THE DOUBLY ADAPTIVE BIASED COIN DESIGN
FOR SEQUENTIAL CLINICAL TRIALS

Jeffrey R. Eisele

MADISON, WISCONSIN
THE DOUBLY ADAPTIVE BIASED COIN DESIGN
FOR SEQUENTIAL CLINICAL TRIALS

By

Jeffrey R. Eisele*

Department of Biostatistics
University of Wisconsin
Madison, WI 53706

ABSTRACT

A biased coin design is proposed for the allocation of subjects to treatments A and B in a clinical trial when the desired allocation proportions are unknown. The design is doubly adaptive in the sense that it takes account of the proportion of subjects assigned to each treatment and the current estimate of the desired allocation proportion. A strong law of large numbers is established for the proportion of subjects assigned to a treatment when subject responses are independent random variables from standard exponential families. The normal case is presented as an application.

AMS Subject Classification: Primary 62L05; Secondary 60F15.

Key words and phrases: Biased coin designs, exponential families, sequential procedure, stopping rule, strong law of large numbers.

* Research supported by the U.S. Army under DAAL-03-88-0122.
1. INTRODUCTION

Suppose subjects arrive sequentially at an experimental site and are assigned immediately to one of two treatments groups A or B. A statistical design problem is how to assign patients to the treatment groups. When balance is desired in the allocation, Efron (1971) and Wei (1978) proposed subject assignment algorithms offering a compromise between complete randomization and perfect balance. These designs achieve balance more quickly than complete randomization, but retain enough randomization to preclude effective guessing of the next treatment to be assigned. Letting +1 and -1 denote assignments to treatments A and B, these algorithms may be described as follows: let \( q \) denote a nonincreasing function from \([-1, 1]\) into \([0, 1]\) for which \( q(0) = 1/2 \); let \( U_1, U_2, \ldots \) denote i.i.d uniformly distributed random variables; and let

\[
Z_k = 2I\left\{ U_k \leq q\left( \frac{S_k-1}{k-1} \right) \right\} - 1, \ \forall \ k = 1, 2, \ldots
\]

where

\[
S_n = Z_1 + \cdots + Z_n, \ \forall \ n = 1, 2, \ldots
\]

\( S_0 = 0 \), 0/0 is to be interpreted as 0, and \( I\{\bullet\} \) denotes the indicator of \( \{\bullet\} \). With these conventions, \( S_n \) is the difference between the number of subjects assigned to treatment group A and the number assigned to treatment group B after \( n \) assignments. Both algorithms are called biased coin designs. They differ in the nature of the function \( q \).

In many problems, balance is not desired. In fact, the desired allocation proportions may be unknown. The purpose of this paper is to present a biased coin design for the allocation of patients to the treatment groups, A and B, in the case where the desired allocation proportion is unknown. The design is doubly adaptive in the
sense that it takes account of both the current proportion of subjects assigned to each
treatment and a current estimate of the desired allocation proportion. It is shown
that the proportion of subjects assigned to a treatment converges almost surely to the
desired proportion to be allocated to that treatment. This design was first presented
in Eisele (1990) in the case where subject responses are normally distributed. In that
paper, Eisele superimposed a biased coin on Robbins, Simons, and Starr’s (1967)
sequential analogue of the Behrens-Fisher problem. Imposition of the biased coin did
not change the asymptotic optimality of the sequential design.

The general model and allocation rule are described in Section 2. Section 3
presents a result on the probability of extreme imbalance which will be needed for
the proof of the strong law. Moment bounds are given in Section 4 and the strong
law of large numbers is proved in Section 5. Section 6 presents as an example the

2. GENERAL MODEL AND ALLOCATION RULE

2.1 General model

Suppose subject responses $X_1, X_2, \ldots$ to treatment $A$ and $Y_1, Y_2, \ldots$ to treatment $B$
are independent random variables from $d$-dimensional standard exponential families.
More formally, assume that

$$X_1, X_2, \ldots \sim f_\theta(x) = \exp(\theta \cdot x - \psi(\theta))$$

and

$$Y_1, Y_2, \ldots \sim g_\omega(y) = \exp(\omega \cdot y - \varphi(\omega)),$$

where $\theta = (\theta_1, \ldots, \theta_d)$, $x = (x_1, \ldots, x_d)$, $\omega = (\omega_1, \ldots, \omega_d)$, $y = (y_1, \ldots, y_d)$, and \ldots
denotes the inner product. Let
\[ \mu = \mathbb{E}_\theta X = \nabla \psi(\theta) \text{ and } \nu = \mathbb{E}_\omega Y = \nabla \varphi(\omega), \]
where \( \nabla \psi \) denotes the gradient vector \( (\partial \psi/\partial \theta_1, \ldots, \partial \psi/\partial \theta_d) \) and \( \nabla \varphi \) denotes the gradient vector \( (\partial \varphi/\partial \omega_1, \ldots, \partial \varphi/\partial \omega_d) \). Let \( m_k \) and \( n_k \) be the number of observations on \( X \) and on \( Y \), respectively, at time \( k \), where \( k = m_k + n_k \), and let
\[
\overline{X}_{m_k} = (m_k)^{-1} \sum_{i=1}^{m_k} X_i \text{ and } \overline{Y}_{n_k} = (n_k)^{-1} \sum_{i=1}^{n_k} Y_i
\]
be the sample means. If the families are steep, then the MLE of \( \mu \) is \( \overline{X}_{m_k} \) and the MLE of \( \nu \) is \( \overline{Y}_{n_k} \), since the likelihood function is not affected by the sequential design. Brown (1986) is recommended for more background on exponential families.

