CENTRAL LIMIT THEOREMS FOR DOUBLY ADAPTIVE BIASED COIN DESIGNS[1]

By Jeffrey R. Eisele and Michael B. Woodroofe

Ciba-Geigy AG and University of Michigan

Asymptotic normality of the difference between the number of subjects assigned to a treatment and the desired number to be assigned is established for allocation rules which use Eisele's biased coin design. Subject responses are assumed to be independent random variables from standard exponential families. In the proof, it is shown that the difference may be magnified by appropriate constants so that the magnified difference is nearly a martingale. An application to the Behrens–Fisher problem in the normal case is described briefly.

1. Introduction. Suppose subjects arrive sequentially at an experimental site and are assigned immediately to one of two treatments groups A or B. A statistical design problem is how to assign subjects to the treatment groups. When balance is desired in the allocation, the complete randomization scheme of making treatment assignments by independent flips of a fair coin could be employed, but this might result in severe imbalance in small experiments. Pocock (1979) recommends using complete randomization only in trials with over 200 subjects. The systematic design (ABAB...) results in perfect balance but unfortunately maximizes the possibility of experimenter bias. Efron (1971) and Wei (1978) proposed subject assignment algorithms offering a compromise between complete randomization and perfect balance. These designs achieve balance more quickly than complete randomization, but retain enough randomization to preclude effective guessing of the next treatment to be assigned.

Letting $+1$ and -1 denote assignments to treatments A and B, these algorithms may be described as follows: let q denote a nonincreasing function from $[-1, 1]$ into $[0, 1]$ for which $q(0) = 1/2$; let $U_1, U_2, \ldots$ denote i.i.d. uniformly distributed random variables; and let

$$Z_k = 2I\left\{U_k \le q\left(\frac{S_{k-1}}{k-1}\right)\right\} - 1 \quad \forall\, k = 1, 2, \ldots,$$

where

$$S_n = Z_1 + \cdots + Z_n \quad \forall\, n = 1, 2, \ldots,$$

$S_0 = 0$; $0/0$ is to be interpreted as 0; and $I\{\cdot\}$ denotes the indicator of $\{\cdot\}$. With these conventions, S_n is the difference between the number of subjects

234

assigned to treatment group A and the number assigned to treatment group B after n assignments. Both algorithms are called biased coin designs. They differ in the nature of the function q.

For Wei's algorithm, q is assumed to have a finite derivative at 0. Under this assumption, Wei (1978) showed that $S_n/\sqrt{n}$ has a limiting normal distribution with mean 0 and variance $1/(1 - 4q'(0))$. Wei proved this by showing the convergence of all moments of $S_n/\sqrt{n}$ to those of the normal distribution. Later, Smith (1984) and Eisele and Woodroofe (1990) showed how this result may be deduced from the martingale central limit theorem. This approach also yields an invariance principle with little additional effort.

Eisele (1994) proposed a randomized subject assignment algorithm for the case when the desired allocation proportions may be unknown. The design is similar in spirit to Efron and Wei's designs, but adapts for both the current proportion of subjects assigned to each treatment and a current estimate of the desired allocation proportion. Eisele (1994) showed that the proportion of subjects assigned to a treatment converges almost surely to the desired proportion.

The motivation for the design originated with an estimation problem of Robbins, Simons and Starr (1967), who developed an algorithm for allocating subjects to treatments in order to minimize the total expected sample size and still obtain a confidence interval of preassigned width and coverage probability in the Behrens–Fisher problem. Eisele (1990) superimposed a biased coin on Robbins, Simons and Starr's design. This is discussed briefly in Section 8. Superimposing biased coin designs on asymptotically optimals ones, such as that of Robbins, Simons and Starr, may make them more appealing to practitioners, especially in long studies where temporal effects may be present. See Hardwick (1989) for a discussion of these issues.

The purpose of this paper is to establish the asymptotic normality of the difference between the number of subjects assigned to a treatment and the desired number. The approach is similar to that of Smith (1984) and Eisele and Woodroofe (1990) for Wei's biased coin design; that is, the process will be decomposed into a martingale plus a remainder term.

The general model and allocation rule are presented in Section 2. Sections 3–6 present some preliminary results, including derivation of the martingale and remainder terms and the asymptotic normality of the martingale. The main result on the asymptotic normality of the difference between the number of subjects assigned to a treatment and the desired number to be assigned to that treatment is established in Section 7. The Behrens–Fisher problem is considered in Section 8.

2. General model and allocation rule.

2.1. *General model.* Suppose subject responses $X_1, X_2, \ldots$ to treatment A and $Y_1, Y_2, \ldots$ to treatment B are independent random variables from d-dimensional standard exponential families, that is,

$$X_1, X_2, \ldots \sim f_\theta(x) = \exp(\theta \cdot x - \psi(\theta))$$

and

$$Y_1, Y_2, \ldots \sim g_\omega(y) = \exp(\omega \cdot y - \varphi(\omega)),$$

where $\theta = (\theta_1, \ldots, \theta_d)'$, $x = (x_1, \ldots, x_d)'$, $\omega = (\omega_1, \ldots, \omega_d)'$, $y = (y_1, \ldots, y_d)'$, prime denotes transpose and "$\cdot$" denotes the inner product. It is assumed throughout that θ and ω are interior points of the natural parameter spaces of the families. Let

$$\mu = \mathbb{E}_\theta X = \nabla\psi(\theta) \quad \text{and} \quad \nu = \mathbb{E}_\omega Y = \nabla\varphi(\omega),$$

where $\nabla\psi$ and $\nabla\varphi$ denote the gradient vectors $(\partial\psi/\partial\theta_1, \ldots, \partial\psi/\partial\theta_d)'$ and $(\partial\varphi/\partial\omega_1, \ldots, \partial\varphi/\partial\omega_d)'$. For a given allocation rule, let m_k and n_k be the number of observations on X and on Y, respectively, at time k, where $k = m_k + n_k$, and let

$$\overline{X}_{m_k} = (m_k)^{-1} \sum_{i=1}^{m_k} X_i \quad \text{and} \quad \overline{Y}_{n_k} = (n_k)^{-1} \sum_{i=1}^{n_k} Y_i$$

be the sample means. If the families are steep and the allocation rule is noninformative, then the MLE of μ is $\overline{X}_{m_k}$ and the MLE of ν is $\overline{Y}_{n_k}$. Brown (1986) is recommended for more background on exponential families.

The goal of the allocation scheme is then to have $m_k/k = \rho$, where $\rho \equiv \rho(\mu, \nu)$: $\mathbb{R}^{2d} \to (0, 1)$ is the desired allocation proportion. To accomplish this, the allocation scheme is designed to sample the X population with probability less than (respectively, greater than) $\hat{\rho}_k$ when $m_k/k > \hat{\rho}_k$ (respectively, $m_k/k < \hat{\rho}_k$), where $\hat{\rho}_k \equiv \rho(\overline{X}_{m_k}, \overline{Y}_{n_k}) \in [0, 1]$ is the current maximum likelihood estimate of ρ. This employs the same idea as Wei's adaptive biased coin design except that he uses $\rho = \frac{1}{2}$.

2.2. *The allocation rule.* Let q be a function from $[0, 1]^2 \to [0, 1]$ such that the following four conditions hold:
   (i) q is jointly continuous;
   (ii) $q(r, r) = r$;
   (iii) $q(p, r)$ is strictly decreasing in p and strictly increasing in r on $(0, 1)^2$;
   (iv) q has bounded derivatives in both arguments and $q_{10}(\rho, \rho) = \partial q(x, y)/\partial x|_{x=\rho, y=\rho} \neq 0$.

Let $\delta_1 = \cdots = \delta_{n_0} = 1$, $\delta_{n_0+1} = \cdots = \delta_{2n_0} = 0$ and

$$\delta_k = I\left\{ U_k \leq q\left(\frac{m_{k-1}}{k-1}, \hat{\rho}_{k-1} \right) \right\} \quad \forall\, k \geq 2n_0 + 1,$$

where the U_k are independent of $X_1, X_2, \ldots$ and $Y_1, Y_2, \ldots$. Then

$$m_k = \delta_1 + \cdots + \delta_k \quad \text{and} \quad n_k = k - m_k.$$

This sampling scheme has the desirable property that it provides some randomization to reduce the possibility of experimenter bias. Although exper-

imenter bias is not completely eliminated, the best guessing strategy at time $k + 1$ has a probability of success equal to $\max(q_k, 1 - q_k)$, where

$$q_k \equiv q\left(\frac{m_{k-1}}{k - 1}, \hat{\rho}_{k-1}\right),$$

for $k \geq 2n_0 + 1$.

