
Volume 137
Issue 3
1 March 2007
ISSN 0378-3758
ELSEVIER
journal of statistical planning and inference
Special Issue on Nonparametric Statistics and Related Topics: In honor of M.L. Puri
Edited by: S. Ralescu
G. Haiman
F. Ruymgaart
For Contents see page iii
Available online at
ScienceDirect
www.sciencedirect.com

Available online at www.sciencedirect.com
ScienceDirect
Journal of Statistical Planning and Inference 137 (2007) 967 – 984
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Non-parametric inference for balanced randomization designs

Andrew L. Rukhin[a, b, *]

[a]Department of Mathematics and Statistics, University of Maryland at Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250, USA
[b]Statistical Engineering Division, National Institute of Standards and Technology, Building 820, Gaithersburg, MD 20899-0001, USA

Available online 21 July 2006

Abstract

This paper compares the properties of two balanced randomization schemes which allocate the known number n of subjects among several treatments. According to the first procedure, the so-called truncated multinomial randomization design, the allocation process starts with the uniform distribution, until a treatment receives the prescribed number of subjects, after which this uniform distribution switches to the remaining treatments, and so on. The second scheme, the random allocation rule, selects at random any assignment of the given number of subjects per treatment. The limiting behavior of these two procedures is shown to be quite different in the sense that for the random allocation rule the instant, at which a treatment gets its prescribed number of subjects, comes much later (after $n - \mathrm{O}(1)$ rather than $n - \mathrm{O}(\sqrt{n})$ subject assignments.) The large sample distribution of standard permutation tests is obtained, and formulas for the accidental bias and for the selection bias of both procedures are derived.

MSC: Primary 60E05; secondary 60C0560G40; 62E20; 62G30; 62P10; 62P15

Keywords: Accidental bias; Clinical trials; Load balancing; Multinomial trials; Normal order statistics; Permutation tests; Probability generating function; Selection bias; Stirling's numbers

1. Introduction

Assume that a given number n of subjects is to be distributed among K treatments at random so that each treatment gets exactly m subjects. Then $n = mK$, which is supposed to be known in advance. A (balanced) randomization design is a probability distribution over the set of $\binom{n}{m \;\cdots\; m}$ sequences of length n which have exactly m subjects per treatment. The two following randomization schemes are commonly used. The *random allocation rule* selects one of these sequences at random. The *truncated multinomial design* uses a randomization scheme which starts with the uniform probability assignment of subjects to treatments until one of the treatments receives m subjects. Then the uniform distribution switches to the remaining $K - 1$ treatments, and the allocation process continues in this way until there is just one treatment with less than m subjects. This treatment then gets all remaining patients. It is of interest to evaluate performance and properties of these two designs. In this paper, the truncated multinomial design and the random allocation rule are compared in terms of selection bias and accidental bias.

* Corresponding author at: Department of Mathematics and Statistics, University of Maryland at Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250, USA. Tel.: +1 410 455 2408; fax: +1 410 455 1066.
E-mail address: rukhin@math.umbc.edu.

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2006.06.020

When $K=2$, Blackwell and Hodges (1957) recommended the former procedure, which they called the truncated binomial design. Indeed, it is the least affected by the selection bias, when the loss is measured by the number of correct guesses concerning the allocation of subjects. Stigler (1969) described the random allocation rule as a balanced Bayes design. He argued that, in many cases, it is superior to the truncated binomial design, which is a minimax procedure. Wei (1978) has provided further support for the use of the random allocation rule by showing that under a different criterion (namely, the squared difference between the number of correct and incorrect guesses) the truncated binomial design loses its minimaxity, and becomes dominated by the random allocation rule.

Equal allocation to each of the treatments seems to be fairly standard practice in comparative studies, especially, in clinical trials. Even if the trial is not designed for the known in advance number of patients, randomization is often applied in blocks or strata of fixed sizes. If, say, $n=(m_1+m_2)K$, and the allocation process is performed in two stages with the quota m_1 at the first stage and m_2 at the second, the formulas obtained in this paper show that for both the truncated multinomial design and the random allocation rule the biases obtained in one compound allocation process are smaller than the sum of biases corresponding to individual stages. Mathematically, this fact is based on sub-additivity of the function $\sqrt{m}$ appearing in Theorems 1 and 2, and in Proposition 4.

Colditz et al. (1989) review several real-life examples demonstrating the dangers of non-randomized assignments in medical studies. Rosenberger and Lachin (2002) give a survey of several sequential randomization schemes (Efron's biased coin design, Wei's adaptive urn design, etc.) which are non-balanced, i.e. they do not guarantee the equal distribution of the subjects per treatment. The selection bias of all these designs cannot be smaller than $n/K=m$, which corresponds to the (non-balanced) complete randomization process in which each treatment is chosen independently with probability $1/K$.

In the presence of the covariates or prognostic factors, the balance should be achieved for each prognostic factor. In this case the minimization rule of Pocock and Simon (1975) is applicable. Sedransk (1973) had compared several allocation schemes, and suggested a technique utilizing information about the inhomogeneity of the subject characteristics.

Another application of randomization designs is in storage of computer files (see, for example, Diekman and Preis, 1999). In a typical load balancing problem one uses a file system in which a computer disk is divided into K equal size zones, so that each zone can hold m fixed size data blocks. A file which is divided into the same size blocks must be stored in a disk whose capacity is $n=mK$ blocks. After a block has been mapped into a zone which already holds $m-1$ other blocks, the disk is considered full. The utility of the disk for given m and K is the number of blocks which have been placed in all the disk zones when a zone starts overflowing.

Clearly, this characteristic coincides with the instant τ_1 defined in Section 2, where distributional characteristics of the truncated multinomial design and asymptotic formulas for the selection bias of both procedures are derived. In Section 3 the different covariance structures of the allocation vectors for the two designs are obtained, and the degree to which the considered rules are susceptible to accidental bias is determined. In Section 4, we determine conditions under which the asymptotic distribution of the standard permutation test is normal. All proofs are collected in the Appendix.

2. Limiting distributions and selection bias

Let $T_1,\ldots,T_n$ be K-nomial random variables representing the allocation of the subjects to K treatments, so that $T_i(\ell)=1$, if at stage i the subject is assigned to treatment ℓ, $\ell=1,\ldots,K$, and $T_i(\ell)=0$, otherwise. The truncated multinomial design starts with the probability of assignment to each of K treatments being $1/K$. For a given number $n=Km$ of trials, denote by τ_1 the stopping rule of the form

$$\tau_1=\min\left\{j:\max_{\ell}\sum_{i=1}^{j}T_i(\ell)\geqslant m\right\}. \tag{1}$$

After the random time τ_1, the probability of assignment to each of the remaining $K-1$ treatments changes to $1/(K-1)$, and the allocation random vectors T_j have the uniform distribution over these treatments. The allocation process continues until the instant τ_2 when the second treatment receives exactly m subjects. If $\sum_{i=1}^{\tau_1}T_i(k)=m$, then formally,

$$\tau_2=\min\left\{j:j>\tau_1,\max_{\ell:\ell\neq k}\sum_{i=1}^{j}T_i(\ell)\geqslant m\right\}.$$

After τ_2, the uniform allocation is performed among the remaining $K-2$ treatments with the uniform distribution, and so on, until the time τ_{K-1}, after which only one uncompleted treatment is left. This treatment receives then all the remaining subjects. Thus, for a fixed sequence $\tau_1 < \tau_2 < \cdots < \tau_{K-1}$, the allocation vectors T_i in the truncated multinomial design have the form: when $i \leqslant \tau_1$, T_i are K-nomial vectors with the uniform distribution on $1, \ldots, K$; for $\tau_1 < i \leqslant \tau_2$, T_i takes $K-1$ treatment values left incomplete after τ_1, each with probability $1/(K-1)$, (and the first completed treatment value with probability 0), and so on, for $\tau_r < i \leqslant \tau_{r+1}$, $r = 1, \ldots, K-1$, T_i takes the available treatment values each with probability $1/(K-r)$. It is convenient to put $\tau_K = n$.

