

EXACT PROPERTIES OF EFRON'S BIASED COIN RANDOMIZATION PROCEDURE

BY TIGRAN MARKARYAN AND WILLIAM F. ROSENBERGER*

Accenture, L.L.P. and George Mason University

Efron (1971) developed a restricted randomization procedure to promote balance between two treatment groups in a sequential clinical trial. He called this the *biased coin design*. He also introduced the concept of *accidental bias*, and investigated properties of the procedure with respect to both accidental and selection bias, balance, and randomization-based inference using the steady-state properties of the induced Markov chain. While many years have passed since the publication of this fundamental paper, more recent work has shown that the procedure has some optimal properties, including being fully randomized with small variability (e.g., Hu, Rosenberger, and Zhang (2006)). In this paper we revisit this procedure, and derive closed-form expressions for the exact properties of the measures derived asymptotically in Efron's paper. In particular, we derive the exact distribution of the treatment imbalance and the variance-covariance matrix of the treatment assignments. These results have application in the design and analysis of clinical trials, by providing exact formulas to determine the role of the coin's bias probability in the context of selection and accidental bias, balancing properties, and randomization-based inference.

1. Introduction. Efron (1971) introduced his famous *biased coin design* as a method that "...tends to balance the experiment, but at the same time is not over vulnerable to various common forms of experimental bias." The primary application is in sequential clinical trials, where balance in the numbers randomly assigned to two treatment groups is sometimes desirable for power considerations. In such cases, it is also desirable to maintain near-balance at intermediate points in the trial, as heterogeneity or time trends in patient characteristics may lead to less comparable treatment arms. Randomization protects from imbalances in unknown covariates related to outcomes (which Efron referred to as *accidental bias*, introduced for the first time in the 1971 paper), selection bias, and provides a basis for inference. Efron explored the balancing properties of the biased coin design, as well

*Research supported by National Science Foundation Grant DMS-0904253 under the 2009 Recovery and Reinvestment Act.

AMS 2000 subject classifications: Primary 62E15, 62K99; secondary 62L05, 62J10

Keywords and phrases: accidental bias, exact distribution theory, randomization test, restricted randomization, selection bias

as its susceptibility to selection and accidental bias, and discussed the implications for randomization-based inference. All of these results were based on studying the steady-state properties of the Markov chain induced by the imbalance process of biased coin randomization.

Let $\mathbf{T}_n = (T_1, \dots, T_n)'$ be a randomization sequence, where $T_i = 1$ if treatment A is assigned, and $T_i = -1$ if treatment B is assigned, $i = 1, \dots, n$. After j assignments, let D_j be the difference in the number of patients assigned to treatments A and B ; i.e., $D_j = \sum_{i=1}^j T_i$. The biased coin design with bias $p \in [0.5, 1]$, denoted BCD(p), is defined by

$$P(T_j = 1) = \begin{cases} 1/2, & \text{when } D_{j-1} = 0, \\ p, & \text{when } D_{j-1} < 0, \\ 1 - p, & \text{when } D_{j-1} > 0. \end{cases} \quad j = 1, 2, 3, \dots,$$

Note that $p = 0.5$ results in complete randomization and $p = 1$ results in a permuted block design with block size of 2, in which case every alternate assignment is deterministic. Efron notes that the $\{|D_n|\}_{n=1}^{\infty}$ process forms a Markov chain of period 2 with states $0, 1, 2, \dots$, and a reflecting barrier at the origin. He then proves that the $|D_n|$ process has stationary probabilities π_j , given by

$$(1.1) \quad \pi_j = \begin{cases} \frac{r^2 - 1}{2r^{j+1}}, & \text{when } j \geq 1, \\ \frac{r - 1}{2r}, & \text{when } j = 0, \end{cases}$$

where $q = 1 - p$, $r = p/q \geq 1$. Efron uses the formulas obtained for stationary probabilities to write the form of the limiting probabilities of perfect balance (n is even) and imbalance of 1 (n is odd) as

$$\begin{aligned} \lim_{n \rightarrow \infty} P(|D_{2m}| = 0) &= 2\pi_0 = \frac{r - 1}{r}, \\ \lim_{n \rightarrow \infty} P(|D_{2m+1}| = 1) &= 2\pi_1 = \frac{r^2 - 1}{r^2}. \end{aligned}$$

Most research on the theory of randomization in recent years has focused on generalizations of Efron's procedure (see, for example, Wei (1978); Soares and Wu (1982); Eisele (1994); Chen (1999); Baldi Antognini and Giovagnoli (2004); Hu and Zhang (2004)) rather than Efron's procedure itself. In particular, Baldi Antognini and Giovagnoli's (2004) "adjustable biased coin design" is stochastically more balanced, and therefore uniformly more powerful, than the other procedures (Baldi Antognini (2008)).

The remainder of Efron's article is devoted to selection bias, as defined by Blackwell and Hodges (1957), accidental bias, and randomization as a basis for inference. Efron notes that the best guessing strategy against the BCD(p) is to always guess the group that has occurred least often up to that point. The probability of correctly guessing at the j th step is

$$(1.2) \quad \frac{1}{2}P(D_{j-1} = 0) + pP(|D_{j-1}| > 0),$$

which asymptotically approaches $1/2 + (r-1)/4r$ and therefore has asymptotic excess selection bias of

$$(1.3) \quad \frac{(r-1)}{4r}.$$

Accidental bias refers to the squared bias of the treatment effect in a linear regression when an unknown covariate \mathbf{z} is left out of the model. Efron derives this bias as

$$E(\mathbf{z}'\mathbf{T}_n)^2 = \mathbf{z}'\boldsymbol{\Sigma}_{\mathbf{T}_n}\mathbf{z},$$

where $\boldsymbol{\Sigma}_{\mathbf{T}_n} = \text{Var}(\mathbf{T}_n)$. He suggests a minimax approach by noting that

$$(1.4) \quad \mathbf{z}'\boldsymbol{\Sigma}_{\mathbf{T}_n}\mathbf{z} \leq \text{maximum eigenvalue of } \boldsymbol{\Sigma}_{\mathbf{T}_n},$$

where the inequality follows from the assumption that $\|\mathbf{z}\| = 1$. Note that the minimum possible value for the maximum eigenvalue is 1, which corresponds to complete randomization. Instead of directly examining $\boldsymbol{\Sigma}_{\mathbf{T}_n}$ (which he acknowledges is difficult), Efron looks at the much simpler process $T_1, T_2, T_3, \dots, T_n$, assuming that it is stationary, and aims at finding the asymptotic covariance structure of the process. He then shows that the asymptotic maximum eigenvalue of the covariance vector $(T_{h+1}, \dots, T_{h+N})$ as $h \rightarrow \infty$, λ_N , is increasing in N and has a finite limit. Based on numerical evidence, Efron conjectures that $\lim_{N \rightarrow \infty} \lambda_N = 1 + (p-q)^2$. This was later proved by Steele (1980). However, Smith (1984) shows by counterexample that Efron's solution may be unsatisfactory when there are short-term dependencies in the data.

In this paper, we derive exact properties of Efron's procedure. In particular, in Section 2, we derive a closed-form expression for the distribution of D_n and give the explicit form of $\boldsymbol{\Sigma}_{\mathbf{T}_n}$. These formulas are remarkably compact for the complexity of the problems. We describe computational considerations in Section 3. In Section 4, we apply these results to deriving an explicit form for the excess selection bias, prove a result on the maximum eigenvalue of $\boldsymbol{\Sigma}_{\mathbf{T}_n}$, and discuss randomization as a basis for inference. We compare the exact results with Efron's for various n and p . In Section 5, we draw conclusions. Finally, all proofs are given in the appendices.