2.3. *Conditions on* ρ. In addition to conditions (i)–(iv) imposed on q, the following two conditions on ρ are needed:

(v) There are positive constants C and r for which

$$\frac{1}{\rho} + \frac{1}{1 - \rho} \leq C\{\|\mu\|^r + \|\nu\|^r\}.$$

(vi) For sufficiently small $\varepsilon > 0$, ρ is twice continuously differentiable on the set

$$R = \{(x, y): \|x - \mu\| \leq 2\varepsilon, \|y - \nu\| \leq 2\varepsilon\}.$$

2.4. *Notation*. The definitions of S_k and Z_k stated below will be used throughout the remainder of the paper. They are different from those in the Introduction. Let

$$Z_k = \delta_k - \rho$$

and

$$S_k = m_k - k\rho = Z_1 + \cdots + Z_k \quad \forall\, k > 2n_0.$$

Also, let

$$\mathscr{F}_k \equiv \sigma\{Z_1, \ldots, Z_k, X_1, \ldots, X_{m_k}, Y_1, \ldots, Y_{n_k}\}$$

be the σ-algebra representing the natural history. Then the conditional mean and variance of Z_k given $\mathscr{F}_{k-1}$ are

$$\mu_k = \mathbb{E}\{Z_k \,|\, \mathscr{F}_{k-1}\} = q_k - \rho$$

and

$$\sigma_k^2 = \mathbb{E}\{(Z_k - \mu_k)^2 \,|\, \mathscr{F}_{k-1}\} = q_k(1 - q_k).$$

3. Preliminary results. The following results can be found in Eisele (1994). It is Theorem 1 that requires condition (v).

THEOREM 1. *There is an* $0 < \varepsilon_0 < 1$ *for which*
$$P\{m_k < \varepsilon_0 k\} = o(k^{-\alpha}) \quad as\ k \to \infty\ \forall\, \alpha > 0.$$

LEMMA 1. *Assume*
$$\mathbb{E}\{(\hat{\rho}_k - \rho)^{2r}\} = O(k^{-r}) \quad \forall\, r > 0\ as\ k \to \infty.$$

PROPOSITION 1. *We have*
$$\mathbb{E}\{S_n^2\} = O(n) \quad as\ n \to \infty.$$

THEOREM 2 (Strong law of large numbers). *Under conditions* (i)–(vi),

$$\lim_{n \to \infty} \frac{S_n}{n} = 0 \quad w.p.1.$$

Lemma 2 presents a calculation that is repeated frequently below, often without comment. Lemma 3 is applied many times throughout the remainder of this paper.

LEMMA 2. *Let*

$$\mathbb{E}\left\{ \left(\overline{X}_{m_k} - \mu \right) \big| \mathscr{F}_{k-1} \right\} = \left(\overline{X}_{m_{k-1}} - \mu \right)\left(1 - \frac{q_k}{m_{k-1} + 1} \right).$$

PROOF. It follows that

$$\mathbb{E}\left\{ \left(\overline{X}_{m_k} - \mu \right) \big| \mathscr{F}_{k-1} \right\} = \mathbb{E}\left\{ \frac{m_{k-1}}{m_k}\left(\overline{X}_{m_{k-1}} - \mu \right) + \frac{\delta_k}{m_k}\left(X_{m_k} - \mu \right) \big| \mathscr{F}_{k-1} \right\}$$

$$= \mathbb{E}\left\{ \frac{m_{k-1}}{m_k}\left(\overline{X}_{m_{k-1}} - \mu \right) \big| \mathscr{F}_{k-1} \right\}$$

$$= \left(\overline{X}_{m_{k-1}} - \mu \right)\left(1 - \mathbb{E}\left\{ \frac{\delta_k}{m_k} \big| \mathscr{F}_{k-1} \right\} \right)$$

$$= \left(\overline{X}_{m_{k-1}} - \mu \right)\left(1 - \frac{q_k}{m_{k-1} + 1} \right). \qquad \square$$

LEMMA 3. *There exists a constant C such that*

$$\int \frac{1}{m_k^p} W \, d\mathbb{P} \leq \frac{C}{k^p} \sqrt{\mathbb{E}\{ W^2 \}},$$

for all $p > 0$ *and all square integrable random variables* W.

PROOF. Let $\varepsilon = \varepsilon_0$ be as in Theorem 1. If W is any square integrable random variable, then

$$\int \frac{1}{m_k^p} W \, d\mathbb{P} = \int_{m_k \leq \varepsilon k} \frac{1}{m_k^p} W \, d\mathbb{P} + \int_{m_k > \varepsilon k} \frac{1}{m_k^p} W \, d\mathbb{P}$$

$$\leq \sqrt{\mathbb{P}\{ m_k \leq \varepsilon k \}} \sqrt{\mathbb{E}\{ W^2 \}} + \frac{1}{\varepsilon^p k^p} \mathbb{E}\{ |W| \}$$

$$\leq \frac{C}{k^p} \sqrt{\mathbb{E}\{ W^2 \}},$$

where the last inequality follows from Theorem 1. $\square$

COROLLARY 1.

$$\mathbb{E}\left\{ \| \overline{X}_{m_k} - \mu \|^{2r} + \| \overline{Y}_{n_k} - \nu \|^{2r} \right\} = O(k^{-r}) \quad as \ k \to \infty \ \forall \ r > 0.$$

Of course, it follows from Corollary 1 and Markov's inequality that

$$\mathbb{P}\big\{\|\overline{X}_{m_k} - \mu\| \geq \varepsilon\big\} + \mathbb{P}\big\{\|\overline{Y}_{n_k} - \nu\| \geq \varepsilon\big\} = O(k^{-r})$$

as $k \to \infty$ for all $r > 0$.

4. The martingale. The idea in this section is to magnify S_k by an appropriate sequence of constants so that this magnified value is nearly a martingale. In Wei's design, there is only one term in the conditional expectation of S_k. Here, there are two terms. The additional term is due to estimating ρ. In order to get a martingale, these terms must be magnified by different sequences of constants. This requires looking at a vector of these two terms and then magnifying this vector by a vector of constants. Propositions 2 and 3 lay the foundation for determining this matrix.

PROPOSITION 2.

$$\mathbb{E}\{S_k \mid \mathscr{F}_{k-1}\} = \left(1 - \frac{\alpha}{k-1}\right)S_{k-1} + \gamma(\hat{\rho}_{k-1} - \rho) + r\left(\frac{m_{k-1}}{k-1}, \hat{\rho}_{k-1}\right),$$

where

$$\alpha = -q_{10}(\rho, \rho), \qquad \gamma = q_{01}(\rho, \rho),$$
$$r(x, y) = q(x, y) - \rho + \alpha(x - \rho) - \gamma(y - \rho)$$

and q_{01} and q_{10} are first partial derivatives of q.

PROOF. It is convenient to write

$$r_{1,k} = r\left(\frac{m_{k-1}}{k-1}, \hat{\rho}_{k-1}\right).$$

Then

$$\begin{aligned}
\mathbb{E}\{S_k \mid \mathscr{F}_{k-1}\} &= \mathbb{E}\{S_{k-1} + Z_k \mid \mathscr{F}_{k-1}\} \\
&= S_{k-1} + q\left(\frac{m_{k-1}}{k-1}, \hat{\rho}_{k-1}\right) - \rho \\
&= S_{k-1} - \alpha\left(\frac{m_{k-1}}{k-1} - \rho\right) + \gamma(\hat{\rho}_{k-1} - \rho) + r_{1,k} \\
&= \left(1 - \frac{\alpha}{k-1}\right)S_{k-1} + \gamma(\hat{\rho}_{k-1} - \rho) + r_{1,k}. \qquad \square
\end{aligned}$$