The joint probability distribution of random variables $\tau_1, \ldots, \tau_{K-1}$ is given in Lemma 1 in the Appendix and the probability generating function of this vector is derived in Proposition 1 there. These results lead to the asymptotic distribution of $(\tau_1, \tau_2, \ldots, \tau_{K-1})$.

Theorem 1. *For $m \to \infty$, the distribution of $m^{-1/2}(\tau_1 - Km, \tau_2 - \tau_1, \ldots, \tau_{K-1} - \tau_{K-2})$ converges weakly to that of $(K(Z_{(1)} - \bar{Z}), (K-1)(Z_{(2)} - Z_{(1)}), \ldots, 2(Z_{(K-1)} - Z_{(K-2)}))$, where $Z_{(1)} < Z_{(2)} < \cdots < Z_{(K)}$ denote the order statistics of standard normal random sample, $Z_1, \ldots, Z_K$, of size K, and $\bar{Z} = (Z_1 + \cdots + Z_K)/K$.*

It follows that for $r = 1, \ldots, K-1$,

$$\frac{n-\tau_r}{\sqrt{m}} \sim \sum_{r+1}^{K} Z_{(j)} - (K-r)Z_{(r)} = \sum_{j=r}^{K} [Z_{(j)} - Z_{(r)}].$$

In particular,

$$n - \tau_1 \sim \sqrt{Km}(Z_{(K)} - \bar{Z}),$$

which follows from the fact that if $\eta_{\max}(N)$ is the maximal coordinate of a multinomial vector with K equiprobable classes in N independent trials, then

$$P(\tau_1 > N) = P(\eta_{\max}(N) < m).$$

The distribution of the extreme deviation $Z_{(K)} - \bar{Z}$ has been used for outlier detection. In the computer storage applications K can be a large number, so that the formula, $EZ_{(K)} = \sqrt{2 \log K} - \log \log K / 2\sqrt{2 \log K} + \mathrm{O}(1/\sqrt{\log K})$, can be used to approximate the average utility. Rukhin (2004) has demonstrated the relationship of the limiting distribution of τ_1 and the skew-normal distribution.

The *selection bias* was defined by Blackwell and Hodges (1957) to be the number of correct guesses of the experimenter who tries to assign subjects with the largest expected response to a particular treatment. The minimaxity of the truncated multinomial design proven by Blackwell and Hodges for $K = 2$ and the loss function, determined as the number of correct guesses, G_{TM}, is almost automatically extended to any K. In this case, in the notation of Theorem 1,

$$\begin{aligned} EG_{\mathrm{TM}} &= \frac{1}{K}E\tau_1 + \frac{1}{K-1}E(\tau_2 - \tau_1) + \cdots + \frac{1}{2}E(\tau_{K-1} - \tau_{K-2}) + E(n - \tau_{K-1}) \\ &= m + \sqrt{m}EZ_{(K)} + \mathrm{o}(\sqrt{m}) \end{aligned} \tag{2}$$

independently of the guessing strategy.

The random allocation rule is described by the assignment vectors S_i, $i = 1, \ldots, n$, with possible values being the unit basis vectors e_r, $r = 1, \ldots, K$, so that for any non-negative integer-valued vector $s = (s_1, \ldots, s_K)^{\mathrm{T}}$,

$$P\left(\sum_{i=1}^{j} S_i = s\right) = \frac{\binom{j}{s_1 \;\cdots\; s_K}\binom{n-j}{m-s_1 \;\cdots\; m-s_K}}{\binom{n}{m \;\cdots\; m}}, \tag{3}$$

if $\sum_k s_k = j$. Stirling's formula shows that as $m \to \infty$, this distribution converges to the multinomial distribution with $p_1 = \cdots = p_K = 1/K$,

$$P\left(\sum_{i=1}^{j} S_i = s\right) \to \begin{pmatrix} & j & \\ s_1 & \dots & s_K \end{pmatrix} \frac{1}{K^j}.$$

The conditional probability distribution of S_{j+1} has the form,

$$P\left(S_{j+1} = e_r \,\middle|\, \sum_{i=1}^{j} S_i = s\right) = \frac{m - s_r}{n - j}.$$

The natural (convergence) strategy under this design is to guess the treatment which has been previously least used. Thus, the experimenter predicts that the subject $j + 1$ will be allocated to any treatment r such that $\sum_{i=1}^{j} S_i(r) = \min_k \sum_{i=1}^{j} S_i(k)$. The convergence strategy maximizes the expected number of the correct guesses, EG_{RA}, and then,

$$EG_{\mathrm{RA}} = \sum_{j=0}^{n-1} \frac{m - E\min_k \sum_{i=1}^{j} S_i(k)}{n - j}.$$

Theorem 2. *For the random allocation rule under the convergence strategy*,

$$EG_{\mathrm{RA}} = m + \frac{\pi\sqrt{m}}{2} EZ_{(K)} + \mathrm{o}(\sqrt{m}).$$

The comparison of Theorem 2 and (2) shows that, as in the case $K = 2$, the truncated multinomial design has smaller selection bias than the random allocation rule. Namely, for any K, $(EG_{\mathrm{RA}} - m)/(EG_{\mathrm{TM}} - m) \sim \pi/2$. Diaconis and Graham (1981) have established a similar result for optimal guessing strategies in a sequential experiment with feedback. Proshan (1991) corrected Theorem 1 of that paper by deriving the asymptotic (Weibull) distribution of G_{RA} when $K = 2$.

For the random allocation rule, define the stopping time κ_1 exactly as in (1), namely,

$$\kappa_1 = \min\left\{ j : \max_{\ell} \sum_{i=1}^{j} S_i(\ell) \geqslant m \right\} \tag{4}$$

with S_i denoting the allocation vectors after this rule. Proposition 2 shows that the limiting behavior of κ_1 is quite different from that of τ_1.

Let $\left\{ {r \atop K} \right\}$ denote Stirling's number of the second kind, i.e. the number of partitions of an r element set into K non-empty subsets. It is well known that for all real x

$$x^r = \sum_i \left\{ {r \atop i} \right\} x(x-1)\dots(x-i+1).$$

Proposition 2. *With* κ_1 *defined by* (4) *for* $n \to \infty$,

$$n - \kappa_1 \to U,$$

where $U = U_K$ *is a discrete random variable with the distribution*

$$P(U = r) = \frac{K!}{K^{r+1}} \left\{ {r \atop K-1} \right\}, \quad r = K-1, K, \dots.$$

For $t < K/(K-1)$,

$$\sum_r t^r P(U=r) = \frac{(K-1)!t^{K-1}}{(K-t)(K-2t)\dots(K-(K-1)t)},$$

$$EU = K\sum_{j=1}^{K-1}\frac{1}{K-j},$$

and

$$\mathrm{Var}(U) = \sum_{j=1}^{K-1}\frac{j^2+j}{(K-j)^2}.$$

For example, when $K=2$, $\{{K \atop 2}\} = 2^{k-1}-1$, and $P(U<k) = 1-\frac{1}{2^{k-1}}$, which gives the geometric distribution on integers $1, 2, \dots$ with $p=\frac{1}{2}$.

For any K and $r = K, K+1, \dots,$

$$P(U<r) = \frac{K!}{K^r}\left\{{r \atop K}\right\} = \sum_{j=0}^{K}(-1)^j\binom{K}{j}\left(1-\frac{j}{K}\right)^r$$

is the probability that in r independent trials with K equally likely outcomes all K outcomes occur (or, in the problem of distribution of r balls in K cells, the probability that all cells are occupied, or that all coupons are collected in the coupons collection problem.) See Feller (1968, Sec. IV.2), for the further relationship of these probabilities to the classical occupancy problem.

