The Annals of Statistics
1978, Vol. 6, No. 1, 92-100

THE ADAPTIVE BIASED COIN DESIGN FOR SEQUENTIAL EXPERIMENTS

BY L. J. WEI

University of South Carolina—Columbia

In comparing two treatments, eligible subjects come to the experiment sequentially and must be treated at once. To reduce experimental bias and to increase the precision of inference about treatment effects, the adaptive biased coin design, which offers a compromise between perfect balance and complete randomization, is proposed and analyzed. This new design has the property that it forces a small-sized experiment to be balanced, but tends toward the complete randomization scheme as the size of the experiment increases.

1. Introduction. In comparing two treatments A and B, eligible subjects come to the experiment sequentially and must be treated at once. A statistical design problem is how to assign subjects to different treatment groups. One usually considers assignment rules which compromise between complete randomization and perfect balance to reduce experimental bias and to increase the precision of inference about treatment difference. Efron (1971) introduced the biased coin design, $BCD(\eta)$, which can be described in the following manner. Suppose that each time an eligible subject arrives, one calculates $D =$ (no. of subjects previously assigned to A) $-$ (no. of subjects previously assigned to B). Then the following rule is used: if $D = 0$, assign this subject to either treatment with probability $\frac{1}{2}$; if $D < 0$, assign this subject to treatment A with probability η; if $D > 0$, assign this subject to treatment B with probability η. A value of η is used so that $\eta \geqq \frac{1}{2}$. The $BCD(\eta)$ indeed forces the experiment to be balanced and also retains some randomization. But neither can it discriminate between large absolute values of D versus small nonzero values, nor can it discriminate between large and small numbers of experimental subjects. For example, consider the following two situations:

(i) $2A$'s and $0B$ have been assigned previously.
(ii) $18A$'s and $16B$'s have been assigned previously.

Although in both cases $D = 2$, the imbalance between treatment groups in (ii) is much less serious that in (i). The treatment difference is not even estimable in (i).

In an experiment, usually there are several factors which are known or thought to affect the subject's ability to respond to treatment. Each factor has several levels. A group of subjects which have a particular combination of such factor

Received July 1976; revised April 1977.
AMS 1970 *subject classifications.* Primary 62L05; Secondary 60K30.
Key words and phrases. Sequential experiment, biased coin design, experimental bias, balance, randomization, clinical trial.

levels is called a stratum. In this paper, we consider treatment assignments which are performed entirely separately within each stratum. We assume that the total number of subjects in this experiment is predetermined, but due to random entry of subjects, the number of subjects who will fall in each stratum is unknown beforehand. Essentially then, we face a design problem of a sequential experiment whose size cannot be predetermined. But no special stopping rule has to be considered.

In Section 2, we propose the adaptive biased coin design which forces an extremely unbalanced or a small-sized experiment to be balanced, but tends toward the complete randomization scheme as the size of the experiment increases.

In order to preserve the continuity in the presentation of this paper, all proofs are placed in the Appendix.

2. The adaptive biased coin design. Because of the symmetry of this design problem, the first treatment assignment is determined by tossing a fair coin. The adaptive biased coin design can best be described by considering an arbitrary stage of the experiment. Suppose after n (> 1) assignments we have N_A A's and N_B B's assigned. Let $D_n = N_A - N_B$ and let p be a nonincreasing function of D_n/n. The possible values of p range from 0 to 1. Then the $(n+1)$th subject is assigned to treatment group A with probability $p = p(D_n/n)$ and to treatment group B with probability $q = q(D_n/n)$, where $p + q = 1$. The function $p(x)$ is chosen to be symmetric with respect to the point $(0, \frac{1}{2})$; i.e., $p(x) = q(-x)$, for $x \in [-1, 1]$.

