FORCING A SEQUENTIAL EXPERIMENT TO BE BALANCED

BY

BRADLEY EFRON

TECHNICAL REPORT NO. 14
OCTOBER 30, 1970

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT GP-15909

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
FORCING A SEQUENTIAL EXPERIMENT TO BE BALANCED

By

Bradley Efron

TECHNICAL REPORT NO. 14

OCTOBER 30, 1970

Prepared Under the Auspices of
National Science Foundation Grant GF-15909

Reproduction in Whole or in Part is Permitted for any Purpose of the United States Government

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
Forcing a Sequential Experiment to be Balanced

By

Bradley Efron

1. The Problem. The random assignment of subjects to "Treatment" and "Control" groups is a canon of good experimental design. In situations where the subjects are not available to the experimenter all at one time but rather arrive sequentially,* as is often the case in medical experimentation, Complete Randomization is achieved by the flip of a coin, that is by assigning each subject randomly to the Treatment or Control group with equal probability, independently of the assignment of the other subjects.

Complete Randomization has several attractive properties (which are listed below) but suffers from the disadvantage that in experiments which are limited to a small number of subjects, the final distribution of Treatments and Controls can be very unbalanced. For example, table 1 shows the distribution of Treatments and Controls in an investigation of a new treatment for Hodgkin's disease.** The patients were divided into four age categories since age was thought to be a possible factor influencing the response. It obviously would be preferable to have the youngest age-group more equitably divided, as would certainly have been done if all 29 patients could have been assigned at once.

*This is what "sequential experiment" means here, not necessarily implying the use of sequential analysis on the results of the experiment.

**Data courtesy of Dr. Henry Kaplan, Dept. of Radiology, Stanford Univ.
Table 1. Distribution of Treatments and Controls in Four Age Categories, Hodgkin's Disease Investigation.

It is simple to devise methods of sequential experimentation which force approximately equal numbers of Treatments and Controls. For example, the completely non-random Systematic Design TCTCTC... ("T" for Treatment, "C" for Control) followed separately in each category of subjects guarantees that the unbalance will never exceed one in any category. A variant is the "Student Sandwich Plan" TCCTCCTT... However, these plans can easily bias the results of the experiment, as shown below, and are not serious competitors to Complete Randomization.)*

The problem then is to compromise between a perfectly balanced experiment and the advantages of Complete Randomization. These advantages are:

(1) Freedom from Selection Bias. If the experimenter knows for certain that the next assignment will be a Treatment (or a Control) he may consciously or unconsciously bias the experiment by such decisions as who is or isn't a suitable experimental subject, which category the subject belongs in, etc. It is obvious that Complete Randomization

*A controversy between Student and Fisher arose on a closely related point in the 1930's, see [8], [1], [9] and also [5].
eliminates selection bias,* and that the Systematic Designs maximize it.

(2) Freedom from Accidental Bias. Known or unknown to the experimenter, there may be nuisance factors systematically affecting the experimental units. Typical examples are time trends, sex-linked differences, differing experimental conditions, etc. Complete Randomization tends to balance out such factors (see section 5) and thus protect the significance level of the usual hypothesis tests. The Systematic Designs mentioned above are quite vulnerable to accidental bias.

(3) Randomization as a Basis for Inference. Probability statements, such as the obtained significance level of the experiment, can be based entirely on the randomness induced by the Complete Randomization between Treatments and Controls. This eliminates the need for probability assumptions on the responses of the individual experimental units and guarantees the validity of the stated significance level. Advantage 3 is closely related to but not identical with Advantage 2.

One compromise between Complete Randomization and balanced Systematic Designs that is used in practice is the Permutted Block Design. This design divides the experiment into blocks of even length, say 2b, and within each block randomly assigns $b$ units to Treatment and $b$ units

---

*Blackwell and Hodges [2] coined the term "selection bias". They advocate using Complete Randomization but continuing the experiment until there is a certain minimum number of both Treatments and Controls. In practice this can be very difficult advice to follow, particularly in an experiment such as the Hodgkin's Disease Investigation where there are many categories of subjects.
to Control, all \( \binom{2b}{b} \) combinations being equally likely. Permutated Blocks can be quite effective in eliminating unbalanced designs but they suffer from the disadvantage that at certain points in the experiment the experimenter knows for certain whether the next subject will be assigned as a Treatment or as a Control. (For example, if \( b=5 \) the probability is \( 1/6 \) that the experimenter will know for certain the assignment of units 8, 9, 10, and \( 4/9 \) that he will know for certain the assignment of units 9, 10.

The Balanced Coin Designs introduced in section 2 are motivated by the desire to achieve balanced experiments without ever giving the experimenter a high probability of guessing the assignment of the next unit. Comparisons are made between these designs and Permutated Blocks in the later sections.

2. **Biased Coin Designs.** Suppose that at a certain stage in the experiment a new subject arrives and is noted to be in a category which has had \( \tilde{D} \) more Treatments than Controls previously assigned to it. We assign the new subject as follows:

\[
\begin{align*}
\text{If } \tilde{D} > 0 \text{ assign Treatment with probability } q \text{ and Control with probability } p \\
\text{If } \tilde{D} = 0 \text{ assign Treatment with probability } \frac{1}{2} \text{ and Control with probability } \frac{1}{2} \\
\text{If } \tilde{D} < 0 \text{ assign Treatment with probability } p \text{ and Control with probability } q
\end{align*}
\]

(2.1)

Here \( p \geq q, p+q=1 \), so that the assignment rule tends to balance the number of Treatments and Controls, the tendency being weakest if \( p=\frac{1}{2} \).
(Complete Randomization) and strongest if \( p = 1 \) (Permuted Block Design with \( b = 1 \)).