3. Covariance matrices and accidental bias

We start this section with the form of the covariance matrix of the random allocation vectors T_i or S_i assigning subjects to treatments according to the truncated multinomial design or the random allocation rule. It turns out that the covariance structures of these two designs are quite different. Denote by $\mathbf{e} = \mathbf{e}_K$ the K-dimensional vector with unit coordinates, and let for $i = 1, \dots, n-1$,

$$H(i) = \frac{1}{K-1}\sum_{r=1}^{K-1}\frac{rP(\tau_r < i \leqslant \tau_{r+1})}{K-r} = \frac{K}{K-1}\sum_{r=1}^{K-1}\frac{P(\tau_r<i)}{(K-r)(K-r+1)}.$$

By Theorem 1,

$$H(n-\sqrt{m}x) \to K\sum_{r=1}^{K-1}\frac{P(\sum_{j=r}^{K}(Z_{(j)}-Z_{(r)})>x)}{(K-1)(K-r)(K-r+1)}.$$

Proposition 3. *For any* $i = 1, \dots, n$,

$$ET_i = \frac{1}{K}\mathbf{e}, \quad \mathrm{Var}(T_i) = \frac{1}{K}\left[\mathbf{I} - \frac{1}{K}\mathbf{e}\mathbf{e}^{\mathrm{T}}\right] = \Sigma_0, \tag{5}$$

and for $i \neq j$,

$$\mathrm{Cov}(T_i, T_j) = H(i\wedge j)\Sigma_0. \tag{6}$$

For the random allocation scheme vectors S_i, $i = 1, \dots, n$,

$$ES_i = \frac{1}{K}\mathbf{e}, \quad \mathrm{Var}(S_i) = \Sigma_0,$$

and for $i \neq j$,

$$\mathrm{Cov}(S_i, S_j) = -\frac{1}{n-1}\Sigma_0.$$

Let Σ_T denote the $Kn \times Kn$ joint covariance matrix of the vector T formed by the stacked vectors $T_1, \ldots, T_n$. Then Σ_T consists of the block matrices (6) as given in Proposition 3, so that $\Sigma_T = \mathbf{C} \otimes \Sigma_0$ with $n \times n$ matrix $\mathbf{C}$ having elements $(1-\delta_{ij})H(i \wedge j) + \delta_{ij}$.

For the random allocation rule, $\Sigma_S = \mathbf{B} \otimes \Sigma_0$, with $n \times n$ matrix $\mathbf{B} = (n-1)^{-1}[n\mathbf{I} - \mathbf{e}_n\mathbf{e}_n^{\mathrm{T}}]$.

Efron (1971) introduced the concept of *accidental bias* which is a measure of the expected bias of the treatment effect from a linear regression model if important covariates are ignored, and Steele (1980) has established a useful formula for this bias for the biased coin design. In our situation, let z be a Kn-dimensional vector of (deterministic) covariates left out of the model such that $z^{\mathrm{T}}(\mathbf{e} \otimes \cdots \otimes \mathbf{e}) = 0$ and $z^{\mathrm{T}}z = 1$. The degree of susceptibility to accidental bias is given by $E(z^{\mathrm{T}}T)^2 = z^{\mathrm{T}}\Sigma_T z$. Clearly, $z^{\mathrm{T}}\Sigma_T z$ cannot exceed the maximum eigenvalue $\lambda_{\max}(\Sigma_T)$.

Thus, define the accidental bias of an allocation rule with the covariance matrix Σ as

$$\rho = \max_{z: z^{\mathrm{T}}(\mathbf{e}\otimes\cdots\otimes\mathbf{e})=0} \frac{z^{\mathrm{T}}\Sigma z}{z^{\mathrm{T}}z}.$$

The eigenvalues of Σ_T are products of eigenvalues of the matrix $\mathbf{C}$ and eigenvalues of Σ_0. The same is true with regard to Σ_S and $\mathbf{B}$. As the eigenvalues of Σ_0 are $1/K$ or 0, $\lambda_{\max}(\Sigma_T) = \lambda_{\max}(\mathbf{C})/K$, and the same conclusion holds for Σ_S. Let $z = u \otimes v$ with n-dimensional vector u, and K-dimensional vector v such that $\Sigma_0 v = v$, $v^{\mathrm{T}}\mathbf{e} = 0$ (the subspace of such vectors has dimension $K-1$.) Then $z^{\mathrm{T}}(\mathbf{e} \otimes \cdots \otimes \mathbf{e}) = (u^{\mathrm{T}}\mathbf{e} \otimes \cdots \otimes \mathbf{e})(v^{\mathrm{T}}\mathbf{e}) = 0$. Therefore, for the truncated multinomial design,

$$\rho \geqslant \frac{(u\otimes v)^{\mathrm{T}}\mathbf{C}\otimes\Sigma_0(u\otimes v)}{(u\otimes v)^{\mathrm{T}}(u\otimes v)} = \frac{(u^{\mathrm{T}}\mathbf{C}u)(v^{\mathrm{T}}\Sigma_0 v)}{(uu^{\mathrm{T}})(v^{\mathrm{T}}v)} = \frac{u^{\mathrm{T}}\mathbf{C}u}{Kuu^{\mathrm{T}}}.$$

It follows that $\lambda_{\max}(\mathbf{C}) \leqslant \rho K$. As $\rho K \leqslant \lambda_{\max}(\Sigma_T)K = \lambda_{\max}(\mathbf{C})$, we see that $\rho = \lambda_{\max}(\mathbf{C})/K$, with a similar conclusion for the random allocation rule. The next result gives the asymptotic behavior of $\lambda_{\max}(\mathbf{C})$ and of $\lambda_{\max}(\mathbf{B})$.

Proposition 4. *The largest eigenvalue $\lambda_{\max}(\mathbf{C})$ has order $n^{1/2}$. More precisely, with the density ϕ_K, defined by* (17) *in the Appendix, for sufficiently large n,*

$$\frac{(K-1)\int_0^\infty\int_0^\infty \min[x,y]\phi_K(x)\phi_K(y)\,\mathrm{d}x\,\mathrm{d}y}{EZ_{(K)}} \leqslant \frac{\lambda_{\max}(\mathbf{C})\sqrt{K}}{\sqrt{n}} \leqslant \frac{KEZ_{(K)}}{K-1}.$$

The largest eigenvalue $\lambda_{\max}(\mathbf{B})$ of $\mathbf{B}$ *is*

$$\lambda_{\max}(\mathbf{B}) = \frac{n}{(n-1)K}.$$

If $Z_{(1)} < Z_{(2)} < \cdots < Z_{(K)}$, $Z^\star_{(1)} < Z^\star_{(2)} < \cdots < Z^\star_{(K)}$, denote the order statistics of two independent standard normal random samples each of size K, then in the lower bound in Proposition 4,

$$\begin{aligned}&\int_0^\infty\int_0^\infty \min[x,y]\phi_K(x)\phi_K(y)\,\mathrm{d}x\,\mathrm{d}y\\ &\quad= \sum_{j,\ell}\frac{K^2E\min[\sum_{i=j}^K(Z_{(i)}-Z_{(j)}),\sum_{i=\ell}^K(Z^\star_{(i)}-Z^\star_{(\ell)})]}{(K-1)^2(K-j)(K-\ell)(K-j+1)(K-\ell+1)}\\ &\quad\geqslant E\min[(Z_{(K)}-Z_{(K-1)}),(Z^\star_{(K)}-Z^\star_{(K-1)})].\end{aligned}$$

Thus, the truncated multinomial design can be susceptible to a high degree of accidental bias. In contrast, for the random allocation rule this bias is close to $1/K$ for large m. Proposition 4 extends Proposition 2 established for $K=2$ in Rosenberger and Rukhin (2003).