The goal of the allocation scheme is then to have \( m_k/k = \rho \), where \( m_k/k \) is the proportion of subjects assigned to treatment \( A \) at time \( k \) and \( \rho \in \mathbb{R}^{2d} \to [0, 1] \) is the desired allocation proportion. To accomplish this, the allocation scheme is designed to sample the \( X \) population with probability less than (respectively, greater than) \( \hat{\rho}_k \) when \( m_k/k > \hat{\rho}_k \) (respectively, \( m_k/k < \hat{\rho}_k \)), where \( \hat{\rho}_k \equiv \rho \left( \overline{X}_{m_k}, \overline{Y}_{n_k} \right) \in [0, 1] \) is the current maximum likelihood estimate of \( \rho \). This employs the same idea as Wei's adaptive biased coin design except that he uses \( \rho = 1/2 \).

2.2 The allocation rule

Let \( q \) be a function from \([0, 1]^2 \to [0, 1]\) such that the following four conditions hold

(i) \( q \) is jointly continuous,

(ii) \( q(r, r) = r \),

(iii) \( q(p, r) \) is strictly \( \searrow \) in \( p \) and strictly \( \nearrow \) in \( r \) on \([0, 1]^2\),

(iv) \( q \) has bounded derivatives in both arguments.
Let \( \delta_1 = \cdots = \delta_{n_0} = 1, \delta_{n_0+1} = \cdots = \delta_{2n_0} = 0 \), and
\[
\delta_k = I \left\{ U_k \leq q \left( \frac{m_{k-1}}{k-1}, \hat{\rho}_{k-1} \right) \right\}, \quad \forall k \geq 2n_0 + 1.
\]
where the \( U_k \) are independent of \( X_1, X_2, \ldots \) and \( Y_1, Y_2, \ldots \). Then
\[
m_k = \delta_1 + \cdots + \delta_k \text{ and } n_k = k - m_k.
\]

This sampling scheme has the desirable property that it provides some randomization in an effort to reduce the possibility of experimenter bias. Although experimenter bias is not completely eliminated, the best guessing strategy at time \( k + 1 \) has a probability of success equal to the \( \max (q_k, 1 - q_k) \), where
\[
q_k \equiv q \left( \frac{m_{k-1}}{k-1}, \hat{\rho}_{k-1} \right),
\]
for \( k \geq 2n_0 + 1 \). It should be noted that \( X \) and \( Y \) are interchangeable (with probability one).

2.3 Conditions on \( \rho \)

In addition to conditions (i)-(iv) imposed on \( q \), the following two conditions on \( \rho \) are needed.

(v) There are constants \( C \) and \( r \) for which
\[
\frac{1}{\rho} + \frac{1}{1 - \rho} \leq C \left\{ \| \mu \|_r + \| \nu \|_r \right\}.
\]

(vi) For sufficiently small \( \epsilon > 0 \), \( \rho \) is continuously differentiable on the set
\[
R = \{(x, y) : \| x - \mu \| \leq \epsilon, \| y - \nu \| \leq \epsilon\}.
\]
2.4 Notation

The following notation and preliminary results will prove useful in the next three sections in showing that

$$\lim_{k \to \infty} \frac{m_k}{k} = \rho \text{ w.p.1.}$$

Let

$$S_k = m_k - k\rho, \ \forall k = 2n_0 + 1, 2n_0 + 2, \ldots.$$ 

Then

$$S_k = Z_1 + \ldots + Z_k.$$ 

where

$$Z_k = \delta_k - \rho.$$ 

$S_k$ is now the difference between the number of subjects assigned to treatment $A$ and the desired number to be assigned to treatment $A$ after $k$ assignments. Let

$$\mathcal{F}_{k-1} \equiv \sigma \big\{Z_1, \ldots, Z_{k-1}, X_1, \ldots, X_{m_{k-1}}, Y_1, \ldots, Y_{n_{k-1}}\}$$

be the $\sigma$-algebra representing the natural history. Then the conditional mean and variance of $Z_k$ given $\mathcal{F}_{k-1}$ are

$$\mu_k = \mathbb{E} \{Z_k / \mathcal{F}_{k-1}\} = q_k - \rho$$

and

$$\sigma_k^2 = \mathbb{E} \{(Z_k - \mu_k)^2 / \mathcal{F}_{k-1}\} = q_k (1 - q_k).$$

3. ON THE PROBABILITY OF EXTREME IMBALANCE

The following theorem on the probability of extreme imbalance is needed for the proof of Lemma 3 in Section 4. Lemmas 1 and 2, following Theorem 1, are part of the proof Theorem 1. It is in the proof of Lemma 1 that condition (v) is used.
\textbf{Theorem 1} There is an $0 < \varepsilon_0 < 1$ for which

$$P \left\{ m_k < \varepsilon k \right\} = o \left( k^{-\alpha} \right), \quad \forall \alpha > 0 \text{ as } k \to \infty,$$

for all $0 < \varepsilon \leq \varepsilon_0$.

