Bias Properties and Nonparametric Inference for Truncated Binomial Randomization

By WILLIAM F. ROSENBERGER and ANDREW. L. RUKHIN

Department of Mathematics and Statistics, University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore, Maryland 21250, U. S. A.; billr@math.umbc.edu, rukhin@math.umbc.edu

Abstract

This note examines certain properties of the truncated binomial randomization scheme for clinical trials. In particular, we derive the degree to which the scheme is susceptible to accidental bias and discuss the large sample properties of standard nonparametric permutation tests under a randomization model.

Keywords: Asymptotic normality; Random allocation rule; Selection bias

1. Introduction

In forcing balance in a clinical trial of $n = 2m$ patients and two treatments (A and B), two randomization schemes arise naturally. The first is the *random allocation rule*, which randomly selects one of the $\binom{2m}{m}$ permutations of exactly m As and m Bs. The other is the *truncated binomial design* (Blackwell and Hodges, 1957), which flips a fair coin (A if heads, B if tails) until either m As are assigned or m Bs are assigned, and then fills out the sequence (deterministically) with all remaining patients on the appropriate treatment. The degree to which the two designs are susceptible to selection bias has been well-studied (e.g., Blackwell and Hodges, 1957; Stigler, 1969; Wei, 1978). In this paper, we explore other properties of the truncated binomial design.

Equal allocation to both treatments has become fairly standard practice in clinical trials today. Even if the trial is not designed for exactly n (known in advance) patients, randomization is often blocked or stratified with known block or strata sizes. Unfortunately, most medical journals and protocols often use imprecise statements, such as "a randomization sequence was generated". Then it is largely impossible to determine exactly what randomization scheme was used used in the clinical trial. But it can be assumed that, at least within blocks in blocked randomization, if not in general, either the random allocation rule or the truncated binomial design would be used to ensure equal sample sizes within blocks. In fact, a programmer with the task of randomly generating a sequence of 50 As and 50 Bs will likely use a truncated binomial design. The nonchalance with which the act of randomization is often treated in clinical trials is deeply disturbing to those who understand the possibilities for subtle biases, such as selection bias and accidental bias, and also to those who promote randomization as a basis for inference.

In this paper, we explore the truncated binomial design further. In Section 2, we describe distributional properties of this design. In Section 3 its accidental bias is determined. In Section 4, we derive the asymptotic distribution of the standard permutation test from a randomization-based analysis.

2. The Truncated Binomial Design

Let $X_1, X_2, \ldots$ represent independent Bernoulli random variables with probability 1/2 of assignment to treatment A and probability 1/2 of assignment to treatment B. For a given number $n = 2m$ of trials, denote by τ the stopping rule of the form

$$\tau = \min\left\{j : \sum_{k=1}^{j} X_k \geq m \ \text{ or } \ \sum_{k=1}^{j}(1 - X_k) \geq m\right\}. \tag{1}$$

Note that

$$\tau = \min\left[\tau_1, \tau_2\right]$$

with

$$\tau_1 = \min\left\{j : \sum_{k=1}^{j} X_k \geq m\right\},$$

$$\tau_2 = \min\left\{j : \sum_{k=1}^{j}(1 - X_k) \geq m\right\}.$$

Clearly both τ_1 and τ_2 have a negative binomial distribution with the parameter m.

The probability distribution of the random variable τ has the form

$$\Pr\left(\tau = m + x\right) = \binom{m + x - 1}{x} \frac{1}{2^{m+x-1}}, \quad x = 0, 1, \ldots, m - 1, \tag{2}$$

(Blackwell and Hodges, 1957). The moment generating function of τ, for example, can be obtained from Proposition 2.1.1 of Uppuluri and Blot (1970). With $I_p(k, \ell)$ denoting the incomplete beta-function, one has for $s < \log 2$

$$Ee^{s\tau} = 2\left(\frac{e^s}{2 - e^s}\right)^m \left[1 - I_{e^s/2}(m, m)\right]. \tag{3}$$

The following lemma, whose proof easily follows from (3) and Stirling formula, gives the asymptotic distribution of τ.