4. Asymptotic distribution of permutation tests

Because the allocation vectors are not independent, standard inference procedures may not be appropriate. A *randomization test* or *permutation test* (Lehmann, 1975, p. 43) assumes that the sequence of subject responses is deterministic and only the sequence of treatment assignments is random. If a_{jn} is a score coefficient of subject j out of n subjects, $\sum_1^n a_{jn}=0$, the basic form of the linear rank statistic employed in a permutation test is $\sum_1^n a_{jn}S_j$ (or $\sum_1^n a_{jn}T_j$). Typical score functions include the simple ranks or van der Waerden scores. The asymptotic distribution of such permutation tests have been derived for various specialized randomization schemes (Rosenberger, 1993; Smythe and Wei, 1983; Wei et al., 1986). Rosenberger and Lachin (2002), Section 7.5, discuss their use for the data from clinical trials when $K=2$.

The permutation test under a random allocation rule is given by a vector rank statistic, $\sum_{j=1}^n a_{jn}(S_j - K^{-1}\mathbf{e})$. Its asymptotic normality under a Lindeberg-type condition on the scores,

$$\max_{1\leqslant j\leqslant n} a_{jn}^2 \Big/ \sum_{j=1}^n a_{jn}^2 \to 0, \tag{7}$$

follows from Hajek et al. (1999, Section 6.1.5, Theorem 1). This is discussed when $K=2$ in Section 14.3 of Rosenberger and Lachin (2002).

According to Proposition 3, the covariance matrix of the permutation test statistic for the truncated multinomial design is

$$\operatorname{Var}\left(\sum_{j=1}^n a_{jn}\left(T_j - \frac{1}{K}\mathbf{e}\right)\right) = \left[\sum_j a_{jn}^2 + 2\sum_{i<j} a_{in}a_{jn}H(i)\right]\Sigma_0.$$

If for any $r=1,\ldots,K-1$,

$$\lim_{n\to\infty} \frac{\sum_{i<j} a_{in}a_{jn}P(\tau_r<i)}{\sum_{j=1}^n a_{jn}^2} = 0, \tag{8}$$

then the definition of $H(i)$ shows that

$$\operatorname{Var}\left(\sum_{j=1}^n a_{jn}\left(T_j - \frac{1}{K}\mathbf{e}\right)\right) \sim \sum_j a_{jn}^2\Sigma_0.$$

Assuming (8), we give now an extension of Proposition 2 in Rosenberger and Rukhin (2003), proven for $K=2$ under different conditions, for the permutation test statistic written in the form,

$$L_n = \left(\sum_j a_{jn}^2\right)^{-1/2} \sum_{j=1}^n a_{jn}\left(T_j - \frac{1}{K}\mathbf{e}\right). \tag{9}$$

Theorem 3. *Assume that when* $n \to \infty$, (7) *and* (8) *hold, and for a sequence* δ_n *such that* $\delta_n<n$ *and* $n^{-1/2}(n-\delta_n)\to\infty$,

$$\lim_{n\to\infty} \frac{\sum_{j\leqslant\delta_n} a_{jn}^2}{\sum_{j=1}^n a_{jn}^2} = 1. \tag{10}$$

Then the limiting distribution of L_n in (9) is K-dimensional normal distribution $N_K(0,\Sigma_0)$.

Theorem 3 justifies the approximate χ^2-distribution of the quadratic statistic, $L_n^{\mathrm{T}} L_n$, for testing the null hypothesis of equivalence of all K treatments. Indeed, $L_n^{\mathrm{T}} \Sigma_0^- L_n$ must have such a distribution with $K-1$ degrees of freedom, and one of the generalized inverses Σ_0^- is known to be Σ_0. As $\Sigma_0(T_j - K^{-1}\mathbf{e}) = T_j - K^{-1}\mathbf{e}$, the statistic, $L_n^{\mathrm{T}} L_n = L_n^{\mathrm{T}} \Sigma_0 L_n = L_n^{\mathrm{T}} \Sigma_0^- L_n$, is approximately χ^2-distributed. Wei et al. (1987) discuss further tests of this hypothesis.

5. Conclusions

In this paper we give formulas for the biases of two most common balanced randomization designs with several treatments. The main optimality property of the truncated multinomial design is minimization of selection bias, which in principle can arise only in unmasked studies. The random allocation rule leads to a "larger degree of uniformity" expressed by the fact that the instant, κ_1, when a treatment gets its prescribed quota, comes much later than the time τ_1 for the truncated multinomial design. Essentially for this reason, the random allocation rule allows easier application of standard permutation tests. Another reflection of the same fact is that according to Proposition 3 the covariance matrix of allocation vectors S_j is a negative multiple of Σ_0 while that of T_j is a positive multiple of Σ_0.

Theorem 1 and Proposition 2 provide a comparison of two balancing strategies for computers with distributed memory, and they clearly demonstrate the advantage in terms of disk utility of the random allocation method. For these reasons the random allocation rule, which is also optimal from the point of view of accidental bias, can be recommended, if its implementation is practical.

Acknowledgments

This research was supported by NSA Grant #MDA904-00-1-0033. The author is grateful to Bill Rosenberger for discussion and helpful comments.

Appendix A

A.1. Lemmas 1 and 2

Lemma 1. *The joint probability distribution of random variables $\tau_1, \ldots, \tau_{K-1}$ has the form,*

$$
\begin{aligned}
&P(\tau_1 = t_1, \tau_2 - \tau_1 = t_2, \ldots, \tau_{K-1} - \tau_{K-2} = t_{K-1}) \\
&\quad = \frac{K!}{K^{t_1}(K-1)^{t_2}\cdots 2^{t_{K-1}}} \sum \binom{t_1 - 1}{m-1 \quad v_2^{(1)} \quad \cdots \quad v_K^{(1)}} \\
&\qquad \cdots \binom{t_k - 1}{m - 1 - \sum_{j=1}^{k-1} v_k^{(j)} \quad v_{k+1}^{(k)} \quad \cdots \quad v_K^{(k)}} \\
&\qquad \cdots \binom{t_{K-1} - 1}{m - 1 - \sum_{j=1}^{K-2} v_{K-1}^{(j)} \quad v_K^{(K-1)}}.
\end{aligned} \tag{11}
$$

Summation in (11) *is over all indices $v_2^{(1)}, \ldots, v_K^{(1)}, v_3^{(2)}, \ldots, v_K^{(K-1)}$ such that for any $k = 2, \ldots, K$, $\sum_{j=1}^{k-1} v_k^{(j)} \leqslant m-1$.*

Proof. To prove (11) it suffices to notice that there are $K!$ ways in which K treatments can be ordered in the assignments completion. The probability of each such ordering is the sum of products of multinomial probabilities corresponding to the regimes between, say, τ_{k-1} and τ_k. Thus, the probability of each assignment sequence there equals $(K-k+1)^{-t_k}$, with

$$\begin{pmatrix} t_k - 1 \\ m-1-\sum_{j=1}^{k-1} v_k^{(j)} \quad v_{k+1}^{(k)} \quad \cdots \quad v_K^{(k)} \end{pmatrix}$$

choices of distributing $K-k+1$ treatments, so that the kth treatment gets all m subjects, and exactly $K-k$ treatments remain incomplete. □

Lemma 2. *With $p_{t_1\cdots t_{K-1}}$ denoting the probabilities in* (11),

$$\sum_{t_1\cdots t_{K-1}} \frac{(Ku_1)^{t_1-1}[(K-1)u_2]^{t_2-1}\cdots[2u_{K-1}]^{t_{K-1}-1}}{(t_1-1)!(t_2-1)!\cdots(t_{K-1}-1)!} p_{t_1\cdots t_{K-1}}$$
$$= \frac{[u_1(u_1+u_2)\ldots(u_1+\cdots+u_{K-1})]^{m-1}}{[\Gamma(m)]^{K-1}} \sum_{k=0}^{m-1} \frac{(u_1+\cdots+u_{K-1})^k}{k!}. \tag{12}$$

