Forcing a Sequential Experiment to be Balanced
Author(s): Bradley Efron
Source: *Biometrika*, Vol. 58, No. 3 (Dec., 1971), pp. 403-417
Published by: Oxford University Press on behalf of Biometrika Trust
Stable URL: http://www.jstor.org/stable/2334377
Accessed: 27-04-2018 17:16 UTC

Biometrika (1971), **58**, 3, *p.* 403
With 2 text-figures
Printed in Great Britain

Forcing a sequential experiment to be balanced

By BRADLEY EFRON
Stanford University

SUMMARY

Subjects arrive sequentially at an experimental site and must be assigned immediately to treatment or control groups. In order to avoid biasing the results of the experiment it is customary to make the assignments by independent flips of a fair coin, but in small-sized experiments this may result in a severe imbalance between the numbers of treatments and controls. This paper discusses a new method of assigning the subjects which tends to balance the experiment, but at the same time is not over vulnerable to various common forms of experimental bias.

1. The problem

The random assignment of subjects to treatment and control groups is a canon of good experimental design. In many situations, 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 data are given by courtesy of Dr Henry Kaplan, Department of Radiology, Stanford University. 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.

Table 1. *Distribution of treatments and controls in four age categories, Hodgkin's disease investigation*

	Age			
	10–19	20–29	30–39	40–49
Treatments	7	5	4	1
Controls	2	5	3	2

It is simple to devise methods of sequential experimentation which force approximately equal numbers of treatments and controls. For example, the completely nonrandom systematic design $TCTCTC\ldots$, 'T' for treatment, 'C' for control, followed separately in each category of subjects guarantees that the imbalance will never exceed one in any category. A variant is the Student sandwich plan $TCCTTCCTT\ldots$. However, these plans can easily bias the results of the experiment, as shown below, and are not serious competitors to complete randomization. A controversy between Student and Fisher arose on a closely related point in the 1930's; see Student (1937), Barbacki & Fisher (1936), Yates (1939) and Greenberg (1951).

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 is not a suitable experimental subject, in which category the subject belongs, etc. It is obvious that complete randomization eliminates selection bias, and that the systematic designs maximize it. Blackwell & Hodges (1957) 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. Selection bias is not a factor in blind experiments where admission to the study and related decisions are made by someone ignorant of the past assignment of treatments and controls. However, even in this situation complete randomization enjoys the two advantages which follow.

(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 § 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 permuted 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 to control, all $\binom{2b}{b}$ combinations being equally likely. Permuted 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 $\frac{1}{6}$ that the experimenter will know for certain the assignment of units 8, 9 and 10, and $\frac{4}{9}$ that he will know for certain the assignment of units 9 and 10.

The biased coin designs introduced in § 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 permuted 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:

$$\begin{array}{ll} \text{If } \tilde{D} > 0, & \text{assign treatment with probability } q \text{ and control with probability } p. \\ \text{If } \tilde{D} = 0, & \text{assign treatment with probability } \tfrac{1}{2} \text{ and control with probability } \tfrac{1}{2}. \\ \text{If } \tilde{D} < 0, & \text{assign treatment with probability } p \text{ and control with probability } q. \end{array} \tag{2·1}$$

Here $p \geqslant q, p+q = 1$, 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 $p = \frac{2}{3}$, which is the author's personal favourite, 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 §§4, 5 and 6. All proofs are deferred until §7.

A note on methodology. Complete randomization is usually implemented by means of a deck of envelopes which the statistician gives the experimenter. After each patient is admitted to the experiment, an envelope is opened which directs him into either the treatment or control group, these choices having been made by the statistician using a random device or random number table. To implement a permuted block design the deck must be ordered in blocks of the proper size, and a separate deck provided for each category of patient. Usually the statistician makes the assignments by means of a table of random permutations.

A biased coin design for a multicategory experiment can be implemented using a single unordered deck of envelopes, each of which contains instructions covering the three cases $\tilde{D} > 0, \tilde{D} = 0$ and $\tilde{D} < 0$ described in (2·1). It is probably best to have the three instructions in separate envelopes within the envelope so that the experimenter does not see the ones he does not use. The assignments can be made with a simple random device, such as a die if $p = \frac{2}{3}$, or a random number table.

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 $0, 1, 2, \ldots$ and transition probabilities

$$\begin{aligned} \mathrm{pr}\,(D_{n+1} = j+1 \mid D_n = j) &= q \quad (j \geqslant 1), \\ \mathrm{pr}\,(D_{n+1} = j-1 \mid D_n = j) &= p \quad (j \geqslant 1), \end{aligned} \tag{3·1}$$

$$\mathrm{pr}(D_{n+1} = 1 \mid D_n = 0) = 1.$$

This is a random walk with a reflecting barrier at the origin (Cox & Miller, 1965, p. 41) and has stationary probabilities π_j given by

$$\pi_0 = \frac{r-1}{2r}, \quad \pi_j = \frac{r-1}{2r}\frac{r+1}{r^j} \quad (j \geqslant 1), \tag{3·2}$$

where $r = p/q$. The first few values of the π_j are given for $r = 2$, 3 and 4 in Table 2.