We will call the rule described in 2.1 the Biased Coin Design with bias \( p \), abbreviated BCD(p) for convenient reference. The rule is meant to be applied separately within each category of the subjects, and so for our purposes we can think of each category as being a separate experiment in which we are trying to balance Treatments and Controls. In what follows we drop all reference to separate categories.

The value \( p = 2/3 \), which is the author's personal favorite, will be seen to yield generally good designs and will be featured in the numerical computations.

Section 3 discusses the balancing properties of the Biased Coin Designs. Selection bias, accidental bias, and randomization as a basis for inference are discussed in sections 4, 5 and 6. All proofs are deferred until section 7.

3. Balancing Properties of the Biased Coin Designs. Define \( D_n \) to be the absolute difference between the number of Treatments and number of Controls after \( n \) assignments have been made (\( D_0 = 0 \)). Under BCD(p) the \( D_n \) form a Markov chain with states 0,1,2,... and transition probabilities

\[
\begin{align*}
P(D_{n+1} = j + 1|D_n = j) &= q \\
P(D_{n+1} = j - 1|D_n = j) &= p & j \geq 1 \\
\text{and} & \\
P(D_{n+1} = 1|D_n = 0) &= 1.
\end{align*}
\]
This is a random walk with a reflecting barrier at the origin, \[3, \text{ page } 41\] and has stationary probabilities \(\pi_j\) given by

\[
\pi_0 = \frac{r-1}{2r}, \quad \pi_j = \frac{r-1}{2r} \frac{r^j}{r^j}, \quad j \geq 1,
\]

where

\[r = \frac{p}{q}.
\]

The first few values of the \(\pi_j\) are given for \(r=2, 3,\), and \(4\) in table 2.

<table>
<thead>
<tr>
<th>(r=2) ((p=\frac{2}{3}))</th>
<th>(\frac{1}{4})</th>
<th>(\frac{3}{8})</th>
<th>(\frac{3}{16})</th>
<th>(\frac{3}{32})</th>
<th>(\frac{3}{64})</th>
<th>(\frac{3}{128})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(r=3) ((p=\frac{3}{4}))</td>
<td>(\frac{1}{3})</td>
<td>(\frac{4}{9})</td>
<td>(\frac{4}{27})</td>
<td>(\frac{4}{81})</td>
<td>(\frac{4}{243})</td>
<td>(\frac{4}{729})</td>
</tr>
<tr>
<td>(r=4) ((p=\frac{4}{5}))</td>
<td>(\frac{3}{8})</td>
<td>(\frac{15}{32})</td>
<td>(\frac{15}{124})</td>
<td>(\frac{15}{512})</td>
<td>(\frac{15}{2048})</td>
<td>(\frac{15}{8192})</td>
</tr>
</tbody>
</table>

Table 2. Values of the First Few Stationary Probabilities.

Since the \(D_n\) can take on only odd or even values as \(n\) is odd or even, the Markov chain has period 2, and the limiting probabilities should be doubled accordingly,

\[
\lim_{m \to \infty} P(D_{2m} = 0) = 2\pi_0 = \frac{r-1}{r}, \quad \lim_{m \to \infty} P(D_{2m+1} = 1) = 2\pi_1 = \frac{r^2-1}{r^2}
\]

etc. Thus with \(p=2/3\) the experiment has asymptotic probability 1/2 of being exactly balanced for \(n\) even, and asymptotic probability 3/4 of being as close as possible to balanced for \(n\) odd.

The experiment starts with \(D_0=0\) so it is natural to expect the limiting distribution of \(D_n\) to be approached "from below". This is made precise as follows:
Theorem 1. If \( h(j) \) is a non-decreasing function of \( j \) for \( j=0,1,2, \ldots \), then \( \mathbb{E} h(D_{n+2}) \geq \mathbb{E} h(D_n) \) for every value of \( n \).

Taking \( h(0)=0, h(j)=1 \) for \( j>0 \), in the theorem shows that \( P(D_{2m+1}=0) \) is a decreasing function of \( m \), and taking \( h(0)=0, h(1)=0, h(j)=1 \) for \( j>1 \) shows that \( P(D_{2m+1}=1) \) is a decreasing function of \( m \). We can write down the total "bonus" probability of having \( D_{2m}=0 \) or \( D_{2m+1}=1 \) explicitly:

Theorem 2.

\[
\sum_{m=1}^{\infty} \left[ P(D_{2m} = 0) - 2\pi(0) \right] = \frac{1}{r(r-1)},
\]

\[
\sum_{m=1}^{\infty} \left[ P(D_{2m+1} = 1) - 2\pi(1) \right] = \frac{2}{r^2(r-1)}.
\]

(Notice that we are not including the trivial cases \( D_0=0 \) and \( D_1=1 \).)

Table 3 shows the actual distribution of \( D_n \) for \( n=2,3, \ldots, 10 \), for the case \( r=2 \). For \( r=2 \) both sums in theorem 2 equal \( 5 \), and it can be seen from the table that .39 of the even-\( n \) bonus and .35 of the odd-\( n \) bonus have occurred by \( n=10 \).