Proof. According to Lemma 1, the sum in the left-hand side of (12) can be written as

$$Q = \sum \frac{u_1^{m-1+\sum_{i=2}^K v_i^{(1)}} \ldots u_k^{m-1-\sum_{j=1}^{k-1} v_k^{(j)} + \sum_{i=k+1}^K v_i^{(k)}} \ldots u_{K-1}^{m-1-\sum_{j=1}^{K-2} v_{K-1}^{(j)} + v_K^{(K-1)}}}{(m-1)! v_2^{(1)}! \ldots v_K^{(1)}! \ldots (m-1-\sum_{j=1}^{K-2} v_{K-1}^{(j)})! v_K^{(K-1)}!},$$

where the range of summation is indicated in Lemma 1. By summing first over $v_2^{(1)}$, then over $v_3^{(1)}$, $v_3^{(2)}$, and so on, with $v_K^{(1)}, \ldots, v_K^{(K-1)}$ forming the last sum, we obtain

$$Q = \frac{u_1^{m-1}(u_1+u_2)^{m-1}\ldots(u_1+\cdots+u_{K-1})^{m-1}}{[\Gamma(m)]^{K-1}} F \sum_{v_K^{(1)}+\cdots+v_K^{(K-1)} \leqslant m-1} \frac{u_1^{v_K^{(1)}} \ldots u_{K-1}^{v_K^{(K-1)}}}{v_K^{(1)}! \ldots v_K^{(K-1)}!},$$

which is equivalent to (12). □

A.2. Proposition 1

Proposition 1. *For any positive $z_1, \ldots, z_{K-1}$*

$$E z_1^{\tau_1} z_2^{\tau_2-\tau_1} \cdots z_{K-1}^{\tau_{K-1}-\tau_{K-2}}$$
$$= \frac{K!}{[\Gamma(m)]^K} \int \cdots \int_{0<x_1<\cdots<x_{K-1}<x_K} (x_1 \cdots x_K)^{m-1}$$
$$\times \exp\left\{-\sum_{r=1}^K \frac{(K-r+1)(x_r - x_{r-1})}{z_r}\right\} \mathrm{d}x_1 \cdots \mathrm{d}x_K \tag{13}$$

with $x_0 = 0$ and $z_K = 1$.

Proof. Multiplying (12) by $\exp\{-\sum_{r=1}^{K-1}(K-r+1)u_r z_r^{-1}\}$ and integrating over positive $u_1,\ldots,u_{K-1}$, leads to (13) via transformation of variables, $x_r = u_1+\cdots+u_r$, and the use of the formula, $(m-1)!\sum_{\ell=0}^{m-1}\frac{x_{K-1}^{\ell}}{\ell!} = e^{x_{K-1}}\int_{x_{K-1}}^{\infty} e^{-x}x^{m-1}\,dx$. □

A.3. Proof of Theorem 1

According to Proposition 1,

$$
\begin{aligned}
&Ez_1^{m^{-1/2}(\tau_1-Km)} z_2^{m^{-1/2}(\tau_2-\tau_1)} \cdots z_{K-1}^{m^{-1/2}(\tau_{K-1}-\tau_{K-2})} \\
&\quad = \frac{K!\,z_1^{-Km^{1/2}}}{[\Gamma(m)]^K}\int\cdots\int_{0<x_1<\cdots<x_{K-1}<x_K}(x_1\cdots x_K)^{m-1} \\
&\qquad \times \exp\left\{-\sum_{r=1}^{K}(K-r+1)(x_r-x_{r-1})z_r^{-1/\sqrt{m}}\right\} dx_1\cdots dx_K.
\end{aligned}
$$

Make the transformation of variables $x_r = m+\sqrt{m}\,y_r$, $r=1,\ldots,K$, so that $x_r^{m-1} = m^{m-1}\exp\{\sqrt{m}\,y_r - y_r^2/2\}[1+|y_r|^3 O(m^{-1/2})]$, and observe that

$$
\begin{aligned}
&\sum_{r=2}^{K}(K-r+1)(x_r-x_{r-1})z_r^{-1/\sqrt{m}} \\
&\quad \sim \sqrt{m}\sum_{r=2}^{K}(K-r+1)(y_r-y_{r-1}) - \sum_{r=2}^{K}(K-r+1)(y_r-y_{r-1})\log z_r.
\end{aligned}
$$

For $r=1$,

$$
Kx_1 z_1^{-1/\sqrt{m}} = mK + \sqrt{m}\,Ky_1 - \sqrt{m}\,K\log z_1 - Ky_1\log z_1 + \frac{K}{2}(\log z_1)^2 + O(m^{-1/2}).
$$

Since the sum of terms in the exponent, with the factor $\sqrt{m}$, vanishes,

$$
\sum_{r=2}^{K}(K-r+1)(y_r-y_{r-1}) + Ky_1 - \sum_{r=1}^{K} y_r = 0,
$$

one obtains by using Stirling's formula for $\Gamma(m)$,

$$
\begin{aligned}
&Ez_1^{m^{-1/2}(\tau_1-Km)} z_2^{m^{-1/2}(\tau_2-\tau_1)} \cdots z_{K-1}^{m^{-1/2}(\tau_{K-1}-\tau_{K-2})} \\
&\quad \to \frac{K!\,e^{-K(\log z_1)^2/2}}{(2\pi)^{K/2}}\int\cdots\int_{-\infty<y_1<\cdots<y_{K-1}<y_K<\infty} \\
&\qquad \times \exp\left\{Ky_1\log z_1 + \sum_{r=2}^{K}(K-r+1)(y_r-y_{r-1})\log z_r - \sum_{1}^{K} y_r^2/2\right\} dy_1\cdots dy_K.
\end{aligned}
$$

To recognize this integral as the probability generating function of $(K(Z_{(1)}-\bar{Z}), (K-1)(Z_{(2)}-Z_{(1)}), \ldots, 2(Z_{(K-1)}-Z_{(K-2)}))$, notice that by independence of this vector and $\bar{Z}$,

$$\frac{K!}{(2\pi)^{K/2}}\int\cdots\int_{-\infty<y_1<\cdots<y_{K-1}<y_K<\infty}$$
$$\times \exp\left\{Ky_1\log z_1+\sum_{r=2}^{K}(K-r+1)(y_r-y_{r-1})\log z_r-\sum_1^K y_r^2/2\right\}\mathrm{d}y_1\cdots\mathrm{d}y_K$$
$$=E\exp\left\{KZ_{(1)}\log z_1+\sum_{r=2}^{K}(K-r+1)(Z_{(r)}-Z_{(r-1)})\log z_r\right\}$$
$$=\exp\{K(\log z_1)^2/2\}$$
$$\times E\exp\left\{K(Z_{(1)}-\bar{Z})\log z_1+\sum_{r=2}^{K}(K-r+1)(Z_{(r)}-Z_{(r-1)})\log z_r\right\}.$$

A.4. Lemma 3 and the Proof of Theorem 2

Theorem 2 is similar to Theorem 3 in Diaconis and Graham (1981). We give our, much shorter, version of the proof based on Lemma 3 below which may be of independent interest.

Let $\pi_{js}=P(\min_k\sum_{i=1}^{j}S_i(k)\geqslant s)$. To get the asymptotics of

$$EG_{\mathrm{RA}}=\sum_{j=0}^{n-1}\frac{m-E\min_k\sum_{i=1}^{j}S_i(k)}{n-j}=\sum_{s=1}^{m}\sum_{j=0}^{n-1}\frac{1-\pi_{js}}{n-j},$$

we derive the probability generating function for π_{js}.

Lemma 3. *For any fixed s, $s=1,\ldots,m$, and z, $0<z<1$,*

$$\sum_{j=0}^{n}\binom{n}{j}z^j(1-z)^{n-j}\pi_{sj}=\left[\sum_{i=s}^{m}\binom{m}{i}z^i(1-z)^{m-i}\right]^K=[I_z(s,m-s+1)]^K \tag{14}$$

and

$$\sum_{j=0}^{n-1}\frac{1-\pi_{js}}{n-j}=\int_0^1\frac{1-[I_z(s,m-s+1)]^K}{1-z}\,\mathrm{d}z \tag{15}$$

with $I_z(i,j)$ denoting the incomplete beta-function.