The D_n form a Markov process with states of all integers, $\cdots, -2, -1, 0, 1, 2, \cdots$, and the transition probabilities from the nth assignment to the $(n+1)$th assignment are as follows:

$$P(D_{n+1} = j + 1 \mid D_n = j) = P(j/n) ,$$

and

$$P(D_{n+1} = j - 1 \mid D_n = j) = q(j/n) , \qquad \text{where} \quad -n \leqq j \leqq n .$$

We note that the $BCD(\eta)$ is a special case of the adaptive biased coin design; i.e., we take

$$p(x) = \eta , \quad \text{for} \quad -1 \leqq x \leqq 0 , \qquad \text{and} \qquad p(0) = \tfrac{1}{2} .$$

If we take the straight line $p(x) = (1 - x)/2$, then this is an urn design which has been shown to yield a generally good design (Wei (1977a)).

3. Experimental bias. If the experimenter is aware of or guesses which treatment a subject will receive before selecting the subject, then he may consciously or unconsciously bias the experiment by his choice of who is or is not a suitable experimental subject. The complete randomization scheme eliminates this bias, but the systematic design ($ABAB\cdots$ or $BABA\cdots$) maximizes it. This kind of bias is called selection bias (Blackwell and Hodges (1957), Stigler (1969)). A natural measure of the selection bias of a sequential design is the expected number of correct guesses of treatment assignments the experimenter can make if he guesses optimally. The best guessing strategy against the adaptive biased coin design is to guess treatment A or B 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 at stage $(n + 1)$ is the expected value of $p(-|D_n|/n)$. The following theorem shows that the selection bias for some adaptive designs is almost eliminated as the size of the experiment increases.

THEOREM 1. *Let D_n be the process which is generated by $p(x)$. If $p(x)$ is continuous at $x = 0$, then $Ep(-|D_n|/n) \to \frac{1}{2}$, as $n \to \infty$.*

With respect to the $BCD(\eta)$, the $Ep(-|D_n|/n) \to \frac{1}{2} + (r - 1)/4r$, where $r = \eta/(1 - \eta)$ (Efron (1971)). The $BCD(\eta)$ balances a large-sized experiment "too excessively" and has a high potential of being biased by the experimenter through his selection of experimental subjects.

Another common kind of bias is accidental bias (Efron (1971)). There may be nuisance factors known or unknown to the experimenter systematically affecting the experimental subjects. For example, age, sex, time trend, etc. The following theorem shows that the design which satisfies the condition in Theorem 1 is almost free of any kind of bias as experimental size increases.

THEOREM 2. *Let $T_j = 1$ or -1 as the jth subject is assigned to treatment group A or B by a design with function $p(x)$, where $j = 1, 2, \cdots$. If $p(x)$ is continuous at $x = 0$, then $\rho_{n,k} = E(T_n T_{n+k}) \to 0$, as $n \to \infty$, for $k = 1, 2, \cdots$.*

We note that for the $BCD(\eta)$, the function $p(x)$ is not continuous at $x = 0$. Efron (1971) has found the limit of $\rho_{n,k}$ for the $BCD(\eta)$. Asymptotically, T_n and T_{n+k} for $BCD(\eta)$ are still correlated. On the other hand, the design with function $p(x) = (1 - x)/2$ satisfies the condition stated in Theorem 2. So, it behaves more and more like the complete randomization scheme as the size of the experiment increases.

4. The balancing property and robustness of the adaptive design. The following theorem shows the asymptotic balancing property of some adaptive designs.

THEOREM 3. *The process D_n generated by a function $p(x)$ which is differentiable at $x = 0$ has the following property: as $n \to \infty$, the distribution of $n^{-\frac{1}{2}}D_n$ converges to a normal distribution with mean 0 and variance $1/(1 - 4p'(0))$, where $p'(0)$ is the first derivative of $p(x)$ at $x = 0$.*

