The Annals of Statistics
1986, Vol. 14, No. 1, 265–274

K-TREATMENT COMPARISONS WITH RESTRICTED RANDOMIZATION RULES IN CLINICAL TRIALS

By L. J. Wei[1], R. T. Smythe and R. L. Smith

George Washington University, George Washington University and Imperial College, London

In the course of conducting a clinical trial to compare K (≥ 2) treatments, it is often desirable to balance the trial with respect to the assignments of patients to treatments. On the other hand, some form of randomization of treatment assignments is essential for reducing experimental bias. In this article, the large-sample approximation to the null distribution of K-sample randomization tests generated from a broad class of restricted randomization rules is derived. The implication of this result for conditional inference is also discussed.

1. Introduction. In most comparative clinical trials patients become available one at a time for treatment and must be assigned to a treatment group upon arrival. One of the most fundamental statistical issues in the evaluation of new treatments is how to allocate patients to treatment groups during the course of the trial. Avoidance of experimental bias may best be achieved by adopting some random mechanism for patient assignments. The most straightforward kind of randomization scheme, simple (or complete) randomization, assigns each patient with probability K^{-1} to one of K possible treatments, assignments being made independently of one another. Simple randomization has the advantage that each treatment assignment is completely unpredictable, and it provides a basis for statistical inference. However, in small-to-moderate sized experiments, simple randomization may result in severe imbalance among the numbers of patients in the treatment groups. Pocock (1979, p. 188) recommends the simple randomization scheme only in large trials with over 200 patients. Even then, if one analyzes early results while the trial is in progress, the scheme may not be satisfactory.

An alternative to complete randomization is the use of a restricted randomization rule to ensure comparability of treatment numbers during the course of the trial. In this article we focus on a class of adaptive treatment assignment rules. The term "adaptive" here indicates that the treatment assignment of the $(n + 1)$st patient may depend upon the assignments of the first n patients, but not upon the observed responses (cf. Simon, 1977). Now, suppose that at the end of the trial we are interested in testing the hypothesis H_0 that there is no difference among K (≥ 2) treatment groups. Suppose further that it is inappropriate to postulate that patients in the trial have been obtained by random sampling from a certain population. Then, experimental randomization of treatments to patients is the basis for inference (see Lehmann, 1975, Chapter 1). The

Received May 1984; revised July 1985.

[1] Research supported by the National Cancer Institute under Grant 1R01-CA 34428.

AMS 1980 *subject classifications.* Primary 62L05; secondary 60K30.

Key words and phrases. Biased coin design, conditional inference, martingale central limit theorem, randomization model, urn design.

significance tests used for testing H_0 must be only those tests generated by the experimental randomization design actually employed. For small trials, the randomization tests may be performed by computer simulation (cf. Simon, 1979, page 508). For large trials, however, this procedure becomes unwieldy.

In the comparison of two treatments, a large-sample approximation to the null randomization distribution of test statistics for testing H_0 has been obtained by Smythe and Wei (1983) for a particular class of adaptive designs. In this paper we study the asymptotic null distribution of test statistics for K-treatment comparisons under the randomization model, when treatment assignments are made by restricted randomization schemes described, with examples, in Section 2. The main limit theorem, which provides the asymptotic null randomization distribution of the K-sample test statistics, is presented in Section 3. In Section 4 we pursue a suggestion of Cox (1982) concerning conditional randomization tests and present a conjecture regarding convergence of conditional distributions. For continuity of presentation, the proof of the main theorem is deferred to the Appendix.

2. Restricted randomization rules. Efron (1971) proposed the "biased coin" design in the two-treatment case as a compromise between a perfectly balanced assignment scheme and a completely randomized one. This was generalized by Wei (1978a) to an "adaptive biased coin" design. Adaptive procedures for $K > 2$ treatments have been proposed by Wei (1978b), Efron (1980), Atkinson (1982), and Smith (1984a). Asymptotically, Wei's procedure balances the experiment completely, i.e., in the limit, each of the K treatments receives a portion $1/K$ of patients. Atkinson's procedure is based on the concepts of D and D_A optimality and allows different limiting proportions of patients assigned to treatment groups. Smith's procedure, which we will consider in this article, generalized Wei's scheme to achieve prespecified limiting proportions $\xi_1, \dots, \xi_K$, where $\xi_j > 0$, $1 \le j \le K$, and $\Sigma_{l=1}^{K} \xi_l = 1$.