\textbf{Proof.} Let

$$M_k = \sum_{j=1}^{k} (\delta_j - q_j) = m_k - \sum_{j=1}^{k} q_j$$

for $k = 1, 2, \ldots$. Then $M_k$, $k = 1, 2, \ldots$, is a martingale. If $0 < \varepsilon < 1/4$ and $1 \leq l < \varepsilon k$, then

$$\{l \leq m_k \leq \varepsilon k\} \subseteq \left\{ l \leq m_k \leq \varepsilon k, \sum_{j=1}^{k} q_j \leq 2\varepsilon k \right\} \cup \{ |M_k| > \varepsilon k \}.$$

Now, $m_j/j \leq 2\varepsilon$ for all $k/2 < j \leq k$ on the event $\{m_k \leq \varepsilon k\}$, so that on $\{m_k \leq \varepsilon k\}$

$$\sum_{j=1}^{k} q_j \geq \sum_{k/2 < j \leq k} q \left( \frac{m_{j-1}}{j-1} , \hat{\rho}_j \right)$$

$$\geq \sum_{k/2 < j \leq k} q \left( 2\varepsilon, \hat{\rho}_j \right) \geq \frac{1}{3} k q \left( 2\varepsilon, \hat{\rho}_l \right)$$

where

$$\hat{\rho}_l = \inf \left\{ \rho \left( \overline{X}_m, \overline{X}_n \right) : m, n \geq l \right\}.$$

So, $l \leq m_k \leq \varepsilon k$ and $\sum_{j=1}^{k} q_j \leq 2\varepsilon k$ require that $q \left( 6\varepsilon, \hat{\rho}_l \right) \leq q \left( 2\varepsilon, \hat{\rho}_l \right) \leq 6\varepsilon$ and, therefore, that $\hat{\rho}_l \leq 6\varepsilon$ for all sufficiently large $k$. Thus,

$$P \left\{ l \leq m_k \leq \varepsilon k \right\} \leq P \left\{ \hat{\rho}_l \leq 6\varepsilon \right\} + P \left\{ |M_k| > \varepsilon k \right\}$$

for all $0 < \varepsilon < 1/4, l < \varepsilon k$, and all sufficiently large $k$. By Lemmas 1 and 2, below, there are constants $C$ and $\eta$ for which

$$P \left\{ \hat{\rho}_l \leq 6\varepsilon \right\} \leq C \exp \left\{ - \left( \frac{1}{C\varepsilon} \right)^{1/4} \eta \sqrt{l} \right\}$$
and\\\[ \mathbb{P} \{ | M_k | \geq \epsilon k \} \leq C \exp \left\{ -\frac{1}{2C^2} \epsilon^2 \right\} \]

for all \( 1 \leq l < \epsilon k \) for all sufficiently large \( k \) and sufficiently small \( \epsilon > 0 \), say \( 0 < \epsilon < \epsilon_0 \).

Setting \( l = 1 \) and \( \epsilon = k^{-1/4} \) shows that

\[
\mathbb{P} \left\{ m_k \leq k^{3/4} \right\} \leq \mathbb{P} \{ \hat{\rho}_l \leq 6\epsilon \} + \mathbb{P} \left\{ | M_k | \geq \epsilon k \right\} \\
= o \left( k^{-\alpha} \right) \text{ as } k \to \infty. \quad \forall \quad \alpha > 0.
\]

Then setting \( l = \left( k^{3/4} \right) \), where \( \lfloor x \rfloor \) denotes the greatest integer which is \( \leq x \), and letting \( 0 < \epsilon < \epsilon_0 \) be independent of \( k \) shows that

\[
\mathbb{P} \{ m_k \leq \epsilon k \} \leq \mathbb{P} \{ m_k \leq 1 \} + \mathbb{P} \{ l \leq m_k \leq \epsilon k \} \\
\leq \mathbb{P} \{ \hat{\rho}_l \geq 6\epsilon \} + \mathbb{P} \left\{ | M_k | \geq \epsilon k \right\} + o \left( k^{-\alpha} \right) \\
= o \left( k^{-\alpha} \right) + o \left( k^{-\alpha} \right) = o \left( k^{-\alpha} \right)
\]

as \( k \to \infty \) for all \( 0 < \alpha < \infty \).

\[ \square \]

**Lemma 1** There are \( 0 < \eta, C < \infty \) and \( 0 < \epsilon_0 < 1 \) for which

\[
\mathbb{P} \{ \hat{\rho}_l \leq 6\epsilon \} \leq C \exp \left\{ -\left( \frac{1}{C\epsilon} \right)^{\frac{1}{2}} \eta \sqrt{l} \right\}
\]

for all \( 0 < \epsilon < \epsilon_0 \) and all \( l = 1, 2, \ldots \).