2. Exact distribution of D_n and Σ_{T_n} . We will assume the following conventions throughout the mathematical developments.

1. For brevity, we adopt the convention to treat a combination $\begin{pmatrix} x \\ y \end{pmatrix}$ as zero, whenever any of the following conditions is true: $x < 0, y < 0, x < y, y$ is not an integer.
2. We treat summations as 0 if the upper limit of the summation is smaller than the lower limit.
3. We treat conditional probabilities, conditional on zero-probability events to be 0.

The distribution of D_n requires determination of the exact distribution of a denumerable homogeneous random walk. The result is given as the first theorem:

THEOREM 2.1. *Let $n = 1, 2, 3, \dots, 0 \leq k \leq n$ and n and k have the same parity. Then, the distribution of D_n of the $BCD(p)$ is given by formulas (2.1) and (2.2).*

For $k > 0$,

$$(2.1) \quad P(D_n = \pm k) = \frac{1}{2} p^{\frac{n-k}{2}} \sum_{l=0}^{\frac{n-k}{2}} \frac{n+k-2l}{n+k+2l} \binom{\frac{n+k}{2}+l}{l} q^{k+l-1}.$$

For $k = 0$,

$$(2.2) \quad P(D_n = 0) = p^{\frac{n}{2}} \sum_{l=0}^{\frac{n}{2}-1} \frac{n-2l}{n+2l} \binom{\frac{n}{2}+l}{l} q^l.$$

PROOF. See Appendix A. □

The compact form of these equations arises from patterns in polynomials of p and q that can be seen developing for small n as n increments. The proof is then by induction. Note that the distribution of $N_A(n)$ follows immediately, since $N_A(n) = (D_n + n)/2$.

Define $t_k = P(T_n = 1 | D_{n-1} = k)$. We now derive the covariance of (T_n, T_m) :

THEOREM 2.2. *Let $1 \leq n < m$. Then the joint distribution of (T_n, T_m) of the $BCD(p)$, $p \in [1/2, 1]$, is given by*

(2.3)

$$P(T_n = 1, T_m = 1) = \sum_{k=-n+1}^{n-1} \left(\left(\frac{1}{2} - t_{k+1} \right) \hat{f}_{k+1,0}^{(m-n-1)} + t_{k+1} \right) d_{n-1,k} t_k,$$

where

$d_{n,k} = P(D_n = k)$ and is given in (2.1) and (2.2) and

$$(2.4) \quad \hat{f}_{k,0}^{(u)} = \sum_{l=|k|}^u f_{k,0}^{(l)} = \begin{cases} \sum_{l=|k|}^u \frac{|k|}{l} \binom{l}{\frac{l+|k|}{2}} p^{\frac{l+|k|}{2}} q^{\frac{l-|k|}{2}}, & \text{when } k \neq 0; \\ 1, & \text{when } k = 0. \end{cases}$$

PROOF. See Appendix B. \square

The form of $\Sigma_{\mathbf{T}_n}$ follows immediately:

COROLLARY 2.1. *Let $\Sigma_{\mathbf{T}_n}$ be the covariance matrix of \mathbf{T}_n of the BCD(p), $p \in [1/2, 1]$. Then the (i, j) th entry of the matrix, σ_{ij} , where $1 \leq i \leq j \leq n$, is given by*

$$(2.5) \quad \sigma_{ij} = \begin{cases} 4 \cdot \sum_{k=-i+1}^{i-1} \left(\left(\frac{1}{2} - t_{k+1} \right) \hat{f}_{k+1,0}^{(j-i-1)} + t_{k+1} \right) d_{i-1,k} t_k - 1, & \text{when } i < j; \\ 1, & \text{when } i = j; \end{cases}$$

where $\hat{f}_{k,0}^{(u)}$ is defined in (2.4).

3. Computational considerations. This section contains some observations on the computation of $P(D_n = k)$ according to formulas (2.1) and (2.2) and the computation of $P(T_n = 1, T_m = 1)$ according to formulas (2.3) and (2.4). These formulas involve terms that are products of large factorials and powers of numbers that are between 0 and 1. The key is to calculate these products in such order that the result does not get too large or too small too quickly. We focus on the computation of (2.1) here, as the other formulas are similar. For $n \leq 100$, calculating the combination and multiplying by powers of p and q directly works well. However, for larger values of n , precision may be lost if the intermediate products become too large or too small.

Formula (2.1) involves $(n-k)/2 + 1$ terms, each of which is a product of powers of p , powers of q , positive integers and reciprocals of positive integers. The generic term of the r.h.s. of (2.1) can be written as:

$$(3.1) \quad \frac{1}{2} \frac{n+k-2l}{n+k+2l} \overbrace{p \cdots p}^{\frac{n-k}{2}} \overbrace{q \cdots q}^{k+l-1} \overbrace{\frac{1}{2} \cdot \frac{1}{3} \cdots \frac{1}{l}}^{l-1} \overbrace{\left(\frac{n+k}{2} + 1 \right) \cdots \left(\frac{n+k}{2} + l \right)}^l.$$

There are $(n+k)/2+2l$ factors less than 1 and l greater than 1. Denote these two groups by $\{a_s\}_{s=1}^{(n+k)/2+2l}$ and $\{b_s\}_{s=1}^l$, respectively. Assume that the a_s are indexed in decreasing order. The following simple algorithm ensures that the running products for calculating (3.1) do not become too large or too small early on in the calculation process.

1. Fix a number M that the running product cannot exceed. Any number that is larger than $2n$ will work.
2. Fix a number m that is close to the machine epsilon. When the running product gets close to m , the algorithm will know that further multiplication by small numbers may result in loss of precision.
3. Start multiplication using numbers in $\{b_s\}_{s=1}^l$ until the running product exceeds M .
4. Multiply the running product with numbers in $\{a_s\}_{s=1}^{(n+k)/2+2l}$ until the running product is less than M .
5. Iterate through Steps 3) and 4) until numbers in $\{b_s\}_{s=1}^l$ are depleted.
6. Continue multiplying the running product with the remaining numbers in $\{a_s\}_{s=1}^{(n+k)/2+2l}$, from the largest to the smallest. Two cases are possible: the final product is larger than m , in which case the algorithm is completed; at some point, the running product becomes less than m , in which case one can save the result as a product of two or more small numbers.

For example, we used the algorithm to compute Table 1, which gives the value of n at which the steady state probabilities are within certain percentages of the exact probability $P(D_n = k)$, for various k and p . The same idea can be used for calculating (2.4).

Finally, for the computation of $\Sigma_{\mathcal{T}_n}$, the following proposition gives a property of the matrix that can facilitate computation. The proof follows from Corollary 2.1 and Lemma B.1 and is omitted.

PROPOSITION 3.1. *If $\Sigma_{\mathcal{T}_n}$ is partitioned into 2×2 submatrices, then all the off-diagonal submatrices are constant (i.e., have the same elements in both rows and columns).*

4. Applications to clinical trials. In this section we apply the results of Section 2 to the study of balancing properties of the randomization procedure, selection and accidental biases, and randomization as a basis for inference. Each of these is a consideration in the appropriate selection of a randomization procedure in clinical trials (see Rosenberger and Lachin (2002)).