Suppose that X and Y are the responses of subjects treated by A and B, with a common variance σ^2 and means μ_A and μ_B, respectively. At any stage n of this sequential experiment, the perfect balance design is robust in the sense that the design is insensitive to wild observations (Box and Draper (1975)). Box and Draper gave a suitable measure of design sensitivity to wild observations. In our case, it can be shown that this measure γ becomes $1/N_A + 1/N_B$, which is proportional to $\mathrm{Var}(\bar{X} - \bar{Y})$, where $\bar{X}$ and $\bar{Y}$ are, respectively, the sample means of X and Y at stage n. If the number of experimental subjects is preassigned, say $2m$, then we can obtain a perfectly balanced experiment and the measure $\gamma(= 2/m)$ is minimized. Now if the number of subjects is not known beforehand, it is interesting to know how many extra observations we need for the adaptive biased coin design to reduce γ to be less than or equal to $2/m$; i.e., we continue taking observations until N_A and N_B satisfy

$$(4.1) \qquad 1/N_A + 1/N_B \leqq 2/m .$$

If we write $N_A + N_B = 2m + U$, then $U \geqq 0$ is the number of additional observations required by the adaptive design to satisfy (4.1). For a given value $U = u$, let $n_A{}^u$ denote the smallest value of N_A for which (4.1) holds; i.e.,

$$1/n_A{}^u + 1/n_B{}^u \leqq 2/m < 1/(n_A{}^u - 1) + 1/(n_B{}^u + 1) ,$$

where $n_A{}^u + n_B{}^u = 2m + u$. Therefore $U \leqq u$ if and only if, at the $(2m + u)$th step, $n_A{}^u \leqq N_A \leqq n_B{}^u$, where

$$n_B{}^u = [(2m + u + (2m + u)u^{\frac{1}{2}})/2]$$

and $[\cdot]$ is the greatest integer function. As $m \to \infty$, U has a simple limit law. By Theorem 3, the distribution of $(2m + u)^{-\frac{1}{2}}D_{2m+u}$ converges to a normal distribution with mean 0 and variance $1/(1 - 4p'(0))$, as $m \to \infty$. Since as $m \to \infty$, $n_B{}^u \approx (m + u/2) + u^{\frac{1}{2}}(2m + u)^{\frac{1}{2}}/2$, it follows that

$$P(U \leqq u) \to_{m\to\infty} \Phi(((1 - 4p'(0))u)^{\frac{1}{2}}) - \Phi(-((1 - 4p'(0))u)^{\frac{1}{2}}) , \qquad u = 0, 1, 2, \cdots ,$$

where Φ is the cdf of $N(0, 1)$. Also, since $\Phi(-3.29) = .0005$, then it is practically certain that the adaptive design does not cost as many as $10.8241/(1 - 4p'(0))$ extra observations in a large-scaled experiment. For example, if we choose $p(x) = (1 - x)/2$, then practically at most four extra observations are needed to satisfy (4.1). We note that this comparison is unfair to the adaptive design, as the inequality (4.1) is usually strict and we are obtaining a somewhat more accurate estimate of $\mu_A - \mu_B$.

5. Remarks.

(1) A direct application of the above adaptive biased coin design is in controlled sequential clinical trials. The main difficulty in treating each stratum as a separate experiment is that the number of strata increases rapidly as the number of factors increases. For example, in some cancer studies, there are many factors that are thought to have influence on the responses of patients to treatment. In this case, very few patients fall in each stratum. Any within-stratum assignment rule may fail to achieve the aim of balance. An example of such a case was given by Zelen (1974). The problem becomes even more serious when one considers multicenter trials. An overall assignment rule should be considered in this situation (Zelen (1974), Pocock and Simon (1975), Freedman and White (1976), Wei (1977b)).

(2) A very important feature of the design problem of sequential experiments is the requirement of simplicity for the assignment rule. In practice, we recommend the adaptive design with $p(x) = (1 - x)/2$, because it is very easy to implement (Wei (1977a)).

APPENDIX

Several lemmas are needed to prove Theorems 1 and 2.