**Proof.** By assumption, there are constants \( C \) and \( r \) for which

\[
\frac{1}{\hat{\rho}_l} \leq C \sup_{m,n \geq l} \left\{ \| \overline{X}_m \|^r + \| \overline{Y}_n \|^r \right\}.
\]

Let \( C'' = 12 \left( 1 + \sqrt{d} \right)^r \). Then for all sufficiently small \( \epsilon > 0 \) and \( l = 1, 2, \ldots \),

\[
\mathbb{P} \{ \hat{\rho}_l \leq 6\epsilon \} \leq \mathbb{P} \left\{ \sup_{m,n \geq l} \left( \| \overline{X}_m \|^r + \| \overline{Y}_n \|^r \right) > \frac{1}{6C\epsilon} \right\}
\]
\[
\begin{align*}
&\leq \mathbb{P}\left\{ \sup_{m \geq l} \| X_m \| > (1 + \sqrt{d}) \left( \frac{1}{C' \epsilon} \right)^{\frac{1}{2}} \right\} \\
&\quad + \mathbb{P}\left\{ \sup_{n \geq l} \| Y_n \| > (1 + \sqrt{d}) \left( \frac{1}{C' \epsilon} \right)^{\frac{1}{2}} \right\} \\
&\leq \mathbb{P}\left\{ \sup_{m \geq l} \| X_m - \mu \| > \sqrt{d} \left( \frac{1}{C' \epsilon} \right)^{\frac{1}{2}} \right\} \\
&\quad + \mathbb{P}\left\{ \sup_{n \geq l} \| Y_n - \nu \| > \sqrt{d} \left( \frac{1}{C' \epsilon} \right)^{\frac{1}{2}} \right\} \\
&\leq \sum_{i=1}^{d} \left[ \mathbb{P}\left\{ \sup_{m \geq l} |X_{m,i} - \mu_i| > \left( \frac{1}{C' \epsilon} \right)^{\frac{1}{2}} \right\} \\
&\quad + \mathbb{P}\left\{ \sup_{n \geq l} |Y_{n,i} - \nu_i| > \left( \frac{1}{C' \epsilon} \right)^{\frac{1}{2}} \right\} \right].
\end{align*}
\]

To bound the latter terms, recall that \( X_m, m = 1, 2, \ldots, \) is a reverse martingale.

Let \( \eta \) be so small that
\[
\mathbb{E}\left\{ e^{t X_1} \right\} < \infty
\]
for all \( \| t \| \leq \eta \). Then, by the submartingale inequality
\[
\mathbb{P}\left\{ \sup_{m \geq l} |X_{m,i} - \mu_i| > \left( \frac{1}{C' \epsilon} \right)^{\frac{1}{2}} \right\} \leq 2 \exp \left\{ - \left( \frac{1}{C' \epsilon} \right)^{\frac{1}{2}} \eta \sqrt{l} \right\} \mathbb{E}\left\{ e^{\eta \sqrt{l} |X_{l,i} - \mu_i|} \right\}
\]
and
\[
\mathbb{E}\left\{ e^{\eta \sqrt{l} |X_{l,i} - \mu_i|} \right\}
\]
is bounded in \( l = 1, 2, \ldots \).

\[\square\]

**Lemma 2**

\[
\mathbb{P}\{ |M_k| \geq \epsilon k \} \leq 2 \exp \left\{ - \frac{1}{4} k \epsilon^2 \right\}
\]

for all sufficiently small \( \epsilon > 0 \) and all \( k = 1, 2, \ldots \).

**Proof.** Let \( X_j = \delta_j - q(m_{j-1}/(j-1), \hat{\rho}_{j-1}) \). Then \( M_k = \sum_{j=2m_0+1}^{k} X_j \). The conditional generating function of \( X_j \) is
\[
\psi_j(t) = \mathbb{E}\left\{ e^{t X_j} / \mathcal{F}_{j-1} \right\}.
\]
Let
\[ \Psi_k(t) = \prod_{j=1}^{k} \psi_j(t). \]

Then
\[ \frac{e^{tM_k}}{\Psi_k(t)} = \frac{e^{tM_{k-1}}}{\Psi_{k-1}(t)} \times \frac{e^{tX_k}}{\psi_k(t)}. \]

So that
\[ \mathbb{E} \left\{ \frac{e^{tM_k}}{\Psi_k(t)} \right\} = 1. \]

Also note that
\[ \psi_j(t) = \mathbb{E} \left\{ \sum_{n=0}^{\infty} \frac{1}{n!} (tX_j)^n / \mathcal{F}_{j-1} \right\} \]
\[ \leq 1 + \sum_{n=2}^{\infty} \frac{1}{n!} |t|^n \]
\[ = 1 + t^2 \sum_{n=2}^{\infty} \frac{1}{n!} |t|^{n-2} \]
\[ \leq 1 + t^2 \]
\[ \leq e^{t^2} \]

for all \( 0 < |t| \leq t_0 = \log 2 \). So
\[ \Psi_k(t) \leq e^{kt^2} \text{ and } \mathbb{E} \left\{ e^{tM_k} \right\} \leq e^{kt^2} \]
for all \( |t| \leq t_0 \) and \( k = 1, 2, \ldots \). So, by Bernstein's inequality (see Chow and Teicher, 1988),
\[ \mathbb{P} \{ |M_k| > k\epsilon \} \leq 2 \inf_{t > 0} \exp \left\{ -ket + kt^2 \right\} = 2 \exp \left\{ -\frac{1}{4} k\epsilon^2 \right\}. \]

for all \( 0 < \epsilon \leq \epsilon_1 = 2t_0 \).