TABLE 1

Values of n starting at which steady state probabilities are within 10%, 5%, 1% and 0.1% of $P(D_n = k)$, $k = 0, 1, 2, 3, 4$.

| k | p | Errors within | | | |
|-----|-----|---------------|-----|-----|------|
| | | 10% | 5% | 1% | 0.1% |
| 0 | 0.6 | 20 | 34 | 74 | 146 |
| | 0.7 | 6 | 8 | 18 | 34 |
| | 0.8 | 2 | 4 | 8 | 14 |
| | 0.9 | 2 | 2 | 4 | 6 |
| 1 | 0.6 | 19 | 33 | 73 | 145 |
| | 0.7 | 5 | 7 | 17 | 33 |
| | 0.8 | 1 | 3 | 7 | 13 |
| | 0.9 | 1 | 1 | 3 | 5 |
| 2 | 0.6 | 14 | 28 | 68 | 140 |
| | 0.7 | 4 | 4 | 8 | 22 |
| | 0.8 | 4 | 4 | 8 | 14 |
| | 0.9 | 2 | 4 | 6 | 8 |
| 25 | 0.6 | 183 | 211 | 279 | 379 |
| | 0.7 | 85 | 93 | 113 | 141 |
| | 0.8 | 53 | 57 | 65 | 77 |
| | 0.9 | 37 | 39 | 43 | 49 |
| 50 | 0.6 | 342 | 380 | 464 | >500 |
| | 0.7 | 158 | 168 | 194 | 226 |
| | 0.8 | 100 | 104 | 116 | 130 |
| | 0.9 | 70 | 72 | 78 | 86 |

4.1. *Balancing properties of the biased coin design.* All finite balancing properties of the biased coin design can be investigated with the help of Theorem 2.1, which provides the means for exact calculations of the probabilities involving $P(D_n = k)$. In particular, the exact variance is given in the following proposition:

PROPOSITION 4.1. *The exact variance of D_n is given by:*

$$(4.1) \quad \text{Var}(D_n) = \sum_{\substack{k=1 \\ n-k \text{ even}}}^n k^2 p^{\frac{n-k}{2}} \sum_{l=0}^{\frac{n-k}{2}} \frac{n+k-2l}{n+k+2l} \binom{\frac{n+k}{2}+l}{l} q^{k+l-1}.$$

The variance of the imbalance of the biased coin design for different values of n and p is provided in Table 2. Also given in the table is the limiting variance based on the steady state distribution of the induced Markov chain. The formulas for odd and even n are given in the following proposition, which follows directly from (1.1).

TABLE 2
 Variance of the imbalance of the BCD(p) for different values of n and p .

| n even | p | | | |
|----------|-------|------|------|------|
| | 0.6 | 0.7 | 0.8 | 0.9 |
| 10 | 5.19 | 2.55 | 1.18 | 0.46 |
| 20 | 7.65 | 2.91 | 1.21 | 0.46 |
| 50 | 10.78 | 3.04 | 1.21 | 0.46 |
| 100 | 12.10 | 3.04 | 1.21 | 0.46 |
| 200 | 12.45 | 3.04 | 1.21 | 0.46 |
| ∞ | 12.48 | 3.04 | 1.21 | 0.46 |
| n odd | p | | | |
| | 0.6 | 0.7 | 0.8 | 0.9 |
| 5 | 3.30 | 2.15 | 1.45 | 1.10 |
| 15 | 6.63 | 2.95 | 1.56 | 1.10 |
| 25 | 8.52 | 3.13 | 1.57 | 1.10 |
| 75 | 11.73 | 3.20 | 1.57 | 1.10 |
| ∞ | 12.52 | 3.21 | 1.57 | 1.11 |

PROPOSITION 4.2. *Under the limiting distribution of the BCD(p), $p \in [1/2, 1]$, the variance of the imbalance is given by:*

$$(4.2) \quad \begin{aligned} & \frac{4r(r^2 + 1)}{(r^2 - 1)^2}, \quad \text{when number of trials is even} \\ & \frac{8r^2}{(r^2 - 1)^2} + 1, \quad \text{when number of trials is odd.} \end{aligned}$$

As can be seen in the table, odd and even n form different patterns. This is due to the differences in the supports of the distributions, in particular, a significant mass is concentrated at 0 when n is even and p is large. Note that both odd and even n form an increasing series for each p . This is expected and follows from Theorem 1 in Efron (1971) with $h(j) = j^2$. It is also the case that $Var(D_n)$ is a decreasing function of p for each n . This is also expected and was proved by Efron as Theorem 3 with $h(j) = j^2$. It is clear that balancing properties stabilize for moderate-sized trials of around 75 to 100. This contrasts to other randomization procedures such as complete randomization and the urn design (Wei, 1978) where the variance of D_n grows at a rate $O(n)$ (Rosenberger and Lachin (2002, Chapter 3)).

4.2. *Selection bias.* Theorem 2.1 allows us to calculate the selection bias for the BCD(p) using Equation (1.2). When n is even, $P(D_{n-1} = 0) = 0$, and therefore the selection bias is p . Obviously, the selection bias when $n = 1$ is $1/2$. When n is an odd number exceeding 1, $n = 2m + 1$ and $m \in \mathbb{N}$, substituting the expression for $P(D_{n-1} = 0)$ from Theorem 2.1, we obtain

TABLE 3
Average excess selection bias of the BCD(p) for different values of n and p .

| n | p | | | |
|----------|-------|-------|-------|-------|
| | 0.6 | 0.7 | 0.8 | 0.9 |
| 5 | 0.058 | 0.107 | 0.146 | 0.177 |
| 10 | 0.070 | 0.129 | 0.178 | 0.217 |
| 15 | 0.072 | 0.129 | 0.173 | 0.207 |
| 20 | 0.075 | 0.136 | 0.183 | 0.220 |
| 25 | 0.076 | 0.135 | 0.179 | 0.213 |
| 50 | 0.080 | 0.140 | 0.186 | 0.221 |
| 75 | 0.081 | 0.140 | 0.185 | 0.219 |
| 100 | 0.081 | 0.141 | 0.187 | 0.222 |
| 200 | 0.082 | 0.142 | 0.187 | 0.222 |
| ∞ | 0.083 | 0.143 | 0.188 | 0.222 |

the expression for the selection bias for this case:

$$(4.3) \quad p - (p - \frac{1}{2})p^m \sum_{l=0}^{m-1} \frac{m-l}{m+l} \binom{m+l}{l} q^l.$$

Now we can formulate a result on the total selection bias in n trials.

PROPOSITION 4.3. *The total amount of selection bias in n , $n \geq 1$, trials for the BCD(p) is given by:*

$$(4.4) \quad \frac{1}{2} + (n-1)p - (p - \frac{1}{2}) \sum_{m=1}^{[(n-1)/2]} p^m \sum_{l=0}^{m-1} \frac{m-l}{m+l} \binom{m+l}{l} q^l,$$

where $[a]$ denotes the integer part of a and we use the adopted convention that the sum is treated as zero when the upper limit of summation is smaller than the lower limit.

One subtracts $n/2$ from (4.4) to obtain the excess selection bias in n trials. The average excess selection bias in n trials (total excess selection bias divided by n) of the BCD(p) for different values of n and p is provided in Table 3.

As expected, the excess selection bias increases with p . Also, note that the average excess selection bias is not a monotonic function of n . Asymptotically, the excess is given in (1.3), and is reported in the table under $n = \infty$. One can see that the asymptotic formula is a good approximation even for sample sizes as small as 50.