<table>
<thead>
<tr>
<th>( n \rightarrow )</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>66.7</td>
<td>59.3</td>
<td>56.0</td>
<td>54.1</td>
<td>53.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>88.9</td>
<td>84.0</td>
<td>81.2</td>
<td>79.5</td>
<td>38.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>33.3</td>
<td>37.0</td>
<td>37.9</td>
<td>38.0</td>
<td>17.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>11.1</td>
<td>14.8</td>
<td>16.5</td>
<td>7.0</td>
<td>7.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>3.7</td>
<td>5.8</td>
<td>2.2</td>
<td>2.9</td>
<td>1.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1.2</td>
<td>0.4</td>
<td>0.8</td>
<td>0.3</td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.0</td>
<td>0.3</td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Percentage Probabilities \( 100 \cdot P(D_n=j) \) for \( r=2 \).
A comparison of BCD(p) with \( p=2/3 \) and the Permuted Block Design with \( b=5 \) is given in table 4, where the two plans are seen to behave rather similarly for the crucial small values of \( n \). Comparisons of the probabilities of some early extreme unbalances \( (D_4=4, D_5=3, \text{etc.}) \) reinforce this impression, though the different natures of the two rules make such comparisons difficult.

<table>
<thead>
<tr>
<th>n</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCD(2/3)</td>
<td>66.7</td>
<td>88.9</td>
<td>59.3</td>
<td>84.0</td>
<td>56.0</td>
<td>81.2</td>
<td>54.1</td>
<td>79.5</td>
<td>53.0</td>
</tr>
<tr>
<td>Permuted Blocks b=5</td>
<td>55.6</td>
<td>83.3</td>
<td>47.6</td>
<td>79.3</td>
<td>47.6</td>
<td>83.3</td>
<td>55.5</td>
<td>100.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Table 4. Percentage Probabilities That Experiment Is Exactly Balanced at Stage \( n \) (within 1 if \( n \) is odd).

It seems obvious that increasing the value of \( p \), or equivalently of \( r \), should shrink the distribution of \( D_n \) toward 0. We conclude this section with a theorem to that effect:

**Theorem 3.** If \( h(j) \) is a non-decreasing function of \( j \) for \( j=0,1,2,\ldots \) then \( E_h(D_n) \) is a non-increasing function of \( r \) for \( r \geq 1 \) and any value of \( n \).

4. **Selection Bias.** A natural measure of the selection bias of a sequential design is the expected number of correct guesses the experimenter can make if he guesses optimally. Every guessing strategy is equally useless against Complete Randomization, yielding an expected \( n/2 \) correct guesses in \( n \) trials. It is intuitively clear (and proved in [2]) that the best strategy against a Permuted Block Design is to guess Treatment
or Control on the basis of which has so far occurred least often in the block. Blackwell and Hodges show that this results in an expected 

\[ 2^{2b}/(2b) - 1 \] 

correct guesses per block of length 2b against the corresponding Permuted Block Design. We can think of this as an "excess selection bias" of 

\[ 2^{2b}/(2b) - (b+1) \]  

(4.1)

expected correct guesses per block for the Permuted Block Design compared to Complete Randomization.

The best guessing strategy against a Biased Coin Design is to guess Treatment or Control on the basis of which has so far occurred least often in the experiment (with no preferred guess if there is a tie).

The probability of guessing correctly on trial \( n \) is 

\[ \frac{1}{2} \left( D_{n-1} = 0 \right) + p \left( D_{n-1} > 0 \right) \]  

(4.2)

which asymptotically approaches  

\[ \frac{1}{2} \sigma_0 + p(1-\sigma_0) = \frac{1}{2} + \frac{r-1}{4r} \]  

(4.3)

The excess selection bias of BCD(p) in 2b trials, ignoring initial effects, is therefore 

\[ \frac{r-1}{2r} b \]  

(4.4)

Figure 1 compares 4.1 with 4.4 as a function of \( b \). The case \( r=2 \) (\( p=2/3 \)) is seen to yield the same excess selection bias as a Permuted Block Design with \( b \) between 8 and 9, that is with block length.
between 16 and 18. (For \( r=2 \) the excess is asymptotically \( (r-1)/4r = 1/8 \) excess correct guesses per trial.)

![Graph showing the excess selection bias in 2b Trials.](image)

Figure 1. Excess Selection Bias in 2b Trials.

5. Accidental Bias. Let \( T_k = +1 \) or \(-1\) as the \( k \)th experimental unit is assigned to the Treatment or Control group, and let \( T = (T_1, T_2, \ldots, T_N) \) be the vector of assignments after some fixed number \( N \) of trials. All of the random designs we have mentioned have \( ET=0 \).

In this section we assess the vulnerability of a design to accidental bias by the magnitude of the largest eigenvalue of the covariance matrix \( \xi_T \) of \( T \). This choice is motivated in the following way.