4. MOMENT BOUNDS

The proof of Proposition 1 in the next section requires Lemmas 3 and 4 below. It is in the proof of Lemma 4 below that condition (vi) is used.
Lemma 3

\[
\mathbb{E} \left\{ \| X_m_k - \mu \|^{2r} + \| Y_m_k - \nu \|^{2r} \right\} = O \left( \frac{1}{k^r} \right), \ \forall \ r > 0 \ \text{as} \ k \to \infty.
\]

Proof. If \( \epsilon > 0 \) is sufficiently small, then

\[
\mathbb{E} \left\{ \| X_m_k - \mu \|^{2r} \right\} = \int_{m_k \leq \epsilon k} \| X_m_k - \mu \|^{2r} \, d\mathbb{P} + \int_{m_k > \epsilon k} \| X_m_k - \mu \|^{2r} \, d\mathbb{P}
\]

\[
\leq \sqrt{\mathbb{P} \{ m_k \leq \epsilon k \}} \sqrt{\mathbb{E} \left\{ \sup_m \| X_m - \mu \|^{4r} \right\}}
\]

\[
+ \frac{1}{\epsilon^{2r} k^{2r}} \int \left\| \sum_{j=1}^{k} \delta_j (X_j - \mu) \right\|^{2r} \, d\mathbb{P}
\]

\[
= O \left( \frac{1}{k^r} \right) + \frac{1}{\epsilon^{2r} k^{2r}} O \left( k^r \right) = O \left( \frac{1}{k^r} \right),
\]

by Theorem 1 and the fact that \( \sum_{j=1}^{k} \delta_j (X_j - \mu) \) is a martingale. \( \square \)

Lemma 4

\[
\mathbb{E} \left\{ (\hat{\rho}_k - \rho)^{2r} \right\} = O \left( \frac{1}{k^r} \right), \ \forall \ r > 0 \ \text{as} \ k \to \infty.
\]

Proof. Let \( \epsilon > 0 \) be so small that \( \rho \) is continuously differentiable on the set

\[
R = \{ (x, y) : \| x - \mu \| \leq \epsilon, \| y - \nu \| \leq \epsilon \}.
\]

Let

\[
B = \{ \| X_{m_k} - \mu \| \leq \epsilon, \| Y_{m_k} - \nu \| \leq \epsilon \}.
\]

Then

\[
C = \sup_{\mu', \nu' \in R} \left\{ \| \rho_{10} (\mu', \nu') \| + \| \rho_{01} (\mu', \nu') \| \right\} < \infty.
\]

So, letting \( \hat{\mu} \) and \( \hat{\nu} \) denote intermediate points

\[
\mathbb{E} \left\{ (\hat{\rho}_k - \rho)^{2r} \right\}
\]

\[
= \int_{B'} (\hat{\rho}_k - \rho)^{2r} \, d\mathbb{P} + \int_{B} (\hat{\rho}_k - \rho)^{2r} \, d\mathbb{P}
\]

11
\[ \leq \mathbb{P}(B') + \int_B \left\{ \rho_{10}(\hat{\mu}, \hat{\nu}) \cdot (\overline{X}_{nk} - \mu) + \rho_{01}(\hat{\mu}, \hat{\nu}) \cdot (\overline{Y}_{nk} - \nu) \right\}^{2r} \, d\mathbb{P} \]

\[ \leq \mathbb{P} \left\{ \| \overline{X}_{nk} - \mu \| \geq \epsilon \right\} + \mathbb{P} \left\{ \| \overline{Y}_{nk} - \nu \| \geq \epsilon \right\} + 2^{2r} C \mathbb{E} \left\{ \| \overline{X}_{nk} - \mu \|^{2r} + \| \overline{Y}_{nk} - \nu \|^{2r} \right\} \]

\[ \leq \left[ 2^{2r} C + \frac{1}{\epsilon^{2r}} \right] \mathbb{E} \left\{ \| \overline{X}_{nk} - \mu \|^{2r} + \| \overline{Y}_{nk} - \nu \|^{2r} \right\} \]

\[ = O \left( \frac{1}{k^{2r}} \right) \text{ by Lemma 3.} \]

\[ \Box \]

5. STRONG LAW OF LARGE NUMBERS

This section begins with a bound on the second moment of \( S_n \), which will be needed for the proof of the strong law of large numbers.