Proposition 3 contains an analogous approximation for $\mathbb{E}\{k(\hat{\rho}_k - \rho) \mid \mathscr{F}_{k-1}\}$. In the proof, it is convenient to let

$$\begin{aligned}
\tilde{\rho}_k &= \rho\left(\overline{X}_{m_{k-1}}, \overline{Y}_{n_{k-1}}\right) + \rho_{10}\left(\overline{X}_{m_{k-1}}, \overline{Y}_{n_{k-1}}\right) \cdot \left(\frac{X_{m_k} - \overline{X}_{m_{k-1}}}{m_{k-1}+1}\right)\delta_k \\
(1) & \\
&\quad + \rho_{01}\left(\overline{X}_{m_{k-1}}, \overline{Y}_{n_{k-1}}\right) \cdot \left(\frac{Y_{n_k} - \overline{Y}_{n_{k-1}}}{n_{k-1}+1}\right)(1 - \delta_k),
\end{aligned}$$

on A_k, where ρ_{10} and ρ_{01} denote the gradient vectors,

$$\rho_{10}(x,y) = \left[\frac{\partial \rho(x,y)}{\partial x_1}, \ldots, \frac{\partial \rho(x,y)}{\partial x_d}\right]',$$

$$\rho_{01}(x,y) = \left[\frac{\partial \rho(x,y)}{\partial y_1}, \ldots, \frac{\partial \rho(x,y)}{\partial y_d}\right]',$$

and

$$(2) \qquad A_k = \left\{\|\bar{X}_{m_{k-1}} - \mu\| \le \varepsilon, \|\bar{Y}_{n_{k-1}} - \nu\| \le \varepsilon\right\}$$

and

$$(3) \qquad B_k = \left\{\|X_{m_k} - \mu\| \le \varepsilon m_k, \|Y_{n_k} - \nu\| \le \varepsilon n_k\right\},$$

for $k = 1, 2, \ldots$. Observe that $\mathbb{P}(A_k^c) = O(k^{-r})$ as $k \to \infty$ for all $r > 0$, by Corollary 1. The same is true of $\mathbb{P}(B_k^c)$, since, for example, $\mathbb{E}\{m_k^{-r}\|X_{m_k} - \mu\|^r\}$ $= \mathbb{E}\{m_k^{-r}\mathbb{E}\{\|X_{m_k} - \mu\|^r \mid \mathscr{F}_{k-1}\}\} = \mathbb{E}\{\|X_1 - \mu\|^r\}\mathbb{E}\{m_k^{-r}\} = O(k^{-r})$ as $k \to \infty$ for all $r > 0$.

LEMMA 4.   *Let $\tilde{\rho}_k = \hat{\rho}_k$ off A_k. Then*

$$\mathbb{E}\left\{(\hat{\rho}_k - \tilde{\rho}_k)^2\right\} = O(k^{-4}) \quad as\ k \to \infty.$$

PROOF.   Let $\varepsilon > 0$ be such that condition (vi) is satisfied, and define A_k and B_k by (2) and (3). Then

$$\mathbb{E}\left\{(\hat{\rho}_k - \tilde{\rho}_k)^2\right\} = \int_{A_k^c \cup B_k^c} (\hat{\rho}_k - \tilde{\rho}_k)^2 \, d\mathbb{P}$$

$$+ \int_{A_k \cap B_k} (\hat{\rho}_k - \tilde{\rho}_k)^2 \, d\mathbb{P},$$

where

$$\int_{A_k^c \cup B_k^c} (\hat{\rho}_k - \tilde{\rho}_k)^2 \, d\mathbb{P} \le C\mathbb{P}(A_k^c \cup B_k^c) \le C\mathbb{P}(A_k^c) + C\mathbb{P}(B_k^c) = O(k^{-r}),$$

for some constant C and all $r > 0$. Next, since ρ is twice differentiable on R [defined in (vi)], there is a constant C for which

$$\int_{A_k \cap B_k} (\hat{\rho}_k - \tilde{\rho}_k)^2 \, d\mathbb{P} \le \int_{A_k \cap B_k} C\left\{\frac{\|X_{m_k} - \bar{X}_{m_{k-1}}\|^2}{m_k^2} + \frac{\|Y_{n_k} - \bar{Y}_{n_{k-1}}\|^2}{n_k^2}\right\}^2 \, d\mathbb{P}$$

$$\le \int_{A_k \cap B_k} C\left\{\frac{\left(\|X_{m_k} - \mu\| + \|\bar{X}_{m_{k-1}} - \mu\|\right)^2}{m_k^2}\right.$$

$$\left. + \frac{\left(\|Y_{n_k} - \nu\| + \|\bar{Y}_{n_{k-1}} - \nu\|\right)^2}{n_k^2}\right\}^2 \, d\mathbb{P}$$

$$= O(k^{-4}),$$

where the last equality follows from Lemma 3. $\square$

LEMMA 5.

$$\mathbb{E}\left\{\left(k\left[\hat{\rho}_k - \mathbb{E}\{\hat{\rho}_k \mid \mathscr{F}_{k-1}\}\right] - k\left[\tilde{\rho}_k - \mathbb{E}\{\tilde{\rho}_k \mid \mathscr{F}_{k-1}\}\right]\right)^2\right\}$$
$$= O(k^{-2}) \quad as \ k \to \infty.$$

PROOF. The left-hand side is at most $k^2\mathbb{E}\{(\hat{\rho}_k - \tilde{\rho}_k)^2\}$. □

PROPOSITION 3.

$$\mathbb{E}\{k(\hat{\rho}_k - \rho) \mid \mathscr{F}_{k-1}\} = (k-1)(\hat{\rho}_{k-1} - \rho) + r_{2,k},$$

where

$$\mathbb{E}\{(r_{2,k})^2\} = o(k^{-1}) \quad as \ k \to \infty.$$

PROOF. With $r_{2,k} = \mathbb{E}\{k(\hat{\rho}_k - \rho) \mid \mathscr{F}_{k-1}\} - (k-1)(\hat{\rho}_{k-1} - \rho)$, as in the proposition,

$$\int_{A_k^c} r_{2,k}^2 \, d\mathbb{P} \leq 4k^2 \mathbb{P}(A_k^c) = O(k^{-2})$$

as $k \to \infty$. As in the proof of Lemma 2,

$$\mathbb{E}\{k(\tilde{\rho}_k - \rho) \mid \mathscr{F}_{k-1}\} = k(\hat{\rho}_{k-1} - \rho)$$
$$+ \frac{kq_k}{m_{k-1}+1}\rho_{10}\left(\overline{X}_{m_{k-1}}, \overline{Y}_{n_{k-1}}\right) \cdot \left(\mu - \overline{X}_{m_{k-1}}\right)$$
$$+ \frac{k(1-q_k)}{n_{k-1}+1}\rho_{01}\left(\overline{X}_{m_{k-1}}, \overline{Y}_{n_{k-1}}\right) \cdot \left(\nu - \overline{Y}_{n_{k-1}}\right)$$

on A_k. So

$$\mathbb{E}\{k(\hat{\rho}_k - \rho) \mid \mathscr{F}_{k-1}\} = (k-1)(\hat{\rho}_{k-1} - \rho)$$
$$+ k\mathbb{E}\left((\hat{\rho}_k - \tilde{\rho}_k) \mid \mathscr{F}_{k-1}\right) + U_k,$$

where

$$U_k = \hat{\rho}_{k-1} - \rho - \left[\frac{kq_k}{m_{k-1}+1}\rho_{10}\left(\overline{X}_{m_{k-1}}, \overline{Y}_{n_{k-1}}\right) \cdot \left(\overline{X}_{m_{k-1}} - \mu\right)\right.$$
$$\left. + \frac{k(1-q_k)}{n_{k-1}+1}\rho_{01}\left(\overline{X}_{m_{k-1}}, \overline{Y}_{n_{k-1}}\right) \cdot \left(\overline{Y}_{n_{k-1}} - \nu\right)\right].$$

Clearly,

$$\mathbb{E}\left|k\mathbb{E}\{(\hat{\rho}_k - \tilde{\rho}_k) \mid \mathscr{F}_{k-1}\}\right|^2 \leq k^2\mathbb{E}\left\{(\hat{\rho}_k - \tilde{\rho}_k)^2\right\} = O(k^{-2})$$

as $k \to \infty$, by Lemma 4. So it suffices to consider U_k. On A_k, U_k may be written in the form

$$U_k = U_{1,k} + U_{2,k} + U_{3,k},$$

where

$$U_{1,k} = \left[\rho_{10}(\mu,\nu) - \frac{kq_k}{m_{k-1}+1}\rho_{10}\left(\overline{X}_{m_{k-1}},\overline{Y}_{n_{k-1}}\right)\right]\cdot\left(\overline{X}_{m_{k-1}} - \mu\right),$$

$$U_{2,k} = \left[\rho_{01}(\mu,\nu) - \frac{k(1-q_k)}{n_{k-1}+1}\rho_{01}\left(\overline{X}_{m_{k-1}},\overline{Y}_{n_{k-1}}\right)\right]\cdot\left(\overline{Y}_{n_{k-1}} - \nu\right),$$

$$U_{3,k} \le C\left[\|\overline{X}_{m_{k-1}} - \mu\|^2 + \|\overline{Y}_{n_{k-1}} - \nu\|^2\right].$$

Here

$$\mathbb{E}\{U_{3,k}^2\} = O(k^{-2})$$

as $k \to \infty$, by Corollary 1. For $U_{1,k}$, one finds from Theorem 2, Corollary 1 and the conditions imposed on q that, w.p.1 as $k \to \infty$,

$$\hat{\rho}_k \to \rho = \rho(\mu,\nu), \qquad \rho_{10}\left(\overline{X}_{m_{k-1}},\overline{Y}_{n_{k-1}}\right) \to \rho_{10}(\mu,\nu), \qquad \frac{m_k}{k} \to \rho$$

and

$$q_k = q\left(\frac{m_{k-1}}{k-1},\hat{\rho}_k\right) \to q(\rho,\rho) = \rho.$$