Smith's procedure works in the following way. Suppose that after i assignments ($i \ge 1$) there are N_{ji} patients in treatment group j, where $1 \le j \le K$ and $\Sigma_{j=1}^{K} N_{ji} = i$. Let $\mathbf{p} = (p_1, p_2, \dots, p_{K-1})'$ be a $(K-1) \times 1$ vector whose jth component is the probability that treatment j will be assigned to the $(i+1)$st patient. We regard $\mathbf{p}$ as a function from Ω to Ω, where $\Omega = \{\mathbf{y} = (y_1, \dots, y_{K-1})'$: $y_i \ge 0$, $\Sigma_{l=1}^{K-1} y_l < 1\}$. This function depends on the assignments of the first i patients through the vector $i^{-1}\mathbf{N}_i$, where $\mathbf{N}_i = (N_{1i}, \dots, N_{K-1,i})$. Given $\mathbf{y}$, let $p_K(\mathbf{y}) \equiv 1 - \Sigma_{l=1}^{K-1} p_l(\mathbf{y})$ and $y_K \equiv 1 - \Sigma_{l=1}^{K-1} y_l$. Then $p_1, \dots, p_K$ are assumed to satisfy:

$$\text{(A.1)} \qquad \text{If} \quad y_j \ge \xi_j \quad \text{then} \quad p_j(\mathbf{y}) \le \xi_j, \qquad 1 \le j \le K.$$

(This formulation differs from the one in Smith (1984a), but is essentially equivalent to it.) Here are some examples of treatment assignment rules satisfying (A.1):

EXAMPLE 2.1. The adaptive biased coin design (Wei, 1978a): Let $K = 2$ and let p_1 be a continuous nonincreasing function from $[0,1]$ into $[0,1]$. Let ξ_1 be the unique value satisfying $p_1(\xi_1) = \xi_1$; then $\mathbf{p} = p_1$ satisfies (A.1). If p_1 is not assumed continuous, but $p_1(x)$ is $\geq \frac{1}{2}$, $= \frac{1}{2}$, or $\leq \frac{1}{2}$ according as $x < \frac{1}{2}$, $x = \frac{1}{2}$, or $x > \frac{1}{2}$, then p_1 satisfies (A.1) with $\xi_1 = \xi_2 = \frac{1}{2}$. Efron's (1971) biased coin model is of this type: $p_1(x) = \lambda$ for $x < \frac{1}{2}$, $p_1(\frac{1}{2}) = \frac{1}{2}$, and $p_1(x) = 1 - \lambda$ for $x > \frac{1}{2}$, where $\frac{1}{2} < \lambda < 1$.

EXAMPLE 2.2. Simple randomization: Take $\xi_j = K^{-1}$ and $p_j(\mathbf{y}) = K^{-1}$, $1 \leq j \leq K - 1$.

EXAMPLE 2.3. Generalized urn design (Wei, 1978b): Let $\xi_j = K^{-1}$ and $p_j(\mathbf{y}) = (1 - y_j)/(K - 1)$, $1 \leq j \leq K - 1$.

Asymmetric designs can easily be constructed by modifying this example: If, say, $K = 3$, $\xi_1 = \frac{1}{3}$, $\xi_2 = \frac{2}{9}$, and $\xi_3 = \frac{4}{9}$, let $p_1(\mathbf{y}) = (1 - y_1)/2$ and $p_2(\mathbf{y}) = \frac{1}{3} - \frac{1}{2}y_2$.

EXAMPLE 2.4. Atkinson's (1982) design: If we again take $\xi_j = K^{-1}, 1 \leq j \leq K$, this design gives

$$p_j(\mathbf{y}) = \left(y_j^{-1} - 1\right) \Big/ \sum_{l=1}^{K} \left(y_l^{-1} - 1\right), \qquad 1 \leq j \leq K - 1.$$

Smith (1984a) showed that (A1) holds in this case.

The following proposition is crucial for the proo' of the main theorem in the next section. It shows that the desired proportions a.e achieved in the limit, for a continuous $\mathbf{p}$ satisfying (A1).

