The Annals of Statistics
1989, Vol. 17, No. 3, 1087–1102

VERY WEAK EXPANSIONS FOR SEQUENTIALLY DESIGNED EXPERIMENTS: LINEAR MODELS[1]

By Michael Woodroofe

University of Michigan

In sequentially designed experiments with linear models, each design variable may depend on previous responses. The use of such sequential designs does not affect the likelihood function or the functional form of the maximum likelihood estimator, but it may affect sampling distributions. In this paper, asymptotic expansions for sampling distributions are obtained. The expansions are very weak ones in which a confidence curve (a function of the unknown parameters) is replaced by a confidence functional defined on a class of prior distributions. The proofs use a version of Stein's identity.

1. Introduction. Consider an adaptive linear model of the form

$$y_k = x_k'\theta + e_k, \qquad k = 1, 2, \ldots, \tag{1}$$

where $e_1, e_2, \ldots$ are i.i.d. standard normal random variables, $\theta = (\theta_1, \ldots, \theta_p)'$ is a (column) vector of unknown parameters, with values in an open subset Ω of R^p and the prime denotes transpose. Here "adaptive" means that x_k may be of the form

$$x_k = x_k(w_1, \ldots, w_k, y_1, \ldots, y_{k-1}) \in R^p, \tag{2}$$

for $k = 1, 2, \ldots$, where $w_1, w_2, \ldots$ are independent of $e_1, e_2, \ldots$ and have a (joint) distribution which is independent of θ. The $w_1, w_2, \ldots$ may represent auxiliary randomization and/or covariates. Thus, if x_k is thought of as a design variable, then the model allows the design variables to depend (measurably) on previous responses, auxiliary randomization and/or covariates. This model is quite broad and there has been substantial interest in it; see, for example, Wei (1979), Lai and Wei (1982) and Wu (1985) and their references.

The primary objective of the paper is to find approximations to the sampling distributions of maximum likelihood estimators for models of the form (1) and (2). Attention is restricted to cases in which these estimators are asymptotically normal, after suitable standardization, but even then, normal approximation may not have good numerical accuracy, due to the adaptive nature of the design. Woodroofe and Keener (1987) provide an example in which optional stopping may have a dramatic effect on sampling distributions. Refined approximations may be obtained in the form of asymptotic expansions.

The expansions presented here are very weak ones, as in Stein (1985) and Woodroofe (1986), in which a confidence curve (a function of θ) is replaced by a confidence functional defined on a class of prior distributions; see Section 5 for

Received October 1987; revised June 1988.

[1]Research supported by National Science Foundation Grant DMS-84-13452.

AMS 1980 *subject classifications*. Primary 62E20; secondary 62F12, 62L05.

Key words and phrases. Martingale convergence theorem, maximum likelihood estimators, posterior distributions, Stein's identity.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics*. www.jstor.org

the precise formulation. In order to derive these expansions, it is necessary to develop asymptotic expansions for posterior distributions which may be integrated with respect to the marginal distribution of the data. This question is of independent interest; it has been considered by Ghosh, Sinha and Joshi (1982), Rehalia (1983) and by Bickel, Goetze and van Zwet (1985). Here it is discussed in Section 4.

Applications to a sequential design proposed by Robbins and Siegmund (1974) are described in Section 6 and compared to simulations in Section 7. Extensions to the case of unknown variability are presented in Section 9.

The objectives of this paper are similar to those of Ghosh, Sinha and Joshi (1982), Bickel, Goetze and van Zwet (1985) and Woodroofe (1986), but the derivations are quite different. Here a version of Stein's identity is used in place of Taylor series approximations. This leads to less restrictive conditions on the prior [for the model (1) and (2)].

2. Stein's identity. A real valued, measurable function f defined on $\mathbb{R}^p$ is said to be almost differentiable iff there is a function ∇f from $\mathbb{R}^p$ into $\mathbb{R}^p$ for which

$$f(x+y) - f(x) = \int_0^1 y' \nabla f[x + ty]\, dt$$

for a.e. $x \in \mathbb{R}^p$ for each $y \in \mathbb{R}^p$ [see Stein (1981)].

Let Φ_p denote the standard normal distribution in $\mathbb{R}^p$ and let Ψ denote a finite signed measure of the form

$$\Psi(dz) = f(z)\Phi_p(dz), \tag{3}$$

where f is an almost differentiable function for which ∇f is integrable and $\nabla f = 0$ a.e. on $\{f = 0\}$. If h is a measurable function on $\mathbb{R}^p$, which is integrable with respect to Φ_p and Ψ, let

$$\Phi_p h = \int h\, d\Phi_p \quad \text{and} \quad \Psi h = \int h\, d\Psi.$$

Here and below integrals extend over the entire space unless otherwise indicated.

If h is Φ_p-integrable, then one may define a function

$$g = (g_1, \dots, g_p)' \colon \mathbb{R}^p \to \mathbb{R}^p$$

by

$$g_j(y, z) = e^{z^2/2} \int_z^\infty [h_j(y, w) - h_{j-1}(y)] e^{-w^2/2}\, dw, \tag{4}$$

$$h_{j-1}(y) = \int_{\mathbb{R}^{p-j+1}} h(y, z)\Phi_{p-j+1}(dz) \tag{5}$$

for $z \in \mathbb{R}$, a.e. $y \in \mathbb{R}^{j-1}$, $j = 1, \dots, p$ and $h_p = h$. Here g_j is regarded as a function on $\mathbb{R}^p$, which is constant in its last $p - j$ variables for $j = 1, \dots, p$. For example, if $h(z) = z_1$, $z \in \mathbb{R}^p$, then $g(z) = (1, 0, \dots, 0)'$ for $z \in \mathbb{R}^p$ and if $h(z) = \|z\|^2$, $z \in \mathbb{R}^p$, then $g(z) = z$ for $z \in \mathbb{R}^p$.