Suppose the responses \( y_k \) of the experimental units are determined by the linear model

\[
y_k = \mu + \alpha t_k + \beta z_k + \epsilon_k, \quad k = 1, 2, \ldots, N
\]  

where \( t_k = +1 \) or \(-1\) as unit \( k \) is in the Treatment or Control group, \( z_k \) is the measurement of some nuisance factor (e.g. age of
subject) on unit \( k \), and the \( \epsilon_k \) are independent \( \mathcal{N}(0, \sigma^2) \) random variables. In vector notation, \( y = \mu + \alpha t + \beta z + \epsilon \), where \( \epsilon = (1, 1, \ldots, 1) \), \( t = (t_1, t_2, \ldots, t_N) \), etc. There is no loss of generality in this model in assuming that

\[
z \cdot \epsilon = 0, \quad \|z\|^2 = 1 .
\] (5.2)

If we test \( \alpha = 0 \) using student’s \( t \) in the usual way, without allowing for the covariate \( z \), there will be a spurious non-centrality parameter of magnitude \( (\beta^2 / \sigma^2) \cdot (z \cdot t)^2 \) in the numerator of student’s statistic when the null hypothesis is true. If we do allow for \( z \) then the non-centrality parameter for testing \( \alpha = 0 \) is

\[
\frac{N \alpha^2}{\sigma^2} \left( 1 - \frac{1}{N} \left[ \left( \frac{e}{\sqrt{N}} \cdot t \right)^2 + (z \cdot t)^2 \right] \right)
\] (5.3)

and we see that the loss of noncentrality due to the nuisance factor is \( (\alpha^2 / \sigma^2) \cdot (z \cdot t)^2 \).

In both cases it is ideal to have \( t \) orthogonal to \( z \), and we are penalized proportionately to \( (z \cdot t)^2 \) as we depart from this ideal. (It is also good to have \( t \) orthogonal to \( e \), i.e. to balance the experiment, which is what we are trying to do in this paper.)

Now if \( t \) is the realization of the random vector \( T \), with mean vector \( 0 \) and covariance matrix \( \Sigma_T \), then for a fixed vector \( z \)

\[
E(z \cdot T)^2 = z \Sigma_T z' .
\] (5.4)

The least favorable vector \( z \) is that vector satisfying 5.2 which maximizes 5.4, and we see that
\[ E(z \cdot T)^2 \leq \text{largest eigenvalue of } \mathcal{Z}_T, \quad (5.5) \]

with equality if the corresponding eigenvector is orthogonal to \( e \).
(This will turn out to be the case for all the designs we are studying.)

For example, under Complete Randomization \( \mathcal{Z}_T = I \) and \( E(z \cdot T)^2 = 1 \) for every \( z \). (Notice that 1 is the smallest possible value for the maximum eigenvalue of \( \mathcal{Z}_T \) since \( \text{tr} \mathcal{T}_k = 0 \) implies \( \text{Var} \mathcal{T}_k = 1 \) and hence trace \( \mathcal{Z}_T = n = \text{sum of eigenvalues of } \mathcal{Z}_T \).) For the Permuted Block Design with block length \( 2b \) we have \( \mathcal{Z}_n = (1 + 1/(2b-1)) I - 1/(2b-1) e'e \) for \( N \) satisfying \( 2 \leq N \leq 2b \). This matrix has largest eigenvalue \( 1 + 1/(2b-1) \), attained for any vector \( z \) orthogonal to \( e \). This same result holds for \( N > 2b \). If we use Permuted Blocks with \( b = 5 \) we therefore increase the maximum vulnerability to accidental bias from 1 to 1.1/9.

We now compute the maximum eigenvalue for the Biased Coin Design. A much simpler, and in some ways more informative, answer is possible if we analyze the process \( T_1, T_2, \ldots, T_N \) as if it were derived from a stationary process. (Specifically, the results obtained below pertain to the sequence \( T_{h+1}, T_{h+2}, \ldots, T_{h+N} \) with both \( h \) and \( N \) approaching infinity.) It then turns out that the maximum value of the spectral density of the process is the maximum eigenvalue we are looking for.

The limiting autocovariance function

\[ \rho_k \equiv \lim_{h \to \infty} E(T_h T_{h+k}) \quad (5.6) \]

has a closed form expression. Define \( B_k(i) \) as the c.d.f. of a binomial random variable with parameters \( k \) and \( p \),

\[ B_k(i) = \sum_{j=0}^{I} \binom{k}{j} p^i q^{k-i}. \quad (5.7) \]
Theorem 4.

\[ \rho_1 = -\psi(r) \frac{1}{2(r+1)} \quad \text{and} \quad \rho_{k+1} - \rho_k = \psi(r) \frac{1}{k} \left\{ \frac{k-2}{2} \sum_{i=0}^{\left[ \frac{k}{2} \right]} B_k(i) + \delta_k B_k(\left[ \frac{k}{2} \right]) \right\} \]

where \( \psi(r) = \frac{(r-1)^2}{r(r+1)} \), \( \delta_k = 0 \) if \( k \) is even and \( =1 \) if \( k \) is odd. (The square brackets indicate the greatest integer function as usual.)

Cor. 1. For \( k \) even and positive \( \rho_{k+1} - \rho_k = \rho_{k+2} - \rho_{k+1} \).

Cor. 2. \( \rho_k \) is negative for \( k \geq 1 \), and \( \rho_{k+1} - \rho_k \) is positive and decreasing.

Cor. 3. \[ \sum_{k=1}^{\infty} \rho_k = -0.5, \quad \sum_{k=1}^{\infty} (-1)^k \rho_k = \frac{(r-1)^2}{r(r+1)} \]

Table 5 gives the first 11 values of \( \rho_k \) for \( r = 2, 3, \) and \( 4 \).