PROPOSITION 2.5. *Let* $\mathbf{p}$ *satisfy* (A.1). *If each* p_j *is continuous, then* $p_j(i^{-1}\mathbf{N}_i) \to_p \xi_j$ *as* $i \uparrow \infty$, *for* $1 \leq j \leq K$.

PROOF. By (A.1), $p_j(\boldsymbol{\xi}) \leq \xi_j$, $1 \leq j \leq K$. If any of these inequalities were strict, we would have, summing both sides from $j = 1$ to $j = K$, that $1 < 1$; hence $p_j(\boldsymbol{\xi}) = \xi_j$, $1 \leq j \leq K$. Lemma 4 of Smith (1984a) gives that $i^{-1}N_{ji} \to_p \xi_j$ for $1 \leq j \leq K - 1$; the result now follows from Slutsky's theorem. □

3. Randomization tests based on restricted randomization. Suppose patients have been assigned to treatments by a restricted randomization rule and at the end of the trial $\{x_1, \dots, x_n\}$ is a sequence of observed responses. Let the corresponding scores of the x's be denoted by $a_{1n}, \dots, a_{nn}$, where a_{in} may be, e.g., a function of the rank of x_i among all x's. Furthermore, let T_{ji} be 1 if the ith patient is assigned to treatment j and 0, otherwise ($j = 1, 2, \dots, K$). We are interested in testing the hypothesis H_0 that there is no difference among the K treatments. In this section the large-sample approximation to the null distribution of K-sample randomization test statistics will be derived under a randomization model.

If simple randomization of patients to treatments is employed (so that $\xi_j = K^{-1}$ for each j) it is known that the following condition on the scores $\{a_{in}\}$ implies asymptotic joint normality (after standardization) of the vector $\langle \sum_{i=1}^{n} a_{in}(T_{ji} - K^{-1})\rangle_{j=1}^{K-1}$:

$$\lim_{n\to\infty} \max_{1\le i\le n} a_{in}^2 / \sum_{i-1}^{n} a_{in}^2 = 0 \tag{3.1}$$

(cf. Theorem 5 of the Appendix of Lehmann (1975)).

We will consider the class of designs satisfying (A.1) and the following additional condition:

(A.2) $\mathbf{p}$ is twice continuously differentiable with bounded second derivatives.

The next theorem will show that condition (3.1) is also sufficient for the asymptotic normality of the standardized $\sum a_{in}T_{ji}$ resulting from designs satisfying both (A.1) and (A.2). First, a lemma and some notation are needed:

LEMMA 3.2. *Let $d_{kj} \equiv \partial p_k(\boldsymbol{\xi})/\partial y_j$. Then $d_{kj} = \delta_{kj}\gamma$, $1 \le k, j \le K-1$, where δ_{kj} is the Kronecker delta and γ is a constant.*

PROOF. The argument is similar to that made in Section 3 of Smith (1984a); our γ and Smith's ρ are related by $\gamma = -\rho$. Recall (Proposition 2.5) that $p_k(\boldsymbol{\xi}) = \xi_k$. For δ sufficiently small, either positive or negative, and $k \ne j$,

$$p_k(\xi_1,\ldots,\xi_{j-1},\xi_j+\delta,\xi_{j+1},\ldots,\xi_{K-1}) \le \xi_k = p_k(\boldsymbol{\xi}). \tag{3.3}$$

Thus $\partial p_k(\boldsymbol{\xi})/\partial y_j = d_{kj} = 0$ if $k \ne j$. Next, observe that for sufficiently small δ,

$$\xi_K \ge p_K(\xi_1+\delta,\xi_2-\delta,\xi_3,\ldots,\xi_{K-1}) = \xi_K + \delta d_{K1} - \delta d_{K2} + o(\delta).$$

This implies that $d_{K1} = d_{K2}$, and similarly that all d_{Kj}, $1 \le j \le K-1$, have the same value. But since $\sum_{j=1}^{K} p_j(\mathbf{y}) \equiv 1$, it follows that $\sum_{j=1}^{K} d_{kj} = 0$, $1 \le k \le K-1$, and hence that $d_{jj} = -d_{Kj} = \gamma$ for $1 \le j \le K-1$. (From (3.3) it is evident that $\gamma \le 0$.) □