LEMMA 1. *Let S_n be the process generated by the function $p(x) = \frac{1}{2}$, for $-1 \leqq x \leqq 1$. Also, let $g(j)$ be an increasing function of j, $j = 0, 1, 2, \cdots$. Then,*

$$E\{g(|D_n|)\} \leqq E\{g(|S_n|)\}, \qquad \textit{for} \quad n = 0, 1, 2, \cdots.$$

PROOF. This lemma is trivially true for $n = 0$. An inductive proof will be given for n. Now,

$$\begin{aligned} Eg(|D_{n+1}|) &= E\{g(|D_n| + 1)p(|D_n|/n) + g(|D_n| - 1)q(|D_n|/n)\} \\ &= \tfrac{1}{2}(Eg(|D_n| + 1) + Eg(|D_n| - 1)) \\ &\quad + E(p(|D_n|/n) - \tfrac{1}{2})(g(|D_n| + 1) - g(|D_n| - 1)). \end{aligned}$$

Since $p(x) \leqq \frac{1}{2}$, for $x \geqq 0$ and since g is increasing, the second term is $\leqq 0$. Since the functions $g(j + 1)$ and $g(j - 1)$ are increasing in j, the induction assumption implies that the first term is bounded by $\frac{1}{2}(Eg(|S_n| + 1) + Eg(|S_n| - 1)) = Eg(|S_{n+1}|)$. □

LEMMA 2. *For any positive real number t and $n = 1, 2, \cdots$,*

$$E|D_n|^t \leqq E|S_n|^t.$$

LEMMA 3. *$D_n/n \to 0$, with probability 1, as $n \to \infty$.*

PROOF. Since $E|D_n|^4 \leqq E|S_n|^4 = 3n^2 - 2n$, and

$$\textstyle\sum_{n=1}^{\infty} E|D_n|^4/n^4 \leqq \sum_{n=1}^{\infty} 3n^{-2} < \infty,$$

it follows that $D_n/n \to_{\text{a.s.}} 0$, as $n \to \infty$. □

PROOF OF THEOREM 1. Since $D_n/n \to 0$, a.s. as $n \to \infty$, $|D_n/n| \to 0$, a.s. as $n \to \infty$. The function $p(x)$ is continuous at $x = 0$, then $p(-|D_n|/n) \to p(0) = \frac{1}{2}$, a.s. as $n \to \infty$. By the dominated convergence theorem, $Ep(-|D_n|/n) \to \frac{1}{2}$, as $n \to \infty$. □

PROOF OF THEOREM 2. Consider $\rho_{n+1,k} = E(T_{n+1}T_{n+k+1})$. Let $N_A = \sum_{j=1}^{n}(T_j + 1)/2$ and $M_A = \sum_{j=n+1}^{n+k}(T_j + 1)/2$. Then, $E(T_{n+1}T_{n+k+1}) = EEE(T_{n+1}T_{n+k+1} \mid M_A, N_A)$. Consider the third expectation

$$\begin{aligned} &E(T_{n+1}T_{n+k+1} \mid M_A = m_A, N_A = n_A) \\ &\quad = E(T_{n+1} \mid M_A = m_A, N_A = n_A)E(T_{n+k+1} \mid M_A = m_A, N_A = n_A) \qquad \text{(A.1)} \\ &\quad = E(T_{n+1} \mid M_A = m_A, N_A = n_A) \\ &\qquad \times \{2p((2(m_A + n_A) - n - k)/(n + k)) - 1\}. \end{aligned}$$

By Lemma 3 and the fact that $p(x)$ is continuous at $x = 0$, for any fixed m_A, where $0 \leqq m_A \leqq k$, $p((2(m_A + N_A) - n - k)/(n + k)) \to_{\text{a.s.}} \frac{1}{2}$, as $n \to \infty$. Also, since the first term of (A.1) is bounded by 1,

$$E(T_{n+1}T_{n+k+1} \mid M_A = m_A, N_A) \to 0 , \quad \text{a.s. as} \quad n \to \infty ,$$

and

$$E(T_{n+1}T_{n+k+1} \mid N_A) \to 0 , \quad \text{a.s. as} \quad n \to \infty .$$

By the dominated convergence theorem,

$$\rho_{n+1,k} = E(T_{n+1}T_{n+k+1}) \to 0 , \qquad \text{as} \quad n \to \infty . \qquad \square$$

Several lemmas are also needed to prove Theorem 3.