<table>
<thead>
<tr>
<th>( k )</th>
<th>( r = 2 )</th>
<th>( r = 3 )</th>
<th>( r = 4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>1</td>
<td>-0.0833</td>
<td>-0.1667</td>
<td>-0.2250</td>
</tr>
<tr>
<td>2</td>
<td>-0.0556</td>
<td>-0.0833</td>
<td>-0.0900</td>
</tr>
<tr>
<td>3</td>
<td>-0.0463</td>
<td>-0.0625</td>
<td>-0.0630</td>
</tr>
<tr>
<td>4</td>
<td>-0.0370</td>
<td>-0.0417</td>
<td>-0.0360</td>
</tr>
<tr>
<td>5</td>
<td>-0.0319</td>
<td>-0.0325</td>
<td>-0.0263</td>
</tr>
<tr>
<td>6</td>
<td>-0.0267</td>
<td>-0.0234</td>
<td>-0.0166</td>
</tr>
<tr>
<td>7</td>
<td>-0.0234</td>
<td>-0.0187</td>
<td>-0.0124</td>
</tr>
<tr>
<td>8</td>
<td>-0.0201</td>
<td>-0.0140</td>
<td>-0.0082</td>
</tr>
<tr>
<td>9</td>
<td>-0.0178</td>
<td>-0.0113</td>
<td>-0.0062</td>
</tr>
<tr>
<td>10</td>
<td>-0.0155</td>
<td>-0.0087</td>
<td>-0.0042</td>
</tr>
</tbody>
</table>

Table 5. Values of the Autocorrelation Function \( \rho_k \).

Since \( \sum_{k=0}^{\infty} |\rho_k| < \infty \) the process has an absolutely continuous spectrum with spectral density

\[ f(\omega) = \sum_{k=-\infty}^{\infty} \rho_k e^{-i\omega k} = 1 + 2 \sum_{k=1}^{\infty} \rho_k \cos \omega k \quad (5.8) \]
(this is a factor of \(2\pi\) times the usual definition [3]). Corollary 3 can be restated in terms of \(f(\omega)\) as

\[
f(0) = 0, \quad f(\pi) = 1 + \left(\frac{r-1}{r+1}\right)^2 = 1 + (p-q)^2.
\]  

(5.9)

Since \(f(-\omega) = f(\omega)\) the abscissa is plotted only from \(\omega=0\) to \(\omega=\pi\).

![Graph](image-url)

**Figure 2.** Spectral Density of Biased Coin Assignments.

We see that the spectral density increases monotonically as \(\omega\) goes from 0 to \(\pi\) and so obtains its maximum value at \(\pi\). By 5.8 this maximum is equal to \(1 + (p-q)^2\). Therefore \(\text{BCD}(2/3)\) has maximum vulnerability to accidental bias of amount \(1+(2/3-1/3)^2 = 1 + \frac{1}{9}\), the same as the Permuted Block Design with \(b=5\).

Although \(f(\omega)\) was monotone for every value of \(r\) numerically investigated, the author was not able to prove the result in general, and so it remains as a conjecture that \(1 + (p-q)^2\) is the maximum value for \(\text{BCD}(p)\).
It is easy to see that the maximum of the spectral density corresponds to the maximum eigenvalue definition given before. Let \( z_1, z_2, \ldots, z_N \) be any sequence satisfying 5.2 and define
\[
\beta_z(\omega) = \sum_{k=1}^{N} z_k e^{-i\omega k}, \quad L_z(\omega) = \frac{1}{2\pi} |\beta_z(\omega)|^2.
\] (5.10)

Then by Parseval's lemma, \( L_z(\omega) \) is a probability density on \(-[\pi, \pi]\),
\[
\int_{-\pi}^{\pi} L_z(\omega) d\omega = 1,
\]
and by the standard theory of linear transforms, [3] page 321,
\[
E(z'|T)^2 = \int_{-\pi}^{\pi} f(\omega)L_z(\omega) d\omega.
\] (5.11)

We see that the integral is bounded by the maximum value of \( f(\omega) \), a "bad" \( z \) being one with \( L_z(\omega) \) concentrated near the maximizing value of \( \omega \). In the cases above the least favorable \( z \) was asymptotically proportional to the sequence \(+1, -1, +1, -1, +1, -1 \ldots\).

The definition of vulnerability to accidental bias we have given is of a minimax nature and so favors Complete Randomization. Figure 2 and equation 5.11 show that for some values of \( z \), namely those concentrated at low frequencies, the Biased Coin Designs are superior to Complete Randomization.

6. Randomization as a Basis for Inference. Suppose that we have a fixed sample size experiment for comparing \( m \) Treatment measurements \( x_1, x_2, \ldots, x_m \) with \( n \) Control measurements \( y_1, y_2, \ldots, y_n \), \( m+n=N \). Let \( U_1, U_2, \ldots, U_N \) be the experimental units and let \( c \) represent the
collection of \( m \) indices corresponding to Treatments (say \( c \in \mathcal{C} \), where \( \mathcal{C} \) is the set of all choices of \( m \) integers from \( 1, 2, \ldots, N \)).