Proof. It follows from (3) that for a fixed s,

$$\sum_{j=0}^{n}\frac{z^j(1-z)^{n-j}\pi_{js}}{j!(n-j)!}=\frac{(m!)^K}{(mK)!}\sum_{s\leqslant i_1,\ldots,i_K\leqslant m}\frac{z^{i_1+\cdots+i_K}(1-z)^{Km-i_1-\cdots-i_K}}{i_1!\cdots i_K!(m-i_1)!\cdots(m-i_K)!}$$
$$=\frac{1}{n!}\left[\sum_{i=s}^{m}\binom{m}{i}z^i(1-z)^{m-i}\right]^K,$$

which implies (14).

As for $s\geqslant 1$, $\pi_{ns}=1$, by subtracting both sides of (14) from 1, dividing by $1-z$, and integrating over (0, 1), one obtains (15). □

Proof of Theorem 2. Lemma 3 shows that

$$EG_{\mathrm{RA}} = \sum_{s=1}^{m} \int_0^1 \frac{1-[I_z(s,m-s+1)]^K}{1-z}\,\mathrm{d}z = \int_0^1 \frac{m-\sum_{s=1}^{m}[I_z(s,m-s+1)]^K}{1-z}\,\mathrm{d}z.$$

Notice that with independent $Bin(m,z)$ random variables $B_1,\ldots,B_K$,

$$\sum_{s=1}^{m}[I_z(s,m-s+1)]^K = \sum_{s=1}^{m} P(B_1 \geqslant s,\ldots,B_K \geqslant s) = E\min(B_1,\ldots,B_K),$$

so that for $V_i = (B_i - mz)/\sqrt{mz(1-z)}$, $i=1,\ldots,K$,

$$EG_{\mathrm{RA}} - m = -\sqrt{m}\int_0^1 \sqrt{\frac{z}{1-z}}\,E\min(V_1,\ldots,V_K)\,\mathrm{d}z.$$

If $z(1-z) \geqslant \log m/m$, the normal approximation error bound implies that

$$E\min(V_1,\ldots,V_K) = EZ_{(1)} + \mathrm{O}\left(\frac{1}{\sqrt{\log m}}\right),$$

where the remainder term can be chosen to be independent of z.

When $x = m(1-z) < m(1/2-\sqrt{1/4-\log m/m})$, the known bound for the Poisson approximation shows that

$$E\min(V_1,\ldots,V_K) = \frac{(E\min(\Pi_1,\ldots,\Pi_K)-x)}{\sqrt{x}}\left[1+(x+0.25)\mathrm{O}\left(\frac{1}{m}\right)\right],$$

with Poisson, parameter x, independent random variables $\Pi_1,\ldots,\Pi_K$. A similar formula holds when $z < 1/2-\sqrt{1/4-\log m/m}$.

Thus,

$$EG_{\mathrm{RA}} - m = -\sqrt{m}\int_0^{m(1/2-\sqrt{1/4-\log m/m})} \sqrt{\frac{z}{1-z}}\,\mathrm{d}z\,EZ_{(1)} - \frac{2}{\sqrt{m}}\int_0^{m(1/2-\sqrt{1/4-\log m/m})}$$
$$\times \frac{[E\min(\Pi_1,\ldots,\Pi_K)-x]}{x}\,\mathrm{d}x + \mathrm{o}(\sqrt{m}) = \frac{\sqrt{m}\pi}{2}EZ_{(K)} + \mathrm{o}(\sqrt{m}). \qquad \square$$

A.5. Proof of Proposition 2

For any fixed k,

$$P(n-\kappa_1 < k) = P\left(\max_{\ell}\sum_{i=1}^{n-k} S_i(\ell) < m\right)$$
$$= \sum_{\substack{j_1+\cdots+j_K=k \\ j_1\geqslant 1,\ldots,j_K\geqslant 1}} \frac{\binom{k}{j_1\ \cdots\ j_K}\binom{n-k}{m-j_1\ \cdots\ m-j_K}}{\binom{n}{m\ \cdots\ m}}$$
$$\to \sum_{\substack{j_1+\cdots+j_K=k \\ j_1\geqslant 1,\ldots,j_K\geqslant 1}} \binom{k}{j_1\ \cdots\ j_K}\frac{1}{K^k} = f_k.$$

As in the proof of Lemma 3, for any z, $0<z<1$,

$$\sum_{j=0}^{n}\binom{n}{j}z^j(1-z)^{n-j}P(\kappa_1>j)=\left[\sum_{i=0}^{m-1}\binom{m}{i}z^i(1-z)^{m-i}\right]^K=(1-z^m)^K. \tag{16}$$

By putting $z=1-x/m$ and letting $m\to\infty$, one obtains,

$$\mathrm{e}^{-xK}\sum_{k=0}^{\infty}f_k\frac{(xK)^k}{k!}=(1-\mathrm{e}^{-x})^K.$$

Thus, for $k=0,1,\ldots$

$$f_k=\frac{1}{K^k}\sum_{j=0}^{K}(-1)^{K+j}\binom{K}{j}j^k=\frac{K!}{K^k}\left\{\begin{matrix}k\\K\end{matrix}\right\}.$$

Clearly, f_k is the probability that in k multinomial trials with K classes, outcomes from each class occur.

With U denoting the random variable determined by f_k, one has

$$P(U\geqslant r)=1-\frac{K!}{K^r}\left\{\begin{matrix}r\\K\end{matrix}\right\},\quad r=K,K+1,\ldots,$$

so that

$$P(U=r)=\frac{K!}{K^{r+1}}\left[\left\{\begin{matrix}r+1\\K\end{matrix}\right\}-K\left\{\begin{matrix}r\\K\end{matrix}\right\}\right]=\frac{K!}{K^{r+1}}\left\{\begin{matrix}r\\K-1\end{matrix}\right\}$$

for $r=K-1,K,\ldots$. The form of the probability generating function follows, and the formulas for the mean and the variance are immediate.

A.6. Proof of Proposition 3

For any i, $P(T_i(k)=1)=K^{-1}$, i.e. $ET_i=K^{-1}\mathbf{e}$, and, as is easy to see,

$$E(T_i-K^{-1}\mathbf{e})(T_i-K^{-1}\mathbf{e})^{\mathrm{T}}=K^{-1}[I-K^{-1}\mathbf{e}\mathbf{e}^{\mathrm{T}}].$$

We evaluate $ET_iT_j^{\mathrm{T}}$ for a given sequence $\tau_1<\tau_2<\cdots<\tau_{K-1}$. When $i<j$ and the kth treatment has been completed, then $T_i(k)=0$, and, necessarily, $T_j(k)=0$. Assume that $\tau_r<i\leqslant\tau_{r+1}$ and $\tau_s<j\leqslant\tau_{s+1}$, $r\leqslant s$. The matrix, $E(T_iT_j^{\mathrm{T}}|\tau_1,\tau_2,\ldots,\tau_{K-1})$, has the diagonal elements, $P(T_i(k)=T_j(k)=1|\tau_1,\tau_2,\ldots,\tau_{K-1})$ of the form

$$\frac{\binom{K-1}{r}\binom{K-r-1}{s-r}}{\binom{K}{r}\binom{K-r}{s-r}(K-r)(K-s)}=\frac{1}{K(K-r)}.$$

Indeed there are $\binom{K}{r}$ choices for coordinates of the vector T_i corresponding to the completed treatments, and these choices are to be combined with $\binom{K-r}{s-r}$ choices of treatments filled out between τ_r and τ_s (i.e. additional zero coordinates of the vector T_j.) These must be counted jointly with $K-r$ ways to choose a vacant treatment for T_i, and $K-s$ similar possibilities for T_j. Out of the obtained, equally likely outcomes, $\binom{K-1}{r}$ correspond to the choice of the completed treatments. Similarly, to choose $s-r$ additional zero coordinates in T_j, one has $\binom{K-r-1}{s-r}$ ways.