PROPOSITION 1. *If h is Φ_p-integrable, then h is Ψ-integrable and*

$$\Psi h - \Phi_p h \cdot \Psi 1 = \int g'\left(\frac{\nabla f}{f}\right) d\Psi.$$

PROOF. If $p = 1$, write Φ for Φ_1 and let φ denote the standard normal density. Then (temporarily letting the prime denote derivative)

$$\begin{aligned}\Psi h - \Phi h \cdot \Psi 1 &= \int_{-\infty}^{\infty}\left\{\int_{-\infty}^{y} f'(x)\,dx\right\}[h(y) - \Phi h]\varphi(y)\,dy \\ &= \int_{-\infty}^{\infty}\left\{\int_{x}^{\infty}[h(y) - \Phi h]\varphi(y)\,dy\right\}f'(x)\,dx \\ &= \int_{-\infty}^{\infty} g(x)f'(x)\Phi(dx),\end{aligned}$$

where the interchange of orders of integration is justified by the assumed integrability of f'. For $p \geq 2$, define f_j by (5) and let π_j denote the projection of $\mathbb{R}^j$ onto $\mathbb{R}^{j-1}$ for $j = 1, \dots, p$. Then for a.e. $y_1, \dots, y_p$,

$$h = \Phi_p h + \sum_{j=1}^{p}(h_j - h_{j-1})$$

and

$$\begin{aligned}\Psi(h_j - h_{j-1}) &= \int_{\mathbb{R}^j}(h_j - h_{j-1}\circ\pi_j)(y)f_j(y)\Phi_j(dy) \\ &= \int_{\mathbb{R}^{j-1}}\left\{\int_{-\infty}^{\infty} g_j(y, z)\frac{\partial}{\partial z}f_j(y, z)\varphi(z)\,dz\right\}\Phi_{j-1}(dy) \\ &= \int_{\mathbb{R}^p} g_j(y_1, \dots, y_j)\frac{\partial}{\partial y_j}f(y_1, \dots, y_p)\Phi_p(dy),\end{aligned}$$

where $(\partial/\partial y_j)f$ denotes the jth component of ∇f for $j = 1, \dots, p$ (and the formal calculations are easily justified). The proposition now follows by summing over $j = 1, \dots, p$. □

It is easily seen that the transformation from h to g is linear and that

$$\sup_y \|g(y)\| \leq \sqrt{(2\pi p)}\,\sup_y |h(y)|, \tag{6}$$

where $\|\cdot\|$ denotes the Euclidean norm of a vector (and later the trace norm of a matrix); see Stein [(1987), Chapter 2]. For later reference, observe that

$$\Phi_p(g) := [\Phi_p(g_1), \dots, \Phi_p(g_p)] = \int y h(y)\Phi_p(dy), \tag{7}$$

provided that the later integral exists and h is Φ_p-integrable.

3. Preliminary lemmas. Let $x_1, y_1, x_2, \ldots$ be as in (1) and (2). Further, let $X_n = (x_1, \ldots, x_n)'$, $Y_n = (y_1, \ldots, y_n)'$ and $\varepsilon_n = (e_1, \ldots, e_n)$, so that the model may be written

$$Y_n = X_n\theta + \varepsilon_n, \qquad n \geq 1.$$

Suppose that $X_n'X_n$ is positive definite w.p. 1 P_θ for all $\theta \in \Omega$ for all sufficiently large n, say $n \geq n_0$, where n_0 is nonrandom and let

$$\hat{\theta}_n = (X_n'X_n)^{-1}X_n'Y_n, \qquad n \geq n_0.$$

It is well known that the likelihood function for the adaptive model (1) and (2) is the same as if the design variables had been predetermined. So, $\hat{\theta}_n$ is the maximum likelihood estimator of θ, whenever it is in Ω and the likelihood function is

$$L_n(\theta) \propto \exp\left\{-\tfrac{1}{2}\|X_n\theta - X_n\hat{\theta}_n\|^2\right\}, \qquad \theta \in \Omega.$$

Now consider a Bayesian model in which there is a random variable Θ with a density ξ, (1) holds conditionally given $\Theta = \theta$ for all $\theta \in \mathbb{R}^p$ and $e_1, e_2, \ldots$, $w_1, w_2, \ldots$ are independent of Θ. Probability and expectation in this model are denoted by P and E, or by P^ξ and E^ξ if there is danger of confusion and conditional expectation given $w_1, x_1, \ldots, w_n, x_n$ is denoted by E^n. Here and below ξ is regarded as a function on $\mathbb{R}^p$ which vanishes off of Ω.

For $n \geq n_0$, $X_n'X_n$ may be written in the form

$$X_n'X_n = B_nB_n',$$

where B_n is a nonsingular $p \times p$ matrix. There are several ways to do this. One is to let B_n be a square root of $X_n'X_n$. Another is to first obtain orthogonal vectors $c_1, \ldots, c_p$ in $\mathbb{R}^n$ from Gram–Schmidt orthogonalization of the columns of X_n. Then $C = (c_1, \ldots, c_p)$ may be written in the form $C = X_nU$, where U is an upper triangular matrix and $X_n'X_n = B_nB_n'$ with $B_n' = U^{-1}$. Let

$$Z_n = B_n'(\Theta - \hat{\theta}_n)$$

and

$$\lambda_n = \text{minimum eigenvalue of } X_n'X_n, \qquad n \geq n_0.$$

If Θ has prior density ξ, then the posterior density of Z_n is

$$\zeta_n(z) \propto \exp\left(-\tfrac{1}{2}\|z\|^2\right)\xi\left[\hat{\theta}_n + B_{n'}^{-1}z\right]$$

for $z \in \mathbb{R}^p$ and $n \geq n_0$. This is of the form (3), so that Stein's identity is applicable.

LEMMA 1. *Suppose that ξ is almost differentiable and $\nabla\xi$ is integrable. If h is Φ_p-integrable, then*

$$E^n[h(Z_n)] = \Phi_p h + E^n\left[g(Z_n)'B_n^{-1}\left(\frac{\nabla\xi}{\xi}\right)(\Theta)\right], \qquad n \geq n_0.$$

Moreover, if

$$\sup_y |h(y)| \le 1,$$

then

$$|E^n[h(Z_n)] - \Phi_p(h)| \le \left(\frac{2\pi p}{\lambda_n}\right)^{1/2} E^n\left[\left\|\frac{\nabla \xi}{\xi}(\Theta)\right\|\right],$$

w.p. 1 *for all* $n \ge n_0$.

PROOF. The first assertion follows directly from Proposition 1. If $|h| \le 1$, then $\|g\| \le (2\pi p)^{1/2}$ by (6) and, therefore,

$$\left|g(Z_n)'B_n^{-1}\left(\frac{\nabla \xi}{\xi}\right)(\Theta)\right| \le \left(\frac{2\pi p}{\lambda_n}\right)^{1/2}\left\|\left(\frac{\nabla \xi}{\xi}\right)(\Theta)\right\|,$$

w.p. 1 for all $n \ge n_0$. The second assertion follows by taking conditional expectations. □

LEMMA 2. *Suppose that ξ is almost differentiable and that $\nabla\xi$ is integrable. If*

$$\lim_n \lambda_n = \infty \quad w.p.\,1\ (P_\theta),\ a.e.\ \theta \in \Omega, \tag{8}$$

then the conditional distribution of Z_n given $w_1, y_1, \dots, w_n, y_n$ converges weakly Φ_p w.p. 1 (P) *as* $n \to \infty$.