So

$$\int_{A_k} U_{1,k}^2\, d\mathbb{P} \le \left[\int_{A_k}\left|\rho_{10}(\mu,\nu) - \frac{kq_k}{m_{k-1}+1}\rho_{10}\left(\overline{X}_{m_{k-1}},\overline{Y}_{n_{k-1}}\right)\right|^4 d\mathbb{P}\right]^{1/2}$$

$$\times\left[\int_{A_k}\|\overline{X}_{m_{k-1}} - \mu\|^4\, d\mathbb{P}\right]^{1/2} = o(k^{-1}),$$

by Corollary 1 and the dominated convergence theorem. The analysis of $U_{2,k}$ is similar, to complete the proof. $\square$

4.1. *An approximating martingale.* Let

$$(4) \qquad\qquad V_k = \begin{pmatrix} S_k \\ k(\hat{\rho}_k - \rho) \end{pmatrix}.$$

Then, by Propositions 2 and 3,

$$(5) \qquad \mathbb{E}\{V_k \mid \mathscr{F}_{k-1}\} = \begin{pmatrix} 1 - \dfrac{\alpha}{k-1} & \dfrac{\gamma}{k-1} \\ 0 & 1 \end{pmatrix} V_{k-1} + \begin{pmatrix} r_{1,k} \\ r_{2,k} \end{pmatrix}.$$

So, if D_k is any 2×2 matrix, then

$$\mathbb{E}\{D_k V_k \mid \mathscr{F}_{k-1}\} = D_k\begin{pmatrix} 1 - \dfrac{\alpha}{k-1} & \dfrac{\gamma}{k-1} \\ 0 & 1 \end{pmatrix} V_{k-1} + D_k r_k,$$

where

$$r_k = \begin{pmatrix} r_{1,k} \\ r_{2,k} \end{pmatrix}.$$

This suggests the recursion

$$D_k = D_{k-1} \begin{pmatrix} 1 - \dfrac{\alpha}{k-1} & \dfrac{\gamma}{k-1} \\ 0 & 1 \end{pmatrix}^{-1} = \prod_{i=m+1}^{k} \begin{pmatrix} 1 - \dfrac{\alpha}{i-1} & \dfrac{\gamma}{i-1} \\ 0 & 1 \end{pmatrix}^{-1},$$

where $m = \max(2n_0, [\alpha] + 1)$ and $[x]$ denotes the greatest integer which is less than or equal to x for $-\infty < x < \infty$. Observe that

$$\begin{pmatrix} 1 - \dfrac{\alpha}{i-1} & \dfrac{\gamma}{i-1} \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} \left(1 - \dfrac{\alpha}{i-1}\right)^{-1} & \dfrac{-\gamma}{i-1}\left(1 - \dfrac{\alpha}{i-1}\right)^{-1} \\ 0 & 1 \end{pmatrix},$$

for $i > m$. Let $c_k = 1$ for $k = 1, \ldots, m$, and let

$$c_k = \prod_{i=m+1}^{k} \left(1 - \dfrac{\alpha}{i-1}\right)^{-1} \quad \forall\, k \geq m + 1.$$

Then

(6) $$c_k \sim \beta k^{\alpha} \quad \text{as } k \to \infty,$$

where $0 < \beta = \beta_\alpha < \infty$, as is easily seen by taking logarithms. It is also easily seen that

$$D_k = \begin{pmatrix} c_k & -\gamma \displaystyle\sum_{j=m+1}^{k} \left(\dfrac{c_j}{j-1}\right) \\ 0 & 1 \end{pmatrix}$$

and

$$D_k V_k = \begin{pmatrix} c_k(m_k - k\rho) - \gamma \displaystyle\sum_{j=m+1}^{k} \left(\dfrac{c_j}{j-1}\right) k(\hat{\rho}_k - \rho) \\ k(\hat{\rho}_k - \rho) \end{pmatrix}.$$

By (5) and the definition of D_k, $\mathbb{E}\{D_k V_k \mid \mathscr{F}_{k-1}\} = D_{k-1} V_{k-1} + D_k r_k$ for $k > m$. So

$$D_n V_n = D_m V_m + \sum_{k=m+1}^{n} (D_k V_k - D_{k-1} V_{k-1})$$

$$= D_m V_m + \sum_{k=m+1}^{n} [D_k V_k - \mathbb{E}\{D_k V_k \mid \mathscr{F}_{k-1}\}] + \sum_{k=m+1}^{n} D_k r_k$$

$$= D_m V_m + M_n + R_n,$$

where

$$M_n = \sum_{k=m+1}^{n} [D_k V_k - \mathbb{E}\{D_k V_k \mid \mathscr{F}_{k-1}\}]$$

and

$$R_n = \sum_{k=m+1}^{n} D_k r_k.$$

Observe that M_n is a martingale and write $M_n = (M_{1,n}, M_{2,n})'$.

5. Conditional variance. In this section, some preliminary results necessary for the calculation of the asymptotic conditional variance will be derived. The conditional variance terms are stated explicitly in Section 6.

PROPOSITION 4. *Let*

$$\sigma^2 = \rho'_{10}(\mu, \nu)\, \nabla^2\psi(\theta)\, \rho_{10}(\mu, \nu)$$

and

$$\tau^2 = \rho'_{01}(\mu, \nu)\, \nabla^2\varphi(\omega)\, \rho_{01}(\mu, \nu),$$

where $\nabla^2\psi(\theta) = \mathrm{Cov}_\theta X$ *denotes the Hessian matrix* $(\partial^2\psi/\partial\theta_i\, \partial\theta_j)_{1 \le i \le d,\, 1 \le j \le d}$ *and* $\nabla^2\varphi(\omega) = \mathrm{Cov}_\omega Y$ *denotes the Hessian matrix* $(\partial^2\varphi/\partial\omega_i\, \partial\omega_j)_{1 \le i \le d,\, 1 \le j \le d}$. *Then the following hold:*

(a)
$$\mathbb{E}\left\{\left(k\tilde{\rho}_k - \mathbb{E}\{k\tilde{\rho}_k \mid \mathscr{F}_{k-1}\}\right)^2 \mid \mathscr{F}_{k-1}\right\}$$
$$\to \frac{\sigma^2}{\rho} + \frac{\tau^2}{1-\rho} \quad w.p.1, \, as \, k \to \infty;$$

(b)
$$\mathbb{E}\left\{\left(k\tilde{\rho}_k - \mathbb{E}\{k\tilde{\rho}_k \mid \mathscr{F}_{k-1}\}\right)^2 \mid \mathscr{F}_{k-1}\right\}$$
$$\to \frac{\sigma^2}{\rho} + \frac{\tau^2}{1-\rho} \quad in \, L^1, \, as \, k \to \infty;$$

(c)
$$\mathbb{E}\left\{\left(k\hat{\rho}_k - \mathbb{E}\{k\hat{\rho}_k \mid \mathscr{F}_{k-1}\}\right)^2 \mid \mathscr{F}_{k-1}\right\}$$
$$\to \frac{\sigma^2}{\rho} + \frac{\tau^2}{1-\rho} \quad in \, L^1, \, as \, k \to \infty;$$

(d)
$$\mathbb{E}\left\{\left(k\hat{\rho}_k - \mathbb{E}\{k\hat{\rho}_k \mid \mathscr{F}_{k-1}\}\right)^2 \mid \mathscr{F}_{k-1}\right\}$$
$$\to \frac{\sigma^2}{\rho} + \frac{\tau^2}{1-\rho} \quad in \, probability, \, as \, k \to \infty.$$