Now define the sequence of modified scores $\{b_{in}\}$ as follows:

$$b_{in} = a_{in} + \gamma \sum_{l=i+1}^{n} \left[\frac{a_{ln}}{l-1} \prod_{j=i}^{l-2}(1+\gamma/j)\right], \qquad i = 1,2,\ldots,n, \tag{3.4}$$

where by convention $\prod_{j=i}^{l} \equiv 1$ if $l < i$, and let $s_n^2 \equiv \sum_{i=1}^{n} b_{in}^2$. Define

$$W_{jn} \equiv \sum_{i=1}^{n} \frac{a_{in}}{s_n}(T_{ji} - \xi_j), \qquad 1 \le j \le K-1.$$

THEOREM 3.5. *Suppose that the sequence $\langle T_{ji}\rangle_{j=1}^{K}$, $i = 1,\ldots,n$ is generated by a design satisfying conditions* (A.1) *and* (A.2). *If* (3.1) *holds, then the random vector $\mathbf{W}_n' = \langle W_{jn}\rangle_{j=1}^{K}$ converges in distribution to a multivariate normal distribution with mean* 0 *and covariance matrix $\boldsymbol{\Sigma} = (\sigma_{kj})$, where $\sigma_{jj} = \xi_j(1-\xi_j)$ and $\sigma_{kj} = -\xi_k\xi_j$, $k \ne j$.*

PROOF. See the appendix. □

A natural statistic for testing H_0 is then $T_n = \mathbf{W}_n'\Sigma^{-1}\mathbf{W}_n$, which has an asymptotic χ^2_{K-1} distribution. (Note that Σ is simply the covariance matrix of a multinomial random vector with cell probabilities $\xi_1, \dots, \xi_{K-1}$.)

We conclude this section with several remarks concerning Theorem 3.5.

REMARK 1. Examples 2.2–2.4 of Section 2 satisfy condition (A.2). For Efron's biased coin model, where $\mathbf{p}$ is not even continuous (cf. Example 2.1), asymptotic normality may fail to hold (cf. Smythe and Wei, 1983). However, restricted randomization designs with the balancing action expressed by conditions different from (A.1) are certainly possible. One seemingly natural alternative is given by

$$\text{(A.1}'\text{)} \qquad y_i \geq \xi_i \quad \text{if and only if} \quad p_i(\mathbf{y}) \leq \xi_i, \qquad 1 \leq i \leq K-1.$$

Here is an example satisfying (A.1′) but not (A.1):

EXAMPLE 3.6. Let $K = 3$, $\xi_1 = \frac{1}{5}$, $\xi_2 = \frac{1}{3}$, $\xi_3 = \frac{7}{15}$, and let $p_1(\mathbf{y}) = \frac{1}{4}(1 - y_1)$, $p_2(\mathbf{y}) = \frac{3}{8}(1 - y_2^2)$. Theorem 3.5 holds for this design. On the other hand, as noted by Smith (1984a), Atkinson's procedure (Example 2.4) satisfies (A.1) but not (A.1′). For designs satisfying (A.1′) but not (A.1), Lemma 3.2 need not hold, so that Theorem 3.5 is not always applicable.

REMARK 2. A real question of interest in inference for these designs concerns the potential loss of accuracy incurred if a significance test is performed assuming simple randomization, when in fact a restricted randomization design has been used. For the designs considered in Theorem 3.5, the accuracy of such an approximation clearly depends on the ratio $\sum_{i=1}^n a_i^2 / \sum_{i=1}^n b_i^2$, which will depend on the value of γ (defined in lemma 3.2). A small simulation study performed for an urn design by Smythe and Wei (1983) when $K = 2$ suggests that if the variance due to simple randomization is used in the analysis, the true significance level is considerably smaller than the nominal level of the test.

REMARK 3. As pointed out in Smythe and Wei (1983), condition (3.1) on the $\{a_{in}\}$ is general enough to permit applications of practical interest. For example, let ϕ be a real-valued function defined on $(0, 1)$ with $\int_0^1 \phi^2(x)\,dx < \infty$, and suppose that either: (a) ϕ is monotonic or (b) ϕ is continuous and monotonic on $(0, \varepsilon)$ and $(1 - \varepsilon, 1)$ for some $\varepsilon > 0$. If $a_{in} \equiv \phi(i/n + 1)$, it is easy to show that (3.1) holds. In particular, if F is a strictly increasing distribution function with a finite second moment, then $\int_0^1 (F^{-1}(x))^2\,dx < \infty$ and $a_{in} = F^{-1}(i/n + 1)$ satisfies (3.1).