4.3. *Accidental bias.* With the help of Corollary 2.1, which provides the exact form of the covariance matrix of the $BCD(p)$, one can compute the accidental bias due to failure to adjust for any covariate \mathbf{z} , given by $\mathbf{z}'\boldsymbol{\Sigma}_{\mathbf{T}_n}\mathbf{z}$. However, the point of the accidental bias is to control the bias of the treatment effect caused by an *unknown* covariate. This leads to Efron's minimax solution of using the maximum eigenvalue of $\boldsymbol{\Sigma}_{\mathbf{T}_n}$, given in inequality (1.4). The maximum eigenvalue of $\boldsymbol{\Sigma}_{\mathbf{T}_n}$ therefore represents maximum susceptibility to accidental bias. At this time we are able to prove the following theorem.

THEOREM 4.1. *One of the eigenvalues of $\boldsymbol{\Sigma}_{\mathbf{T}_n}$ of the $BCD(p)$ is $2p$, for all $n \geq 2$ and $p \in [1/2, 1]$.*

PROOF. See Appendix C. □

REMARK. The theorem affirms that the maximum eigenvalue of $\boldsymbol{\Sigma}_{\mathbf{T}_n}$ exceeds $1+(p-q)^2$. This shows that the maximum eigenvalue of the asymptotic covariance structure studied by Efron (1971) and Steele (1980) is strictly less than the maximum eigenvalue of $\boldsymbol{\Sigma}_{\mathbf{T}_n}$.

We conjecture, based on vast numeric evidence, that the maximum eigenvalue of $\boldsymbol{\Sigma}_{\mathbf{T}_n}$ of the $BCD(p)$ does not depend on n and is equal to $2p$ for all $n \geq 2$ and $p \in [1/2, 1]$. Note that this leads to a maximum eigenvalue of 1 for $p = 0.5$, which is the maximum eigenvalue for complete randomization, and 2 for $p = 1$, which is the maximum eigenvalue for the permuted block design with block size 2 (Rosenberger and Lachin (2002, Chapter 4)).

4.4. *Randomization tests.* The final application of these results is to randomization-based inference procedures. Rosenberger and Lachin (2002, Chapters 7, 11) discuss randomization tests in the context of linear rank statistics. Let $\mathbf{Y}_n = (Y_1, Y_2, \dots, Y_n)$ be the responses based on some primary outcome variable and let \mathbf{y}_n be the realization. The responses, \mathbf{y}_n , are treated as fixed quantities and under the randomization null hypothesis, \mathbf{y}_n is assumed to be unaffected by treatment assignments. The observed difference between Groups A and B , then only depends on the manner the n patients were randomized. The general form of linear rank statistic is $W_n = \mathbf{a}_n' \mathbf{T}_n$, where $\mathbf{a}_n = (a_{1n}, a_{2n}, \dots, a_{nn})'$, is a score function of the ranks of \mathbf{y}_n . The scores $(a_{1n}, a_{2n}, \dots, a_{nn})'$ are usually centered by subtracting the mean. Most standardly used test statistics in clinical trials have an analogous formulation as a linear rank test.

Smythe and Wei (1983) and Hollander and Peña (1988) noted that, unlike for most other restricted randomization procedures, the test W_n is not

asymptotically normal for the biased coin design. Therefore the computation of the test requires either the exact distribution or a Monte Carlo approximation. While our results do not give the exact distribution of the test statistic, we can compute its exact variance, as $Var(W_n) = \mathbf{a}'_n \Sigma_{\mathbf{T}_n} \mathbf{a}_n$ can be computed using Corollary 2.1. For example, using outcome data from a diabetes trial given in Table 7.4 of Rosenberger and Lachin (2002), we generate a sequence of 50 treatment assignments from Efron's BCD($p = 2/3$) and obtain $W_n = -31$ with exact standard deviation 100.52. The latter computation required computing a 50×50 matrix using Corollary 2.1.

5. Conclusions. Despite the favorable properties depicted in Efron's original paper, the biased coin design is sparsely used in clinical trials. The majority of clinical trials use a permuted block design, which forces balance at regular intervals in the trial, and achieves perfect balance unless there is an unfilled final block. However, in permuted blocks, some patients are assigned to treatment with probability 1, which can contribute to a vulnerability to selection bias, particularly in unmasked trials. We believe that Efron's procedure should be used regularly in clinical trials where balance in treatments is desirable, both for its simplicity and for the reason that Efron suggested: it promotes balance with minimal susceptibility to experimental biases. We now have quantified the distribution of balance and the susceptibility to biases in closed-form formulas for any p and n , and this should aid the clinical trialist in designing the trial appropriately.

The selection of p has always been an interesting question. Efron used $p = 2/3$ in some of his examples. At one extreme, $p = 1/2$, we have complete randomization, which has minimal selection and accidental biases, but maximum variability. At the other extreme, $p = 1$, we have a deterministic sequence with maximum selection and accidental biases, but no variability. Formally, the selection should be a trade-off between the degree of randomness desired (as reflected in selection bias), accidental bias (which is linear) and $Var(D_n)$, which are competing objectives. Such multi-objective problems can be solved through a compound optimality criterion, with weights reflecting the relative importance of the criteria to the investigator. We now provide exact formulas for these criteria in (4.1) and (4.4).

We note that these results may have applicability beyond clinical trials, as they form the basis of exact distribution theory for a general asymmetric random walk. While the theorems are proved for $p \geq 0.5$, they can be generalized for any p (Markaryan, 2009).

APPENDIX A: PROOF OF THEOREM 2.1

The following proposition follows immediately from the definition of the $BCD(p)$ and is used without explicit mention in the proof of Theorem 2.1.

PROPOSITION A.1. *Let $n = 1, 2, 3, \dots, k \in \mathbb{Z}$ and $q = 1 - p$. The following hold for the $BCD(p)$:*

1. $P(D_n = k) > 0 \iff |k| \leq n$ and n and k have the same parity,
2. $P(D_n = k) = P(D_n = -k)$,
3. $P(D_{n+1} = 0) = 2pP(D_n = 1)$,
4. $P(D_{n+1} = 1) = \frac{1}{2}P(D_n = 0) + pP(D_n = 2)$,
5. $P(D_{n+1} = k) = (1-p)P(D_n = k-1) + pP(D_n = k+1)$, for $2 \leq k \leq n$,
6. $P(D_{n+1} = n+1) = (1-p)P(D_n = n)$.

Next we formulate and prove two lemmas.