PROOF. Since $\nabla\xi$ is integrable, $E^n\{(\nabla\xi/\xi)(\Theta)\}$, $n \ge 1$, is a convergent martingale and, therefore, $\lim_n E^n\|(\nabla\xi/\xi)(\Theta)\|/(\lambda_n)^{1/2} = 0$. So, Lemma 2 follows easily from Lemma 1. □

LEMMA 3. *If* (8) *holds, then* $\hat{\theta}_n \to \theta$ *w.p.* 1 (P_θ) *for a.e.* $\theta \in \Omega$.

PROOF. If ξ is any almost differentiable density for which $\theta\xi$ and $\nabla\xi$ are integrable and ξ vanishes off of Ω, then Z_n, $n \ge n_0$ are stochastically bounded by Lemma 2. It follows that $\hat{\theta}_n \to \Theta$ in probability as $n \to \infty$ and, therefore, that Θ is almost measurable with respect to the sigma-algebra generated by $w_1, y_1, w_2, \dots$. So, $E^n(\Theta) \to \Theta$ w.p. 1 (P) as $n \to \infty$, by the martingale convergence theorem, since $E|\Theta| = \int|\theta|\xi(\theta)\,d\theta$ is finite, by assumption. Moreover, by Lemma 1,

$$E^n(\Theta) - \hat{\theta}_n = (X_n'X_n)^{-1}E^n\left[\left(\frac{\nabla\xi}{\xi}\right)(\Theta)\right],$$

which tends to zero w.p. 1 (P) since $\lambda_n \to \infty$ and $E^n\{(\nabla\xi/\xi)(\Theta)\}$ is a convergent martingale. The lemma follows easily, in view of the arbitrariness of ξ. □

4. Integrable expansions for posterior distributions. Let $t = t_a$, $a \ge 1$, be a family of stopping times with respect to $w_1, y_1, w_2, y_2, \dots$. That is, t_a is a positive integer or ∞ valued random variable for which $t_a < \infty$ w.p. 1 (P_θ) for a.e. $\theta \in \Omega$ and the event $\{t_a = n\}$ is determined by $w_1, \dots, y_n$ for all $n = 1, 2, \dots$ and $a \geq 1$. Suppose that

$$t_a \geq n_0, \qquad \forall\, a \geq 1, \qquad t_a \to \infty \quad \text{w.p. } 1\ (P_\theta) \tag{9}$$

and

$$Q_a := \sqrt{a}\, B_t^{-1} \to Q(\theta) \quad \text{w.p. } 1\ (P_\theta) \tag{10}$$

as $a \to \infty$, for a.e. $\theta \in \Omega$ where $Q(\theta)$, $\theta \in \Omega$, are $p \times p$ matrices. Here and below t is written for t_a to avoid higher order subscripts. Let ξ denote a density on $\mathbb{R}^p$ which vanishes off of Ω and suppose further that

$$a/\lambda_t,\ a \geq 1, \text{ are uniformly integrable w.r.t. } P = P^\xi. \tag{11}$$

The fixed sample size case, $t_a = a = n$, is not excluded.

THEOREM 1. *Let ξ denote an almost differentiable density (prior) with finite Fisher information, that is,*

$$\int \left\| \frac{\nabla \xi}{\xi} \right\|^2 \xi \, d\theta < \infty. \tag{12}$$

Suppose that conditions (8), (9), (10) *and* (11) *are satisfied. Let*

$$R(h, \theta) = \Phi_p(g)' Q(\theta) \left(\frac{\nabla \xi}{\xi} \right)(\theta), \qquad \theta \in \Omega, \tag{13}$$

for bounded measurable functions h defined on $\mathbb{R}^p$, where g and h are related by (4); *see also* (7). *Then*

$$\lim_{a} \operatorname*{ess\,sup}_{|h| \leq 1} \sqrt{a} \left| E^t[h(Z_t)] - \Phi_p(h) - \frac{1}{\sqrt{a}} R(h, \Theta) \right| = 0$$

w.p. 1 *and in the first mean* [*that is, in $L^1(P)$*], *where E^t denotes conditional expectation given $w_1, \dots, y_t$.*

PROOF. Of course, $E^t\{w(\hat{\theta}_t, \Theta)\} = E^n\{w(\hat{\theta}_n, \Theta)\}$ a.e. on $\{t = n\}$ for all $n = 1, 2, \dots$ and all bounded measurable functions w on Ω^2, since t is a stopping time. For $|h| \leq 1$ and $a \geq 1$, let

$$\begin{aligned} R_a(h) &= E^t\left[g(Z_t)' Q_a \left(\frac{\nabla \xi}{\xi} \right)(\Theta) \right] \\ &= E^t[g(Z_t)]' Q_a E^t\left[\left(\frac{\nabla \xi}{\xi} \right)(\Theta) \right] \\ &\quad + E^t\left\{ g(Z_t)' Q_a \left[\frac{\nabla \xi}{\xi}(\Theta) - E^t\left(\frac{\nabla \xi}{\xi}(\Theta) \right) \right] \right\}. \end{aligned} \tag{14}$$