Whatever test we are using for the hypothesis of no difference between Treatment and Control responses will have a significance level \( \alpha_c \) depending* on the choice of \( c \in \mathcal{C} \). If we use a level \( \alpha \) permutation test, such as the conditional t-test, Wilcoxon's test, etc., then

\[
\frac{1}{\binom{N}{m}} \sum_{c \in \mathcal{C}} \alpha_c = \alpha,
\]

which is just another way of saying that for each set of \( N \) experimental responses the test rejects for \( \alpha \frac{N}{m} \) of the \( \binom{N}{m} \) possible assignments of these responses to the Treatment and Control groups.

Under Complete Randomization if we condition on the number of Treatments which have occurred then \( c \) is in fact randomly selected from \( \mathcal{C} \) with equal probability for all \( \binom{N}{m} \) members, and equation 6.1 allows one to claim an exact significance level of \( \alpha \) for the test without any probability model on the responses of the experimental units. This argument is distinct from that of Section 5, which said that under a reasonable probability model Complete Randomization (and also BCD(p)) was likely to yield a \( c \) with \( \alpha_c \) nearly equal to the nominal \( \alpha \) level. In practice the two points of view conflict only if the randomly selected \( c \) happens to look particularly nonrandom, hence prone to accidental bias, for example if it selects most of the Treatments early in the experiment.

* \( \alpha_c \) will also depend, of course, on the particular experimental conditions, but these are considered fixed in this discussion. Notice that we are assuming that the test statistic is invariant under permutations of the \( x \) values and permutations of the \( y \) values separately, which is the usual case.
The Biased Coin Designs do not give the same conditional distribution of \( c \) and so 6.1 does not apply directly. Theoretically we could redefine the rejection region of any permutation test to give level \( \alpha \) with respect to the distribution of \( c \) under BCD\((p)\) but this is hard work and probably unnecessary as the following asymptotic argument shows.

Assume we are using BCD\((p)\) and that \( N \) is large. We can assume \( m=n=\frac{N}{2} \) without affecting the rough calculation which follows. Let \( w_1, w_2, \ldots, w_N \) be the observed responses on units \( 1,2,\ldots,N \), with the common mean subtracted off so that \( \bar{w}=0 \), and let \( \sigma_w^2 = \frac{\sum w_i^2}{N} / N \).

Suppose our test is based on the conditional student's statistic, that is we reject if the observed value of \( |\bar{x}-\bar{y}| \) is among the \( a(N/2) \) largest of the \( \binom{N}{N/2} \) possible values of \( |\bar{x}-\bar{y}| \) given \( w \). If \( n \) is large this test is actually carried out by a normal approximation: if \( c \) is chosen with equal probability from among the elements of \( C \) then \( \bar{x}-\bar{y} \) is approximately \( \mathcal{N}(0, \frac{4}{N} \sigma_w^2) \).

If \( c \) has been selected by BCD\((p)\) then (conditioned on the observed value of \( w \)) \( \bar{x}-\bar{y} \) has mean 0 and variance approximately

\[
\frac{4}{N} \sigma_w^2 \int_{-\pi}^{\pi} f(\omega) L_z(\omega) d\omega \quad (6.2)
\]

where \( z_k = w_k / (\sqrt{N} \sigma_w) \), as in equations 5.10 and 5.11. For \( p=2/3 \) the "permutation" standard deviation* of \( \bar{x}-\bar{y} \) therefore cannot exceed \( \sqrt{1 \frac{1}{2}} = 1.055 \) times its value under Complete Randomization, so we will not make a seriously anticonservative error using the latter value. On the other hand, the integral in 6.2 can be arbitrarily small, as commented

---

*The standard deviation rather than the variance is the crucial factor for the significance test.
before so the Complete Randomization variance can conceivably be very
conservative from this point of view. In most cases we would expect the
\( w \) sequence to be "noisy", that is to have \( \mathbb{E}_w(\omega) \) spread rather evenly
over \([-\pi, \pi]\), in which case the two answers tend to coincide since
\[
\frac{1}{2\pi} \int_{-\pi}^{\pi} f(\omega) d\omega = 1.
\]
The situation is similar for the Permuted Block Designs. If \( N=2bI \)
then the variance of \( \bar{X} - \bar{Y} \) (conditioned on \( w \)) is
\[
\frac{4}{N} \sigma_w^2 (1 + \frac{1}{2b-1}(1 - \frac{1}{\tilde{w}_1^2} \sum_{i=1}^{I} \bar{w}_i^2))
\]
where \( \tilde{w}_1 \) is the average of the \( w_k \) within the \( i \)th block. Again the
factor multiplying \( \frac{4}{N} \sigma_w^2 \) can range from a high of \( 1 + \frac{1}{2b-1} (=1 \frac{1}{9} ) \) if \( b=5 \)
down to zero.

We have ignored the question of asymptotic normality in this dis-
cussion for which reasonable conditions on the \( w_k \) are necessary [7].

7. Proofs. Define \( \pi_n(j) = \text{Prob}(D_n=j) \) and \( E_k h(D_n) = E(h(D_n)|D_0=k) \),
the expected value of \( h(D_n) \) given that we start the Markov chain 3.1
with \( D_0=k \).

Lemma 0: \( \pi_n(0) = p\pi_{n-1}(1), \pi_n(1) = \pi_{n-1}(0) + p\pi_{n-1}(2), \) and
\( \pi_n(j) = q\pi_{n-1}(j-1) + p\pi_{n-1}(j+1) \) for \( j \geq 2, n \geq 2 \).