LEMMA A.1. *Let n be a positive even integer and let l be an integer satisfying $0 < l < n/2$. Then the following holds*

$$(A.1) \quad \frac{n-2l}{n+2l} \binom{\frac{n}{2}+l}{l} + \frac{n-2l+4}{n+2l} \binom{\frac{n}{2}+l}{l-1} = \frac{n+2-2l}{n+2+2l} \binom{\frac{n}{2}+1+l}{l}.$$

PROOF. First, we make a substitution $u = n/2$ in (A.1) to obtain an equivalent expression

$$(A.2) \quad \frac{u-l}{u+l} \binom{u+l}{l} + \frac{u-l+2}{u+l} \binom{u+l}{l-1} = \frac{u+1-l}{u+1+l} \binom{u+1+l}{l}.$$

Using easily checked identities,

$$\binom{u+l}{l} = \frac{u+1}{u+1+l} \binom{u+1+l}{l}$$

and

$$\binom{u+l}{l-1} = \frac{l}{u+1+l} \binom{u+1+l}{l},$$

the l.h.s. of (A.2) can be re-written as

$$\frac{(u-l)(u+1) + l(u-l+2)}{(u+l)(u+1+l)} \binom{u+1+l}{l},$$

and the lemma follows from noting that

$$\frac{(u-l)(u+1) + l(u-l+2)}{(u+l)} = u+1-l.$$

□

LEMMA A.2. *Let n be a positive integer, k be an integer satisfying $2 \leq k \leq n$, l be an integer satisfying $1 \leq l \leq \frac{n-k+1}{2}$ and n and k have opposite parities. Then the following holds:*

$$(A.3) \quad \frac{n+k-2l+3}{n+k+2l-1} \binom{\frac{n+k+1}{2}+l-1}{l-1} + \frac{n+k-2l-1}{n+k+2l-1} \binom{\frac{n+k-1}{2}+l}{l} \\ = \frac{n+k-2l+1}{n+k+2l+1} \binom{\frac{n+k+1}{2}+l}{l}.$$

PROOF. We first make a substitution $u = (n+k+1)/2$ in (A.3) and obtain an equivalent expression

$$(A.4) \quad \frac{u-l+1}{u+l-1} \binom{u+l-1}{l-1} + \frac{u-l-1}{u+l-1} \binom{u+l-1}{l} = \frac{u-l}{u+l} \binom{u+l}{l}.$$

Using easily verified identities

$$\binom{u+l-1}{l-1} = \frac{l}{u+l} \binom{u+l}{l} \quad \text{and} \quad \binom{u+l-1}{l} = \frac{u}{u+l} \binom{u+l}{l},$$

and dividing both sides of (A.4) by

$$\frac{1}{(u+l-1)(u+l)} \binom{u+l}{l},$$

the result follows. \square

Before we prove the theorem, note that in the light of Proposition A.1, the assumptions on n and k are for the purpose of identifying the nonzero probability events. Also, due to symmetry, we can restrict the proof to the case of nonnegative k . The proof is by induction and involves a series of straightforward calculations. The theorem is trivially true for the cases $n = 1$ and $n = 2$. We assume the theorem is true for all positive integers up to and including n and prove that it is true for $n + 1$. The proof is broken out into four cases: $k = 0$, $k = 1$, $2 \leq k \leq n$ and $k = n + 1$.

PROOF OF THEOREM 2.1.

Case $k = 0$:

$$\begin{aligned} P(D_{n+1} = 0) &= 2pP(D_n = 1) \\ &= 2p \cdot \frac{1}{2} p^{\frac{n-1}{2}} \sum_{l=0}^{\frac{n-1}{2}} \frac{n+1-2l}{n+1+2l} \binom{\frac{n+1}{2}+l}{l} q^{1+l-1} \\ &= p^{\frac{n+1}{2}} \sum_{l=0}^{\frac{n-1}{2}} \frac{n+1-2l}{n+1+2l} \binom{\frac{n+1}{2}+l}{l} q^l, \end{aligned}$$

which is exactly (2.2) with n replaced by $n + 1$.

Case $k = 1$: We need to show that

$$(A.5) \quad P(D_{n+1} = 1) = \frac{1}{2} p^{\frac{n}{2}} \sum_{l=0}^{\frac{n}{2}} \frac{n+2-2l}{n+2+2l} \binom{\frac{n}{2}+1+l}{l} q^l.$$

Then

$$\begin{aligned} P(D_{n+1} = 1) &= \frac{1}{2} P(D_n = 0) + p P(D_n = 2) \\ &= \frac{1}{2} p^{\frac{n}{2}} \sum_{l=0}^{\frac{n}{2}-1} \frac{n-2l}{n+2l} \binom{\frac{n}{2}+l}{l} q^l \\ &\quad + p \cdot \frac{1}{2} p^{\frac{n-2}{2}} \sum_{l=0}^{\frac{n-2}{2}} \frac{n+2-2l}{n+2+2l} \binom{\frac{n+2}{2}+l}{l} q^{2+l-1}. \end{aligned}$$

Now we shift the summation index in the second term, $l := l + 1$, and then collect the terms under a single summation:

$$\begin{aligned} P(D_{n+1} = 1) &= \frac{1}{2} p^{\frac{n}{2}} \sum_{l=0}^{\frac{n}{2}-1} \frac{n-2l}{n+2l} \binom{\frac{n}{2}+l}{l} q^l \\ &\quad + \frac{1}{2} p^{\frac{n}{2}} \sum_{l=1}^{\frac{n}{2}} \frac{n+2-2(l-1)}{n+2+2(l-1)} \binom{\frac{n+2}{2}+l-1}{l-1} q^l \\ (A.6) \quad &= \frac{1}{2} p^{\frac{n}{2}} \left\{ \sum_{l=1}^{\frac{n}{2}-1} \left[\frac{n-2l}{n+2l} \binom{\frac{n}{2}+l}{l} + \frac{n-2l+4}{n+2l} \binom{\frac{n}{2}+l}{l-1} \right] q^l \right\} \\ &\quad + \frac{1}{2} p^{\frac{n}{2}} \left\{ 1 + \frac{2}{n} \binom{n}{\frac{n}{2}-1} q^{\frac{n}{2}} \right\}. \end{aligned}$$

Similar to the r.h.s. of (A.5), the expression obtained in (A.6) is a product of $p^{n/2}/2$ and a $(n/2)$ th order polynomial in q . Therefore it remains to show that the polynomial inside the curly braces in (A.6) is the same as the polynomial in the r.h.s. of (A.5). We will show term by term equality. First, the constant term in (A.6) is 1 which is the same as the constant term in (A.5). To show that the coefficients of $q^{\frac{n}{2}}$ are equal we need to show the following equality:

$$\frac{2}{n} \binom{n}{\frac{n}{2}-1} = \frac{1}{n+1} \binom{n+1}{\frac{n}{2}}.$$

We transform the l.h.s. to obtain the right hand side as follows:

$$\begin{aligned} \frac{2}{n} \binom{n}{\frac{n}{2}-1} &= \frac{1}{\frac{n}{2}} \cdot \frac{n!}{(\frac{n}{2}-1)!(\frac{n}{2}+1)!} = \frac{n!}{(\frac{n}{2})!(\frac{n}{2}+1)!} \\ &= \frac{1}{n+1} \cdot \frac{(n+1)!}{(\frac{n}{2})!(\frac{n}{2}+1)!} = \frac{1}{n+1} \binom{n+1}{\frac{n}{2}}. \end{aligned}$$

To complete the proof for the case $k = 1$ it remains to show that the coefficients of q^l are equal for $0 < l < n/2$. This is contained in Lemma A.1.