Lemma 1. For $m \to \infty$, $\tau/n \to 1$ in probability, and the distribution of $(\tau - 2m)/(2m)^{1/2}$ converges weakly to that of $-|Z|$ with Z denoting the standard normal random variable. With the standard normal density φ,

$$\Pr\left(\tau = m + x\right) \sim \frac{2}{\sqrt{2m}} \varphi\left(\frac{m - x}{\sqrt{2m}}\right).$$

Lemma 1 also can be derived from the fact that both τ_1 and τ_2 are asymptotically normal and $corr(\tau_1, \tau_2) \to -1$, so that the limiting normal joint distribution of $\big((\tau_1 - 2m)/(2m)^{1/2}, (\tau_2 - 2m)/(2m)^{1/2}\big)$ is concentrated on the line $x_1 = -x_2$.

3. Accidental Bias

Using the notation of Section 1, define for $i = 1, ..., n$

$$\begin{aligned} T_i &= 2X_i - 1, \text{ if } i \leq \tau; \\ &= -1, \text{ if } \tau = \tau_1; \\ &= 1, \text{ if } \tau = \tau_2. \end{aligned}$$

Let $T = (T_1, ..., T_n)^T$ Then $E(T) = 0$, and denote $Var(T) = \Sigma$.

Efron (1971) introduced the concept of *accidental bias* which is a measure of the expected bias of the treatment effect from a standard linear model when important covariates are ignored. Let $e = (1, \ldots, 1)^T$, and let z be a vector of (deterministic) covariates left out of the model such that $z^T e = 0$, $z^T z = 1$. The degree to which one is subject to accidental bias is given by the term $E(z^T T)^2 = z^T \Sigma z$.

In order to find a randomization scheme which would minimize the accidental bias, Efron proposed a minimax strategy. Since $z^T \Sigma z$ cannot exceed the maximum eigenvalue of Σ, he suggested that the maximum eigenvalue $\lambda_{\max}$ be examined to determine the degree of susceptibility to accidental bias. Smith (1984) argued that this minimax solution may not be satisfactory, and suggested using

$$AB = \max_{z: z^T \mathbf{e} = 0} \frac{z^T \Sigma z}{z^T z}.$$

To investigate the accidental bias for the truncated binomial design, we derive first the form of the covariance matrix Σ.

Lemma 2. For $i < j$, $cov(T_i, T_j) = \Pr(\tau < i)$.

Proof. It is more convenient to work with $Y_i = (T_i + 1)/2$ as these variables take values 0 and 1 only. Condition first on $\tau = \tau_1$. For $i < j$,

$$E\left(Y_i Y_j \middle| \tau = \tau_1\right) = \Pr(Y_i = 1, Y_j = 1 | \tau = \tau_1) = 1/4 \Pr(\tau \geq j | \tau = \tau_1).$$

Similarly, given that $\tau = \tau_2$, we obtain

$$E\left(Y_iY_j\middle|\tau = \tau_2\right) = 1/4\,\Pr(\tau \geq j|\tau = \tau_2) + 1/2\,\Pr(i \leq \tau < j|\tau = \tau_2) + \Pr(\tau < i|\tau = \tau_2).$$

Unconditioning, we see that

$$\begin{aligned} E(Y_iY_j) &= 1/4\,\Pr(\tau \geq j) + 1/4\,\Pr(i \leq \tau < j) + 1/2\,\Pr(\tau < i) \\ &= 1/4 + 1/4\,\Pr(\tau < i). \end{aligned}$$

Consequently,

$$cov(T_i, T_j) = \Pr(\tau < i).\ \square$$

Let, for $k = 1, \ldots, m$, $p_k = \Pr(\tau = m+k-1)$, as given by (2), and $H(k) = \sum_{\ell=1}^{k} p_\ell$. As $cov(T_i, T_j) = 0$ if $i \leq m$, the matrix Σ has the block structure

$$\begin{pmatrix} I & 0 \\ 0 & C \end{pmatrix},$$

where C is an $m \times m$ matrix with elements $(1 - \delta_{ij})H(i \wedge j) + \delta_{ij}$.