PROOF. (a) $\mathbb{E}\{\tilde{\rho}_k \mid \mathscr{F}_{k-1}\}$ was computed in the proof of Proposition 3. Combining this with (2) gives

$$\tilde{\rho}_k - \mathbb{E}\{\tilde{\rho}_k \mid \mathscr{F}_{k-1}\}$$

$$= \rho_{10}\left(\overline{X}_{m_{k-1}}, \overline{Y}_{n_{k-1}}\right) \cdot \left[\delta_k(X_{m_k} - \mu) - (\delta_k - q_k)\left(\overline{X}_{m_{k-1}} - \mu\right)\right]\left(\frac{1}{m_{k-1}+1}\right)$$

$$(7) \quad + \rho_{01}\left(\overline{X}_{m_{k-1}}, \overline{Y}_{n_{k-1}}\right) \cdot \left[(1 - \delta_k)(Y_{n_k} - \nu) - (q_k - \delta_k)\left(\overline{Y}_{n_{k-1}} - \nu\right)\right]$$

$$\times \left(\frac{1}{n_{k-1}+1}\right)$$

on A_k. Then

$$\mathbb{E}\left\{\left(k\tilde{\rho}_k - \mathbb{E}\{k\tilde{\rho}_k \mid \mathscr{F}_{k-1}\}\right)^2 \mid \mathscr{F}_{k-1}\right\}$$

$$= \rho'_{10}\left(\overline{X}_{m_{k-1}}, \overline{Y}_{n_{k-1}}\right)\nabla^2\psi(\theta)\,\rho_{10}\left(\overline{X}_{m_{k-1}}, \overline{Y}_{n_{k-1}}\right)\left(\frac{k}{m_{k-1}+1}\right)^2 q_k$$

$$+ \left\{\rho'_{10}\left(\overline{X}_{m_{k-1}}, \overline{Y}_{n_{k-1}}\right)\left(\overline{X}_{m_{k-1}} - \mu\right)\right\}^2\left(\frac{k}{m_{k-1}+1}\right)^2 q_k(1-q_k)$$

$$+ \rho'_{01}\left(\overline{X}_{m_{k-1}}, \overline{Y}_{n_{k-1}}\right)\nabla^2\varphi(\omega)\,\rho_{01}\left(\overline{X}_{m_{k-1}}, \overline{Y}_{n_{k-1}}\right)\left(\frac{k}{n_{k-1}+1}\right)^2 (1-q_k)$$

$$+ \left\{\rho'_{01}\left(\overline{X}_{m_{k-1}}, \overline{Y}_{n_{k-1}}\right)\left(\overline{Y}_{n_{k-1}} - \nu\right)\right\}^2\left(\frac{k}{n_{k-1}+1}\right)^2 q_k(1-q_k)$$

$$- 2\rho'_{10}\left(\overline{X}_{m_{k-1}}, \overline{Y}_{n_{k-1}}\right)\left(\overline{X}_{m_{k-1}} - \mu\right)\rho'_{01}\left(\overline{X}_{m_{k-1}}, \overline{Y}_{n_{k-1}}\right)\left(\overline{Y}_{n_{k-1}} - \nu\right)$$

$$\times \left(\frac{k}{m_{k-1}+1}\right)\left(\frac{k}{n_{k-1}+1}\right)q_k(1-q_k)$$

on A_k. As in the proof of Proposition 3, the following hold with probability 1:

$$\rho_{10}\left(\overline{X}_{m_{k-1}}, \overline{Y}_{n_{k-1}}\right) \to \rho_{10}(\mu,\nu), \qquad \rho_{01}\left(\overline{X}_{m_{k-1}}, \overline{Y}_{n_{k-1}}\right) \to \rho_{01}(\mu,\nu),$$

$m_k/k \to \rho$, $n_k/k \to 1-\rho$ and $q_k \to \rho$. Result (a) now follows easily since $(\overline{X}_{m_{k-1}} - \mu)^2$ and $(\overline{Y}_{n_{k-1}} - \nu)^2$ are $o(1)$.

(b) Let

$$\tilde{W}_k = \mathbb{E}\left\{\left(k\tilde{\rho}_k - \mathbb{E}\{k\tilde{\rho}_k \mid \mathscr{F}_{k-1}\}\right)^2 \mid \mathscr{F}_{k-1}\right\}.$$

It is sufficient to show uniform integrability of $|\tilde{W}_k|$. Uniform integrability follows if $\sup_k \mathbb{E}\{|\tilde{W}_k|^2\} < \infty$. To see this, let $\varepsilon > 0$ be such that condition (vi) is satisfied. Let A_k be defined as in (2). Then

$$\mathbb{E}\{\tilde{W}_k^2\} = \int_{A_k^c} \tilde{W}_k^2\, d\mathbb{P} + \int_{A_k} \tilde{W}_k^2\, d\mathbb{P} = \mathrm{I} + \mathrm{II}.$$

Here

$$\int_{A_k^c} \tilde{W}_k^2\, d\mathbb{P} \le k^4 \mathbb{P}(A_k^c) = o(1)$$

as $k \to \infty$. Next, since ρ is differentiable on R [defined in (vi)], there is a constant C for which

$$\int_{A_k} \tilde{W}_k^2\, d\mathbb{P} \le C\int_{A_k}\left\{\left(\frac{k}{m_k}\right)^2\left[1 + \|\overline{X}_{m_{k-1}} - \mu\|^2\right]\right.$$

$$\left. + \left(\frac{k}{n_k}\right)^2\left[1 + \|\overline{Y}_{n_{k-1}} - \nu\|^2\right]\right\}^2 d\mathbb{P}$$

$$= O(1)$$

as $k \to \infty$, where the last equality follows from Lemma 3.

Let

(8) $$\hat{U}_k = k\hat{\rho}_k - \mathbb{E}\{k\hat{\rho}_k \mid \mathscr{F}_{k-1}\},$$

(9) $$\tilde{U}_k = k\tilde{\rho}_k - \mathbb{E}\{k\tilde{\rho}_k \mid \mathscr{F}_{k-1}\}.$$

Then

$$\left| \mathbb{E}\{\hat{U}_k^2 \mid \mathscr{F}_{k-1}\} - \mathbb{E}\{\tilde{U}_k^2 \mid \mathscr{F}_{k-1}\} \right|$$

$$= \left| \mathbb{E}\left\{ 2\tilde{U}_k(\hat{U}_k - \tilde{U}_k) + (\hat{U}_k - \tilde{U}_k)^2 \mid \mathscr{F}_{k-1} \right\} \right|$$

$$\leq 2\sqrt{\mathbb{E}\{\tilde{U}_k^2 \mid \mathscr{F}_{k-1}\}} \sqrt{\mathbb{E}\left\{ (\hat{U}_k - \tilde{U}_k)^2 \mid \mathscr{F}_{k-1} \right\}} + \mathbb{E}\left\{ (\hat{U}_k - \tilde{U}_k)^2 \mid \mathscr{F}_{k-1} \right\}$$

by the Schwarz inequality. So, again using the Schwarz inequality,

$$\mathbb{E}\left| \mathbb{E}\{\hat{U}_k^2 \mid \mathscr{F}_{k-1}\} - \mathbb{E}\{\tilde{U}_k^2 \mid \mathscr{F}_{k-1}\} \right|$$

$$\leq 2\sqrt{\mathbb{E}\{\tilde{U}_k^2\}} \sqrt{\mathbb{E}\left\{ (\hat{U}_k - \tilde{U}_k)^2 \right\}} + \mathbb{E}\left\{ (\hat{U}_k - \tilde{U}_k)^2 \right\}$$

$$\to 0$$

as $k \to \infty$, by (b) and Lemma 5. Part (c) now follows from (b), and (d) follows from the Markov inequality. □

LEMMA 6.

$$\lim_{k \to \infty} \mathbb{E}\left\{ (Z_k - \mu_k)(k\tilde{\rho}_k - \mathbb{E}\{k\tilde{\rho}_k \mid \mathscr{F}_{k-1}\}) \mid \mathscr{F}_{k-1} \right\} = 0 \quad w.p.1$$

and

$$\lim_{k \to \infty} \mathbb{E}\left| \mathbb{E}\left\{ (Z_k - \mu_k)(k\hat{\rho}_k - \mathbb{E}\{k\hat{\rho}_k \mid \mathscr{F}_{k-1}\}) \mid \mathscr{F}_{k-1} \right\} \right| = 0.$$

PROOF. Using (8) and $\mathbb{E}\{X_{m_k} - \mu \mid \delta_k, \mathscr{F}_{k-1}\} = 0 = \mathbb{E}\{Y_{n_k} - \nu \mid \delta_k, \mathscr{F}_{k-1}\}$,

$$\mathbb{E}\left\{ (Z_k - \mu_k)(k\tilde{\rho}_k - \mathbb{E}\{k\tilde{\rho}_k \mid \mathscr{F}_{k-1}\}) \mid \mathscr{F}_{k-1} \right\}$$

$$= \mathbb{E}\left\{ (\delta_k - q_k)\left[\rho_{10}(\overline{X}_{m_{k-1}}, \overline{Y}_{n_{k-1}}) \cdot \left(\frac{k}{m_{k-1} + 1} \right)(q_k - \delta_k)(\overline{X}_{m_{k-1}} - \mu) \right.\right.$$

$$+ \rho_{01}(\overline{X}_{m_{k-1}}, \overline{Y}_{n_{k-1}})$$

$$\left.\left. \times \left(\frac{k}{n_{k-1} + 1} \right)(\delta_k - q_k)(\overline{Y}_{n_{k-1}} - \nu) \right] \mid \mathscr{F}_{k-1} \right\}$$

$$= \rho_{10}(\overline{X}_{m_{k-1}}, \overline{Y}_{n_{k-1}}) \cdot \left(\frac{-k}{m_{k-1} + 1} \right)q_k(1 - q_k)(\overline{X}_{m_{k-1}} - \mu)$$

$$+ \rho_{01}(\overline{X}_{m_{k-1}}, \overline{Y}_{n_{k-1}}) \cdot \left(\frac{k}{n_{k-1} + 1} \right)q_k(1 - q_k)(\overline{Y}_{m_{k-1}} - \nu)$$

on A_k. The result now follows as in the proof of part (a) and by noting that the terms of $(\overline{X}_{m_{k-1}} - \mu)$ and $(\overline{Y}_{n_{k-1}} - \nu)$ are $o(1)$. This establishes the first assertion, and the second follows from Lemma 5, as in the proof of parts (c) and (d). □

COROLLARY 2.

$$\mathbb{E}\{(Z_k - \mu_k)(k\hat{\rho}_k - \mathbb{E}\{k\hat{\rho}_k \mid \mathscr{F}_{k-1}\}) \mid \mathscr{F}_{k-1}\} \to^p 0.$$

6. Central limit theorem for the martingale. Asymptotic normality of M_n will be proved using the Cramér–Wold device and the martingale central limit theorem.