Case $2 \leq k \leq n$: We need to show that

$$(A.7) \quad P(D_{n+1} = k) = \frac{1}{2} p^{\frac{n-k+1}{2}} \sum_{l=0}^{\frac{n-k+1}{2}} \frac{n+k-2l+1}{n+k+2l+1} \binom{\frac{n+k+1}{2}+l}{l} q^{k+l-1}.$$

When $k = n, n+1$ and k have opposite parities; therefore we can assume that $2 \leq k \leq n-1$. We have

$$\begin{aligned} P(D_{n+1} = k) &= pP(D_n = k+1) + qP(D_n = k-1) \\ &= p \cdot \frac{1}{2} p^{\frac{n-k-1}{2}} \sum_{l=0}^{\frac{n-k-1}{2}} \frac{n+k-2l+1}{n+k+2l+1} \binom{\frac{n+k+1}{2}+l}{l} q^{k+l} \\ &\quad + q \cdot \frac{1}{2} p^{\frac{n-k+1}{2}} \sum_{l=0}^{\frac{n-k+1}{2}} \frac{n+k-2l-1}{n+k+2l-1} \binom{\frac{n+k-1}{2}+l}{l} q^{k+l-2}. \end{aligned}$$

Now we shift the summation index in the first term, $l := l+1$, and then collect the terms under a single summation to obtain

$$\begin{aligned} (A.8) \quad &P(D_{n+1} = k) \\ &= \frac{1}{2} p^{\frac{n-k+1}{2}} q^{k-1} \sum_{l=1}^{\frac{n-k+1}{2}} \frac{n+k-2l+3}{n+k+2l-1} \binom{\frac{n+k+1}{2}+l-1}{l-1} q^l \\ &\quad + \frac{1}{2} p^{\frac{n-k+1}{2}} q^{k-1} \sum_{l=0}^{\frac{n-k+1}{2}} \frac{n+k-2l-1}{n+k+2l-1} \binom{\frac{n+k-1}{2}+l}{l} q^l \\ &= c \left\{ \sum_{l=1}^{\frac{n-k+1}{2}} \left[\frac{n+k-2l+3}{n+k+2l-1} \binom{\frac{n+k+1}{2}+l-1}{l-1} \right] \right\} \\ &\quad + c \left\{ \sum_{l=1}^{\frac{n-k+1}{2}} \left[\frac{n+k-2l-1}{n+k+2l-1} \binom{\frac{n+k-1}{2}+l}{l} \right] q^l + 1 \right\}, \end{aligned}$$

where $c = p^{(n-k+1)/2}q^{k-1}/2$. Comparing (A.8) with (A.7) we immediately see that the terms corresponding to $l = 0$ are equal to c . To complete the proof for the case $2 \leq k \leq n$ all that remains is an application of Lemma A.2.

Case $k = n + 1$: This follows immediately from the fact that

$$P(D_{n+1} = n + 1) = \frac{1}{2}q^n.$$

The theorem is proved. \square

APPENDIX B: PROOF OF THEOREM 2.2

The following proposition follows immediately from the Markovian property and time homogeneity of the $BCD(p)$ process.

PROPOSITION B.1. *Let $n = 0, 1, 2, 3, \dots$, $m = 1, 2, 3, \dots$ and $m \geq n$. Define $\sigma(\mathbf{T}_n)$ to be the sigma-algebra generated by T_1, \dots, T_n . The following hold for the $BCD(p)$:*

1. $P(T_m = \pm 1 | D_n, \sigma(\mathbf{T}_n)) = P(T_m = \pm 1 | D_n)$.
2. $P(T_m = \pm 1 | D_n = k) = P(T_{m+l} = \pm 1 | D_{n+l} = k)$, for any $l \geq 0$.

Next we state and prove three lemmas that are used in the proof of Theorem 2.2.

LEMMA B.1. *Let $1 \leq n < m$. Then the following holds for the $BCD(p)$, $p \in [0, 1]$,*

$$(B.1) \quad P(T_n = 1, T_m = 1) = \sum_{k=-n+1}^{n-1} P(T_m = 1 | D_n = k + 1) d_{n-1,k} t_k.$$

PROOF. Before providing the proof, note that because Theorem 2.1 gives the form of $d_{n,k}$, the lemma reduces the finding of $P(T_n = 1, T_m = 1)$ to finding conditional probabilities of the form $P(T_m = 1 | D_n = k)$. By conditioning on D_{n-1} we obtain

$$(B.2) \quad P(T_n = 1, T_m = 1) = \sum_{k=-n+1}^{n-1} P(T_n = 1, T_m = 1 | D_{n-1} = k) \cdot P(D_{n-1} = k).$$

Note that Equation (B.2) holds for $n = 1$ as well, because we defined $P(D_0 = 0) = 1$. Also, instead of requiring $n - k$ be odd so that $P(D_{n-1} = k) > 0$, we follow the adopted convention that the probabilities of events conditional on zero-probability event are treated as 0.

Now we make use of an easily verified identity

$$P(A \cap B|C) = P(A|B \cap C) \cdot P(B|C),^1$$

to transform the conditional probabilities in the r.h.s. of (B.2):

$$(B.3) \quad \begin{aligned} P(T_n = 1, T_m = 1|D_{n-1} = k) \\ = P(T_m = 1|T_n = 1, D_{n-1} = k)P(T_n = 1|D_{n-1} = k). \end{aligned}$$

Now we use the fact that the following two events are equal:

$$\{D_{n-1} = k \text{ and } T_n = 1\} \text{ and } \{D_n = k + 1 \text{ and } T_n = 1\},$$

and that T_m is conditionally independent of T_n given D_n to write

$$\begin{aligned} P(T_m = 1|T_n = 1, D_{n-1} = k) &= P(T_m = 1|T_n = 1, D_n = k + 1) \\ &= P(T_m = 1|D_n = k + 1). \end{aligned}$$

Substituting this last expression into the r.h.s. of (B.3) we obtain

$$P(T_n = 1, T_m = 1|D_{n-1} = k) = P(T_m = 1|D_n = k + 1)P(T_n = 1|D_{n-1} = k).$$

The result follows from substitution into (B.2). \square

The next lemma is devoted to finding the first visit probabilities of the imbalance process into the 0 state. We define τ_i to be the number of steps the imbalance process makes to visit state 0 for the first time from the i th state.

LEMMA B.2. *For the imbalance process of the BCD(p), $p \in [0, 1]$, the probabilities of the first visits from state k , $k = \pm 1, \pm 2, \dots$, into state 0 in exactly l steps, $l \geq |k|$, is given by the following formula*

$$(B.4) \quad f_{k,0}^{(l)} = P(\tau_k = l) = \frac{|k|}{l} \binom{l}{\frac{l+|k|}{2}} p^{\frac{l+|k|}{2}} q^{\frac{l-|k|}{2}},$$

where, according to the adopted convention, the combination is to be treated as 0 when $(l + |k|)/2$ is not an integer.

¹This identity still holds when either B or C have zero probability when used with the adopted convention.

PROOF. First, due the symmetry, $f_{k,0}^{(l)} = f_{-k,0}^{(l)}$, for any $k \in \mathbb{N}$. Therefore without loss of generality, we can assume that k is positive. Thus we are concerned with finding first visit probabilities from state $k, k \in \mathbb{N}$ into state 0 in exactly $l, l \geq k$, steps.

We can treat this problem as a random walk on the nonnegative integers with an absorbing barrier at 0 and use well-known results in the classical gambler's ruin problem where the gambler plays with infinitely reach adversary and at each step wins one unit with probability q and loses one unit with probability p . The question is equivalently formulated as: what is the probability that a gambler with initial capital of $k, k \in \mathbb{N}$, is ruined in exactly $l, l \geq k$, steps? These probabilities are well known and can be found in equation (4.14) of Feller (1968). One needs to reverse the roles of p and q and replace z with k and n with l . \square

Lemma B.2 provides all the nontrivial probabilities for $f_{k,0}^{(l)}$. To complete the remaining cases, we note that $f_{k,0}^{(0)} = 0$ when $k \neq 0$, and $f_{0,0}^{(0)} = 1$.

The next lemma provides probabilities for the imbalance process to ultimately reach the 0 state from any other state.