Lemma 1: For any function \( h \) and \( n \geq 2 \),
\[
E(h(D_n)) = p h(0) \pi_{n-1}(1) + h(1) \pi_{n-1}(0) + q \sum_{j=0}^{\infty} h(j+1) \pi_{n-1}(j) + p \sum_{j=2}^{\infty} h(j-1) \pi_{n-1}(j).
\]

Lemma 2: If \( h(j) \) is a non-decreasing function of \( j \) then \( E_k h(D_n) \geq
E_k h(D_n) \) for \( k \geq 0, n \geq 1 \).
Lemma 0 follows immediately from the definition 3.1 of the Markov chain. Lemma 1 is the expression \( E_n(h_n) = \sum_{j=0}^{\infty} \pi_{n}(j) h(j) \) with the \( \pi_{n}(j) \) expressed in terms of the \( \pi_{n-1}(j) \) via Lemma 1. Lemma 2 is obviously true for \( n=1 \), and assume by induction it is true for the case \( n-1 \). Then, making use of 3.1 again,

\[
E_{k+2} h(D_n) = pE_{k+1} h(D_{n-1}) + qE_{k+3} h(D_{n-1}) \geq E_{k+1} h(D_{n-1}) \quad (7.1)
\]

by the induction hypothesis, and likewise (if \( k > 0 \))

\[
E_{k} h(D_n) = pE_{k-1} h(D_{n-1}) + qE_{k+1} h(D_{n-1}) \leq E_{k+1} h(D_{n-1}) \quad , \quad (7.2)
\]

while for \( k=0 \)

\[
E_{0} h(D_n) = E_{1} h(D_{n-1}) \quad . \quad (7.3)
\]

Combining 7.1 with 7.2 (or 7.3) gives Lemma 2.

Proof of Theorem 1: \( E_{0} h(D_{n+2}) = pE_{0} h(D_{n}) + qE_{2} h(D_{n}) \geq E_{0} h(D_{n}) \) by Lemma 2.

Proof of Theorem 2: Let \( E_{n} = E_{0} + \sum_{n=0}^{\infty} (j+1) \pi_{n-1}(j) + \sum_{j=0}^{\infty} (j+1) \pi_{n-1}(j) \)

by Lemma 1, which simplifies to

\[
E_{n} - E_{n-1} = 1 - 2p + 2p \pi_{n-1}(0) = \left[ \pi_{n-1}(0) - \frac{r-1}{2r} \right] \frac{2r}{r+1} . \quad (7.4)
\]

Writing \( E_{2n} = (E_{n} - E_{2n-2}) + (E_{2n-1} - E_{2n-2}) + \ldots + E_{1} \) and remembering that \( \pi_{j}(0)=0 \) for \( j \) odd gives

\[
E_{2n} = \frac{2r}{r+1} \sum_{j=0}^{n-1} \left[ \pi_{2j}(0) - \frac{r-1}{r} \right] . \quad (7.5)
\]
But \( \lim_{\frac{r}{2r-1}} E_{2j} = \sum_{j=1}^{\infty} \frac{(r-1)(r+1)}{r} (2j)(1/r)^{2j} \) by 3.2 and 3.3, which is evaluated as \( 2r/(r^2-1) \) (the interchange of limit and expectation being easily justified from theorem 3.1 and standard theorems, see [6] page 183).

Passing to the limit in 7.5 gives the first half of theorem 2 upon subtraction of the term for \( j=0 \). The second half is proved in the same way starting from the function \( h(0)=h(1)=0, h(j)=j-1 \) \( j \geq 1 \).

Proof of Theorem 3: The theorem is trivially true for \( n=1 \), and assume as an induction hypothesis that it is true for the case \( n-1 \).

In the proof for case \( n \) we can assume that \( h(0)=h(1)=0 \). For example, if \( n \) is odd we can subtract \( h(1) \) from every value \( h(j) \) without affecting the validity of the result, and then we can assume \( h(0)=0 \) since \( \pi_n(0)=0 \), a similar trick working for \( n \) even. Under these conditions Lemma 1 becomes

\[
E_h(D_n) = qE_h'(D_{n-1}) + pE_h''(D_{n-1}),
\]

(7.6)

where \( h'(j) \) takes values \( 0, h(2), h(3), \ldots \) and \( h''(j) \) takes values \( 0, 0, 0, h(2), h(3), \ldots \) for \( j=0, 1, 2, \ldots \).

By the induction hypothesis \( E_h'(D_{n-1}) \) and \( E_h''(D_{n-1}) \) are non-increasing functions of \( r \). Also \( E_h'(D_{n-1}) \geq E_h''(D_{n-1}) \) since \( h'(j) \geq h''(j) \) for all \( j \). Since \( q \) is a decreasing function of \( r \) and \( p \) an increasing function of \( r \), the theorem now follows from 7.6.

Corollaries 1 and 2 are easily derived from Theorem 4 by calculating that

\[
\frac{d}{dp} \frac{1}{j}[ \sum_{i=0}^{[j-2/2]} B_j(1) + \delta_j B_j([(j/2)]) ] = -B_{2j-1}( [j/2] - 1 )
\]

(7.7)
for \( j \geq 2 \) odd or even. (For \( j=1 \) the derivative equals \(-\frac{1}{2}\).) Letting \( \Delta_j(p) \) represent the quantity being differentiated in 7.7, this says that \( d\Delta_j(p)/dp = d\Delta_{j+1}(p)/dp \) for \( j \) even and \( \geq 2 \). Since \( \Delta_j(1)=0 \) for all \( j \) by the definition of \( B_j(1) \), corollary 1 follows by integration of this equality from 1 to \( p \). It is easily shown from 7.7 that

\[
0 > -d\Delta_{j+2}(p)/dp > -d\Delta_j(p)/dp \quad \text{for} \quad \frac{1}{2} < p < 1 \quad \text{and} \quad j \geq 1, \quad \text{and}
\]
corollary 2 also follows by integration.