REMARK 4. Smith (1984b) has shown that the range of values of $s_n^2 (= \sum_{i-1}^n b_{in}^2)$ provides an indication of the vulnerability of the design to experimental bias. Theorem 3.5 makes possible a more rigorous statement of some results in Sections 5, 6, and 9 of that paper.

4. Conditional inference. Cox (1982) introduced the idea of a conditional randomization test, whereby the significance level is computed conditionally on $N_{1n} - N_{2n}$ (in the case $K = 2$) or some other indicator of the balance of the design. The question therefore arises whether the results of Section 3 may be used to construct conditional tests.

For the sake of simplicity, we consider only the case $K = 2$, $\xi_1 = \xi_2 = \frac{1}{2}$, and $D_n = N_{1n} - N_{2n}$, though the same discussion applies to $K > 2$ and to other conditioning variables. Assume the $\{a_{in}\}$ are rescaled so that $\sum_{i=1}^n b_{in}^2 = 1$ and define an array $\{\bar{b}_{in}\}$ from (3.4) with $n^{-1/2}$ in place of a_{in}. For any real α and β, the linear combination $\alpha W_n + \beta n^{-1/2} D_n$ converges to a normal distribution, by Theorem 3.5. Specifically,

$$\alpha W_n + \beta n^{-1/2} D_n = \sum_i \left(\alpha \frac{a_{in}}{s_n} + 2\beta n^{-1/2} \right) \left(T_{1i} - \frac{1}{2} \right).$$

If $\tilde{b}_{in} \equiv \alpha b_{in}/s_n + 2\beta \bar{b}_{in}$ and $\tilde{s}_n^2 \equiv \sum_i^n \tilde{b}_n^2$, then Theorem 3.5 shows that $\alpha W_n + \beta n^{-1/2} D_n$ has approximately, for large n, a $N(0, \frac{1}{4}\tilde{s}_n^2)$ law, with

$$\tilde{s}_n^2 = (\alpha, \beta) \begin{pmatrix} 1 & 2\sum b_{in}\bar{b}_{in} \\ 2\sum b_{in}\bar{b}_{in} & 4\sum \bar{b}_{in}^2 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix}.$$

The Cramér–Wold device (cf. Billingsley, 1968, page 48) then implies that $(W_n, n^{-1/2} D_n)$ is asymptotically approximated by a bivariate normal with mean 0 and covariance matrix

$$\begin{pmatrix} \frac{1}{4} & \frac{1}{2}\sum b_{in}\bar{b}_{in} \\ \frac{1}{2}\sum b_{in}\bar{b}_{in} & \sum_i \bar{b}_{in}^2 \end{pmatrix}.$$

This suggests the

CONJECTURE 4.1. *The conditional distribution of W_n given D_n can be approximated asymptotically by a normal distribution with mean $n^{-1/2} D_n (\sum_i b_{in}\bar{b}_{in} / 2\sum_i \bar{b}_{in}^2)$ and variance $\frac{1}{4}\{1 - (\sum_i b_{in}\bar{b}_{in})^2 / (\sum_i \bar{b}_{in}^2)\}$.*

The conjecture is not a corollary of Theorem 3.5, because asymptotic conditional normality does not follow automatically from asymptotic joint normality. Steck (1957) and Holst (1981), among others, have considered such questions for sums of independent random vectors, but there is no obvious way to extend their technique here. If D_n were treated as a categorical variable with J categories, the conditional inference implied by the conjecture could be rigorously justified.

APPENDIX

PROOF OF THEOREM 3.5. Let

$$Y_n = \left\langle \sum_{i=1}^{n} (b_{in}/s_n)\left(T_{ji} - p_j(\mathbf{N}_{i-1}/(i-1))\right) \right\rangle_{j=1}^{K-1}, \tag{B.1}$$

where b_{in} is defined in (3.4). We first show that under the condition

$$\max_{1 \le i \le n} b_{in}^2 / \sum_{i=1}^{n} b_{in}^2 \to 0 \quad \text{as } n \to \infty, \tag{B.2}$$

Y_n converges in distribution to $N_{K-1}(0, \Sigma)$.

