = g(x^*)$. So $g(\cdot)$ has a single downcrossing at x^* ■

5 The special case of balance

It is well-known that in many cases the 'optimal' proportion is $\pi_0 = 1/2$, thus we want to show first of all that the various biased coin designs that have been suggested in the literature are asymptotically balanced.

5.1 Wei's Designs

Let $D_n = N_1(n) - N_0(n)$ be the difference between the two groups after n assignments. Let $f : [-1, 1] \to [0, 1]$ be a decreasing function such that $f(-x) = 1 - f(x)$. Wei's designs are defined by:

$$\Phi_{\frac{1}{2}}\left(\delta^{(k)}\right) = f\left(\frac{D_k}{k}\right) \quad \forall k \geq 1. \tag{5.1}$$

A special cases is Efron' s Biased Coin Design BCD(p), a sequential allocation rule randomized by means of the hypothetical tossing of a biased coin with probability $p \in \left[\frac{1}{2}; 1\right]$ which at each step favours the treatment so far under-represented. Efron suggests:

$$\Phi_{\frac{1}{2}}\left(\delta^{(k)}\right) = \begin{cases} p & \text{if} \quad D_k < 0 \\ \frac{1}{2} & \text{if} \quad D_k = 0 \\ 1-p & \text{if} \quad D_k > 0 \end{cases} \qquad \forall k \geq 1.$$

Thus BCD(p) is obtained putting $f(x) = \frac{1}{2} + sgn(x)(\frac{1}{2} - p)$ in (5.1). Special cases are also the two "coin" designs proposed by Atkinson (1982):

$$\begin{aligned} \text{Atkinson}^1 \quad f(x) &= \frac{1-x}{2} \\ \text{Atkinson}^2 \quad f(x) &= \frac{(1-x)^2}{(1-x)^2 + (1+x)^2} \end{aligned}$$

Proposition 6 *If the set of discontinuities of $f(\cdot)$ in (5.1) is nowhere dense, then:*

$$\lim_{n\to\infty} \frac{D_n}{n} = 0 \quad a.s.$$

Proof. Now we let $g = f \circ k : [0; 1] \to [0; 1]$ with $k : [0; 1] \to [-1; 1]$ and d^* defined as above: $k(x) = 2x - 1$ and $d^* = 0$. Since $k(\cdot)$ is continuous and strictly increasing, $g(\cdot)$ is non increasing with a set of discontinuities nowhere dense. Then by Corollary 5

$$\lim_{n\to\infty} \frac{N_1(n)}{n} = \frac{1}{2} \quad a.s.$$

■

Corollary 7 *Efron's Biased Coin Design is asymptotically desirable for $\pi_0 = \frac{1}{2}$.*

Proof. This follows since $f(x) = \frac{1}{2} + sgn(x)(\frac{1}{2} - p)$ has just one point of discontinuity ■

6 Possible extensions to a general desirable proportion

Suppose now that the desirable proportion is any constant $\pi_0 \in (0; 1)$ and consider the problem of finding a randomized design which converges to π_0 almost certainly

and also forces the procedure towards the desirable proportion for small samples too. The designs proposed by Efron and by Wei can be extended as follows.

π_0-**BCD**(a,b)

Possible extensions of Efron's coin are:

$$\Phi^{\star}_{\pi_0}\left(\delta^{(k)}\right)=\begin{cases} a & \text{if} \quad \dfrac{D_k}{k}<2\pi_0-1 \\ \pi_0 & \text{if} \quad \dfrac{D_k}{k}=2\pi_0-1 \\ b & \text{if} \quad \dfrac{D_k}{k}>2\pi_0-1 \end{cases},$$

where $0\leq b\leq \pi_0\leq a\leq 1$.

Proposition 8 *π_0-BCD(a,b) is asymptotically desirable.*

Proof. By Corollary 5 ■

π_0-**Wei's design**

Let $f:[-1,1]\rightarrow[0,1]$ be a non-increasing function with the set of discontinuities nowhere dense and such that $f(-x)=1-f(x)$. Define a new function $f^{\star}:[-1,1]\rightarrow[0,1]$ as:

$$f^{\star}(x)=\begin{cases} 2(1-\pi_0)f\left(\dfrac{x-(2\pi_0-1)}{2\pi_0}\right)+2\pi_0-1 & \text{if} \quad -1\leq x\leq 2\pi_0-1 \\ 2\pi_0 f\left(\dfrac{x-(2\pi_0-1)}{2(1-\pi_0)}\right) & \text{if} \quad 2\pi_0-1\leq x\leq 1 \end{cases}$$

and let:

$$\Phi_{\pi_0}\left(\delta^{(k)}\right)=f^{\star}\left(\frac{D_k}{k}\right) \quad \forall k\geq 1.$$

Observe however that by taking $f(x)=\frac{1}{2}+sgn(x)(\frac{1}{2}-p)$ we obtain a special case of the previously defined π_0-BCD(a,b)

$$f^{\star}(x)=\begin{cases} 2\pi_0(1-p)+2p-1 & \text{if} \quad -1\leq x\leq 2\pi_0-1 \\ 2\pi_0(1-p) & \text{if} \quad 2\pi_0-1\leq x\leq 1 \end{cases}$$

Proposition 9 *π_0-Wei's design is asymptotically desirable.*

Proof. The set of discontinuities of $f^{\star}(\cdot)$ is nowhere dense, $f^{\star}(\cdot)$ is non increasing and satisfying: $f^{\star}(2\pi_0-1)=\pi_0 \ \ \forall\pi_0\in(0;1)$. Then the point π_0 is the only downcrossing of $g(t)=f^{\star}(2t-1)\ \forall t\in[0,1]$ and we can apply Corollary 5 ■

7 The Adjustable Biased Coin Design (ABCD)

The ABCD was proposed in Baldi Antognini and Giovagnoli (2003) and consists in making $\Phi_{\pi_0}\left(\delta^{(n)}\right)$ a decreasing function of D_n, so that the tendency towards balance is stronger the more we move away from it.