Then $E^t[h(Z_t)] - \Phi_p h = (1/\sqrt{a})R_a(h)$ for fixed h, by Lemma 1, since t is a stopping time. So, the theorem asserts that $\operatorname{ess\,sup}_{|h|\le 1}|R_a(h) - R(h,\Theta)| \to 0$ w.p. 1 and in the first mean. Convergence w.p. 1 is clear, since ($|x+y-z| \le |y| + |x-v| + |v-w| + |w-z|$),

$$
\begin{aligned}
|R_a(h) - R(h,\Theta)| \le{}& \left|E^t\left\{g(Z_t)'Q_a\left[\frac{\nabla\xi}{\xi}(\Theta) - E^t\left(\frac{\nabla\xi}{\xi}(\Theta)\right)\right]\right\}\right| \\
&+ \left|E^t\left[g(Z_t) - \Phi_p(g)\right]'Q_a E^t\left(\frac{\nabla\xi}{\xi}(\Theta)\right)\right| \\
&+ \left|\Phi_p(g)'(Q_a - Q(\Theta))E^t\left(\frac{\nabla\xi}{\xi}(\Theta)\right)\right| \\
&+ \left|\Phi_p(g)'Q(\Theta)\left[E^t\left(\frac{\nabla\xi}{\xi}(\Theta)\right) - \frac{\nabla\xi}{\xi}(\Theta)\right]\right| \\
\le{}& (2\pi p)^{1/2}\left(\frac{a}{\lambda_t}\right)^{1/2}\left\|\frac{\nabla\xi}{\xi}(\Theta) - E^t\left(\frac{\nabla\xi}{\xi}(\Theta)\right)\right\| \\
&+ \left\|E^t\left[g(Z_t) - \Phi_p(g)\right]\right\|\left(\frac{a}{\lambda_t}\right)^{1/2}\left\|E^t\left(\frac{\nabla\xi}{\xi}(\Theta)\right)\right\| \\
&+ (2\pi p)^{1/2}\|Q_a - Q(\Theta)\|\left\|E^t\left(\frac{\nabla\xi}{\xi}(\Theta)\right)\right\| \\
&+ (2\pi p)^{1/2}\|Q(\Theta)\|\left\|\frac{\nabla\xi}{\xi}(\Theta) - E^t\left(\frac{\nabla\xi}{\xi}(\Theta)\right)\right\|
\end{aligned}
\tag{15}
$$

for all $a \ge 1$. In fact, $E^t[(\nabla\xi/\xi)(\Theta)] \to (\nabla\xi/\xi)(\Theta)$ w.p. 1 (P), as $a \to \infty$, by the martingale convergence theorem and Lemma 3, $\operatorname{ess\,sup}_{|h|\le 1}\|E^t[g(Z_t) - \Phi_p(g)]\| \to 0$ w.p. 1, by Lemma 1 and $\|Q_a - Q(\Theta)\| \to 0$ w.p. 1, by assumption. For mean convergence, it suffices to show that the last four lines of (15) are uniformly integrable. This follows easily, from Schwarz's inequality, since a/λ_t, $a \ge 1$, are uniformly integrable, by assumption, $\|E^n[g(Z_n) - \Phi_p(g)]\|$, $n \ge n_0$, are bounded and $E^n\|(\nabla\xi/\xi)(\Theta)\|^2$, $n \ge 1$, is a uniformly integrable martingale, by (12). □

5. Average confidence levels. With the notation of the last section, consider confidence sets of the form

$$\mathscr{C}_a\colon \theta \in \hat{\theta}_t + B_t'^{-1}C_t,$$

where C_n, $n \ge 1$, are measurable subsets of $\mathbb{R}^p$ or (X_n, Y_n)-sections of a product measurable subset. The confidence curve of such a procedure is then

$$\gamma_a(\theta) = P_\theta\{Z_t \in C_t\}, \qquad \theta \in \Omega,\ a \ge 1$$

and approximations to γ_a are of interest. In the present context, it is easier to approximate averages of γ_a than γ_a itself. If ξ is a density on $\mathbb{R}^p$ which vanishes off of Ω, then the average confidence levels under ξ are defined by

$$\bar{\gamma}_a(\xi) = \int \gamma_a(\theta)\xi(\theta)\,d\theta, \qquad a \geq 1.$$

If one imagines repeated applications of the confidence procedure, by a sequence of users say, then it seems reasonable to suppose that the parameter values will vary among users. If one supposes further that the parameter values are drawn from a density ξ, then $\bar{\gamma}_a(\xi)$ represents the long run relative frequency of coverage in many replications of the experiment. So, having frequentist confidence γ means that $\bar{\gamma}_a(\xi) \geq \gamma$. Here ξ is unknown and perhaps unknowable, since estimating ξ requires access to others' data sets and, even then, there is only indirect information about it. Requiring $\bar{\gamma}_a(\xi) \geq \gamma$, for all ξ is equivalent to requiring $\gamma_a(\theta) \geq \gamma$ for all θ, but if some smoothness is assumed of ξ and if the inequality is replaced by an approximation, then the two conditions may be different; see below. The approach taken here is to regard the confidence functional $\bar{\gamma}_a$ as a frequentist measure of confidence.

For a fixed ξ, let P and E denote probability and expectation in the Bayesian model of Section 3. Then $\bar{\gamma}_a(\xi) = P\{Z_t \in C_t\}$, for all $a \geq 1$. So, Theorem 1 is relevant.

Theorem 2. *Suppose that* (8), (9) *and* (10) *hold and that there is a fixed measurable subset C of $\mathbb{R}^p$ and a function r on Ω for which*

$$\int_{C_t \Delta C} (1 + \|y\|)\Phi_p(dy) \to 0 \tag{16}$$

and

$$r_a := \sqrt{a}\left[\Phi_p(C_t) - \Phi_p(C)\right] \to r(\theta)$$

as $a \to \infty$ in P_θ-probability for a.e. $\theta \in \Omega$, where Δ denotes symmetric difference. Let Ξ denote the class of all almost differentiable densities ξ for which ξ vanishes off of Ω, r_a, $a \geq 1$, are uniformly integrable with respect to $P = P^\xi$ and (11) *and* (12) *hold. Then, for every $\xi \in \Xi$, as $a \to \infty$,*

$$\bar{\gamma}_a(\xi) = \Phi_p(C) + \frac{1}{\sqrt{a}}\int \left[r(\theta)\xi(\theta) + u'Q(\theta)\nabla\xi(\theta)\right] d\theta + o\left(\frac{1}{\sqrt{a}}\right),$$

where

$$u = \int_C y\Phi_p(dy).$$

Proof. Let C, r_a and r be as in the statement of the theorem and let $\xi \in \Xi$. Then, for all $a \geq 1$,

$$\begin{aligned}\bar{\gamma}_a(\xi) - \Phi_p(C) &= P\{Z_t \in C_t\} - \Phi_p(C) = E\left\{P^t[Z_t \in C_t] - \Phi_p(C)\right\}\\ &= \frac{1}{\sqrt{a}}E\left\{r_a + R_a(I_{C_t})\right\},\end{aligned}$$

where R_a is as in (14) and I_C denotes the indicator of C. Let g_t and g be the functions defined in (4) when $h = I_{C_t}$ and $h = \mathrm{I}_C$. Then $\Phi_p g_t \to \Phi_p g$ in probability by (7) and (16). So

$$E\{r_a + R_a(I_{C_t})\} \to E\left\{r(\Theta) + u'Q(\Theta)\frac{\nabla\xi}{\xi}(\Theta)\right\}$$

as $a \to \infty$, by (16), the uniform integrability of r_a, $a \geq 1$ and Theorem 1. □

In the corollaries below, $\xi^*(\theta)$ denotes the vector whose jth component is

$$\xi_j^*(\theta) = \int_{\{|w|>|\theta_j|\}} \left|\frac{\partial}{\partial w}\xi(\theta_1,\ldots,\theta_{j-1},w,\theta_{j+1},\ldots,\theta_p)\right| dw$$

for all $\theta \in \mathbb{R}^p$ and $j = 1,\ldots,p$.