The formula for the off-diagonal elements is

$$\frac{\binom{K-2}{r}\binom{K-r-1}{s-r}}{\binom{K}{r}\binom{K-r}{s-r}(K-r)(K-s)} = \frac{K-r-1}{K(K-1)(K-r)}.$$

Remarkably, these entries do not depend on s, so that

$$\begin{aligned}&\mathrm{Cov}(T_i, T_j|\tau_r < i \leqslant \tau_{r+1})\\ &= \left[\frac{r}{K(K-1)(K-r)}\mathbf{I} + \frac{K-r-1}{K(K-1)(K-r)}\mathbf{ee}^{\mathrm{T}}\right] - \frac{1}{K^2}\mathbf{ee}^{\mathrm{T}},\end{aligned}$$

and (6) holds.

To complete the proof, we give the corresponding formulas for the random allocation rule. Clearly, $\mathrm{Var}(S_i) = K^{-1}\mathbf{I} - K^{-2}\mathbf{ee}^{\mathrm{T}}$. For $i \neq j$ and $k = 1, \ldots, K$,

$$P(S_i(k) = S_j(k) = 1) = \frac{\binom{n-2}{m-2 \quad m \quad \cdots \quad m \quad m}}{\binom{n}{m \quad \cdots \quad m}} = \frac{m(m-1)}{n(n-1)}.$$

When $k \neq \ell$,

$$P(S_i(k) = S_j(\ell) = 1) = \frac{\binom{n-2}{m-1 \quad m-1 \quad \cdots \quad m \quad m}}{\binom{n}{m \quad \cdots \quad m}} = \frac{m^2}{n(n-1)}.$$

Thus, the matrix, $ES_iS_j^{\mathrm{T}}$, formed by the elements, $P(S_i(k) = S_j(\ell) = 1)$, $k, \ell = 1, \ldots, K$, has the form

$$ES_iS_j^{\mathrm{T}} = -\frac{1}{K(n-1)}\mathbf{I} + \frac{m}{K(n-1)}\mathbf{ee}^{\mathrm{T}},$$

and the last formula of Proposition 3 follows.

A.7. Lemma 4

Define the weights,

$$w_r = \frac{K}{(K-1)(K-r)(K-r+1)}, \quad r = 1, \ldots, K-1,$$

(which add up to one). Then for $i = 1, \ldots, n-1$,

$$H(i) = \sum_{r=1}^{K-1} w_r P(\tau_r < i).$$

With the order statistics of a standard normal random sample, $Z_{(1)} < Z_{(2)} < \cdots < Z_{(K)}$, let for $r = 1, \ldots, K-1$, $\phi_{rK}(t)$, $t > 0$, be the density of $\sum_{j=r}^{K}(Z_{(j)} - Z_{(r)})$. The mixture of these densities with weights w_k will be denoted by

$$\phi_K(t) = \sum_{r=1}^{K-1} w_r \phi_{rK}(t). \tag{17}$$

Lemma 4. *Assume that* $u_i \sim \frac{1}{\sqrt{m}}\varphi(\frac{n-i}{\sqrt{m}})$ *with an integrable function* φ. *Then for any positive* x,

$$\sum_{1\leqslant j\leqslant n-\sqrt{m}x} u_j H(j) \sim \int_x^\infty \varphi(t)\int_t^\infty \phi_K(y)\,\mathrm{d}y\,\mathrm{d}t.$$

Proof of Lemma 4 is based on the fact that

$$H(n-\sqrt{m}x) \to \int_x^\infty \phi_K(t)\,\mathrm{d}t.$$

A.8. Proof of Proposition 4

The matrix $\mathbf{C}$ has the block structure,

$$\begin{pmatrix} I & O \\ O^T & \mathbf{C}_1 \end{pmatrix},$$

where $\mathbf{C}_1$ is an $(n-m)\times(n-m)$ matrix with elements $(1-\delta_{ij})H(i\wedge j)+\delta_{ij}$ and O is $m\times(n-m)$ zero matrix. Clearly, $\lambda_{\max}(\mathbf{C}) = \max[\lambda_{\max}(\mathbf{C}_1), 1]$, and by the known properties of the Perron–Frobenius eigenvalue of a positive matrix for any vector $u=(u_{m+1}\ldots,u_n)^{\mathrm{T}}$ with positive coordinates,

$$\min_i \frac{\sum_{j=m+1}^{i-1} H(j)u_j + H(i)\sum_{j=i+1}^n u_j}{u_i} \leqslant \lambda_{\max}(\mathbf{C}_1) - 1 \leqslant \max_i \frac{\sum_{j=m+1}^{i-1} H(j)u_j + H(i)\sum_{j=i+1}^n u_j}{u_i}.$$

For any m, the corresponding eigenvector, $u=(u_{m+1}\ldots,u_n)^{\mathrm{T}}$, with $\sum_j u_j = 1$, can be found from the condition,

$$\sum_{j=m+1}^{i-1} H(j)u_j + H(i)\left(1-\sum_{j=m+1}^{i-1} u_j\right) = (\lambda_{\max}(\mathbf{C}_1)+H(i))u_i, \quad i=m+1,\ldots,n.$$

To find its behavior when $m\to\infty$, take the sequence u_j to have the form from Lemma 4, i.e. $u_i \sim \frac{1}{\sqrt{m}}\varphi(\frac{n-i}{\sqrt{m}})$ with a positive function φ, $\int_0^\infty \varphi(t)\,\mathrm{d}t = 1$. According to this Lemma for sufficiently large m with $\mu\sqrt{m} = \lambda_{\max}(\mathbf{C}_1)$,

$$\min_x \frac{\int_x^\infty \varphi(t)\int_t^\infty \phi_K(y)\,\mathrm{d}y\,\mathrm{d}t + \int_x^\infty \phi_K(y)\,\mathrm{d}y\int_0^x \varphi(t)\,\mathrm{d}t}{\varphi(x)}$$
$$= \min_x \frac{\int_x^\infty \phi_K(y)\int_0^y \varphi(t)\,\mathrm{d}t\,\mathrm{d}y}{\varphi(x)} \leqslant \mu \leqslant \max_x \frac{\int_x^\infty \phi_K(y)\int_0^y \varphi(t)\,\mathrm{d}t\,\mathrm{d}y}{\varphi(x)}.$$

It follows that the corresponding eigenfunction φ_0, $\int_0^\infty \varphi_0(t)\,\mathrm{d}t = 1$, satisfies the equation

$$\mu\varphi_0(x) = \int_x^\infty \phi_K(y)\int_0^y \varphi_0(t)\,\mathrm{d}t\,\mathrm{d}y,$$

so that

$$\mu\varphi_0(x) \leqslant \int_x^\infty \phi_K(y)\,\mathrm{d}y.$$

Integration over x shows that

$$\mu \leqslant \int_0^\infty \int_x^\infty \phi_K(y)\,\mathrm{d}y\,\mathrm{d}x = \int_0^\infty y\phi_K(y)\,\mathrm{d}y,$$

which proves the second inequality of Proposition 4. Indeed,

$$\int_0^\infty y\phi_K(y)\,dy = \sum_{1\leqslant r\leqslant j\leqslant K} w_r E(Z_{(j)} - Z_{(r)})$$

$$= EZ_{(K)} + \sum_{j=1}^{K-1} EZ_{(j)}\left[\sum_{r=1}^{j} w_r - (K-j+1)w_j\right]$$

$$= EZ_{(K)} - \frac{1}{K-1}\sum_{j=1}^{K-1} EZ_{(j)} = \frac{KEZ_{(K)}}{K-1}.$$

The first inequality follows by putting

$$\varphi(x) = \hat\varphi(x) = \frac{\int_x^\infty \phi_K(y)\,dy}{\int_0^\infty\int_x^\infty \phi_K(y)\,dy\,dx} = \frac{\int_x^\infty \phi_K(y)\,dy}{\int_0^\infty y\phi_K(y)\,dy}.$$

For a fixed x, one has

$$\int_x^\infty \phi_K(y)\int_0^y \hat\varphi(t)\,dt\,dy = \left[\hat\varphi(x)\int_0^x \hat\varphi(y)\,dy + \int_x^\infty \hat\varphi^2(y)\,dy\right]\int_0^\infty y\phi_K(y)\,dy.$$

Notice that $\int_x^\infty \phi_K(y)\int_0^y \hat\varphi(t)\,dt\,dy/\hat\varphi(x)$ is an increasing function of positive x . Indeed, its derivative has the form $-\hat\varphi^{\mathrm{T}}(x)\int_x^\infty \hat\varphi^2(y)\,dy/\hat\varphi^2(x) > 0$. Therefore,

$$\mu \geqslant \frac{\int_0^\infty \hat\varphi^2(x)\,dx}{\hat\varphi(0)} = \int_0^\infty x\phi_K(x)\,dx\int_0^\infty \hat\varphi^2(x)\,dx,$$

and

$$\int_0^\infty \hat\varphi^2(x)\,dx = \frac{\int_0^\infty\int_0^\infty \min[x,y]\phi_K(x)\phi_K(y)\,dx\,dy}{[\int_0^\infty y\phi_K(y)\,dy]^2},$$

which concludes the proof of Proposition 4.