Table 2. *Values of the first few stationary probabilities*

	π_0	π_1	π_2	π_3	π_4	π_5
$r = 2\ (p = \frac{2}{3})$	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{3}{16}$	$\frac{3}{32}$	$\frac{3}{64}$	$\frac{3}{128}$
$r = 3\ (p = \frac{3}{4})$	$\frac{1}{3}$	$\frac{4}{9}$	$\frac{4}{27}$	$\frac{4}{81}$	$\frac{4}{243}$	$\frac{4}{729}$
$r = 4\ (p = \frac{4}{5})$	$\frac{3}{8}$	$\frac{15}{32}$	$\frac{15}{124}$	$\frac{15}{512}$	$\frac{15}{2048}$	$\frac{15}{8192}$

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,

$$\lim_{m\to\infty} \operatorname{pr}(D_{2m} = 0) = 2\pi_0 = (r-1)/r, \quad \lim_{m\to\infty} \operatorname{pr}(D_{2m+1} = 1) = 2\pi_1 = (r^2-1)/r^2, \tag{3·3}$$

etc. Thus, with $p = \frac{2}{3}$, the experiment has asymptotic probability $\frac{1}{2}$ of being exactly balanced for n even, and asymptotic probability $\frac{3}{4}$ 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 nondecreasing function of j $(j = 0, 1, \ldots)$ and D_n is the absolute difference between the numbers of treatments and controls after n assignments have been made, then $E\{h(D_{n+2})\} \geqslant E\{h(D_n)\}$ for every value of n.*

Taking $h(0) = 0$, $h(j) = 1$ for $j > 0$, in the theorem shows that $\operatorname{pr}(D_{2m} = 0)$ is a decreasing function of m, and taking $h(0) = 0$, $h(1) = 0$ and $h(j) = 1$ $(j > 1)$, shows that $\operatorname{pr}(D_{2m+1} = 1)$ 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 as

THEOREM 2. *The following relations hold:*

$$\sum_{m=1}^{\infty} \{\operatorname{pr}(D_{2m} = 0) - 2\pi(0)\} = \frac{1}{r(r-1)},$$

$$\sum_{m=1}^{\infty} \{\operatorname{pr}(D_{2m+1} = 1) - 2\pi(1)\} = \frac{2}{r^2(r-1)}.$$

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 $(n = 2, \ldots, 10)$ for the case $r = 2$. For $r = 2$ both sums in Theorem 2 equal 0·5, and it can be seen from the table that 0·391 of the even-n bonus and 0·336 of the odd-n bonus have occurred by $n = 10$.

A comparison of $BCD(p)$ with $p = \frac{2}{3}$ 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 imbalances, $D_4 = 4$, $D_5 = 3$, etc., reinforce this impression, though the different natures of the two rules make such comparisons difficult.

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.

Table 3. *Percentage probabilities* 100 $(D_n = j)$ *for* $r = 2$

$j \backslash n$	2	3	4	5	6	7	8	9	10
0	66·7	—	59·3	—	56·0	—	54·1	—	53·0
1	—	88·9	—	84·0	—	81·2	—	79·5	—
2	33·3	—	37·0	—	37·9	—	38·0	—	38·0
3	—	11·1	—	14·8	—	16·5	—	17·3	—
4	—	—	3·7	—	5·8	—	7·0	—	7·7
5	—	—	—	1·2	—	2·2	—	2·9	—
6	—	—	—	—	0·4	—	0·8	—	1·2
7	—	—	—	—	—	0·1	—	0·3	—
8	—	—	—	—	—	—	0·0	—	0·1
9	—	—	—	—	—	—	—	0·0	—
10	—	—	—	—	—	—	—	—	0·0

Table 4. *Percentage probabilities that experiment is exactly balanced at stage n, within* 1 *if n is odd*

	n								
	2	3	4	5	6	7	8	9	10
$BCD(\frac{2}{3})$	66·7	88·9	59·3	84·0	56·0	81·2	54·1	79·5	53·0
Permuted blocks $b = 5$	55·6	83·3	47·6	79·3	47·6	83·3	55·5	100·0	100·0

Theorem 3. *If $h(j)$ is a nondecreasing function of j $(j = 0, 1, \ldots)$, then $E\{h(D_n)\}$ is a non-increasing function of r for $r \geqslant 1$ 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 $\frac{1}{2}n$ correct guesses in n trials. It is intuitively clear, and proved by Blackwell & Hodges (1957), 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 $2^{2b}\Big/\binom{2b}{b} - 1$ correct guesses per block of length $2b$ against the corresponding permuted block design. We can think of this as an excess selection bias of

$$2^{2b}\Big/\binom{2b}{b} - (b+1) \qquad (4\cdot1)$$

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

$$\tfrac{1}{2}\,\mathrm{pr}\,(D_{n-1} = 0) + p\,\mathrm{pr}\,(D_{n-1} > 0), \qquad (4\cdot2)$$

which asymptotically approaches

$$\tfrac{1}{2}\pi_0 + p(1-\pi_0) = \tfrac{1}{2} + \frac{r-1}{4r}. \tag{4·3}$$

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

$$\frac{r-1}{2r} b. \tag{4·4}$$

Figure 1 compares (4·1) with (4·4) as a function of b. The case $r = 2$ and $p = \frac{2}{3}$ 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 correct guesses per trial are asymptotically $(r-1)/(4r) = \frac{1}{8}$.