LEMMA B.3. *For the imbalance process of the BCD(p), $p \in [0.5, 1]$, the probability of ultimately reaching state 0 from state $k, k = \pm 1, \pm 2, \dots$, is 1.*

PROOF. The proof of the lemma is similar to that of Lemma B.2. Again, without loss of generality, it can be assumed that k is positive. The problem is equivalent to computing the probability of ultimate ruin in the classical gambler's ruin problem when the gambler, having an initial capital k , plays with infinitely reach adversary and at each step wins one unit with probability q and loses one unit with probability p . These probabilities can be found in equation (2.18) of Feller (1968). One needs to reverse the roles of p and q and replace z with k . \square

Note that Lemma B.3 implies that $f_{k,0}^{(l)}$ is a probability mass function when $p \in [0.5, 1]$.

Before starting the proof of Theorem 2.2, note that only about half of the summands in the r.h.s. of (2.3) will be nonzero because $d_{n,k} = 0$ whenever $n - k$ is not even.

PROOF OF THEOREM 2.2. The essence of the proof is in evaluating conditional probabilities of the form $P(T_m = 1 | D_n = k)$. We will show that for $1 \leq n < m$, $|k| \leq n$ and $n - k$ even, the following holds:

$$(B.5) \quad P(T_m = 1 | D_n = k) = \left(\frac{1}{2} - t_k \right) \hat{f}_{k,0}^{(m-n-1)} + t_k.$$

This equation is of independent interest as it provides the form of probabilities of treatment assignments conditional on a past value of imbalance. Note that when $k = 0$, equation (B.5) simply states that $P(T_m = 1|D_n = 0) = 1/2$, as expected. The case when $m = n + 1$ is the definition of the BCD(p).

To prove (B.5), we use a conditioning argument and condition on the first visit events into the 0 state, as follows:

$$(B.6) \quad P(T_m = 1|D_n = k) = \sum_{l=0}^{m-n-1} P(T_m = 1|D_n = k, \tau_k = l)P(\tau_k = l|D_n = k) \\ + P(T_m = 1|D_n = k, \tau_k \notin [0, m-n-1])P(\tau_k \notin [0, m-n-1]|D_n = k).$$

We first evaluate $P(T_m = 1|D_n = k, \tau_k = l)$ for the case $0 \leq l \leq m - n - 1$. We only need to look at the cases when $(l - |k|)/2$ is a nonnegative integer, because in all other cases $P(\tau_k = l|D_n = k) = 0$.

$$P(T_m = 1|D_n = k, \tau_k = l) \\ = P(T_m = 1|D_n = k, D_{n+1} \neq 0, D_{n+2} \neq 0, \dots, D_{n+l-1} \neq 0, D_{n+l} = 0) \\ = P(T_m = 1|D_{n+l} = 0) = P(T_{m-n-l} = 1) = 1/2.$$

The first equality, in the chain of equalities above, is a consequence of the following equality of events:

$$\{D_n = k, \tau_k = l\} = \{D_n = k, D_{n+1} \neq 0, D_{n+2} \neq 0, \dots, D_{n+l-1} \neq 0, D_{n+l} = 0\}.$$

The second equality is just the Markovian property of the imbalance process (see Proposition B.1.1). The third equality follows from time-homogeneity property formulated in Proposition B.1.2. Thus we have proved that when $1 \leq n < m$, $|k| \leq n$, $n - k$ is even, $0 \leq l \leq m - n - 1$ and $(l - |k|)/2$ is a nonnegative integer, then

$$(B.7) \quad P(T_m = 1|D_n = k, \tau_k = l) = 1/2.$$

Now we turn to the case when $\tau_k \notin [0, m - n - 1]$. As before, we have $1 \leq n < m$, $|k| \leq n$ and $n - k$ is even. We look at three sub-cases.

Case $k > 0$:

$$P(T_m = 1|D_n = k, \tau_k \geq m - n) \\ = P(T_m = 1|D_n = k, D_{m-1} > 0, \tau_k \notin [0, m - n - 1]) \\ = P(T_m = 1|D_{m-1} > 0) = q.$$

The first equality above follows from equality of $\{D_n = k, \tau_k \geq m - n\}$ and $\{D_n = k, D_{m-1} > 0, \tau_k \notin [0, m - n - 1]\}$. The second equality follows from Proposition B.1.1.

Case $k < 0$:

$$\begin{aligned} P(T_m = 1 | D_n = k, \tau_k \geq m - n) \\ &= P(T_m = 1 | D_n = k, D_{m-1} < 0, \tau_k \notin [0, m - n - 1]) \\ &= P(T_m = 1 | D_{m-1} < 0) = p. \end{aligned}$$

The first equality above follows from equality of $\{D_n = k, \tau_k \geq m - n\}$ and $\{D_n = k, D_{m-1} < 0, \tau_k \notin [0, m - n - 1]\}$. The second equality follows from Proposition B.1.1.

Case $k = 0$:

$$P(T_m = 1 | D_n = 0, \tau_0 \notin [0, m - n - 1]) = 0,$$

because of impossibility of the event $\{\tau_0 \geq 1\}$.

Substituting (B.7) and the expressions obtained in the above three cases into (B.6), we obtain

$$(B.8) \quad P(T_m = 1 | D_n = k) = \sum_{l=0}^{m-n-1} \frac{1}{2} f_{k,0}^{(l)} + t_k P(\tau_k \notin [0, m - n - 1] | D_n = k).$$

According to Lemma B.3, when $p \in [1/2, 1]$, we have

$$(B.9) \quad P(\tau_k \notin [0, m - n - 1] | D_n = k) = 1 - P(\tau_k \in [0, m - n - 1] | D_n = k).$$

Substituting (B.9) into (B.8), we obtain

$$\begin{aligned} P(T_m = 1 | D_n = k) &= \sum_{l=0}^{m-n-1} \frac{1}{2} f_{k,0}^{(l)} + t_k (1 - \hat{f}_{k,0}^{(m-n-1)}) \\ &= \frac{1}{2} \hat{f}_{k,0}^{(m-n-1)} + t_k (1 - \hat{f}_{k,0}^{(m-n-1)}) \\ &= \left(\frac{1}{2} - t_k\right) \hat{f}_{k,0}^{(m-n-1)} + t_k. \end{aligned}$$

Thus (B.5) is proved. To complete the proof of the theorem, it remains to use Lemma B.1 and substitute (B.5) with $k := k + 1$ into (B.1). \square

APPENDIX C: PROOF OF THEOREM 4.1

We will show that $2p$ is an eigenvalue of $\Sigma_{\mathbf{T}_n}$ with the corresponding normalized eigenvector $\mathbf{a}^n = (a_1^n, a_2^n, a_3^n, \dots, a_n^n)' = (\sqrt{2}/2, -\sqrt{2}/2, 0, \dots, 0)'$. The proof proceeds by induction. The theorem is trivially true for the case $n = 2$. For the case $n = 2$, one can actually show that $2p$ is the maximum eigenvalue.² The two eigenvalues of $\Sigma_{\mathbf{T}_2}$ are $2p$ and $2 - 2p$ and $2p \geq 2 - 2p$ when $p \geq 1/2$.

²The same can be shown for $n = 3$ and $n = 4$ by solving for the zeroes of the characteristic polynomials.