The second part of corollary 3 follows from corollary 1 by writing

\[
\sum_{k=1}^{\infty} (-1)^k \rho_k = -\frac{1}{2} [\rho_1 - (\rho_2 - \rho_1) + (\rho_3 - \rho_2) - \ldots] \quad \text{for} \quad \sum_{k=1}^{\infty} \rho_k = \frac{1}{2} (\phi_1 - (2(r-1))) \quad \text{and} \quad \rho_2 - \rho_1 = \phi_r/(2(r+1)) \quad \text{from theorem 4.}
\]

The first part of corollary 3 can be obtained by direct summation in theorem 4, but it is simpler to note that

\[
1 + 2 \sum_{k=1}^{\infty} \rho_k = \lim_{N \to \infty} \text{Var}(\sum_{k=1}^{N} \frac{T_k}{\sqrt{N}}) = 0, \quad \text{for} \quad \sum_{k=1}^{\infty} \rho_k = \frac{1}{2} \phi_r/(2(r+1)) \quad \text{from theorem 4.}
\]

The last equality following from

\[
\text{Var}\left(\sum_{k=1}^{N} \frac{T_k}{\sqrt{N}}\right) = \text{Var}\left(\frac{\sum_{k=0}^{N} Y_k}{\sqrt{N}}\right) \leq 4 \text{Var}(D_0)/N. \quad \text{for} \quad \sum_{k=1}^{\infty} \rho_k = \frac{1}{2} \phi_r/(2(r+1)) \quad \text{from theorem 4.}
\]

Finally, for the proof of theorem 4 itself we need two definitions,

\[
e_{kj} = E_k(T_j) = E(T_j | D_0 = k) \quad k \geq 0, j > 1
\]

and

\[
b_{kj} = \text{Prob.} \left\{ \sum_{i=1}^{k} Y_i > k \quad \text{for} \quad k = 0, 1, 2, \ldots, j-1 \right\} \quad k \geq 0, j \geq 1
\]
where the $Y_i$ are i.i.d. random variables taking values +1 and -1 with probabilities $q$ and $p$ respectively. That is $b_{kj}$ is the probability that a random walk with negative drift starts from the origin and stays above the level $-k$ at least until step $j$. (By definition $b_{0j} = 0$ for all $j$.)

**Lemma 4:** $e_{kj} = (q-p)b_{kj}$.

**Proof:** $e_{0j} = 0$, $e_{k,1} = q-p$, and $e_{kj} = q e_{k-1,j-1} + p e_{k-1,j-1}$ for $k > 0$, $j > 1$. An elementary calculation shows that $b_{kj}/(p-q)$ satisfies the same relationships.

**Lemma 5:** $b_{k,j} - b_{k,j+1} = \frac{k}{j} \left( j-k \right) \left( \frac{j-k}{2} \right) q^{(j-k)/2} q^{(j-k)/2}$ if $(j-k)/2$ is a non-negative integer, and equals zero otherwise.

**Proof:** By 7.12, $b_{kj} - b_{k,j+1} = \text{Prob.} \left\{ \sum_{i=1}^{j} Y_i > -k, \sum_{i=1}^{j} Y_i = -k \right\} = \text{Prob.} \left\{ \sum_{i=1}^{j} Y_i = -k \right\}$. Unless $(j-k)/2$ is a non-negative integer the first factor is zero, while if it is a non-negative integer it equals $\left( \frac{j-k}{2} \right) q^{(j-k)/2} q^{(j-k)/2}$. The second factor then equals $k/j$ by the ballot theorem (page 66 [4]).

**Lemma 6:** $\rho_j = (p-q) \sum_{k=1}^{\infty} \pi_k e_{kj} = \left( \frac{r-1}{r+1} \right)^2 \sum_{k=1}^{\infty} \pi_k b_{kj}$.

**Proof:** The second statement follows from the first by Lemma 4. We have $\rho_j = E(T_{j+1} | D_1 = k)$, (where we have taken advantage of the symmetry of the process about zero). But $E(T_{j+1} | D_1 = k) = E(T_1 | D_1 = k) E(T_{j+1} | D_1 = k) = e_{jk} E(T_1 | D_1 = k)$ by the Markov property of the chain $\tilde{D}_n = \tilde{D}_0 + \sum_{j=1}^{n} T_j$. Hence $\tilde{D}_0$ is given the stationary distribution $\tilde{\pi}_0 = \pi_0, \tilde{\pi}_k = \frac{1}{2} \pi_k \left\{ k \neq 0 \right\}$ which makes the entire chain stationary.

Finally, $E(T_1 | D_1 = k) = 0$ if $k=0$, and equals $p-q$ if $k > 0$. 

22
Theorem 4 follows by substituting the values of $b_{kj}$ implicitly given in Lemma 5 into Lemma 6.
REFERENCES