The proof employs the Cramér–Wold device. Given a set of constants $\alpha_1, \dots, \alpha_{K-1}$, consider the random variable

$$Q_n \equiv \sum_{j=1}^{K-1} \alpha_j \Delta^{-1/2} Y_{nj}, \tag{B.3}$$

where Y_{nj} is the jth component of Y_n and $\Delta = \sum_{j=1}^{K-1} \alpha_j^2 \xi_j - (\sum_{j=1}^{K-1} \alpha_j \xi_j)^2$. The variable Q_n can be rewritten as

$$\sum_{i=1}^{n} Z_{ni}, \quad \text{where} \quad Z_{ni} = \frac{b_{in}}{s_n} \Delta^{-1/2} \sum_{j=1}^{K-1} \alpha_j \left(T_{ji} - p_j \left(\frac{\mathbf{N}_{i-1}}{i-1} \right) \right).$$

Since $E[T_{ji} | \mathbf{N}_{i-1}] = p_j(\mathbf{N}_{i-1}/(i-1))$, it follows that the array $\{Z_{ni}\}$ is a martingale difference array with respect to the σ-fields $\mathscr{F}_i$ generated, for each n, by $\{T_{jk}\colon\ j = 1, \dots, K-1;\ k = 1, \dots, i\}$ (the σ-fields do not depend on n). We appeal to Theorem 2.3 of McLeish (1974) to prove asymptotic normality of Q_n. Conditions (a) and (b) of this theorem are trivially satisfied by (B.2) and the fact that $\Delta^{-1/2} \sum_{j=1}^{K-1} \alpha_j (T_{ji} - p_j(\mathbf{N}_{i-1}/(i-1)))$ is bounded. To check condition (c), which is that $\sum_{i=1}^{n} Z_{ni}^2 \to_p 1$ as $n \to \infty$, we define $U_{ni}^2 = \sum_{j=1}^{i} Z_{nj}^2$, $V_{ni}^2 = \sum_{j=1}^{i} E(Z_{nj}^2 | \mathscr{F}_{j-1})$. It follows from (B.2) that the conditional Lindeberg condition holds: For all $\varepsilon > 0$,

$$\sum_{i=1}^{n} E\big(Z_{ni}^2 I(|Z_{ni}| > \varepsilon) | \mathscr{F}_{i-1}\big) \to_p 0,$$

where $I(\cdot)$ is the indicator function. Also, it is easily seen that $\sup_n P(V_{nn}^2 > \lambda) \to 0$ as $\lambda \to \infty$, so it follows from Theorem 2.23 of Hall and Heyde (1980, page 44) that $\max_i |U_{ni}^2 - V_{ni}^2| \to_p 0$ as $n \to \infty$. So if we can prove that $V_{nn}^2 \to_p 1$, it will follow that $U_{nn}^2 = \sum_{i=1}^{n} Z_{ni}^2 \to_p 1$.

But $E(Z_{ni}^2 | \mathscr{F}_{i-1}) = (b_{in}^2 / \Delta s_n^2) D_i$, where

$$\begin{aligned} D_i &= E\left(\left[\sum_{l=1}^{K-1} \alpha_l \big(T_{lj} - p_l(\mathbf{N}_{i-1}/(i-1)) \big) \right]^2 | \mathscr{F}_{i-1} \right) \\ &= \sum_{l=1}^{K-1} \alpha_l^2 \big[p_l(\mathbf{N}_{i-1}/(i-1)) - p_l^2(\mathbf{N}_{i-1}(i-1)) \big] \\ &\quad - \sum_{l=1}^{K-1} \sum_{\substack{m=1 \\ m \ne l}}^{K-1} \alpha_l \alpha_m p_l(\mathbf{N}_{i-1}/(i-1)) p_m(\mathbf{N}_{i-1}/(i-1)), \end{aligned}$$

and using Proposition 2.5, D_i converges in probability to Δ as $i \to \infty$. Thus $V_{nn}^2 \to_p 1$, and we conclude from McLeish's theorem that $\sum_{i=1}^{n} Z_{ni} \to_d N(0,1)$.

Since $\alpha_1, \dots, \alpha_{K-1}$ are arbitrary constants, it follows that Y_n converges in distribution to $N_{K-1}(0, \Sigma)$.