**Proposition 1**

\[ \mathbb{E} \left\{ S_n^2 \right\} = O(n) \text{ as } n \to \infty. \]

**Proof.**

\[ \mathbb{E} \left[ S_k^2 / \mathcal{F}_{k-1} \right] = S_{k-1}^2 + 2S_{k-1} \left[ q \left( \frac{m_{k-1}}{k-1}, \hat{\rho}_{k-1} \right) - \rho \right] + \mathbb{E} \left[ Z_k^2 / \mathcal{F}_{k-1} \right] \]

\[ \leq S_{k-1}^2 + 2S_{k-1} \left[ \left( q \left( \frac{m_{k-1}}{k-1}, \rho \right) - q (\rho, \rho) \right) \right. \]

\[ + \left. \left( q \left( \frac{m_{k-1}}{k-1}, \hat{\rho}_{k-1} \right) - q \left( \frac{m_{k-1}}{k-1}, \rho \right) \right) \right] + 1 \]

\[ \leq S_{k-1}^2 + 2S_{k-1} \left[ q \left( \frac{m_{k-1}}{k-1}, \hat{\rho}_{k-1} \right) - q \left( \frac{m_{k-1}}{k-1}, \rho \right) \right] + 1 \]

\[ = S_{k-1}^2 + 2S_{k-1} q_{01} \left( \frac{m_{k-1}}{k-1}, \hat{\rho} \right) (\hat{\rho} - \rho) + 1, \]

for some intermediate point \( \hat{\rho} \), where the next to last inequality follows from

\[ S_{k-1} \left[ q \left( \frac{m_{k-1}}{(k-1), \rho} - q (\rho, \rho) \right) \right] \leq 0 \text{ and last from the Mean Value Theorem. Taking expectations and using the assumption that } q \text{ has bounded derivatives,} \]

\[ \mathbb{E} \left[ S_k^2 \right] \leq \mathbb{E} \left[ S_{k-1}^2 \right] + C_1 \mathbb{E} \left| S_{k-1} \left( \hat{\rho} - \rho \right) \right| + 1 \]
\[ \leq \mathbb{E} \left[ S_{k-1}^2 \right] + C_1 \sqrt{\mathbb{E} \left[ S_{k-1}^2 \right]} \sqrt{\mathbb{E} \left[ (\hat{\rho}_k - \rho)^2 \right]} + 1 \]

for some constant \( C_1 > 1 \). By Lemma 4, \( \sqrt{\mathbb{E} \left[ (\hat{\rho}_k - \rho)^2 \right]} = O \left( 1/\sqrt{k} \right) \). Therefore,

\[ \mathbb{E} \left[ S_k^2 \right] \leq \mathbb{E} \left[ S_{k-1}^2 \right] + \frac{C_2}{\sqrt{k}} \sqrt{\mathbb{E} \left[ S_{k-1}^2 \right]} + 1 \]

for some other constant \( C_2 > 1 \). Let \( B = 4C_2^2 \). Clearly, \( \mathbb{E} \left( S_1^2 \right) \leq 1 \leq B \). Suppose \( \mathbb{E} \left[ S_{k-1}^2 \right] \leq B(k - 1) \). Then

\[ \mathbb{E} \left[ S_k^2 \right] \leq B(k - 1) + \frac{C_2}{\sqrt{k}} \sqrt{B(k - 1)} + 1 \]

\[ \leq Bk - B + C_2 \sqrt{B} + 1 \]

\[ = Bk - 4C_2^2 + 2C_2^2 + 1 \]

\[ \leq Bk. \]

The result now follows by induction. \( \Box \)

**Theorem 2** *(Strong Law of Large Numbers)* Under conditions (i)–(vi).

\[ \lim_{n \to \infty} \frac{S_n}{n} = 0 \text{ w.p.1.} \]

**Proof.** By Proposition 1, there exists a constant \( C \) such that \( \mathbb{E} \{ S_n^2 \} \leq Cn \). So, \( \mathbb{E} \left\{ (S_n^2/n^2)^2 \right\} \leq C/n^2 \), and therefore

\[ \lim_{n \to \infty} \frac{S_n^2}{n^2} = 0 \text{ w.p.1,} \]

by the Borel–Cantelli lemmas. The almost sure convergence of \( S_k/k \) follows by noting that

\[ \max_{n^2 \leq k \leq (n+1)^2} \left| \frac{S_k}{k} - \frac{S_{n^2}}{n^2} \right| = \max_{n^2 \leq k \leq (n+1)^2} \left| \frac{n^2 S_k - k S_{n^2}}{n^2 k} \right| \]

\[ \leq \max_{n^2 \leq k \leq (n+1)^2} \left\{ \left| \frac{n^2 (S_k - S_{n^2})}{n^4} \right| + \frac{(k - n^2) S_{n^2}}{n^4} \right\} \]

13
\[
\begin{align*}
\leq & \max_{n^2 \leq k \leq (n+1)^2} \left\{ \frac{n^2(k-n^2)}{n^4} + \frac{(k-n^2)n^2}{n^4} \right\} \\
\leq & \frac{2n^2[(n+1)^2 - n^2]}{n^4} \\
= & \frac{4n^3 + 2n^2}{n^4} \to 0.
\end{align*}
\]

6. EXAMPLE: NORMAL RESPONSES

Suppose it is desired to design a sequential procedure, with a randomized allocation scheme, for the fixed width interval estimation of the difference of the means of two populations. Minimizing the total size of the experiment can be accomplished by designing the sequential procedure so that subjects are allocated to the two treatments in the correct proportions.