Lemma 3. We have the following results:

$$\sum_{i=1}^{m} p_i^2 \sim \frac{1}{\sqrt{2\pi m}},$$

$$\sum_{i=1}^{m} p_i H(i)[1 - H(i)] \to \frac{1}{6}.$$

and

$$m^{-1/2} \sum_{i=1}^{m} H(i)[1 - H(i)] \to 2\int_0^\infty [2\Phi(u) - 1][1 - \Phi(u)]\, du = \frac{2-\sqrt{2}}{\sqrt{\pi}}.$$

Proof. The local limit theorem for the probabilities (2) in Lemma 1 shows that for $0 < x \leq m$,

$$p_x \sim \frac{2}{\sqrt{2m}} \varphi\left(\frac{m-x}{\sqrt{2m}}\right).$$

Also

$$\sum_{i=1}^{m} p_i^2 \sim \frac{2}{\pi\sqrt{2m}} \int_0^\infty e^{-u^2}\, du = \frac{1}{\sqrt{2\pi m}}.$$

As for $u > 0$, $H(m - \sqrt{2mu}) \to 2\int_u^\infty \varphi(t)\, dt$,

$$\sum_{i=1}^{m} p_i H(i)[1 - H(i)] \to \int_0^\infty \varphi(t) \left(\int_t^\infty \varphi(u)\, du\right) \left(\int_0^t \varphi(v)\, dv\right)\, dt = \frac{1}{6}.$$

A similar argument proves the last formula of Lemma 3. □

It is clear that AB cannot exceed the largest eigenvalue $\lambda_{\max}$ of matrix C. We show now that both $\lambda_{\max}$ and AB are of order $m^{1/2}$.

PROPOSITION 1. Both $\lambda_{\max}$ and AB are of order $m^{1/2}$.

PROOF. Let $\bar{p} = (p_1 - m^{-1}, \ldots, p_m - m^{-1})^T$. Then the vector $z = \bar{p}/\sqrt{\sum p_i^2 - m^{-1}}$ has unit norm, $z\mathbf{e}^T = 0$, and by Lemma 3

$$z^T C z = 1 + \frac{2\sum_{i=1}^{m}[p_i - m^{-1}]H(i)[1 - H(i)]}{\sum_{i=1}^{m} p_i^2 - m^{-1}} \sim \frac{\sqrt{2\pi m}}{3}.$$

Therefore, for sufficiently large m

$$\frac{\sqrt{2\pi m}}{3} \leq AB \leq \lambda_{\max}.$$

The fact that

$$\lambda_{\max} \leq \sqrt{m}\,[1 + o(1)]\,,$$

can be derived from the Cholesky decomposition, $C = R^T R$ with the upper diagonal matrix R,

$$R = \begin{pmatrix} r_{11} & r_1 & \cdots & r_1 \\ 0 & r_{22} & \cdots & r_2 \\ \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \cdots & r_{mm} \end{pmatrix}.$$