THEOREM 3.   *Under conditions* (i)–(vi),

$$\left(\frac{1}{n^{\alpha+1/2}}M_{1,n}, \frac{1}{\sqrt{n}}M_{2,n}\right)' \Rightarrow \mathscr{N}(\mathbf{0}, \Sigma_1) \quad as\ n \to \infty,$$

where

$$\Sigma_1 = \begin{pmatrix} \dfrac{\beta^2}{2\alpha+1}\rho(1-\rho) + \dfrac{\gamma^2\beta^2}{\alpha^2(2\alpha+1)}\left(\dfrac{\sigma^2}{\rho} + \dfrac{\tau^2}{1-\rho}\right) & \dfrac{-\gamma\beta}{\alpha(\alpha+1)}\left(\dfrac{\sigma^2}{\rho} + \dfrac{\tau^2}{1-\rho}\right) \\ \dfrac{-\gamma\beta}{\alpha(\alpha+1)}\left(\dfrac{\sigma^2}{\rho} + \dfrac{\tau^2}{1-\rho}\right) & \dfrac{\sigma^2}{\rho} + \dfrac{\tau^2}{1-\rho} \end{pmatrix}$$

and $\Rightarrow$ *denotes convergence in distribution* $[\rho = \rho(\mu, \nu),$ *and* $\alpha, \beta, \gamma, \sigma^2$ *and* τ^2 *are defined in Proposition* 2, (6) *and Proposition* 4].

PROOF.   By the Cramér–Wold device it suffices to show that

(10)
$$W_n = t_1\frac{M_{1,n}}{n^{1/2+\alpha}} + t_2\frac{M_{2,n}}{\sqrt{n}}$$

is asymptotically normal with mean zero and variance $\eta^2 = t_1^2\sigma_{11} + 2t_1t_2\sigma_{12} + t_2^2\sigma_{22}$ for arbitrary $t_1, t_2 \in \mathbb{R}$, where the σ_{ij} denote the entries of Σ_1; since (10) is a martingale, asymptotic normality follows from the martingale central limit theorem, if the conditions of that theorem are satisfied. Writing

$$\Delta M_{1,k} = M_{1,k} - M_{1,k-1} = c_k(Z_k - \mu_k) - \left(\gamma\sum_{j=m+1}^{k}\frac{c_j}{j-1}\right)\hat{U}_k$$

and

$$\Delta M_{2,k} = M_{2,k} - M_{2,k-1} = \hat{U}_k,$$

for $k \geq t + 1$, where $\hat{U}_k$ is defined in (8), these conditions may be written

(11)
$$\sum_{k=m+1}^{n}\mathbb{E}\left\{\left(t_1\frac{\Delta M_{1,k}}{n^{1/2+\alpha}} + t_2\frac{\Delta M_{2,k}}{\sqrt{n}}\right)^2 \middle| \mathscr{F}_{k-1}\right\} \to^p \eta^2$$

and

$$
(12) \quad \sum_{k=m+1}^{n} \mathbb{E}\Biggl\{ \Biggl(t_1 \frac{\Delta M_{1,k}}{n^{1/2+\alpha}} + t_2 \frac{\Delta M_{2,k}}{\sqrt{n}} \Biggr)^2
$$

$$
\times I\Biggl\{ \Biggl| t_1 \frac{\Delta M_{1,k}}{n^{1/2+\alpha}} + t_2 \frac{\Delta M_{2,k}}{\sqrt{n}} \Biggr| > \delta \Biggr\} \Bigg| \mathscr{F}_{k-1} \Biggr\} \to^p 0
$$

as $n \to \infty$ for all $\delta > 0$, where $I\{\cdot\}$ denotes indicator of $\{\cdot\}$. [See, e.g., Hall and Heyde (1980), pages 58–63.] The following relation, a direct consequence of (6), is used in the verification of (11) and (12):

$$
(13) \quad \gamma \sum_{j=m+1}^{k} \frac{c_j}{j-1} \sim \frac{\gamma\beta}{\alpha} k^\alpha.
$$

Relation (12) is easy to verify. In fact, $|Z_k - \mu_k| \le 1$ w.p.1, for all $k = 1, 2, \ldots,$ and $\mathbb{E}\{\hat{U}_k^2 | \mathscr{F}_{k-1}\}$, $k \ge m$, are uniformly integrable by Proposition 4(c). Moreover, $c_n/n^{\alpha+1/2} \to 0$ by (6) and $\sum_{j=m+1}^{k} c_j/(j-1) = O(k^\alpha)$ by (13). Relation (12) follows easily from these observations, by taking the expectation of the left-hand side of (12).

For (11), observe first that, by Proposition 4, Corollary 2 and (6) and (13),

$$
\frac{1}{n^{1+2\alpha}} \sum_{k=m+1}^{n} \mathbb{E}\{\Delta M_{1,k}^2 | \mathscr{F}_{k-1}\}
$$

$$
= \frac{1}{n^{1+2\alpha}} \sum_{k=m+1}^{n} \mathbb{E}\Biggl\{ \Biggl(c_k(Z_k - \mu_k) - \Biggl(\gamma \sum_{j=m+1}^{k} \frac{c_j}{j-1} \Biggr) \hat{U}_k \Biggr)^2 \Bigg| \mathscr{F}_{k-1} \Biggr\}
$$

$$
= \frac{1}{n^{1+2\alpha}} \sum_{k=m+1}^{n} \Biggl(c_k^2 q_k(1-q_k)
$$

$$
+ \Biggl(\gamma \sum_{j=m+1}^{k} \frac{c_j}{j-1} \Biggr)^2 \mathbb{E}\{\hat{U}_k^2 | \mathscr{F}_{k-1}\} \Biggr) + o_p(1)
$$

$$
\to^p \frac{\beta^2}{2\alpha+1} \rho(1-\rho) + \Biggl(\frac{\gamma\beta}{\alpha} \Biggr)^2 \Biggl(\frac{1}{2\alpha+1} \Biggr) \Biggl(\frac{\sigma^2}{\rho} + \frac{\tau^2}{1-\rho} \Biggr),
$$

$$
\frac{1}{n^{1+\alpha}} \sum_{k=m+1}^{n} \mathbb{E}\{\Delta M_{1,k} \Delta M_{2,k} | \mathscr{F}_{k-1}\}
$$

$$
= \frac{1}{n^{1+\alpha}} \sum_{k=m+1}^{n} \Biggl(\mathbb{E}\{c_k(Z_k - \mu_k)\hat{U}_k | \mathscr{F}_{k-1}\}
$$

$$
- \Biggl(\gamma \sum_{j=m+1}^{k} \frac{c_j}{j-1} \Biggr) \mathbb{E}\{\hat{U}_k^2 | \mathscr{F}_{k-1}\} \Biggr)
$$

$$
\to^p \frac{-\gamma\beta}{\alpha(\alpha+1)} \Biggl(\frac{\sigma^2}{\rho} + \frac{\tau^2}{1-\rho} \Biggr)
$$

and

$$\frac{1}{n} \sum_{k=m+1}^{n} \mathbb{E}\{\Delta M_{2,k}^2 \mid \mathscr{F}_{k-1}\} = \frac{1}{n} \sum_{k=m+1}^{n} \mathbb{E}\{\hat{U}_k^2 \mid \mathscr{F}_{k-1}\}$$

$$\to^p \frac{\sigma^2}{\rho} + \frac{\tau^2}{1-\rho}. \qquad \square$$

In fact, the proof just given establishes a stronger assertion.