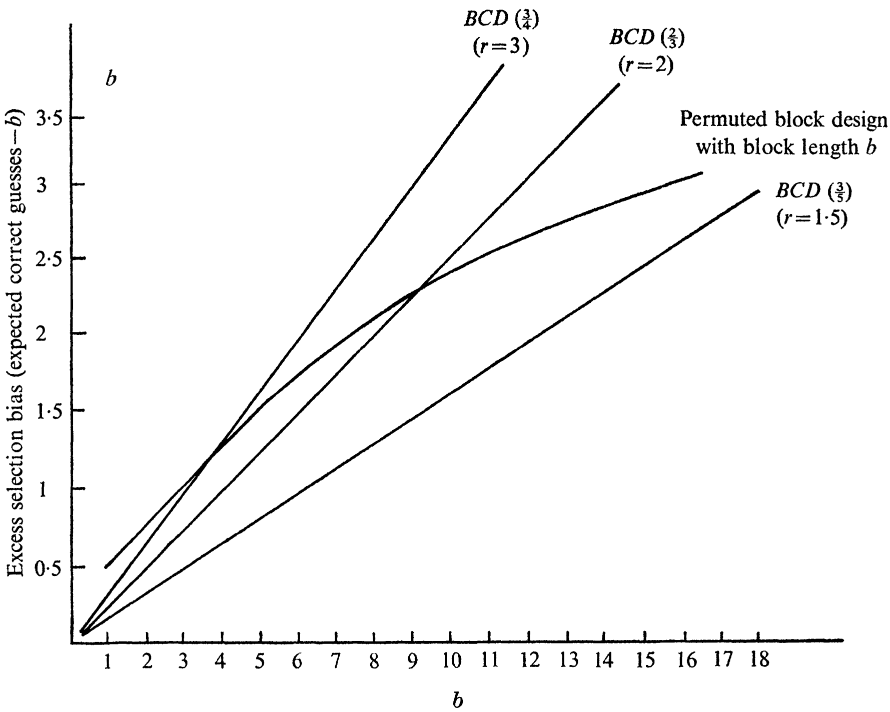

Fig. 1. Excess selection bias in $2b$ trials. In a long sequence of assignments $BCD(\frac{2}{3})$ would have approximately the same selection bias as the permuted block design with $2b = 18$.

5. ACCIDENTAL BIAS

Let $T_k = +1$ or -1 as the kth experimental unit is assigned to the treatment or control group, and let $\mathbf{T} = (T_1, \ldots, T_N)$ be the vector of assignments after some fixed number N of trials. All of the random designs we have mentioned have $E(\mathbf{T}) = \mathbf{0}$. In this section we assess the vulnerability of a design to accidental bias by the magnitude of the largest eigenvalue of the covariance matrix $\boldsymbol{\Sigma}_{\mathbf{T}}$ of $\mathbf{T}$. This choice is motivated in the following way.

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

$$y_k = \mu + \alpha t_k + \beta z_k + \epsilon_k \quad (k = 1, \ldots, N), \tag{5·1}$$

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, for example, age of subject on unit k, and the ϵ_k are independent $\mathcal{N}(0, \sigma^2)$ random variables. In vector notation, $\mathbf{y} = \mu\mathbf{e} + \alpha\mathbf{t} + \beta\mathbf{z} + \boldsymbol{\epsilon}$, where $\mathbf{e}' = (1, \ldots, 1)$, $\mathbf{t}' = (t_1, \ldots, t_N)$, etc. There is no loss of generality in this model in assuming that

$$\mathbf{z}'\mathbf{e} = 0, \quad \|\mathbf{z}\|^2 = 1. \tag{5·2}$$

The nuisance factor $\mathbf{z}$ can harmfully affect our inferences about α in two ways: if we test $\alpha = 0$ using Student's t in the usual way, without allowing for the covariate $\mathbf{z}$, there will be a spurious noncentrality parameter of magnitude $(\beta^2/\sigma^2)(\mathbf{z}'\mathbf{t})^2$ in the numerator of Student's statistic when the null hypothesis is true. If the null hypothesis is false, and if we do allow for the factor $\mathbf{z}$ in our analysis, then the noncentrality parameter for testing $\alpha = 0$ is

$$\frac{N\alpha^2}{\sigma^2}\left[1 - \frac{1}{N}\left\{\left(\frac{\mathbf{e}'}{\sqrt{N}}\mathbf{t}\right)^2 + (\mathbf{z}'\mathbf{t})^2\right\}\right] \tag{5·3}$$

and we see that the loss of noncentrality due to the nuisance factor is $(\alpha^2/\sigma^2)(\mathbf{z}'\mathbf{t})^2$.