LEMMA 4. *The D_n is a symmetric process in the sense that*

$$P(D_n = j) = P(D_n = -j) ,$$

where $j = \cdots, -1, 0, 1, \cdots$ and $n = 0, 1, 2, \cdots$.

PROOF. The result is trivially true for $n = 0$. An inductive proof will be given for n. Now, for any integer j,

$$\begin{aligned} P(D_{n+1} = j) &= P(D_{n+1} = j \mid D_n = j + 1)P(D_n = j + 1) \\ &\quad + P(D_{n+1} = j \mid D_n = j - 1)P(D_n = j - 1) \\ &= q((j + 1)/n)P(D_n = j + 1) + p((j - 1)/n)P(D_n = j - 1) . \end{aligned}$$

We assume that $p(x) = 0$ and $q(x) = 0$ for $x \notin [-1, 1]$. Since $q((j + 1)/n) = p(-(j + 1)/n)$ and $p((j - 1)/n) = q(-(j - 1)/n)$, by the induction assumption, it follows that $P(D_{n+1} = j) = p(-(j + 1)/n)P(D_n = -j - 1) + q(-(j - 1)/n)P(D_n = -j + 1) = P(D_{n+1} = -j)$. $\square$

LEMMA 5. $ED_n^{2k+1} = 0$, *for $k = 0, 1, \cdots$, and $n = 1, 2, \cdots$.*

LEMMA 6.[1] *For any ε, such that $0 < \varepsilon < 1$, and $n\varepsilon^4 \to \infty$, as $n \to \infty$,*

$$\begin{aligned} &\textstyle\sum_{j=-n}^{-\varepsilon n} E(D_{n+1}^{2k+2} \mid D_n = j)P(D_n = j) \\ &\qquad = \textstyle\sum_{j=\varepsilon n}^{n} E(D_{n+1}^{2k+2} \mid D_n = j)P(D_n = j) = o(n^k) . \end{aligned}$$

PROOF.

$$\begin{aligned} &(n + 1)^{-k} \textstyle\sum_{j=\varepsilon n}^{n} E(D_{n+1}^{2k+2} \mid D_n = j)P(D_n = j) \\ &\quad = \textstyle\sum_{j=\varepsilon n}^{n} E(D_{n+1}^{2k+2}/(n + 1)^{k+1} \mid D_n = j)(n + 1)P(D_n = j) \\ &\quad = E\{E(D_{n+1}^{2k+2}/(n + 1)^{k+1} \mid D_n)(n + 1)I_{[D_n/n \geqq \varepsilon]}\} \\ \text{(A.2)} \qquad &\quad \leqq E^{\frac{1}{3}}(E(D_{n+1}^{2k+2}/(n + 1)^{k+1} \mid D_n))^3 \cdot E^{\frac{2}{3}}I_{[D_n/n \geqq \varepsilon]}(n + 1) . \end{aligned}$$

The second term of (A.2) $= (n + 1)E^{\frac{2}{3}}I_{[D_n/n \geqq \varepsilon]}$. By Lemma 1 this quantity

$$\begin{aligned} &\leqq (n + 1)E^{\frac{2}{3}}I_{[S_n/n \geqq \varepsilon]} \leqq (n + 1)(P(S_n/n) \geqq \varepsilon)^{\frac{2}{3}} \\ &\leqq (n + 1)(ES_n^4/n^4\varepsilon^4)^{\frac{2}{3}} = (n + 1)((3n^2 - 2n)/n^4\varepsilon^2)^{\frac{2}{3}} = o(1) . \end{aligned}$$

[1] When we put a noninteger εn in a position where there "obviously" should be an integer, then εn is interpreted as its integral part. Empty summation yields 0.

The first term of (A.2) $\leqq E^{\frac{1}{3}}E(D_{n+1}^{6k+6}/(n+1)^{3k+3} \mid D_n)$

$$= E^{\frac{1}{3}}(D_{n+1}^{6k+6}/(n+1)^{3k+3}) \leqq E^{\frac{1}{3}}(S_{n+1}^{6k+6}/(n+1)^{3k+3}) = O(1) .$$

Therefore, (A.2) $= o(1)$. □

LEMMA 7. $\sum_{j=\varepsilon n}^{n} j^{2l}P(D_n=j) = \sum_{j=-n}^{-\varepsilon n} j^{2l}P(D_n=j) = o(n^{l-1})$, *for* $l = 1, 2, \cdots$.