Let $\overline{F}(\cdot) : \mathbb{R} \to [0,1]$ be a non-decreasing function such that

$$\overline{F}(-x) = 1 - \overline{F}(x).$$

Such a function $\overline{F}(\cdot)$ generates an ABCD letting:

$$\Phi_{\frac{1}{2}}\left(\delta^{(k)}\right) = \bar{F}\left(D_k\right) \quad \forall k \geq 1.$$

Efron's Biased Coin Design is obviously a special case. The Adjustable Biased Coin Design possesses many desirable properties, for instance because of its greater flexibility an ABCD can be chosen that turns out, for all n, to be less predictable and more balanced than the other existing procedures (Baldi Antognini, Bodini and Giovagnoli, 2002; Baldi Antognini and Giovagnoli, 2003). The ABCD is asymptotically desirable for balance; in fact, a stronger result holds.

Theorem 10 *Define* $\Phi_{\frac{1}{2}}\left(\delta^{(n)}\right) = \bar{F}\left(D_n\right)$ *as above, then:*

$$\lim_{n\to\infty} \frac{D_n}{\sqrt{n}} = 0 \quad a.s.$$

Proof. See Baldi Antognini and Giovagnoli (2003). ■

8 Conclusions

The present paper deals with optimal (in Kiefer's sense) sequential randomized designs for parametric inference on two treatments. The problem is finding a randomized design which converges to the desirable one almost certainly and also forces the procedure towards the desirable proportion even for small samples. When balance is optimal we show that Efron's Biased Coin Design (1971), the class of Wei's designs (1978) and the ABCD (Baldi Antognini and Giovagnoli, 2003) are asymptotically desirable. I propose some extensions of the above mentioned algorithms that converge almost surely to any desired proportion, when the value is known. A natural extension of the Adjustable Biased Coin Design seems to be to take $\Phi_{\pi_0}\left(\delta^{(k)}\right)$ a decreasing function of the distance between D_k and $2\pi_0 - 1$. It is still unclear whether the asymptotic desiderability of this design holds and what the performance for small samples of the π_0-ABCD is. These and other connected topics seems to be worth investigating.

Acknowledgment

The author received a "Young Researchers" grant for the year 2002 from the University of Bologna and financial support also from the 2000-2002 Research Project of the Department of Statistical Sciences, University of Bologna: "Equilibrium distributions of Markov chains: methodological tools and biostatistical and industrial applications". The author wishes to thank Prof. Alessandra Giovagnoli for helpful advice.

References

[1] Atkinson, A.C. (1982): Optimum biased coin designs for sequential clinical trials with prognostic factors. *Biometrika*, **69**, 61-67.

[2] Baldi Antognini, A., Bodini, A., and Giovagnoli, A. (2002): Nuovi indicatori di sbilanciamento e prevedibilità nei disegni sequenziali randomizzati: confronti fra "biased coin designs" diversi. *Statistica,* in press.

[3] Baldi Antognini, A. and Giovagnoli, A. (2003): A new "biased coin design" for the sequential allocation of two treatments. *Journal of the Royal Statistical Society* D (in press).

[4] Bandyopadhyay, U. and Biswas, A. (2000): A class of adaptive designs. *Sequential Analisys,* **19**, 45-62.

[5] Berry, D.A. and Fristedt, B. (1985): *Bandit problems: sequential allocation of experiments.* London: Chapman and Hall.

[6] Cox, D.R. and Hinkley, D.V. (1975): *Theoretical Statistics.* London: Chapman & Hall.

[7] Efron, B. (1971): Forcing sequential experiments to be balanced. *Biometrika*, **58**, 403-17.

[8] Eisele, J.R. (1994): The doubly adaptive biased coin design for sequential clinical trials. *Journal of Statistical Planning and Inference*, **38**, 249-62.

[9] Hill, B.M., Suddert, W., and Lane, D. (1980): A strong law for some generalized urn processes. *Annals of Probability*, **8**, 214-26.

[10] Melfi, V., Page, C., and Geraldes, M. (2001): An adaptive randomized design with application to estimation. *Canadian Journal of Statistics*, **29**, 107-16.

[11] Robbins, H., Simons, G., and Starr, N. (1967): A sequential analogueof the Behrens-Fisher problem. *Annals of Mathematical Statistics*, **38**, 1384-91.

[12] Silvey, S.D. (1980): *Optimal Designs.* London: Chapman and Hall.

[13] Smith, R. (1984a): Properties of biased coin designs in sequential clinical trials. *Annals of Mathematical Statistics*, **12**, 1018-34.

[14] Smith, R. (1984b): Sequential treatment allocation using biased coin designs. *Journal of the Royal Statistical Society* B, **46**, 519-43.

[15] Wei, L.J. (1978): The adaptive biased coin design for sequential experiments. *Annals of Statistics*, **6**, 92-100.

[16] Wei, L.J. and Durham, S. (1978): The randomized play-the-winner rule in medical trials. *Journal of the American Statistical Association*, **73**, 840-43.

[17] Zelen, M. (1969): Play the winner rule and the controlled clinical trials. *Journal of the American Statistical Association*, **64**, 131-46.