In both cases it is ideal to have $\mathbf{t}$ orthogonal to $\mathbf{z}$, and we are penalized proportionately to $(\mathbf{z}'\mathbf{t})^2$ as we depart from this ideal. It is also good to have $\mathbf{t}$ orthogonal to $\mathbf{e}$, i.e. to balance the experiment, which is what we are trying to do in this paper.

Now if $\mathbf{t}$ is the realization of the random vector $\mathbf{T}$, with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{T}}$, then for a fixed vector $\mathbf{z}$

$$E\{(\mathbf{z}'\mathbf{T})^2\} = \mathbf{z}'\boldsymbol{\Sigma}_{\mathbf{T}}\mathbf{z}. \tag{5·4}$$

The least favourable vector $\mathbf{z}$ is that vector satisfying (5·2) which maximizes (5·4), and we see that

$$E\{(\mathbf{z}'\mathbf{T})^2\} \leqslant \text{largest eigenvalue of } \boldsymbol{\Sigma}_{\mathbf{T}}, \tag{5·5}$$

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

For example, under complete randomization $\boldsymbol{\Sigma}_{\mathbf{T}} = \mathbf{I}$ and $E\{(\mathbf{z}'\mathbf{T})^2\} = 1$ for every $\mathbf{z}$. Notice that 1 is the smallest possible value for the maximum eigenvalue of $\boldsymbol{\Sigma}_{\mathbf{T}}$ since $E(T_k) = 0$ implies $\operatorname{var}(T_k) = 1$ and hence $\operatorname{tr}(\boldsymbol{\Sigma}_{\mathbf{T}}) = n =$ sum of eigenvalues of $\boldsymbol{\Sigma}_{\mathbf{T}}$. For the permuted block design with block length $2b$, we have $\boldsymbol{\Sigma}_{\mathbf{T}} = \{1 + 1/(2b-1)\}\mathbf{I} - 1/(2b-1)\,\mathbf{e}\mathbf{e}'$ for N satisfying $2 \leqslant N \leqslant 2b$. This matrix has largest eigenvalue $1 + 1/(2b-1)$, attained for any vector $\mathbf{z}$ orthogonal to $\mathbf{e}$. 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 $1\frac{1}{9}$.

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 $T_1, \ldots, T_N$ as if it were derived from a stationary process. Specifically, the results obtained below pertain to the sequence $T_{h+1}, \ldots, T_{h+N}$ 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

$$\rho_k \equiv \lim_{h\to\infty} E(T_h T_{h+k}) \tag{5·6}$$

has a closed form expression. Define $B_k(i)$ as the cumulative distribution function of a binomial random variable with parameters k and p,

$$B_k(i) = \sum_{i'=0}^{i} \binom{k}{i'} p^{i'} q^{k-i'}. \tag{5·7}$$

THEOREM 4. *The limiting autocovariances ρ_k are given by*

$$\rho_1 = -\phi(r)\frac{1}{2(r-1)}, \quad \rho_{k+1} - \rho_k = \phi(r)\frac{1}{k}\left\{\sum_{i=0}^{[\frac{1}{2}(k-2)]} B_k(i) + \delta_k B_k([\tfrac{1}{2}k])\right\},$$

where

$$\phi(r) = \frac{(r-1)^3}{r(r+1)},$$

$\delta_k = 0$ *if k is even and* $\delta_k = 1$ *if k is odd. The square brackets indicate the greatest integer function.*

COROLLARY 1. *For k even and positive* $\rho_{k+1} - \rho_k = \rho_{k+2} - \rho_{k+1}$.

COROLLARY 2. *For $k \geqslant 1$, ρ_k is negative and $\rho_{k+1} - \rho_k$ is positive and decreasing.*

COROLLARY 3. *Also*

$$\sum_{k=1}^{\infty} \rho_k = -0·5, \quad \sum_{k=1}^{\infty} (-1)^k \rho_k = \frac{1}{2}\left(\frac{r-1}{r+1}\right)^2.$$

Table 5 gives the first 11 values of ρ_k for $r = 2$, 3 and 4.