Corollary 1. *If $\Omega = \mathbb{R}^p$ and the entries of $Q(\theta)$ are almost differentiable, then as $a \to \infty$,*

$$\bar{\gamma}_a(\xi) = \Phi_p(C) + \frac{1}{\sqrt{a}}\int\left[r(\theta) - u'Q^{\#}(\theta)\mathbf{1}\right]\xi(\theta)\,d\theta + o\left(\frac{1}{\sqrt{a}}\right),$$

where

$$q_{ij}^{\#}(\theta) = \frac{\partial}{\partial\theta_j}q_{ij}(\theta), \qquad \forall\, i, j, \theta,$$

and $\mathbf{1} = (1,\ldots,1)'$, for all $\xi \in \Xi$ for which $\|Q^{\#}(\theta)\|\,\|\xi^*(\theta)\|$ is integrable over $\mathbb{R}^p$.

Corollary 2. *For $\theta \in \mathbb{R}^p$ and $a \geq 1$, let*

$$\gamma_a^*(\theta) = \Phi_p(C) + \frac{1}{\sqrt{a}}\left[r(\theta) - u'Q^{\#}(\theta)\mathbf{1}\right].$$

Then

$$\lim_a \int\sqrt{a}\left[\gamma_a(\theta) - \gamma_a^*(\theta)\right]\xi(\theta)\,d\theta = 0,$$

for all $\xi \in \Xi$ for which $\|Q^{\#}(\theta)\|\,\|\xi^*(\theta)\|$ is integrable over $\mathbb{R}^p$.

Proofs. Corollary 1 follows from an integration by parts; Corollary 2 is then clear. □

Remarks. 1. For the frequentist interpretation of $\bar{\gamma}_a$ to be reasonable, it is necessary for the class Ξ to be sufficiently large—for example, for its weak closure to contain all distributions.

2. If $\bar{\gamma}_a$ is regarded as a functional on the class Ξ, then the conclusion of Corollary 2 may be regarded as a form of weak convergence. It is the motivation for the term "very weak expansion" in the title.

3. Corollary 2 does not assert that $\gamma_a^*(\theta) - \gamma_a(\theta) = o(1/\sqrt{a})$ for any fixed θ. An example where this fails is described by Woodroofe (1986). In this example, however, $\gamma_a^*(\theta)$ provides a good numerical approximation to $\gamma_a(\theta)$ for those fixed θ for which simulations are available.

If the coverage probabilities depend on θ in a sufficiently smooth manner, then it may be possible to deduce that $\gamma_a^*(\theta) - \gamma_a(\theta) = o(1/\sqrt{a})$ for fixed θ, as in Bickel, Goetze and van Zwet (1985). This is an interesting possibility which deserves further exploration (and will require further conditions on the stopping times and/or design). However, the point is not a crucial one, since the very weak expansions admit their own frequentist interpretation.

4. The terms $Q(\theta)\nabla\xi(\theta)$ and $Q^\sharp(\theta)\mathbf{1}$, which appear in Theorem 2 and Corollary 1, may be regarded as bias terms, since

$$\sqrt{a}\,E(Z_t) = \sqrt{a}\,E\left[E^t(Z_t)\right] = E\left[Q_a\frac{\nabla\xi}{\xi}(\Theta)\right]$$

$$\to E\left[Q(\Theta)\frac{\nabla\xi}{\xi}(\Theta)\right] = \int Q(\theta)\nabla\xi(\theta)\,d\theta,$$

under the conditions of Theorem 2.

5. In some cases it is possible to choose C_n in such a manner that $r(\theta) = u'Q^\sharp(\theta)\mathbf{1}$ for all θ. For example, this may be accomplished by subtracting a correction for bias in some cases. Then $\bar{\gamma}_a(\xi) = \Phi(C) + o(1/\sqrt{a})$ for all $\xi \in \Xi$, so that the speed of convergence of $\bar{\gamma}_a(\xi)$ to its limit is increased.

6. While the derivations of γ_a^* and the correction term $u'Q^\sharp(\theta)\mathbf{1}$ use Bayesian ideas, γ_a and γ_a^* do not depend on the prior.

7. In fact, the convergence in Lemma 2 is in total variation.

8. It may be amusing to observe that no use of Taylor's theorem was made in the proofs of Theorems 1 and 2. An alternative approach is to expand $\xi(\hat{\theta}_n + B_n'^{-1}z)$ in a Taylor series about $\hat{\theta}_n$. This leaves a remainder term involving $\nabla\xi(\hat{\theta}_n)/\xi(\hat{\theta}_n)$, which need not be integrable. The use of Stein's identity avoids this problem.

9. The normality assumption has been used in a crucial way. For some nonnormal models, it may be possible to define a measure of distance Z_n between the parameter and the maximum likelihood estimator, so that the likelihood function is proportional to $\exp(-\|Z_n\|^2/2)$. Then the posterior density of Z_n is of the form (3), but the function f contains some Jacobian terms.

10. While the hypotheses of Theorems 1 and 2 are quite specific about the error distribution, they are quite general about the nature of the adaptive design.

6. An example of Robbins and Siegmund. Suppose there are two treatments A and B which produce normally distributed responses with means μ and ν and unit variances. Let $z = 0$ or 1 accordingly as treatment A or B is given. Then the response to either treatment may be written in the form $y = x'\theta + e$, where $x = (1, z)$, $\theta_1 = \mu$, $\theta_2 = \nu - \mu$ and e is a standard normal random variable. So, the model (1) and (2) allows the treatment given to the nth subject to depend on the results from the first $n - 1$.