Here $r_{11} = 1, r_1 = \sqrt{p_1} = 2^{-(m-1)/2}$; $r_{ii}, i = 2, \ldots, m$ and $r_i, i = 2, \ldots, m-1$ are determined from recursive relations

$$r_{ii}^2 = 1 - \sum_{j=1}^{i-1} r_j^2,$$

and

$$r_{ii} r_i = H(i) - \sum_{j=1}^{i-1} r_j^2.$$

In particular,

$$\sum_{j=1}^{m-1} r_j^2 + r_{mm}^2 = 1.$$

With $||C||_1$ denoting the maximum column sum matrix norm, one has

$$\lambda_{\max} \leq ||C||_1 \leq ||R^T||_1 ||R||_1 = (1 + (m-1)p_1)(\sum_{j=1}^{m-1} r_j + r_{mm})$$

$$\leq \sqrt{m}(1 + (m-1)p_1) \left(\sum_{j=1}^{m} r_j^2 + r_{mm}^2 \right)^{1/2} = \sqrt{m} \left(1 + \frac{m-1}{2^{m-1}} \right),$$

and the conclusion of Proposition 1 follows. □

We can conclude that the truncated binomial design can be susceptible to a high degree of accidental bias. In contrast, the maximal eigenvalue for the random allocation rule is

$$\lambda_{\max} = 1 + \frac{1}{n-1},$$

which tends to one.

4. Asymptotic Distribution of Permutation Tests

Because patients from clinical trials are not randomly sampled from a population, standard inference procedures on population parameters may be inappropriate. A *randomization test* or *permutation test* (e.g., Lehmann, 1975; Lachin, 1988) assumes that the sequence of patient responses is deterministic and only the sequence of treatment assignments is random. If a_{jn} is a (deterministic) score function of some outcome variable measured for patient j out of n patients, the basic form of the permutation test is $\sum_j a_{jn} T_j$, where the a_{jn}s are centered. A randomization p-value is then determined with respect to the reference set of all possible randomization sequences $T_1, ..., T_n$ and their attendant probabilities of realization. Typical score functions include the simple ranks and van der Waerden scores. The asymptotic distribution of such permutation tests have been derived for various specialized randomization schemes (e.g., Smythe and Wei, 1983; Wei, Smythe, and Smith, 1986; Smythe, 1988; Rosenberger, 1993).

The asymptotic normality of the permutation test, when a random allocation rule is used, follows from Hájek (1969, Theorem 4a) under a Lindeberg-type condition (4) on the scores. We now derive the asymptotic distribution of a permutation test when the truncated binomial design is used.

Let $a_{jn}, j = 1, \ldots, n$, denote the score coefficients, such that $\sum_j a_{jn} = 0$. Our goal is to study the behavior of the test statistic

$$S_n = \left[\sum_j a_{jn}^2 + n a_{nn}^2\right]^{-1/2} \sum_{j=1}^{n} a_{jn} T_j$$

$$= \left[\sum_j a_{jn}^2 + n a_{nn}^2\right]^{-1/2} \left[\sum_{j=1}^{\tau} a_{jn}(2X_j - 1) \pm \sum_{j=\tau+1}^{n} a_{jn}\right],$$

where the sign $+$ corresponds to the case when $\tau = \tau_2$ and the sign $-$ arises when $\tau = \tau_1$. In other terms,

$$S_n = \left[\sum_j a_{jn}^2 + n a_{nn}^2\right]^{-1/2} \left[\sum_{j=1}^{\tau} a_{jn}(2X_j - 1) - sign(\tau_1 - \tau_2) \sum_{j=\tau+1}^{n} a_{jn}\right].$$

Proposition 2. Assume that when $n \to \infty$

$$\max_{1 \le j \le n} a_{jn}^2 / \sum_{j=1}^{n} a_{jn}^2 \to 0, \tag{4}$$

$$\frac{1}{n} \sum_{j=1}^{n} a_{jn}^2 \to \sigma_1^2 > 0, \tag{5}$$

and with some σ_2 for any positive δ

$$\lim_{n \to \infty} \max_{j : j \ge n - \delta\sqrt{n}} |a_{jn} - \sigma_2| = 0. \tag{6}$$

Then the limiting distribution of S_n is standard normal.