Table 5. *Values of the autocorrelation function ρ_k*

$k \backslash r$	2	3	4
0	1·000	1·000	1·000
1	−0·0833	−0·1667	−0·2250
2	−0·0556	−0·0833	−0·0900
3	−0·0463	−0·0625	−0·0630
4	−0·0370	−0·0417	−0·0360
5	−0·0319	−0·0325	−0·0263
6	−0·0267	−0·0234	−0·0166
7	−0·0234	−0·0187	−0·0124
8	−0·0201	−0·0140	−0·0082
9	−0·0178	−0·0113	−0·0062
10	−0·0155	−0·0087	−0·0042

Since $\Sigma |\rho_k| < \infty$, the process has an absolutely continuous spectrum with spectral density

$$f(\omega) \equiv \sum_{k=-\infty}^{\infty} \rho_k e^{-i\omega k} = 1 + 2\sum_{k=1}^{\infty} \rho_k \cos(\omega k); \tag{5·8}$$

this is a factor of 2π times the usual definition (Cox & Miller 1965, p. 315). Corollary 3 can be restated in terms of $f(\omega)$ as

$$f(0) = 0, \quad f(\pi) = 1 + \left(\frac{r-1}{r+1}\right)^2 = 1 + (p-q)^2. \tag{5·9}$$

Since $f(-\omega) = f(\omega)$ the abscissa is plotted only from $\omega = 0$ to $\omega = \pi$ in Fig. 2. We see that the spectral density increases monotonically as ω goes from 0 to π and so obtains its maximum value at π. By (5·9) this maximum is equal to $1 + (p-q)^2$. Therefore $BCD(\frac{2}{3})$ has maximum vulnerability to accidental bias of amount $1+(\frac{2}{3}-\frac{1}{3})^2 = 1\frac{1}{9}$, 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 $BCD(p)$.

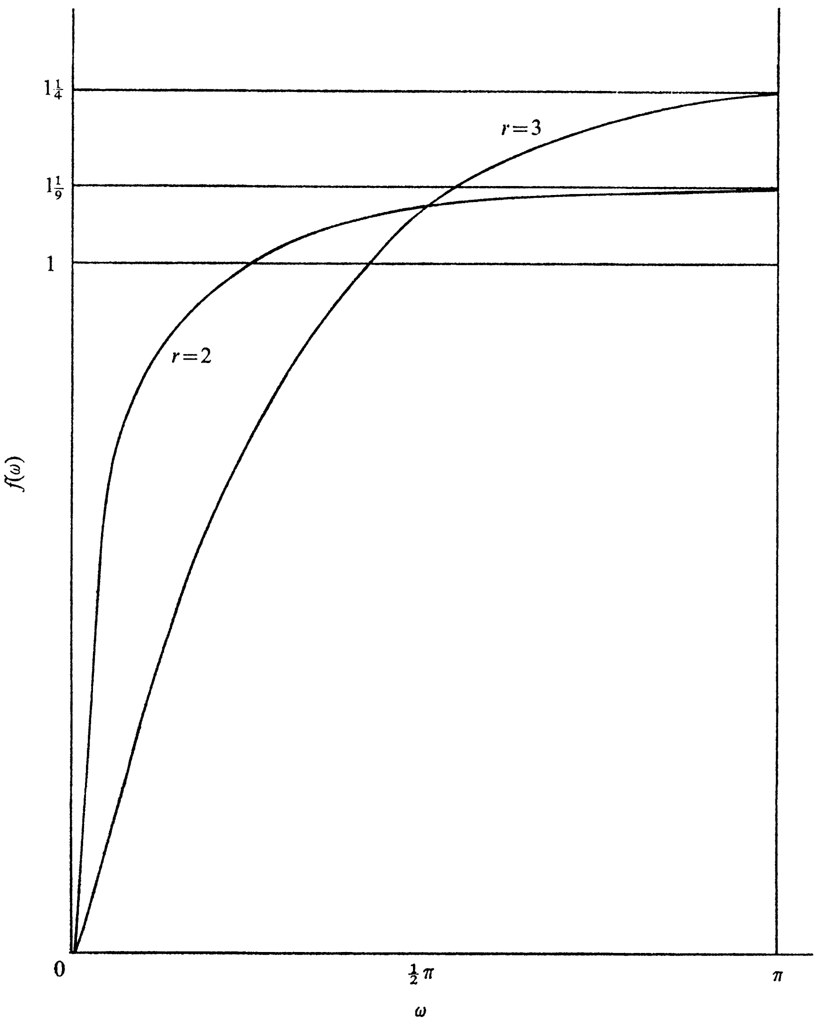

Fig. 2. Spectral density of biased coin assignments, showing maximum value at $\omega = \pi$. The spectral density for complete randomization is the line $f(\omega) = 1$.