More formally, assume that \(X_1, X_2, \ldots\) and \(Y_1, Y_2, \ldots\) are independent random variables for which

\[
X_1, X_2, \ldots \sim N(\mu, \sigma^2) \quad \text{and} \quad Y_1, Y_2, \ldots \sim N(\nu, \tau^2)
\]

where the four parameters \(\mu, \nu, \sigma, \) and \(\tau\) are unknown. Here, \(X_1, X_2, \ldots\) denote responses to treatment A and \(Y_1, Y_2, \ldots\) denote responses to treatment B. These could be, for example, blood pressure readings. Then, the correct allocation proportions for minimizing the total sample size and retaining preassigned coverage probability and interval width are (see Robbins, Simons, and Starr (1967) or Eisele (1990))

\[
\frac{\sigma}{\sigma + \tau} \times k \quad \text{to treatment A}
\]

and

\[
\frac{\tau}{\sigma + \tau} \times k \quad \text{to treatment B}.
\]
Thus,
\[ \rho \left( \sigma^2, \tau^2 \right) = \frac{\sigma}{\sigma + \tau}. \]

Taking
\[ \hat{\sigma}_{m_k}^2 = (m_k - 1)^{-1} \sum_{i=1}^{m_k} \left( X_i - \bar{X}_{m_k} \right)^2 \quad \text{and} \quad \hat{\tau}_{n_k}^2 = (n_k - 1)^{-1} \sum_{i=1}^{n_k} \left( Y_i - \bar{Y}_{n_k} \right)^2 \]
to be the usual estimates of \( \sigma^2 \) and \( \tau^2 \), gives
\[ \hat{\rho}_k = \frac{\hat{\sigma}_{m_k}}{\hat{\sigma}_{m_k} + \hat{\tau}_{n_k}}. \]

The sequential procedure can now be described as follows: to start, take \( n_0 \geq 2 \) observations on \( X \) and on \( Y \). Then, if at any stage there are \( m_k \) observations on \( X \) and \( n_k \) on \( Y \), with \( k = m_k + n_k \geq 2n_0 \), take observation \( k + 1 \) on \( X \) if
\[ U_{k+1} \leq q_{k+1} \equiv q \left( \frac{m_k}{k}, \hat{\rho}_k \right). \]
Otherwise, take observation \( k + 1 \) on \( Y \).

If the desired width and coverage probability of the confidence interval for \( \theta = \mu - \nu \) are \( 2h \) and \( \alpha \), respectively, and if the constant \( a \) is defined by \( 2\Phi(a) - 1 = \alpha \), where \( \Phi \) denotes the \( N(0,1) \) distribution function, then a possible stopping rule for the sequential procedure is: stop after \( N \) observations, where
\[ N = \inf \left\{ k \geq 2n_0 : \frac{\hat{\sigma}_{m_k}^2}{m_k} + \frac{\hat{\tau}_{n_k}^2}{n_k} \leq \left( \frac{h}{a_k} \right)^2 \right\} \]
and \( \{a_k\} \) is a given sequence of positive constants such that \( a_k \to a \) as \( k \to \infty \).

The following simulation results illustrate the asymptotic properties of the doubly adaptive biased coin design for the case of normally distributed responses. For the values
\[ \alpha = .95, \quad a = 1.96, \quad n_0 = 5, \quad a_k^2 = \left( \frac{k + 4}{k - 4} \right) a^2, \]
the optimal sample size becomes

$$n^* = \left\{1.96 \frac{\tau}{h} \left(\frac{\sigma}{\tau} + 1\right)\right\}^2.$$  

The $q$-function, $q(x, y) = [1 - (1/y - 1)x]_+$, was selected for its simplicity. Then,

$$q\left(\frac{m_k}{k}, \hat{p}_k\right) = \left[1 - \left(\frac{\hat{r}_{nk}}{\hat{\sigma}_{mk}}\right)\left(\frac{m_k}{k}\right)\right]_+.$$
Table 1: Simulation Results for 2,000 Trials for Normally Distributed Subject Responses: Allocation Ratios, Expected Sample Sizes and Coverage Probabilities for $\sigma/\tau = 1, 1/2$, and 1/4.