PROOF.

$$n^{-l+1}\sum_{j=\varepsilon n}^{n} j^{2l}P(D_n=j) = E((D_n^{2l}/n^l)nI_{[D_n/n\geqq\varepsilon]}) \leqq E((S_n^{2l}/n^l)nI_{[S_n/n\geqq\varepsilon]})$$
$$\leqq E^{\frac{1}{3}}(S_n^{6l}/n^{3l})nE^{\frac{2}{3}}I_{[S_n/n\geqq\varepsilon]} = o(1) . \qquad \square$$

LEMMA 8. *Let* v_n, a_n *be real numbers for* $n \geqq 0$ *with* $v_{n+1} = a_n v_n + b_n$, *where* $a_n = 1 + [a'/(b+cn)]$, $b_n \approx \phi n^d$ *with* $\phi \neq 0$ *and* $d > c^{-1}a' - 1$, *then*

$$v_n \approx [\phi/(d - (a'/c) + 1)]n^{d+1} .$$

PROOF. Cf. D. A. Freedman (1965), pages 969–970.

LEMMA 9. *For each nonnegative integer* k, $\lim_{n\to\infty} E(D_n^{2k}/n^k) = \mu(2k)$, *where* $\mu(2k)$ *is the* 2kth *moment of* $N(0, 1/(1-4p'(0)))$.

PROOF. The result is trivial for $k = 0$, and $\mu(0) = 1$. An inductive proof will be given for k. We let $\varepsilon = n^{-\frac{1}{8}}$ in this proof so that the condition in Lemma 6 is satisfied. Now,

$$ED_{n+1}^{2k+2} = \sum_{j=-n}^{n} E(D_{n+1}^{2k+2} \mid D_n = j)P(D_n = j) .$$

By Lemma 6, the above quantity becomes

$$\sum_{j=-\varepsilon n}^{\varepsilon n} E(D_{n+1}^{2k+2} \mid D_n = j)P(D_n = j) + o(n^k)$$
$$= \sum_{j=-\varepsilon n}^{\varepsilon n} ((j+1)^{2k+2}p(j/n) + (j-1)^{2k+2}q(j/n))P(D_n = j) + o(n^k)$$
$$= \sum_{j=-\varepsilon n}^{\varepsilon n} \left(\sum_{l=0}^{k+1} \binom{2k+2}{2l} j^{2l} + \sum_{l=0}^{k} \binom{2k+2}{2l+1} j^{2l+1}(2p(j/n) - 1)\right)P(D_n = j)$$
$$+ o(n^k) .$$

By Lemma 4, the above quantity becomes

$$\text{(A.3)} \qquad \sum_{l=0}^{k+1} \binom{2k+2}{2l} \sum_{j=-\varepsilon n}^{\varepsilon n} j^{2l}P(D_n = j)$$
$$+ \sum_{l=0}^{k} \binom{2k+2}{2l+1} \sum_{j=-\varepsilon n}^{\varepsilon n} j^{2l+1}(2p(j/n) - 1)P(D_n = j) + o(n^k) .$$

By Lemma 7, the first term of (A.3) becomes

$$\sum_{l=0}^{k+1} \binom{2k+2}{2l} ED_n^{2l} + o(n^k) .$$

Since we assume that the $p(x)$ is differentiable at $x = 0$,

$$p(x) = \tfrac{1}{2} + xp'(0) + xr(x) ,$$

where $r(x) \to 0$, as $x \to 0$. The second term of (A.3) can be expressed as

$$\sum_{l=0}^{k} \binom{2k+2}{2l+1} \sum_{j=-\varepsilon n}^{\varepsilon n} j^{2l+1}(2p'(0)j/n + 2jr(j/n)/n)P(D_n = j) .$$