It is easy to see that the maximum of the spectral density corresponds to the maximum eigenvalue definition given before. Let $z_1, \ldots, z_N$ be any sequence satisfying (5·2) and define

$$l_{\mathbf{z}}(\omega) = \sum_{k=1}^{N} z_k e^{-i\omega k}, \quad L_{\mathbf{z}}(\omega) = \frac{1}{2\pi} |l_{\mathbf{z}}(\omega)|^2. \tag{5·10}$$

Then by Parseval's lemma $L_{\mathbf{z}}(\omega)$ is a probability density on $[-\pi, \pi]$,

$$\int_{-\pi}^{\pi} L_{\mathbf{z}}(\omega)\, d\omega = 1,$$

and by the standard theory of linear transforms (Cox & Miller, 1965, p. 321)

$$E\{(\mathbf{z}'\mathbf{T})^2\} = \int_{-\pi}^{\pi} f(\omega)\, L_{\mathbf{z}}(\omega)\, d\omega. \tag{5·11}$$

We see that the integral is bounded by the maximum value of $f(\omega)$, a bad $\mathbf{z}$ being one with $L_{\mathbf{z}}(\omega)$ concentrated near the maximizing value of ω. In the cases above the least favourable $\mathbf{z}$ is asymptotically proportional to the sequence $+1, -1, +1, -1, \ldots$.

The definition of vulnerability to accidental bias we have given is of a minimax nature and so favours complete randomization. Figure 2 and (5·11) show that for some values of $\mathbf{z}$, 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 $x_1, \ldots, x_m$ with n control measurements $y_1, \ldots, y_n$ $(m+n=N)$. Let $U_1, \ldots, U_N$ be the experimental units and let c represent the collection of m indices corresponding to treatments, say $c \in \mathscr{C}$, where $\mathscr{C}$ is the set of all choices of m integers from $1, \ldots, N$.

Whatever test we are using for the hypothesis of no difference between treatment and control responses will have a significance level α_c depending on the choice of $c \in \mathscr{C}$. The value of α_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. If we use a level α permutation test, such as the conditional t test, Wilcoxon's test, etc., then

$$\frac{1}{\binom{N}{m}} \sum_{c \in C} \alpha_c = \alpha, \tag{6·1}$$

which is just another way of saying that for each set of N experimental responses the test rejects for $\alpha\binom{N}{m}$ 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 $\mathscr{C}$ with equal probability for all $\binom{N}{m}$ members, and (6·1) allows one to claim an exact significance level of α for the test without any probability model on the responses of the experimental units. This argument is distinct from that of § 5, which said that under a reasonable probability model complete randomization, and also $BCD(p)$, were likely to yield a c with α_c nearly equal to the nominal α 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.

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 α 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 $m = n = \frac{1}{2}N$ without affecting the rough calculation which follows. Let $w_1, \ldots, w_N$ be the observed responses on units $1, \ldots, N$, with the common mean subtracted off so that $\bar{w} = 0$, and let

$$\sigma_w^2 = \sum_{k=1}^{N} w_k^2/N.$$

Suppose our test rejects if the observed value of $|\bar{x} - \bar{y}|$ is among the $\alpha\binom{N}{\frac{1}{2}N}$ largest of the $\binom{N}{\frac{1}{2}N}$ possible values of $|\bar{x} - \bar{y}|$ given w. This is the conditional t test for the treatment and control measurements having identical distributions. 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 $\mathscr{C}$ then $\bar{x} - \bar{y}$ is approximately $\mathscr{N}\left(0, \frac{4}{N}\sigma_w^2\right)$.

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

$$\frac{4}{N}\sigma_w^2 \int_{-\pi}^{\pi} f(\omega)\, L_{\mathbf{z}}(\omega)\, d\omega, \tag{6·2}$$

where $z_k = w_k/(\sigma_w \sqrt{N})$, as in (5·10) and (5·11). For $p = \frac{2}{3}$ the permutation standard deviation of $\bar{x} - \bar{y}$ therefore cannot exceed $\sqrt{(1\frac{1}{9})} = 1{\cdot}055$ times its value under complete randomization, so we will not make a seriously anticonservative error using the latter value. The standard deviation rather than the variance is the crucial factor for the significance test. On the other hand, the integral in (6·2) can be arbitrarily small, as commented 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 $L_{\mathbf{z}}(\omega)$ 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

$$\frac{4}{N}\sigma_w^2\left(1 + \frac{1}{2b-1}\right)\left(1 - \sum_{i=1}^{I} \bar{w}_i^2/I\sigma_w^2\right), \tag{6·3}$$

where $\bar{w}_i$ is the average of the w_k within the ith block. Again the factor multiplying $(4/N)\,\sigma_w^2$ can range from a high of

$$1 + \frac{1}{2b-1},$$

which equals $1\frac{1}{9}$ if $b = 5$, down to zero.

We have ignored the question of asymptotic normality in this discussion for which reasonable conditions on the w_k are necessary (Serfling, 1968).