Proof. We use the asymptotic normality of the random vector

$$n^{-1/2} \left(\sum_{1}^{\tau} a_{jn} T_j, \tau_1 - n, \tau_2 - n\right),$$

whose first coordinate is asymptotically independent of two other coordinates. Indeed, for a fixed $\tau_2, \tau = \tau_2$, the distribution of $\tau_1 - \tau_2$ is negative binomial with the parameter

$n - \tau_2$, as under this condition, $\sum_{j=1}^{\tau_2} X_j = \tau_2 - m$, so that $\sum_{j=\tau_2+1}^{\tau_1} X_j = 2m - \tau_2$. Also, under this condition, $\tau_1 - \tau_2$ is independent of $T_1, \ldots, T_{\tau_2}$, and

$$E\left((\tau_1 - n)\sum_1^\tau a_{jn}T_j \Big| \tau_2, \tau = \tau_2\right)$$

$$= (\tau_1 - \tau_2)E\left(\sum_1^\tau a_{jn}T_j \Big| \tau_2, \tau = \tau_2\right) + E\left((\tau_2 - n)\sum_1^{\tau_2} a_{jn}T_j \Big| \tau_2, \tau = \tau_2\right),$$

so that

$$E\left((\tau_1 - n)\sum_1^\tau a_{jn}T_j \Big| \tau = \tau_2\right)$$

$$= 2E\left((n - \tau_2)\sum_1^\tau a_{jn}T_j \Big| \tau = \tau_2\right) + E\left((\tau_2 - n)\sum_1^\tau a_{jn}T_j \Big| \tau = \tau_2\right)$$

$$= E\left((n - \tau)\sum_1^\tau a_{jn}T_j \Big| \tau = \tau_2\right).$$

Similarly,

$$E\left((\tau_1 - n)\sum_1^\tau a_{jn}T_j \Big| \tau = \tau_1\right) = E\left((\tau - n)\sum_1^\tau a_{jn}T_j \Big| \tau = \tau_1\right),$$

and

$$E(\tau_1 - n)\left(\sum_1^\tau a_{jn}T_j\right) = 0.$$

Thus, the first component is uncorrelated with two other coordinates.

Anscombe's Theorem (see Woodroofe, 1982), implies convergence in distribution,

$$\frac{1}{\sqrt{n}}\sum_1^\tau a_{jn}T_j \to \sigma_1 Z_1.$$

Indeed, under assumptions (4) and (5), the sequence $a_{jn}T_j$ is uniformly continuous in probability, and $\tau/n \to 1$ in probability.

Condition (6) implies that

$$\frac{1}{\sqrt{n}}\sum_{j=\tau+1}^{n} a_{jn} \sim \frac{(n-\tau)\sigma_2}{\sqrt{n}} \to \sigma_2|Z_2|.$$

Here Z_1 and Z_2 are two independent standard normal random variables. Also

$$sign(\tau_1 - \tau_2) = sign((\tau_1 - n) - (\tau_2 - n))/\sqrt{n}) \to sign(Z_2),$$

so that one obtains the following convergence in distribution:

$$n^{-1/2}\left[\sum_{j=1}^{\tau} a_{jn}\,(2X_j-1) - sign(\tau_1-\tau_2)\sum_{j=\tau+1}^{n} a_{jn}\right] \to \sigma_1 Z_1 - \sigma_2 Z_2. \;\square$$

In practice, a_{jn} is a function of the integer ranks s_{jn} of the primary outcome of patient j. Let $a_{jn} = h\,(s_{jn}/(n+1))$ be the scores defined by a continuous function h. For instance, if h is the identity function, then the centered score function a_{jn} will lead to a randomization-based Wilcoxon rank-sum test. If h is the probit function, then the a_{jn}s are the van der Waerden scores, and the resulting test is asymptotically equivalent to the normal scores test. Condition (6) holds if for some constant $\rho, 0<\rho<1$, for all positive δ

$$\lim_{n\to\infty}\max_{j:j\geq n-\delta\sqrt{n}}\left|\frac{s_{jn}}{n+1}-\rho\right| = 0.$$

Then $\sigma_2 = h(\rho)$. Heuristically, this condition means that the relative ranks of the last $\delta\sqrt{n}$ patients are approximately equal.