<table>
<thead>
<tr>
<th>$n^*$</th>
<th>$m_N$</th>
<th>$n_N$</th>
<th>$\frac{m_N}{n_N}$</th>
<th>$\alpha$</th>
<th>$m_N$</th>
<th>$n_N$</th>
<th>$\frac{m_N}{n_N}$</th>
<th>$\alpha$</th>
<th>$m_N$</th>
<th>$n_N$</th>
<th>$\frac{m_N}{n_N}$</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>7.6</td>
<td>7.8</td>
<td>.974</td>
<td>.979</td>
<td>6.1</td>
<td>9.7</td>
<td>.629</td>
<td>.979</td>
<td>5.1</td>
<td>11.2</td>
<td>.455</td>
<td>.975</td>
</tr>
<tr>
<td>20</td>
<td>12.1</td>
<td>12.6</td>
<td>.960</td>
<td>.957</td>
<td>8.4</td>
<td>16.5</td>
<td>.509</td>
<td>.958</td>
<td>5.8</td>
<td>19.4</td>
<td>.299</td>
<td>.960</td>
</tr>
<tr>
<td>30</td>
<td>17.0</td>
<td>17.4</td>
<td>.977</td>
<td>.950</td>
<td>11.3</td>
<td>22.9</td>
<td>.493</td>
<td>.950</td>
<td>7.1</td>
<td>27.8</td>
<td>.255</td>
<td>.954</td>
</tr>
<tr>
<td>40</td>
<td>22.0</td>
<td>22.4</td>
<td>.982</td>
<td>.948</td>
<td>14.3</td>
<td>29.6</td>
<td>.483</td>
<td>.948</td>
<td>8.6</td>
<td>35.9</td>
<td>.240</td>
<td>.951</td>
</tr>
<tr>
<td>50</td>
<td>27.0</td>
<td>27.4</td>
<td>.985</td>
<td>.947</td>
<td>17.6</td>
<td>36.5</td>
<td>.482</td>
<td>.947</td>
<td>10.2</td>
<td>43.9</td>
<td>.232</td>
<td>.950</td>
</tr>
<tr>
<td>60</td>
<td>32.1</td>
<td>32.5</td>
<td>.988</td>
<td>.949</td>
<td>20.7</td>
<td>43.7</td>
<td>.474</td>
<td>.948</td>
<td>12.0</td>
<td>51.6</td>
<td>.233</td>
<td>.917</td>
</tr>
<tr>
<td>70</td>
<td>36.9</td>
<td>37.5</td>
<td>.984</td>
<td>.948</td>
<td>24.2</td>
<td>49.7</td>
<td>.487</td>
<td>.947</td>
<td>13.7</td>
<td>60.2</td>
<td>.228</td>
<td>.949</td>
</tr>
<tr>
<td>80</td>
<td>42.4</td>
<td>42.7</td>
<td>.993</td>
<td>.950</td>
<td>27.9</td>
<td>56.6</td>
<td>.493</td>
<td>.950</td>
<td>16.1</td>
<td>68.8</td>
<td>.234</td>
<td>.951</td>
</tr>
<tr>
<td>90</td>
<td>47.2</td>
<td>47.6</td>
<td>.992</td>
<td>.950</td>
<td>30.9</td>
<td>63.9</td>
<td>.484</td>
<td>.950</td>
<td>17.8</td>
<td>76.3</td>
<td>.233</td>
<td>.919</td>
</tr>
<tr>
<td>100</td>
<td>52.7</td>
<td>52.8</td>
<td>.998</td>
<td>.951</td>
<td>34.2</td>
<td>70.5</td>
<td>.490</td>
<td>.950</td>
<td>19.9</td>
<td>83.8</td>
<td>.237</td>
<td>.919</td>
</tr>
<tr>
<td>125</td>
<td>65.0</td>
<td>65.2</td>
<td>.997</td>
<td>.951</td>
<td>42.7</td>
<td>87.2</td>
<td>.490</td>
<td>.951</td>
<td>25.0</td>
<td>104.4</td>
<td>.290</td>
<td>.950</td>
</tr>
<tr>
<td>150</td>
<td>77.8</td>
<td>77.8</td>
<td>1.000</td>
<td>.951</td>
<td>51.1</td>
<td>103.8</td>
<td>.492</td>
<td>.951</td>
<td>30.1</td>
<td>124.5</td>
<td>.242</td>
<td>.950</td>
</tr>
<tr>
<td>175</td>
<td>90.1</td>
<td>90.2</td>
<td>.999</td>
<td>.952</td>
<td>59.3</td>
<td>120.5</td>
<td>.492</td>
<td>.951</td>
<td>35.3</td>
<td>144.7</td>
<td>.241</td>
<td>.951</td>
</tr>
<tr>
<td>200</td>
<td>102.8</td>
<td>103.9</td>
<td>.989</td>
<td>.952</td>
<td>67.9</td>
<td>137.1</td>
<td>.495</td>
<td>.951</td>
<td>40.2</td>
<td>164.6</td>
<td>.244</td>
<td>.951</td>
</tr>
</tbody>
</table>
Remarks

1. Although the \textit{q-function} selected is not strictly increasing in the second argument on \((0,1)^2\), the proof of the strong law of large numbers given in Eisele (1990) for the case of normally distributed subject responses only requires \(q\) to be strictly increasing near the diagonal.

2. For normally distributed observations,

\[
P \left( \left| \overline{X}_m - \overline{Y}_n - \theta \right| \leq h \right) = 2 \Phi \left( \frac{h}{\left( \frac{s^2_m}{m_0} + \frac{s^2_n}{n_0} \right)^{1/2}} \right) - 1.
\]

Estimates of the coverage probabilities are found by estimating the above expectation using the simulated values of \(N\).

3. Another possible method for estimating the coverage probabilities is to simulate Bernoulli random variables associated with "coverage" and "non-coverage". The method given in remark 2 has less variability than the Bernoulli method and thus provides more accurate estimates of the coverage probabilities.

4. In terms of allocation proportions, the doubly adaptive biased coin appears to be performing as desired for the three examples given. The total sample size is roughly between 4 and 6 observations larger than the optimal total sample size, resulting in high coverage probabilities for small values of \(n^*\) where there is oversampling.

5. For more details on this sequential procedure, including derivations, other stopping rules, asymptotic properties, and simulation results, see Eisele (1990). A different proof of the strong law of large numbers is also given in Eisele (1990) for the case of normally distributed subject responses.
ACKNOWLEDGEMENTS

I would like to thank my thesis advisor, Professor Michael Woodroofe, for introducing this problem and for his guidance in this research.

REFERENCES