7. PROOFS

Define $\pi_n(j) = \mathrm{pr}\,(D_n = j)$ 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. *For $j \geqslant 2$ and $n \geqslant 2$,*

$$\pi_n(0) = p\pi_{n-1}(1), \quad \pi_n(1) = \pi_{n-1}(0) + p\pi_{n-1}(2),$$
$$\pi_n(j) = q\pi_{n-1}(j-1) + p\pi_{n-1}(j+1).$$

LEMMA 1. *For any function h and $n \geqslant 2$,*

$$E\{h(D_n)\} = p\{h(0)\,\pi_{n-1}(1) + h(1)\,\pi_{n-1}(0)\} + q\sum_{j=0}^{\infty} h(j+1)\,\pi_{n-1}(j) + p\sum_{j=2}^{\infty} h(j-1)\,\pi_{n-1}(j).$$

LEMMA 2. *If $h(j)$ is a nondecreasing function of j then*

$$E_{k+2}\{h(D_n)\} \geqslant E_k\{h(D_n)\} \quad (k \geqslant 0,\ n \geqslant 1).$$

Lemma 0 follows immediately from the definition (3·1) of the Markov chain. Lemma 1 is the expression

$$E\{h(D_n)\} = \sum_{j=0}^{\infty} \pi_n(j)\,h(j)$$

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,

$$E_{k+2}\{h(D_n)\} = pE_{k+1}\{h(D_{n-1})\} + qE_{k+3}\{h(D_{n-1})\} \geqslant E_{k+1}\{h(D_{n-1})\} \tag{7·1}$$

by the induction hypothesis, and likewise, if $k > 0$,

$$E_k\{h(D_n)\} = pE_{k-1}\{h(D_{n-1})\} + qE_{k+1}\{h(D_{n-1})\} \leqslant E_{k+1}\{h(D_{n-1})\}, \tag{7·2}$$

while for $k = 0$

$$E_k\{h(D_n)\} = E_{k+1}\{h(D_{n-1})\}. \tag{7·3}$$

Combining (7·1) with (7·2), or (7·3), gives Lemma 2.

Proof of Theorem 1. By Lemma 2,

$$E_0\{h(D_{n+2})\} = pE_0\{h(D_n)\} + qE_2\{h(D_n)\} \geqslant E_0\{h(D_n)\}.$$

Proof of Theorem 2. Let

$$E_n = E(D_n) = p\pi_{n-1}(0) + q\sum_{j=0}^{\infty}(j+1)\,\pi_{n-1}(j) + p\sum_{j=2}^{\infty}(j-1)\,\pi_{n-1}(j)$$

by Lemma 1, which simplifies to

$$E_n - E_{n-1} = 1 - 2p + 2p\pi_{n-1}(0) = \left\{\pi_{n-1}(0) - \frac{r-1}{2r}\right\}\frac{2r}{r+1}. \tag{7·4}$$

Writing $E_{2n} = (E_{2n} - E_{2n-1}) + (E_{2n-1} - E_{2n-2}) + \ldots + E_1$ and remembering that $\pi_j(0) = 0$ for j odd, we have

$$E_{2n} = \frac{2r}{r+1}\sum_{j=0}^{n-1}\left\{\pi_{2j}(0) - \frac{r-1}{r}\right\}. \tag{7·5}$$

But

$$\lim_{l\to\infty} E_{2l} = \sum_{j=1}^{\infty}\frac{(r-1)\,(r+1)}{r}\,(2j)\,(1/r)^{2j}$$

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 Loève (1963, p. 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

$$h(0)=h(1)=0, \quad h(j)=j-1 \ (j \geqslant 1).$$

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

$$E\{h(D_n)\}=qE\{h'(D_{n-1})\}+pE\{h''(D_{n-1})\}, \tag{7·6}$$

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

By the induction hypothesis $E\{h'(D_{n-1})\}$ and $E\{h''(D_{n-1})\}$ are nonincreasing functions of r. Also $E\{h'(D_{n-1})\} \geqslant E\{h''(D_{n-1})\}$ since $h'(j) \geqslant h''(j)$ 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

$$\frac{d}{dp}\frac{1}{j}\left\{\sum_{i=0}^{[\frac{1}{2}(j-2)]} B_j(i)+\delta_j B_j([\tfrac{1}{2}j])\right\}=-B_{2[\frac{1}{2}j]-1}([\tfrac{1}{2}j]-1) \tag{7·7}$$

for $j \geqslant 2$ 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 greater than or equal to 2. Since $\Delta_j(1)=0$ for all j by the definition of $B_j(i)$, Corollary 1 follows by integration of this equality from 1 to p. It is easily shown from (7·7) that $0>-d\Delta_{j+2}(p)/dp>-d\Delta_j(p)/dp$ for $\frac{1}{2}<p<1$ and $j \geqslant 1$, and Corollary 2 also follows by integration.