Next, using condition (A.2), expand $p_l(y_1, \dots, y_{K-1})$ about the point $(\xi_1, \dots, \xi_{K-1})$:

$$\text{(B.4)} \quad \begin{aligned} p_l(y_1, \dots, y_{K-1}) &= p_l(\xi_1, \dots, \xi_{K-1}) + \sum_{j=1}^{K-1} d_{lj}(y_j - \xi_j) \\ &\quad + \tfrac{1}{2} \sum_{j=1}^{K-1} \sum_{m=1}^{K-1} \frac{\partial^2 p_l(\eta_1, \dots, \eta_{K-1})}{\partial y_j \, \partial y_m} (y_j - \xi_j)(y_m - \xi_m), \end{aligned}$$

where $(\eta, \dots, \eta_{K-1})$ lies on the line segment joining $\mathbf{y}$ to $\boldsymbol{\xi}$. We will show that

$$\text{(B.5)} \quad \sum_{i=1}^{n} \frac{b_{in}}{s_n} \left[\sum_{j=1}^{K-1} \sum_{l=1}^{K-1} \left(\frac{N_{j,i-1}}{i-1} - \xi_j \right) \left(\frac{N_{l,i-1}}{i-1} - \xi_l \right) \right] \to_p 0 \quad \text{as } n \uparrow \infty.$$

For this it suffices to prove that

$$\text{(B.6)} \quad \sum_{i=1}^{n} \frac{|b_{in}|}{s_n} \left(\frac{N_{1,i-1}}{i-1} - \xi_1 \right)^2 \to_p 0 \quad \text{as } n \to \infty.$$

But from Lemma 4 of Smith (1984a), $E(N_{1,i-1}/(i-1) - \xi_1)^2 \le K^2 (i-1)^{-1}$, so that the expectation of (B.6) is bounded by

$$\frac{K^2}{s_n} \sum_{i=1}^{n} \frac{|b_{in}|}{i-1} \le \frac{K^2}{s_n} \left(\sum_{i=1}^{n} \frac{|b_{in}|^2}{(i-1)^{2/3}} \right)^{1/2} \left(\sum_{i=1}^{n} (i-1)^{-4/3} \right)^{1/2},$$

using the Cauchy–Schwarz inequality. In view of (B.2), it follows easily that this converges to 0 as $n \to \infty$. Thus (B.5) is established. By Lemma 3.2, $d_{lj} = \gamma \delta_{lj}$ where δ_{lj} is the Kronecker delta, $1 \le j, l \le K-1$. By (B.4), (B.5), and condition (A.2), Y_n is asymptotically equivalent to

$$\left\langle \sum_{i=1}^{n} \frac{b_{in}}{s_n} \left[T_{ji} - \xi_j - \gamma (i-1)^{-1} \sum_{r=1}^{i-1} (T_{jr} - \xi_j) \right] \right\rangle_{j=1}^{K-1},$$

(with the convention that $0/0 = 0$), which can be rewritten using (3.4) as

$$\left\langle \sum_{i=1}^{n} \frac{a_{in}}{s_n} (T_{ji} - \xi_j) \right\rangle_{j=1}^{K-1} = \mathbf{W}_n'.$$

To complete the proof of Theorem 3.5 we will show that (3.1) implies (B.2). To do this we first show that $\max_i b_{in} \le C_1 \le \max_i a_{in}$, and then that $\sum_i a_{in}^2 \le C_2 \sum_i b_{in}^2$.

From (3.4) we have

$$\text{(B.7)} \quad \begin{aligned} |b_{in}| &\le |a_{in}| + |\gamma| \Big(\max_{1 \le i \le n} |a_{in}| \Big) \sum_{l=i+1} \left[(l-1)^{-1} \prod_{j=i}^{l-2} (1 + \gamma j^{-1}) \right] \\ &\le \max_{1 \le i \le n} |a_{in}| \{ 1 + |\gamma| G_{in} \}, \end{aligned}$$

where $G_{in} = \sum_{l=i+1}^{n}[(l-1)^{-1}|\prod_{j=i}^{l-2}(1+\gamma j^{-1})|]$. It is easily seen that for $|\gamma| < i$, $\prod_{j=i}^{l-2}(1+\gamma j^{-1}) \le C((l-2)/i)^{\gamma}$ for $l-2 > i$, with C constant. From this it follows that $G_{in} \le M < \infty$ for all i and n, proving that $|b_{in}| \le C \max_i |a_{in}|$. For the second part, observe that

$$a_{in} = b_{in} - \gamma \sum_{j=i+1}^{n} b_{jn}/(j-1) \equiv b_{in} - \gamma c_{in}. \tag{B.8}$$

Hence

$$\sum_i a_{in}^2 = \sum_i (b_{in} - \gamma c_{in})^2 \le \left[s_n - \gamma\left(\sum_i c_{in}^2\right)^{1/2}\right]^2,$$

by Schwarz's inequality.