THEOREM 4. *Let*

$$\mathbb{M}_n(t) = \begin{pmatrix} \dfrac{1}{n^{\alpha+1/2}} M_{1,[nt]} \\ \dfrac{1}{\sqrt{n}} M_{2,[nt]} \end{pmatrix}, \qquad 0 \le t \le 1, n \ge m,$$

and

$$\Sigma_t = \begin{bmatrix} \sigma_{11} t^{2\alpha+1} & \sigma_{12} t^{\alpha+1} \\ \sigma_{12} t^{\alpha+1} & \sigma_{22} t \end{bmatrix},$$

for $0 \le t \le 1$, where σ_{ij} are the entries in Σ_1 and $[x]$ denotes the greatest integer which is less than or equal to x. Let $\mathbb{M}(t)$, $0 \le t \le 1$, be a process with independent increments for which $\mathbb{M}(t)$ is normal with mean 0 and covariance matrix Σ_t for $0 \le t \le 1$. Then $\mathbb{M}_n \Rightarrow \mathbb{M}$ as $n \to \infty$ in $D^2[0,1]$.

PROOF. Write $\mathbb{M}_n(t) = [\mathbb{M}_{1,n}(t), \mathbb{M}_{2,n}(t)]'$, $0 \le t \le 1$, $n \ge 1$. For fixed $\alpha_1, \alpha_2 \in \mathbb{R}$, it follows from (11), (12) and Durrett and Resnick [(1978), Theorem 2.5] that $a_1 \mathbb{M}_{1,n} + a_2 \mathbb{M}_{2,n}$ converges in distribution to $a_1 \mathbb{M}_1 + a_2 \mathbb{M}_2$ in the topology of $D[0,1]$. That the finite-dimensional distributions of $\mathbb{M}_n$ converge to those of $\mathbb{M}$ follows directly. Moreover, by setting $a_1 = 1$ and $a_2 = 0$ (respectively, $a_1 = 0$ and $a_2 = 1$), it follows that $\mathbb{M}_{1,n}$, $n \ge 1$, and $\mathbb{M}_{2,n}$, $n \ge 1$, are tight in $D[0,1]$. That $\mathbb{M}_n = [\mathbb{M}_{1,n}, \mathbb{M}_{2,n}]$ is tight in $D^2[0,1]$ follows easily. $\square$

7. A central limit theorem. The magnitude of the remainder term, R_n, is determined next.

LEMMA 7.

$$\max_{k \le n} \left| \left(\frac{1}{n^{\alpha+1/2}}, \frac{1}{\sqrt{n}} \right) \cdot R_k \right| \to 0 \quad \text{in probability as } n \to \infty,$$

where $R_n = \sum_{k=m+1}^{n} D_k r_k$ (and the components of r_k are defined in Propositions 2 and 3).

PROOF. It suffices to show that the left-hand side approaches zero in L^1. Recall from (13) that $\gamma \sum_{j=m+1}^{k} c_j/(j-1) = O(k^\alpha)$, and observe that

$$\left(\frac{1}{n^{\alpha+1/2}}, \frac{1}{\sqrt{n}} \right) \cdot R_n = \frac{1}{n^{\alpha+1/2}} \sum_{k=m+1}^{n} \left(c_k r_{1,k} - \left(\gamma \sum_{j=m+1}^{k} \frac{c_j}{j-1} \right) r_{2,k} \right)$$

$$+ \frac{1}{\sqrt{n}} \sum_{k=m+1}^{n} r_{2,k}.$$

That

$$\mathbb{E} \left\{ \sum_{k=m+1}^{n} |r_{2,k}| \right\} = o_p(\sqrt{n})$$

and

$$\mathbb{E} \left\{ \sum_{k=m+1}^{n} k^\alpha |r_{2,k}| \right\} = o(n^{\alpha+\frac{1}{2}}) \quad \text{as } n \to \infty$$

follow directly from Proposition 3. For $r_{1,k}$, let $0 < \varepsilon < 1$ be given. Then there is a $\delta > 0$ for which $r(x,y) \le \varepsilon \|(x-\rho, y-\rho)\|$ for all $\|(x-\rho, y-\rho)\| \le \delta$. Since $|r(x,y)| \le \alpha + \gamma + 1$ for all x and y, it follows that

$$\mathbb{E} \left\{ \sum_{k=m+1}^{n} |k^\alpha r_{1,k}| \right\} \le \sum_{k=m+1}^{n} k^\alpha \left[\varepsilon \mathbb{E} \left\{ \left\| \left(\frac{S_{k-1}}{k-1}, \hat{\rho}_{k-1} - \rho \right) \right\| \right\} \right.$$

$$\left. + (\alpha + \gamma + 1) \mathbb{P} \left\{ \left\| \left(\frac{S_{k-1}}{k-1}, \hat{\rho}_{k-1} - \rho \right) \right\| > \delta \right\} \right]$$

$$\le \sum_{k=m}^{n} (k+1)^\alpha \left[\varepsilon \mathbb{E} \left\{ \left(\left(\frac{S_k}{k} \right)^2 + (\hat{\rho}_k - \rho)^2 \right)^{1/2} \right\} \right.$$

$$\left. + \frac{(\alpha + \gamma + 1)}{\delta^2} \mathbb{E} \left\{ \left(\frac{S_k}{k} \right)^2 + (\hat{\rho}_k - \rho)^2 \right\} \right]$$

$$= \sum_{k=m}^{n} (k+1)^\alpha \left[\varepsilon \frac{C_1}{\sqrt{k}} + \frac{C_2}{\delta^2 k} \right]$$

$$\sim \frac{C_1 \beta \varepsilon}{\alpha + 1} n^{\alpha+1/2},$$

as $n \to \infty$, for some constants C_1 and C_2, where the last equality follows from Lemma 1 and Proposition 1. The lemma follows. $\square$

Asymptotic normality of $(1/\sqrt{n})V_n$ may be established from Lemma 7 and Theorem 3 using Slutsky's theorem.

THEOREM 5. *Under conditions* (i)–(vi),

$$\frac{1}{\sqrt{n}} V_n \Rightarrow \mathcal{N}(0, \Sigma) \quad \text{as } n \to \infty,$$

where

$$\Sigma = \begin{pmatrix} \dfrac{\rho(1-\rho)}{2\alpha+1} + \dfrac{2\gamma^2}{(\alpha+1)(2\alpha+1)}\left(\dfrac{\sigma^2}{\rho} + \dfrac{\tau^2}{1-\rho}\right) & \dfrac{\gamma}{\alpha+1}\left(\dfrac{\sigma^2}{\rho} + \dfrac{\tau^2}{1-\rho}\right) \\[4mm] \dfrac{\gamma}{\alpha+1}\left(\dfrac{\sigma^2}{\rho} + \dfrac{\tau^2}{1-\rho}\right) & \dfrac{\sigma^2}{\rho} + \dfrac{\tau^2}{1-\rho} \end{pmatrix}$$

and $\Rightarrow$ *denotes convergence in distribution.*

PROOF.

$$(14) \qquad \frac{V_n}{\sqrt{n}} = \frac{D_n^{-1}D_n V_n}{\sqrt{n}} = \frac{D_n^{-1}}{\sqrt{n}}(D_m V_m + M_n + R_n).$$

Here

$$\frac{D_n^{-1}}{\sqrt{n}} = \begin{pmatrix} \dfrac{1}{\sqrt{n}\,c_n} & \dfrac{\gamma}{\sqrt{n}\,c_n}\displaystyle\sum_{j=m+1}^{n}\left(\dfrac{c_j}{j-1}\right) \\[4mm] 0 & \dfrac{1}{\sqrt{n}} \end{pmatrix}.$$

By Lemma 7 and equations (6) and (13), the limiting distribution of $V_n/\sqrt{n}$ is the same as the limiting distribution of

$$(15) \qquad \begin{pmatrix} 1 & \gamma \\ \dfrac{1}{\beta} & \alpha \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \dfrac{1}{n^{\alpha+1/2}}M_{1,n} \\[3mm] \dfrac{1}{\sqrt{n}}M_{2,n} \end{pmatrix}.$$

By Theorem 3, the limiting distribution of (15) is bivariate normal with mean **0** and covariance matrix

$$\begin{pmatrix} 1 & \gamma \\ \dfrac{1}{\beta} & \alpha \\ 0 & 1 \end{pmatrix}\Sigma_1\begin{pmatrix} \dfrac{1}{\beta} & 0 \\[2mm] \dfrac{\gamma}{\alpha} & 1 \end{pmatrix} = \Sigma. \qquad\qquad \square$$

COROLLARY 3.

$$\frac{S_n}{\sqrt{n}} = \frac{1}{\sqrt{n}}(m_n - n\rho) \Rightarrow \mathcal{N}\!\left(0, \frac{\rho(1-\rho)}{2\alpha+1} + \frac{2\gamma^2}{(\alpha+1)(2\alpha+1)}\left(\frac{\sigma^2}{\rho} + \frac{\tau^2}{1-\rho}\right)\right).$$

THEOREM 6. *Let*

$$\mathbb{V}_n(t) = \frac{1}{\sqrt{n}}V_{[nt]}, \qquad 0 \le t \le 1, n \ge 1,$$

and

$$\mathbb{V}(t) = \begin{pmatrix} \dfrac{1}{\beta t^\alpha} & \dfrac{\gamma}{\alpha} \\ 0 & 1 \end{pmatrix} \mathbb{M}(t), \qquad 0 \le t \le 1,$$

where $\mathbb{M}(t)$ is as in Theorem 4. Then $\mathbb{V}_n \Rightarrow \mathbb{V}$ in $D^2[0,1]$.