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

$$\begin{aligned}\sum_{k=1}^{\infty}(-1)^k\rho_k &= -\tfrac{1}{2}\{\rho_1-(\rho_2-\rho_1)+(\rho_3-\rho_2)-(\rho_4-\rho_3)+\ldots\}\\ &= -\tfrac{1}{2}\{\rho_1-(\rho_2-\rho_1)\},\end{aligned} \tag{7·8}$$

and evaluating $\rho_1=-\phi_r/\{2(r+1)\}$ and $\rho_2-\rho_1=\phi_r/\{2(r+1)\}$ 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

$$1+2\sum_{k=1}^{\infty}\rho_k=\lim_{N\to\infty}\operatorname{var}\left(\sum_{k=1}^{N}T_k/\sqrt{N}\right)=0, \tag{7·9}$$

the last equality following from

$$\operatorname{var}\left(\sum_{k=1}^{N}T_k/\sqrt{N}\right)=\operatorname{var}\{(D_N-D_0)/\sqrt{N}\} \leqslant 4\operatorname{var}(D_0)/N. \tag{7·10}$$

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

$$e_{kj} \equiv E_k(T_j) \equiv E(T_j|D_0=k, \text{ control in excess}) \quad (k \geqslant 0, j>1), \tag{7·11}$$

$$b_{kj}=\operatorname{pr}\left\{\sum_{i=1}^{l}Y_i>-k \ (l=0,1,\ldots,j-1)\right\} \quad (k \geqslant 0, j \geqslant 1), \tag{7·12}$$

where the Y_i are independently and identically distributed 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. For $k > 0, j > 1, e_{0,j} = 0, e_{k,1} = q-p$, and $e_{kj} = qe_{k+1,j-1} + pe_{k-1,j-1}$. An elementary calculation shows that the right hand side satisfies the same relationships.

LEMMA 5. *If* $\frac{1}{2}(j-k)$ *is a nonnegative integer,*

$$b_{k,j} - b_{k,j+1} = (k/j)\left\{\binom{j}{\frac{1}{2}(j-k)} p^{\frac{1}{2}(j+k)} q^{\frac{1}{2}(j-k)}\right\},$$

and equals zero otherwise.

Proof. By (7·12),

$$\begin{aligned} b_{kj} - b_{k,j+1} &= \mathrm{pr}\left\{\sum_{i=1}^{l} Y_i > -k \quad (l = 0, 1, \ldots, j-1), \quad \sum_{i=1}^{j} Y_i = -k\right\} \\ &= \mathrm{pr}\left(\sum_{i=1}^{j} Y_i = -k\right) \mathrm{pr}\left\{\sum_{i=1}^{l} Y_i > -k \quad (l = 0, 1, \ldots, j-1) \,\middle|\, \sum_{i=1}^{j} Y_i = -k\right\}. \end{aligned}$$

Unless $\frac{1}{2}(j-k)$ is a nonnegative integer the first factor is zero, while if it is a nonnegative integer it equals $\binom{j}{\frac{1}{2}(j-k)} p^{\frac{1}{2}(j+k)} q^{\frac{1}{2}(j-k)}$. The second factor then equals k/j by the ballot theorem (Feller, 1968, p. 66).

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

$$\rho_j = E(T_1 T_{j+1}) = \sum_{k=0}^{\infty} \pi_k E(T_1 T_{j+1} | D_1 = k),$$

where we have taken advantage of the symmetry of the process about zero. But

$$E(T_1 T_{j+1} | D_1 = k) = E(T_1 | D_1 = k)\, E(T_{j+1} | D_1 = k) = e_{jk} E(T_1 | D_1 = k)$$

by the Markov property of the chain

$$\tilde{D}_n = \tilde{D}_0 + \sum_{j=1}^{n} T_j.$$

Here $\tilde{D}_0$ is given the stationary distribution $\tilde{\pi}_0 = \pi_0$, $\tilde{\pi}_k = \frac{1}{2}\pi_{|k|}$ $(k \neq 0)$ which makes the entire chain stationary. Finally, $E(T_1 | D_1 = k) = 0$ if $k = 0$, and equals $p-q$ if $k > 0$.

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

This work was done under the auspices of a National Science Foundation Grant.

REFERENCES

BARBACKI, S. & FISHER, R. A. (1936). A test of the supposed precision of systematic arrangements. *Ann. Eugen.* **7**, 189.

BLACKWELL, D. & HODGES, J. L. (1957). Design for the control of selection bias. *Ann. Math. Statist.* 28, 449–60.

COX, D. R. & MILLER, H. D. (1965). *The Theory of Stochastic Processes.* New York: Wiley.

FELLER, W. (1968). *An Introduction to Probability Theory and Its Applications*, **1**, 3rd edition. New York: Wiley.

GREENBERG, B. G. (1951). Why randomize? *Biometrics* **7**, 309–22.

LOÈVE, M. (1963). *Probability Theory*, 3rd edition. Princeton: Van Nostrand.

SERFLING, R. J. (1968). Contributions to central limit theory for dependent variables. *Ann. Math. Statist.* **39**, 1158–75.

SUDENT. (1937). Comparison between balanced and random arrangements of field plots. *Biometrika* **29**, 363–79.

YATES, F. (1939). The comparative advantages of systematic and randomized arrangements in the design of agricultural and biological experiments. *Biometrika* **30**, 441–64.

[*Received November* 1970. *Revised April* 1971]

Some key words: Randomization in experimental design; Balance; Two treatment sequential experiments; Permutation tests.