One can replace condition (6), by a more general one:

$$\lim_{n\to\infty}\max_{1\leq j\leq C\sqrt{n}}|a_{n-j,n}/t_n - \sigma_2| = 0$$

for some sequence t_n. However, then the limiting variance of $\sum_{j=1}^{n} a_{jn}T_j$ behaves like $\sum_{j=1}^{n} a_{jn}^2 + t_n^2 n\sigma_2^2$, and the limiting distribution may be determined by just one of the terms.

Clearly, conditions (4) and (5), do not guarantee that $\sum_{j=1}^{n} a_{jn}T_j$ has a limiting distribution, as $\sum_{\tau+1}^{n} a_{jn}$, does not need to converge at all. For example, if

$$\begin{aligned} a_{jn} &= 1 \text{ when } n=2k, j\geq k \text{ or } n=2k+1, j\leq k,\\ &= -1 \text{ otherwise,}\end{aligned}$$

the normalized sum, $n^{-1/2}\sum_{j=\tau+1}^{n} a_{jn}$, converges weakly to $|Z_2|$ for $n=2k$, and the limit is $-|Z_2|$, when $n=2k+1$.

5. Conclusions

In this note, we have shown that the truncated binomial design can lead to a high potential for accidental bias, and we have also demonstrated difficulties with implementation of the large sample permutation test. Some care should be taken in selecting a randomization scheme to force balance, and particular care should be used in randomization-based inference following truncated binomial randomization.

Acknowledgements

William F. Rosenberger is also affiliated with the Department of Epidemiology and Preventive Medicine, University of Maryland Medical School. His work was supported by a grant from the National Cancer Institute. A. L. Rukhin research was supported by a grant from the National Security Agency.

References

Blackwell, D. and Hodges, J. L., Jr. (1957). Design for the control of selection bias. *Annals of Mathematical Statistics* **28** 449-460.

Efron, B. (1971). Forcing a sequential experiment to be balanced. *Biometrika* **58** 403-417.

Hajek, J. (1969). *A Course in Nonparametric Statistics.* Holden-Day, San Francisco.

Lachin, J. M. (1988). Statistical properties of randomization in clinical trials. *Controlled Clinical Trials.* **9** 289-311.

Lehmann, E. L. (1975). *Nonparametrics: Statistical Methods Based on Ranks.* Holden-Day, San Francisco.

Rosenberger, W. F. (1993). Asymptotic inference with response-adaptive treatment allocation designs. *Annals of Statistics* **21** 2098-2107.

Smith, R. L. (1984). Sequential treatment allocation using biased coin designs. *Journal of the Royal Statistical Society B* **46** 519-543.

SMYTHE, R. T. (1988). Conditional inference for restricted randomization designs. *Annals of Statistics* **16** 1155-1161.

SMYTHE, R. T. AND WEI, L. J. (1983). Significance tests with restricted randomization design. *Biometrika* **70** 496-500.

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

UPPULURI, R. R. AND BLOT, J. (1970). A probability distribution arising in a riff-shuffle. In PATIL, G. P., ED. *Random Counts in Scientific Work*. Pennsylvania State University Press, University Park.

WEI, L. J. (1978). On the random allocation design for the control of selection bias in sequential experiments. *Biometrika* **65** 79-84.

WEI, L. J., SMYTHE, R. T., AND SMITH, R. L. (1986). K-treatment comparisons with restricted randomization rules in clinical trials. *Annals of Statistics* **14** 265-274.

WOODROOFE, M. (1982). *Nonlinear Renewal Theory in Sequential Analysis*. SIAM, Philadelphia.