Now, since

$$\frac{2}{n} \sum_{j=-\varepsilon n}^{\varepsilon n} j^{2l+2} r(j/n) P(D_n = j) = \frac{2}{n} ED_n^{2l+2} r(D_n/n) I_{[|D_n| \leqq \varepsilon n]}$$
$$\leqq \frac{2}{n} \sup_{|x| \leqq \varepsilon} |r(x)| ED_n^{2l+2}$$
$$\leqq \text{constant} \cdot n^k \cdot \sup_{|x| \leqq \varepsilon} |r(x)|$$

and $\sup_{|x| \leqq \varepsilon} |r(x)| \to 0$, as $n \to \infty$ (we note that $\varepsilon = n^{-\frac{1}{3}}$). It follows that

$$ED_{n+1}^{2k+2} = \sum_{l=0}^{k+1} \binom{2k+2}{2l} ED_n^{2l} + 2p'(0) n^{-1} \sum_{l=0}^{k} \binom{2k+2}{2l+1} ED_n^{2l+2} + o(n^k) .$$

Let $V_n(2k) = ED_n^{2k}$, then we have

$$V_{n+1}(2k+2) = (1 + 2p'(0)n^{-1}(2k+2)) V_n(2k+2) + b_n ,$$

where

$$b_n = \sum_{l=0}^{k} \binom{2k+2}{2l} ED_n^{2l} + 2p'(0) n^{-1} \sum_{l=0}^{k-1} \binom{2k+2}{2l+1} ED_n^{2l+2} + o(n^k) .$$

By the induction assumption

$$b_n \approx \binom{2k+2}{2k} \mu(2k) n^k .$$

By Lemma 8, we have $V_n(2k+2) \approx \binom{2k+2}{2} \mu(2k) n^{k+1} / ((k+1)(1-4p'(0)))$. Therefore, $\mu(2k+2) = \binom{2k+2}{2} \mu(2k) / ((k+1)(1-4p'(0)))$. It can be shown that $\mu(2k)$ is the $2k$th moment of $N(0, 1/(1-4p'(0)))$. □

PROOF OF THEOREM 3. Since the sequence of moments of a normal distribution uniquely determines the distribution, it follows that the distribution of $n^{-\frac{1}{2}} D_n$ tends to a normal distribution with mean 0 and variance $1/(1-4p'(0))$, as $n \to \infty$.

Acknowledgment. The author is grateful to the Associate Editor and the referee for a number of valuable comments, and especially for indicating a subtlety in the original proof of Theorem 3. Appreciation is also extended to Professors I. R. Savage and S. M. Stigler for many helpful suggestions.

REFERENCES

[1] BLACKWELL, D. and HODGES, J. L. (1957). Design for the control of selection bias. *Ann. Math. Statist.* **28** 449-460.

[2] BOX, G. E. P. and DRAPER, N. R. (1975). Robust designs. *Biometrika* **62** 347-352.

[3] EFRON, B. (1971). Forcing a sequential experiment to be balanced. *Biometrika* **58** 403-417.

[4] FREEDMAN, D. A. (1965). Bernard Friedman's urn. *Ann. Math. Statist.* **36** 956-970.

[5] FREEDMAN, L. S. and WHITE, S. J. (1976). On the use of Pocock and Simon's method for balancing treatment numbers over prognostic factors in the controlled clinical trial. *Biometrics* **32** 691-694.

[6] POCOCK, S. and SIMON, R. (1975). Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics* **31** 103-115.

[7] STIGLER, S. M. (1969). The use of random allocation for the control of selection bias. *Biometrika* **56** 553-560.

[8] WEI, L. J. (1977a). A class of designs for sequential clinical trials. *J. Amer. Statist. Assoc.* To appear.

[9] WEI, L. J. (1977b). An application of an urn model to the design of sequential controlled clinical trials. Unpublished manuscript.

[10] ZELEN, M. (1974). The randomization and stratification of patients to clinical trials. *J. Chronic Disease* **27** 365–375.

DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE
UNIVERSITY OF SOUTH CAROLINA
COLUMBIA, SOUTH CAROLINA 29208