In clinical trials, it may be desirable to use a design which reduces the number of subjects who receive an inferior treatment. Robbins and Siegmund (1974) proposed the following design for such purposes. Let $z_1 = 0$ and $z_2 = 1$. For $n \geq 2$, let

$$z_{n+1} = 1 \quad \text{iff} \quad i_n^2 \cdot \hat{\theta}_{n,2} > \left(\frac{2s_n - n}{n}\right)(1+\varepsilon)a,$$

where

$$s_n = z_1 + \cdots + z_n, \qquad i_n^2 = s_n(n - s_n)/n,$$

$\varepsilon > 0$ and $a \geq 1$ are design parameters and $\hat{\theta}_n = (\hat{\theta}_{n,1}, \hat{\theta}_{n,2})$ for all $n \geq 2$. This sequential design is to be used in conjunction with the stopping time

$$t = \inf\{n \geq 3 \colon |i_n^2 \cdot \hat{\theta}_{n,2}| > a\}.$$

Then

$$X_n' X_n = \begin{bmatrix} n & s_n \\ s_n & s_n \end{bmatrix}, \qquad n \geq 2,$$

and

$$Q_a = \frac{\sqrt{a}}{i_t}\begin{bmatrix} i_t/\sqrt{t} & 0 \\ -s_t/t & 1 \end{bmatrix}, \qquad a \geq 1.$$

Now,

$$\frac{1}{a}s_t \to \frac{1}{\theta_2}\frac{2+2\varepsilon}{\varepsilon}, \qquad \frac{t}{a} \to \frac{1}{\theta_2}\frac{4(1+\epsilon)^2}{\varepsilon(2+\varepsilon)}$$

and

$$\frac{a}{i_t^2} \to \theta_2,$$

w.p. 1 (P_θ) for all θ for which $\theta_2 > 0$, by (the proof of) Theorem 1 of Robbins and Siegmund (1974) and a dual result holds for $\theta_2 < 0$. It follows easily that $Q_a \to Q(\theta)$ w.p. 1 (P_θ) for a.e. θ, where $Q(\theta) = \sqrt{|\theta_2|}\,M$ and M is one constant value for $\theta_2 > 0$ and another for $\theta_2 < 0$. The lower right-hand entry of M is 1 in both cases.

In this example, there is natural interest in $\theta_2 = \nu - \mu$, the mean difference in responses. Let $\hat{b}_n = b_n(\hat{\theta}_n)$, where b_n, $n \geq 1$, are bounded continuous functions which converge continuously to a limit b a.e. on $\mathbb{R}^2$ and consider upper confidence bounds of the form $\mathscr{C}_a$: $\theta_2 \leq \hat{\theta}_{t,2} + c_t/i_t$, where $c_n = c + \hat{b}_n/i_n$, $n \geq 3$. This is of the form considered in Section 5 with $C_n = \{z \colon z_2 \leq c_n\}$, for $n \geq 3$. It is easily seen that

$$u = -\varphi(c)(0,1)', \qquad u'Q^\#(\theta)\mathbf{1} = -\frac{\operatorname{sign}\theta_2}{2\sqrt{|\theta_2|}}\varphi(c)$$

and

$$r_a = \sqrt{a}\left[\Phi(c_t) - \Phi(c)\right]^{1/2} \to \varphi(c)b(\theta)\sqrt{|\theta_2|}$$

in P_θ-probability for a.e. θ. Let

$$\gamma_a^*(\theta) = \Phi(c) + \frac{1}{\sqrt{a}}\varphi(c)\left[\sqrt{|\theta_2|}\,b(\theta) + \frac{\operatorname{sign}\theta_2}{2\sqrt{|\theta_2|}}\right] \tag{17}$$

for $\theta \in \mathbb{R}^2$ and $a \geq 1$. Then $\sqrt{a}[\gamma_a - \gamma_a^*] \to 0$ very weakly, by Corollary 2.

The size of Ξ depends on b_n, $n \geq 1$, and there is special interest in $b(\theta) = -\operatorname{sign}(\theta_2)/2|\theta_2|$, $\theta_2 \neq 0$. The next result determines Ξ for this case.

PROPOSITION 2. *Suppose that ξ has finite Fisher information and that $\theta_2\xi$ is integrable. Then condition* (11) *is satisfied and if $|\hat{b}_n| \leq K/|\hat{\theta}_{n,2}|$ for all $n \geq 3$ for some constant K, then r_a, $a \geq 1$, are uniformly integrable.*

PROOF. Since the minimal eigenvalue of a 2×2 matrix is at least as big as the determinant divided by the trace, $\lambda_t \geq i_t^2/2 \geq \frac{1}{4}$ and, therefore,

$$\frac{a}{\lambda_t} \leq \frac{2a}{i_t^2} \leq 2|\hat{\theta}_{t,2}| \leq 2|\theta_2| + 8\|Z_t\|,$$

which is uniformly integrable by Lemma 4, below.

Neglecting the excess over the boundary, as in Wald (1947), Robbins and Siegmund show that $E_\theta(i_t) \simeq a/|\theta_2|$ for $\theta_2 \neq 0$. A variation on this argument, as in the Appendix to Wald's book, shows that there is a $C > 1$ for which

$$E_\theta\left(\frac{a}{i_t^2}\right) \leq C + C/|\theta_2|, \qquad \forall\, a,\ \theta_2 \neq 0.$$

So,

$$\sup_a E\left[\left|\frac{a}{i_t^2}\right|^\alpha\right] \leq \int C^\alpha[1 + |\theta_2|^{-\alpha}]\xi(\theta)\,d\theta < \infty$$

for $0 < \alpha < 1$ and, therefore $|i_t^2/a|^\alpha$, $a \geq 1$, are uniformly integrable for $0 < \alpha < 1$. So, if $|\hat{b}_n| \leq K/|\hat{\theta}_{n,2}|$ for all $n \geq 3$ for some K, then r_a, $a \geq 1$, are uniformly integrable, since then

$$|r_a| \leq \left(\frac{\sqrt{a}}{i_t}\right)|\hat{b}_t| \leq K\frac{i_t}{\sqrt{a}}, \qquad \forall\, a \geq 1. \qquad \square$$

7. Simulations. To assess the accuracy of the approximations presented in the last section, a Monte Carlo study was conducted. For $a = 6$, $\varepsilon = 1$ and selected values of θ, 10,000 replications of the Robbins–Siegmund procedure were generated. From this Monte Carlo estimates of the distribution functions of $Z_{t,2}$ were computed and compared with both the approximation (17) and direct normal approximation. The results are presented in Table 1.