PROOF. By (14),

$$\mathbb{V}_n(t) = \frac{1}{\sqrt{n}} D_{[nt]}^{-1}\big(D_m V_m + M_{[nt]} + R_{[nt]}\big),$$

for all $0 < t \le 1$ and $n \ge 2m/t$. So, by Lemma 7 and equations (6), (13) and (14),

$$\lim_{n \to \infty} \mathbb{E}\left\{ \sup_{\delta \le t < 1} \left| \mathbb{V}_n(t) - \begin{pmatrix} \dfrac{1}{\beta t^\alpha} & \dfrac{\gamma}{\alpha} \\ 0 & 1 \end{pmatrix} \mathbb{M}_n(t) \right| \right\} = 0 \quad \forall\, 0 < \delta \le 1.$$

Also,

$$(16) \qquad \sup_k \mathbb{E}\left\{ \sup_{0 \le t \le \delta} |\mathbb{V}_k(t)| \right\} = O(\sqrt{\delta}) \quad \text{as } \delta \searrow 0.$$

To verify (16), let r be the least integer for which $2^r \ge n$. Then there is a constant C for which

$$\mathbb{E}\left\{ \max_{k \le n} \|D_k^{-1} R_k\| \right\} \le C \sum_{k=m}^{n} \mathbb{E}\{k^{-\alpha}|r_{1,k}| + |r_{2,k}|\} = O(\sqrt{n})$$

and

$$\mathbb{E}\left\{ \max_{k \le n} \|D_k^{-1} M_k\| \right\} \le C \sum_{j=1}^{r} \mathbb{E}\left\{ \max_{k \le 2^j} 2^{-\alpha j} |M_{1,k}| + |M_{2,k}| \right\} = O(\sqrt{n}),$$

as $n \to \infty$, by Propositions 2 and 3 and Doob's (1953) maximal inequality. Relation (16) and Theorem 6 follow. $\square$

8. Example: normal responses.

Suppose it is desired to design a sequential procedure, with a randomized allocation scheme, for the fixed width interval estimation of the difference of the means of two populations. Minimizing the total size of the experiment can be accomplished by designing the sequential procedure so that subjects are allocated to the two treatments in the correct proportions.

More formally, assume that $X_1, X_2, \ldots$ and $Y_1, Y_2, \ldots$ are independent random variables for which

$$X_1, X_2, \ldots \sim \mathcal{N}(\mu, \sigma^2) \quad \text{and} \quad Y_1, Y_2, \ldots \sim \mathcal{N}(\nu, \tau^2),$$

where the four parameters μ, ν, σ and τ are unknown. Here, $X_1, X_2, \ldots$ denote responses to treatment A, and $Y_1, Y_2, \ldots$ denote responses to treat-

ment B. These could be, for example, blood pressure readings. Then the correct allocation proportions for minimizing the total sample size and retaining preassigned coverage probability and interval width are [see Robbins, Simons and Starr (1967) or Eisele (1990)] $\sigma/(\sigma + \tau) \times k$ to treatment A and $\tau/(\sigma + \tau) \times k$ to treatment B. Thus,

$$\rho(\sigma^2, \tau^2) = \frac{\sigma}{\sigma + \tau}.$$

Taking

$$\hat{\sigma}_{m_k}^2 = (m_k - 1)^{-1} \sum_{i=1}^{m_k} \left(X_i - \overline{X}_{m_k}\right)^2 \quad \text{and} \quad \hat{\tau}_{n_k}^2 = (n_k - 1)^{-1} \sum_{i=1}^{n_k} \left(Y_i - \overline{Y}_{n_k}\right)^2$$

to be the usual estimates of σ^2 and τ^2 gives

$$\hat{\rho}_k = \frac{\hat{\sigma}_{m_k}}{\hat{\sigma}_{m_k} + \hat{\tau}_{n_k}}.$$

The sequential procedure can now be described as follows: to start, take $n_0 \geq 2$ observations on X and on Y. Then, if at any stage there are m_k observations on X and n_k on Y, with $k = m_k + n_k \geq 2n_0$, take observation $k + 1$ on X if

$$U_{k+1} \leq q_{k+1} \equiv q\left(\frac{m_k}{k}, \hat{\rho}_k\right).$$

Otherwise, take observation $k + 1$ on Y.

If the desired width and coverage probability of the confidence interval are $2h$ and α, respectively, and if the constant a is defined by $2\Phi(a) - 1 = \alpha$, where Φ denotes the $\mathcal{N}(0, 1)$ distribution function, then a possible stopping rule for the sequential procedure is: stop after N observations, where

$$N = \inf\left\{k \geq 2n_0 : \frac{\hat{\sigma}_{m_k}^2}{m_k} + \frac{\hat{\tau}_{n_k}^2}{n_k} \leq \left(\frac{h}{a_k}\right)^2\right\}$$

and $\{a_k\}$ is a given sequence of positive constants such that $a_k \to a$ as $k \to \infty$.

THEOREM 7 (Central limit theorem). *Under conditions* (i)–(vi), *as* $h \to 0$, *for all* μ, ν, σ *and* τ,

$$\frac{1}{\sqrt{N}}(m_N - N\rho)$$

$$\Rightarrow \mathcal{N}\left(0, \frac{\sigma\tau}{(\sigma + \tau)^2(2\alpha + 1)} + \frac{2\gamma^2}{(\alpha + 1)(2\alpha + 1)}\left(\frac{\sigma^3 + \tau^3}{4\sigma\tau(\sigma + \tau)^3}\right)\right).$$

PROOF. Theorem 7 is a special case of Corollary 3. To find the variance, note that

$$\rho_{10}(\sigma^2, \tau^2) = \frac{\tau}{2\tau(\sigma + \tau)^2}$$

and

$$\rho_{01}(\sigma^2, \tau^2) = \frac{-\sigma}{2\tau(\sigma + \tau)^2}.$$

Now apply Corollary 3 to get the variance. □

For more details on this sequential procedure, including derivations, other stopping rules, asymptotic properties and simulation results, see Eisele (1990).

REFERENCES

BROWN, L. (1986). *Fundamentals of Statistical Exponential Families*. IMS, Hayward, CA.

DOOB, J. L. (1953). *Stochastic Processes*. Wiley, New York.

DURRETT, R. and RESNICK, S. (1978). Functional central limit theorems for dependent random variables. *Ann. Probab.* **6** 829–847.

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

EISELE, J. (1990). An adaptive biased coin design for the Behrens–Fisher problem. *Sequential Anal.* **9** 343–359.

EISELE, J. (1994). The doubly adaptive biased coin design for sequential clinical trials. *J. Statist. Plann. Inference* **38** 249–262.

EISELE, J. and WOODROOFE, M. (1990). A martingale proof of normality for Wei's biased coin. Technical Report 183, Dept. Statistics, Univ. Michigan.

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

HARDWICK, J. (1989). Recent progress in clinical trial designs that adapt for ethical purposes. *Statist. Sci.* **4** 327–336.

POCOCK, S. J. (1979). Allocation of patients to treatment in clinical trials. *Biometrics* **35** 183–197.

ROBBINS, H., SIMONS, G. and STARR, N. (1967). A sequential analogue of the Behrens–Fisher problem. *Ann. Math. Statist.* **38** 1384–1391.

SMITH, R. L. (1984). Properties of biased coin designs in sequential clinical trials. *Ann. Statist.* **12** 1018–1034.

WEI, L. J. (1978). The adaptive biased coin design for sequential experiments. *Ann. Statist.* **6** 92–100.

BIOMETRICS
CIBA-GEIGY AG
CH-4002, BASEL
SWITZERLAND

DEPARTMENT OF STATISTICS
UNIVERSITY OF MICHIGAN
ANN ARBOR, MICHIGAN 48109