Now

$$\begin{aligned}\sum_i c_{in}^2 &= \sum_i \sum_{j>i} \sum_{l>i} b_{jn} b_{ln}/\{(j-1)(l-1)\}\\ &= \sum_i \sum_{j>i} b_{jn}^2/(j-1)^2 + 2\sum_i \sum_{j>i} \sum_{l>j} b_{jn} b_{ln}/\{(j-1)(l-1)\}\\ &= \sum_j b_{jn}^2/(j-1) + 2\sum_j \sum_{l>j} b_{jn} b_{ln}/(l-1)\\ &\le s_n^2 + 2\sum_j b_{jn} c_{jn} \le s_n^2 + 2 s_n \left(\sum_i c_{in}^2\right)^{1/2},\end{aligned}$$

where we have used Schwarz's inequality in the last inequality. If $c^2 \equiv \sum_{i=1}^{n} c_{in}^2$, we get $s_n^2 + 2s_n c - c^2 \ge 0$, whence we must have $s_n \ge (\sqrt{2} - 1)c$. It now follows from (B.8) that $\sum_i a_{in}^2 \le (1 + |\gamma|[1 + \sqrt{2}])s_n^2$ completing the proof of Theorem 3.5.

Acknowledgments. We are grateful to the referee for helpful comments and to the editor for references related to the conditional distributions problem.

REFERENCES

ATKINSON, A. C. (1982). Optimum biased coin designs for sequential clinical trials with prognostic factors. *Biometrika* **69** 61–67.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

COX, D. R. (1982). A remark on randomization in clinical trials. *Utilitas Math.* **21A** 245–252.

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

EFRON, B. (1980). Randomizing and balancing a complicated sequential experiment. In *Biostatistics Casebook* (R. G. Miller, B. Efron, B. W. Brown, and L. E. Moses, eds.) 19–30. Wiley, New York.

HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Applications.* Academic, New York.

HOLST, L. (1981). Some conditional limit theorems in exponential families. *Ann. Probab.* **9** 818–830.

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

MCLEISH, D. L. (1974). Dependent central limit theorems and invariance principles. *Ann. Probab.* **2** 620–628.

POCOCK, S. J. (1979). Allocation of patients to treatment in clinical trials. *Biometrics* **35** 183–197.
SIMON, R. (1977). Adaptive treatment assignment methods and clinical trials. *Biometrics* **33** 743–749.
SIMON, R. (1979). Restricted randomization designs in clinical trials. *Biometrics* **35** 503–512.
SMITH, R. L. (1984a). Properties of biased coin designs in sequential clinical trials. *Ann. Statist.* **12** 1018–1034.
SMITH, R. L. (1984b). Sequential treatment allocation using biased coin designs. *J. Roy. Statist. Soc. B* **46** 519–543.
SMYTHE, R. T. and WEI, L. J. (1983). Significance tests with restricted randomization design. *Biometrika* **70** 496–500.
STECK, G. P. (1957). Limit theorems for conditional distributions. *Univ. California Publ. Statist.* **2** 237–284.
WEI, L. J. (1977). A class of designs for sequential clinical trials. *J. Amer. Statist. Assoc.* **72** 382–386.
WEI, L. J. (1978a). The adaptive biased coin design for sequential experiments. *Ann. Statist.* **6** 92–100.
WEI, L. J. (1978b). A class of treatment assignment rules for sequential experiments. *Comm. Statist. A—Theory Methods* **7** 283–295.

L. J. WEI
R. T. SMYTHE
DEPARTMENT OF STATISTICS
GEORGE WASHINGTON UNIVERSITY
WASHINGTON DC 20052

R. L. SMITH
DEPARTMENT OF MATHEMATICS
IMPERIAL COLLEGE
LONDON, ENGLAND