The most striking aspect of Table 1 is the amount by which direct normal approximation underestimates the simulations, as much as 0.11 when $c = 0$ and $\theta_2 = 0.5$. The refined approximations of (17) correct by too much in all but three cases, but are much closer to the simulations than direct normal approximation. They are generally better for $c > 0$ than for $c < 0$.

TABLE 1
Simulated and approximate values of $P_\theta\{Z_{t,2} \le c\}$

	$\theta = 0.5$			$\theta = 1.0$		
c	γ^*	MC	Normal	MC	γ^*	SE
−2.00	0.0384	0.0337	0.0228	0.0332	0.0338	0.002
−1.75	0.0650	0.0575	0.0401	0.0566	0.0577	0.002
−1.50	0.1042	0.0935	0.0668	0.0889	0.0932	0.003
−1.25	0.1579	0.1448	0.1056	0.1444	0.1429	0.0035
−1.00	0.2286	0.2114	0.1587	0.1987	0.2080	0.004
−0.75	0.3135	0.2992	0.2266	0.2881	0.2881	0.0045
−0.50	0.4102	0.3957	0.3085	0.3729	0.3804	0.005
−0.25	0.5129	0.5047	0.4013	0.4682	0.4806	0.005
0.00	0.6152	0.6154	0.5000	0.5738	0.5814	0.005
0.25	0.7103	0.7087	0.5987	0.6694	0.6776	0.0045
0.50	0.7932	0.7903	0.6915	0.7554	0.7633	0.004
0.75	0.8603	0.8581	0.7735	0.8296	0.8348	0.0035
1.00	0.9112	0.9057	0.8413	0.8853	0.8907	0.003
1.25	0.9471	0.9423	0.8944	0.9250	0.9316	0.003
1.50	0.9706	0.9663	0.9332	0.9541	0.9596	0.002
1.75	0.9848	0.9822	0.9599	0.9741	0.9776	0.0015
2.00	0.9928	0.9990	0.9772	0.9866	0.9883	0.001

Note: The third and fifth columns are Monte Carlo estimates, based on 10,000 replications; the second and sixth columns are the very weak approximation, at $\theta = 0.5$ and $\theta = 1$; the fourth column is direct normal approximation; and the last column is the standard deviation of the Monte Carlo, rounded to the nearest multiple of 0.0005.

An alternative form in which to write the approximations is to let

$$\gamma_a^{**}(\theta) = \Phi\left\{c + \frac{1}{\sqrt{a}}\left[\sqrt{|\theta_2|}\, b(\theta) + \frac{\operatorname{sign}\theta_2}{2\sqrt{|\theta_2|}}\right]\right\}.$$

This is asymptotically equivalent to γ_a^* and has the advantage of being a distribution function in c. In this form the approximation is substantially poorer for $c < 0$, but slightly better for $c > 0$.

For the sequential probability ratio test with i.i.d. normal observations, a similar but simpler problem, the very weak expansions of Corollary 2 may be compared with the simulations of Woodroofe and Keener (1987). For this problem the agreement is slightly better than that of Table 1.

8. A lemma. The following lemma was used in Section 6 and is needed again in Section 9.

LEMMA 4. *Suppose that* (8) *and* (9) *are satisfied and let* ξ *be a density for which* (12) *holds. Then* $\|Z_t\|^2 I\{\lambda_t \ge \varepsilon\}$ $a \ge 1$, *are uniformly integrable for any* $\varepsilon > 0$.

PROOF. By Lemma 2, $Z_t I\{\lambda_t \ge \varepsilon\}$ has a limiting standard normal distribution as $a \to \infty$, under P. So, it suffices to show that $\lim_a E\{\|Z_t\|^2 I\{\lambda_t \ge \varepsilon\}\} = p$

and for this it suffices to show that $E^t\{\|Z_t\|^2\}I\{\lambda_t \geq \varepsilon\}$, $a \geq 1$, are uniformly integrable. By Lemma 1 and Schwarz's inequality,

$$E^t[\|Z_t\|^2] - p = E^t\left[Z_t'B_t^{-1}\left(\frac{\nabla\xi}{\xi}\right)(\Theta)\right] \leq \frac{1}{\sqrt{\lambda_t}}\left(E^t\|Z_t\|^2\right)^{1/2}\left(E^t\left\|\frac{\nabla\xi}{\xi}(\Theta)\right\|^2\right)^{1/2}.$$

So,

$$\left(E^t\|Z_t\|^2\right)^{1/2} \leq 2\sqrt{p} + \frac{1}{\sqrt{\lambda_t}}\left(E^t\left\|\frac{\nabla\xi}{\xi}(\Theta)\right\|^2\right)^{1/2}$$

and the right side is uniformly square integrable on $\lambda_t \geq \varepsilon$ for any $\varepsilon > 0$. □

9. Unknown variability. If the model is changed by replacing (1) by

$$y_k = x_k'\theta + \sigma e_k, \qquad k = 1, 2, \dots,$$

where $\sigma > 0$ is another unknown parameter, then there is a simple extension of Theorem 2. Suppose that $n_0 > p$ and let

$$\hat{\sigma}_n^2 = \left(\frac{1}{n-p}\right)\|Y_n - X_n\hat{\theta}_n\|^2$$

and

$$\hat{Z}_n = Z_n/\hat{\sigma}_n, \qquad n \geq n_0.$$

Some additional conditions are required of the stopping times $t = t_a$, $a \geq 1$. For a fixed ξ and σ, suppose that there are a.e. positive functions $\delta = \delta_\xi$ and η for which

$$\int \sqrt{\delta}(\theta)\xi(\theta)\,d\theta < \infty, \tag{18}$$

$$P\left\{\frac{a}{t} > \delta(\Theta)\right\} + P\{t \leq \sqrt{a}\} = o\left(\frac{1}{\sqrt{a}}\right) \tag{19}$$

and

$$\frac{a}{t} \to \eta(\Theta) \quad \text{in probability.} \tag{20}$$

THEOREM 3. *Suppose that condition* (9) *and* (10) *are satisfied and that there is a measurable subset C of $\mathbb{R}^p$ for which*

$$\int_{(\hat{\sigma}_t C_t)\Delta(\sigma C)} (1 + \|y\|)\Phi_p(dy) \to 0$$

and

$$\hat{r}_a := \sqrt{a}\left[\Phi_p(C_t\hat{\sigma}_t) - \Phi_p(C\hat{\sigma}_t)\right] \to \sigma r(\theta)$$

P_θ-probability for a.e. $\theta \in \Omega$. Let ξ denote a density for which $\hat{r}_a$, $a \geq 1$ are

uniformly integrable and (11), (12), (18), (19) *and* (20) *hold. Then*

$$P\{\hat{Z}_t \in C_t\} = \Phi_p(C) + \left(\frac{\sigma}{\sqrt{a}}\right)\int [r(\theta)\xi(\theta) - u'Q(\theta)\nabla\xi(\theta)]\, d\theta + o\left(\frac{1}{\sqrt{a}}\right)$$

as $a \to \infty$.