PROOF. We assume the theorem is true for all positive integers $n \geq 2$, and prove that it is true for $n + 1$. We partition $\Sigma_{T_{n+1}}$ as follows:

$$\left[\begin{array}{c|c} \Sigma_{T_n} & \mathbf{b} \\ \hline \mathbf{b}' & 1 \end{array} \right],$$

where $\mathbf{b} = (\sigma_{1,n+1}, \sigma_{2,n+1}, \dots, \sigma_{n,n+1})'$, with $\sigma_{ij} = \text{Cov}(T_i, T_j)$. Denote

$$\mathbf{x} = (\mathbf{x}_1, 0)'$$

where \mathbf{x}_1 is the n -dimensional vector:

$$\mathbf{x}_1 = (\sqrt{2}/2, -\sqrt{2}/2, 0, \dots, 0)'$$

We need to show that:

$$(C.1) \quad \left[\begin{array}{c|c} \Sigma_{T_n} & \mathbf{b} \\ \hline \mathbf{b}' & 1 \end{array} \right] \left[\begin{array}{c} \mathbf{x}_1 \\ 0 \end{array} \right] = 2p \left[\begin{array}{c} \mathbf{x}_1 \\ 0 \end{array} \right].$$

By the induction assumption, we have that $\Sigma_{T_n} \mathbf{x}_1 = 2p \mathbf{x}_1$. To prove Equation (C.1), it remains to show that $\mathbf{b}' \mathbf{x}_1 = 0$. This is equivalent to $\sqrt{2}/2(\sigma_{1,n+1} - \sigma_{2,n+1}) = 0$, which in turn is equivalent to (see Corollary 2.1):

$$(C.2) \quad P(T_1 = 1, T_{n+1} = 1) = P(T_2 = 1, T_{n+1} = 1).$$

From Theorem 2.2 we have the forms of $P(T_1 = 1, T_{n+1} = 1)$ and $P(T_2 = 1, T_{n+1} = 1)$:

$$\begin{aligned} P(T_1 = 1, T_{n+1} = 1) &= \left(\left(\frac{1}{2} - q \right) \hat{f}_{1,0}^{(n-1)} + q \right) \frac{1}{2} = \frac{1}{2} \left(\frac{1}{2} - q \right) \hat{f}_{1,0}^{(n-1)} + \frac{1}{2} q, \\ P(T_2 = 1, T_{n+1} = 1) &= \left(\left(\frac{1}{2} - \frac{1}{2} \right) \hat{f}_{0,0}^{(n-2)} + \frac{1}{2} \right) \frac{1}{2} p + \left(\left(\frac{1}{2} - q \right) \hat{f}_{2,0}^{(n-2)} + q \right) \frac{1}{2} q \\ &= \frac{1}{4} p + \frac{1}{2} q \left(\frac{1}{2} - q \right) \hat{f}_{2,0}^{(n-2)} + \frac{1}{2} q^2. \end{aligned}$$

Thus, in order to show (C.2), we need to show that:

$$(C.3) \quad \frac{1}{2} \left(\frac{1}{2} - q \right) \hat{f}_{1,0}^{(n-1)} + \frac{1}{2} q = \frac{1}{4} p + \frac{1}{2} q \left(\frac{1}{2} - q \right) \hat{f}_{2,0}^{(n-2)} + \frac{1}{2} q^2.$$

Using an easily verified identity:

$$\hat{f}_{1,0}^{(n-1)} = p + q \hat{f}_{2,0}^{(n-2)},$$

and substituting it into (C.3), we obtain:

$$(C.4) \quad \frac{1}{2} \left(\frac{1}{2} - q \right) \left(p + q \hat{f}_{2,0}^{(n-2)} \right) + \frac{1}{2} q = \frac{1}{4} p + \frac{1}{2} q \left(\frac{1}{2} - q \right) \hat{f}_{2,0}^{(n-2)} + \frac{1}{2} q^2.$$

The term $\frac{1}{2} q \left(\frac{1}{2} - q \right) \hat{f}_{2,0}^{(n-2)}$ appears in both sides of (C.4); subtracting it from both sides we require:

$$\frac{1}{2} p \left(\frac{1}{2} - q \right) + \frac{1}{2} q = \frac{1}{4} p + \frac{1}{2} q^2.$$

This last equality is trivially checked. \square

REFERENCES

- [1] BALDI ANTOGNINI, A. (2008). A theoretical analysis of the power of biased coin designs. *J. Statist. Plann. Inf.* **138** 1792–1798.
- [2] BALDI ANTOGNINI, A. AND GIOVAGNOLI, A. (2004). A new ‘biased coin design’ for the sequential allocation of two treatments. *J. Roy. Statist. Soc. – Ser. C* **53** 651–664.
- [3] BLACKWELL, D. AND HODGES, J. L. (1957). Design for the control of selection bias. *Ann. Math. Statist.* **28** 449–460.
- [4] CHEN, Y-P. (1999). Biased coin design with imbalance intolerance. *Commun. Statist. – Stochastic Models*, **15**: 953–975.
- [5] EFRON, B. (1971). Forcing a sequential experiment to be balanced. *Biometrika* **58** 403–417.
- [6] EISELE, J. R. (1994). The doubly adaptive biased coin design for sequential clinical trials. *J. Statist. Plann. Inf.* **38** 249–261.
- [7] FELLER, W. (1968). *An Introduction to Probability Theory and Its Applications, Volume I*. Wiley, New York.
- [8] HOLLANDER, M. AND PEÑA, E. (1988). Nonparametric tests under restricted treatment-assignment rules. *J. Amer. Statist. Assoc.* **83** 1141–1151.
- [9] HU, F., ROSENBERGER, W. F., AND ZHANG, L-X. (2006). Asymptotically best response-adaptive randomization procedures. *J. Statist. Plann. Inf.* **136** 1911–1922.
- [10] HU, F. AND ZHANG, L.-X. (2004). Asymptotic properties of doubly adaptive biased coin designs for multi-treatment clinical trials. *Ann. Statist.* **32** 268–301.
- [11] MARKARYAN, T. (2009). *Exact Distributional Properties of Efron’s Biased Coin Design with Applications to Clinical Trials*. George Mason University, Fairfax (doctoral dissertation).
- [12] ROSENBERGER, W. F. AND LACHIN, J. L. (2002). *Randomization in Clinical Trials: Theory and Practice*. Wiley, New York.
- [13] SMITH, R. L. (1984). Sequential treatment allocation using biased coin designs. *J. Roy. Statist. Soc. – Ser. B* **46** 519–543.
- [14] SMYTHE, R. T. AND WEI, L. J. (1983). Significance tests with restricted randomization. *Biometrika* **70** 496–500.
- [15] SOARES, J. F. AND WU, C. F. J. (1982). Some restricted randomization rules in sequential designs. *Commun. Statist. – Theor. Meth.* **12**: 2017–2034.
- [16] STEELE, J. M. (1980). Efron’s conjecture on vulnerability to bias in a method for balancing sequential trials. *Biometrika* **67** 503–504.
- [17] WEI, L. J. (1978). The adaptive biased coin design for sequential experiments. *Ann. Statist.* **6** 92–100.

ACCENTURE, L.L.P.
 ONE FREEDOM SQUARE
 11951 FREEDOM DRIVE
 RESTON, VA 20190
 E-MAIL: tigran.markaryan@accenture.com

DEPARTMENT OF STATISTICS
 GEORGE MASON UNIVERSITY
 4400 UNIVERSITY DRIVE, MS 4A7
 FAIRFAX, VA 22030-4444
 E-MAIL: wrosenbe@gmu.edu