The numerical evaluation of the bounds of Proposition 4 shows that for $K = 2$, $0.663 \leqslant \frac{\lambda_{\max}(\mathbf{C})}{\sqrt{m}} \leqslant 0.727$, and when $K = 3$, $0.524 \leqslant \frac{\lambda_{\max}(\mathbf{C})}{\sqrt{m}} \leqslant 0.853$. The quality of these bounds deteriorates as K increases.

If the bounded integral operator $\mathscr{K}$ is determined by the kernel $K(x, y) = \phi_K(x \wedge y)$, then μ from the proof of Proposition 4 is merely the largest eigenvalue of this operator, and the second inequality there is an immediate corollary of this fact.

A.9. Proof of Theorem 3

Put

$$\left[\sum_j a_{jn}^2\right]^{1/2} L_n = W_1 + W_2$$

$$= \sum_{j=1}^{\tau_1} a_{jn}\left(T_j - \frac{1}{K}\mathbf{e}\right) + \sum_{j=\tau_1+1}^{n} a_{jn}\left(T_j - \frac{1}{K}\mathbf{e}\right).$$

An application of Proposition 3 shows that

$$\mathrm{Var}(W_2) = \left[\sum_j a_{jn}^2 P(\tau_1 < j) + 2 \sum_{i<j} a_{in} a_{jn} H(i) \right] \Sigma_0,$$

and

$$\mathrm{Var}(W_1) = \sum_j a_{jn}^2 P(\tau_1 \geqslant j) \Sigma_0.$$

In particular,

$$\mathrm{Var}(W) = \mathrm{Var}(W_1) + \mathrm{Var}(W_2),$$

and sums W_1 and W_2 are non-correlated. Condition (10) implies that

$$\mathrm{Var}(W_1)/\mathrm{Var}(W) \to 1.$$

Indeed for any positive ε for all sufficiently large n, $P(\tau_1 \geqslant \delta_n) \geqslant 1 - \varepsilon$, and

$$\sum_j a_{jn}^2 P(\tau_1 \geqslant j) \geqslant P(\tau_1 \geqslant \delta_n) \sum_{j \leqslant \delta_n} a_{jn}^2 \geqslant (1-\varepsilon) \sum_{j \leqslant \delta_n} a_{jn}^2.$$

Thus, $[\sum_j a_{jn}^2]^{-1/2} W_2 \to 0$ in probability, since $[\sum_j a_{jn}^2]^{-1} \mathrm{Var}(W_2) \to 0$. Therefore, L_n and the rescaled version of W_1 must have the same limiting distribution.

Let T_j', $j = 1, \ldots, n$, denote independent and identically distributed multinomial vectors. Under condition (7), the statistic L_n' with T_j replaced by T_j' has asymptotic normal distribution, whereas under condition (10),

$$\left[\sum_j a_{jn}^2 \right]^{-1} \mathrm{Var} \left(\sum_{j=\tau_1+1}^{n} a_{jn} \left(T_j' - \frac{1}{K} \mathbf{e} \right) \right) = \left[\sum_j a_{jn}^2 \right]^{-1} \sum_j a_{jn}^2 P(\tau_1 > j) \Sigma_0 \to 0.$$

It follows that the rescaled version of W_1, whose distribution coincides with that of $\sum_{j \leqslant \tau_1} a_{jn} (T_j' - \frac{1}{K} \mathbf{e})$, has the same limiting normal distribution.

One can replace condition (8) by

$$\max_{k:1 \leqslant k \leqslant n} \frac{[\sum_{j=1}^{k} a_{jn}]^2}{\sum_j a_{jn}^2} \to 0. \tag{18}$$

Indeed (10) shows that

$$\frac{\sum_i a_{in}^2 P(\tau_1 < i)}{\sum_j a_{jn}^2} \to 0,$$

and (8) holds if and only if

$$\frac{\sum_{i,j} a_{in} a_{jn} P(\tau_1 < i \wedge j)}{\sum_j a_{jn}^2} \to 0.$$

It follows that (8) is valid if

$$\frac{\sum_k P(\tau_1 = k)[\sum_{j \geqslant k} a_{jn}]^2}{\sum_j a_{jn}^2} \to 0,$$

and this is implied by (18). Actually it suffices to take the maximum in (18) only over the region, $k \geqslant \delta_n$.

References

Blackwell, D., Hodges Jr., J.L., 1957. Design for the control of selection bias. Ann. Math. Statist. 28, 449–460.

Colditz, G.A., Miller, A., Mosteller, F., 1989. How study design affects outcomes in comparison of therapy. I: medical. Statist. Med. 8, 441–454.

Diaconis, P., Graham, R.L., 1981. The analysis of sequential experiments with feedback to subjects. Ann. Statist. 9, 3–23.

Diekman, R., Preis, R., 1999. Load balancing strategies for distributed memory machines. In: Topping, B.H.V. (Ed.), Parallel and Distributed Processing for Computational Mechanics: Systems and Tools. Saxe-Coburg Publications, Edinburgh, UK. (Chapter 7).

Efron, B., 1971. Forcing a sequential experiment to be balanced. Biometrika 58, 403–417.

Feller, W., 1968. An introduction to Probability Theory and Its Applications, vol. 1. Wiley, New York.

Hajek, J., Sidak, Z., Sen, P.K., 1999. Theory of Rank Tests. Academic Press, San Diego.

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

Pocock, S.J., Simon, R., 1975. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. Biometrics 31, 103–115.

Proshan, M., 1991. A note on Blackwell and Hodges (1957) and Diaconis and Graham (1981). Ann. Statist. 19, 1106–1109.

Rosenberger, W.F., 1993. Asymptotic inference with response-adaptive treatment allocation designs. Ann. Statist. 21, 2098–2107.

Rosenberger, W.F., Lachin, J.M., 2002. Randomization in Clinical Trials: Theory and Practice. Wiley, New York.

Rosenberger, W.F., Rukhin, A.L., 2003. Bias properties and nonparametric inference for truncated binomial randomization. J. Nonparametric Statist. 15, 455–465.

Rukhin, A.L., 2004. Limiting distributions in sequential occupancy problems. Sequential Anal. 23, 141–158.

Sedransk, N., 1973. Allocation of sequentially available units to treatment groups. Proceedings of the 39th International Statistical Institute, vol. 2, pp. 393–400.

Smythe, R.T., Wei, L.J., 1983. Significance tests with balanced randomization design. Biometrika 70, 496–500.

Steele, M., 1980. Efron's conjecture on vulnerability to bias in a method for balancing sequential trials. Biometrika 67, 503–504.

Stigler, S.M., 1969. The use of random allocation for the control of selection bias. Biometrika 56, 553–560.

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., Smith, R.L., 1986. K-treatment comparisons with restricted randomization rules in clinical trials. Ann. Statist. 14, 265–274.