Proof. It suffices to prove the theorem in the special case that $\sigma = 1$, by a simple reparameterization (in which $y_k^* = y_k/\sigma$ and $x_k^* = x_k/\sigma$, $k \geq 1$). Let $\hat{R}_a = R_a(I_{C_t\hat{\sigma}_t})$ where R_a is as in (14). Then

$$\sqrt{a}\left[P\{\hat{Z}_t \in C_t\} - \Phi_p(C)\right] = \sqrt{a}\,E\left[\Phi_p(C\hat{\sigma}_t) - \Phi_p(C)\right] + E\{\hat{r}_a + \hat{R}_a\}$$

and

$$E\{\hat{r}_a + \hat{R}_a\} \to \int [r(\theta)\xi(\theta) - u'Q(\theta)\dot{\nabla}\xi(\theta)]\, d\theta,$$

as $a \to \infty$, as in the proof of Theorem 2. So it suffices to show that $\sqrt{a}\,E[\Phi_p(C\hat{\sigma}_t) - \Phi_p(C)] \to 0$ as $n \to \infty$.

Let $W_a = \sqrt{a}\,[\Phi_p(C\hat{\sigma}_t) - \Phi_p(C)]$ for $a \geq 1$. Then, W_a has a limiting distribution with mean zero as $a \to \infty$, a mixture of normal distributions, by Anscombe's (1952) theorem, applied conditionally given θ. So, it suffices to show that W_a, $a \geq 1$, are uniformly integrable. There is a constant K for which $|W_a| \leq K\sqrt{a}\,|\hat{\sigma}_t^2 - 1|$ for all $a \geq 1$, since $\partial\Phi_p(C\sigma)/\partial\sigma^2$ is bounded on compacts (in σ) and $|W_a| \leq 2\sqrt{a}$ for all a. Moreover, with $\varepsilon_n = (e_1, \dots, e_n)'$,

$$\hat{\sigma}_n^2 - 1 = \left(\frac{1}{n-p}\right)\left\{\left(\|\varepsilon_n\|^2 - n\right) - \left(\|Z_n\|^2 - p\right)\right\}, \qquad n \geq n_0.$$

Let

$$A = \left\{\frac{a}{t} \leq \delta(\Theta)\right\}, \qquad B = \left\{t \geq \sqrt{a}\,, \lambda_t \geq 1\right\}.$$

Then

$$\int_{A'\cup B'} |W_a|\, dP \leq 2\sqrt{a}\,P(A' \cup B') \to 0$$

as $a \to \infty$ by (11) and (19) and $\sqrt{a}(1/(t-p))(\|Z_t\|^2 - p)I_B$, $a \geq 1$, are uniformly integrable, by Lemma 4. Let $u_k = e_k^2 - 1$ for all $k = 1, 2, \dots$. Then, for any event F,

$$\sup_{a\geq 1}\int_{F\cap A}\left|\left(\frac{1}{t-p}\right)\left(\|\varepsilon_t\|^2 - t\right)\right| dP \leq p\int \sup_{a\geq 1} E_\theta\left\{\sup_{n > a/\delta(\theta)} \sqrt{a}\,|\bar{u}_n| I_F\right\}\xi(\theta)\, d\theta.$$

Since $\bar{u}_n$, $n \geq 1$, is a reverse martingale given θ, the supremum on the right is bounded by $4\sqrt{\delta(\theta)}$ for all $\theta \in \Omega$ and it converges to zero in measure (in θ) as $P(F) \to 0$ for the same reason. So, the integrand on the left is uniformly integrable. □

Acknowledgments. It is a pleasure to acknowledge helpful conversations with L. J. Wei and a helpful referee's report.

REFERENCES

ANSCOMBE, F. J. (1952). Large sample theory of sequential estimation. *Proc. Cambridge Philos. Soc.* **48** 600–607.

BICKEL, P. J., GOETZE, F. and VAN ZWET, W. R. (1985). A simple analysis of third-order efficiency of estimates. In *Proc. of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer* (L. M. Le Cam and R. A. Olshen, eds.) **2** 749–768. Wadsworth, Monterey, Calif.

GHOSH, J. K., SINHA, B. and JOSHI, S. (1982). Expansions for posterior probabilities and integrated Bayes risk. In *Statistical Decision Theory and Related Topics —III* (S. S. Gupta and J. O. Berger, eds.) **1** 403–456. Academic, New York.

LAI, T. L. and WEI, C. Z. (1982). Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *Ann. Statist.* **10** 154–166.

REHALIA, M. (1983). L^p asymptotic expansions for posterior moments with applications to sequential analysis. Ph.D. dissertation, Univ. Michigan.

ROBBINS, H. and SIEGMUND, D. (1974). Sequential tests involving two populations. *J. Amer. Statist. Assoc.* **69** 132–139.

STEIN, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151.

STEIN, C. (1985). On the coverage probability of confidence sets based on a prior distribution. In *Sequential Methods in Statistics* (R. Zieliński, ed.) 485–514. PWN-Polish Scientific Publishers, Warsaw.

STEIN, C. (1987). *Approximate Computation of Expectations*. IMS, Hayward, Calif.

WALD, A. (1947). *Sequential Analysis*. Wiley, New York.

WEI, L. J. (1979). The generalized Pólya's urn design for sequential medical trials. *Ann. Statist.* **7** 291–296.

WOODROOFE, M. (1986). Very weak expansions for sequential confidence levels. *Ann. Statist.* **14** 1049–1067.

WOODROOFE, M. and KEENER, R. (1987). Asymptotic expansions in boundary crossing problems. *Ann. Probab.* **15** 102–114.

WU, C. F. J. (1985). Asymptotic inference from sequential design in a nonlinear situation. *Biometrika* **72** 553–558.

DEPARTMENT OF STATISTICS
UNIVERSITY OF MICHIGAN
ANN ARBOR, MICHIGAN 48109