Statistica Sinica **7**(1997), 53-74

CORRECTED CONFIDENCE SETS FOR SEQUENTIALLY DESIGNED EXPERIMENTS

Michael Woodroofe and D. Stephen Coad

University of Michigan and University of Sussex

Dedicated to Herbert Robbins on the occasion of his 80th birthday.

Abstract: Consider a linear model, $y_k = x_k'\theta + e_k$, $k = 1, 2, \ldots$, in which the current design variable x_k may be a function of the previous responses $y_1, \ldots, y_{k-1}$ and auxiliary randomization. Here the x's and θ are p-dimensional, $'$ denotes transpose, and the errors e_k are taken to be i.i.d standard normal variables. The goal is to construct confidence sets for θ which are asymptotically valid to a high order. This is accomplished by obtaining very weak asymptotic expansions for the distributions of an appropriate pivotal quantity. The accuracy of the approximation is assessed by simulation experiments for two sequential tests proposed by Siegmund (1980, 1993).

Key words and phrases: Asymptotic expansions, average confidence levels, contrasts, Martingale Convergence Theorem, posterior distributions, sequential allocation, Stein's Identity.

1. Introduction

The purpose of this article is to show how to construct asymptotically valid confidence regions for the parameters of a linear model when the design variables may depend on previous responses. There is a substantial and growing list of models of this type. These include control problems, as in Lai and Wei (1982), Wu's (1985) adaptive designs for estimating non-linear functions, a sequential allocation rule due to Robbins and Siegmund (1974), Siegmund's (1980, 1993) tests for comparing three treatments, and Eisele's (1994) adaptive biased coin designs.

To fix ideas, let $e_1, e_2, \ldots$ and $u_1, u_2, \ldots$ denote independent random variables for which $e_1, e_2, \ldots$ are i.i.d. standard normal random variables, and consider a statistical model in which the data are of the form $y_k = x_k'\theta + e_k$, for $k = 1, 2, \ldots$, where $\theta = (\theta_1, \ldots, \theta_p)' \in \Re^p$ is unknown, $'$ denotes transpose, and $x_k = (x_{k1}, \ldots, x_{kp})'$ may depend (measurably) on previous responses, say $x_k = x_k(u_1, \ldots, u_k, y_1, \ldots, y_{k-1})$ for all $k = 1, 2, \ldots$ The $u's$ are included in the model to accommodate auxiliary randomization as in Eisele's (1994) biased coin designs; their distributions do not depend on θ. If $y_1, \ldots, y_n$ are observed, then the model may be written in the familiar form $\mathbf{y}_n = X_n\theta + \mathbf{e}_n$, where $\mathbf{y}_n = (y_1, \ldots, y_n)'$, $\mathbf{e}_n = (e_1, \ldots, e_n)'$, and $X_n = [x_1, \ldots, x_n]'$. Throughout the paper it is assumed that there is a possibly random integer n_0 for which X_n is of rank p w.p.1 (P_θ) for all $\theta \in \Re^p$ and $n \geq n_0$, and consideration is restricted to $n \geq n_0$. Then the log likelihood function and maximum likelihood estimator are $\ell_n(\theta) = -\|\mathbf{y}_n - X_n\theta\|^2/2$ and $\hat{\theta}_n = (X_n'X_n)^{-1}X_n'\mathbf{y}_n$ for all $n \geq n_0$. In fact, ℓ_n and $\hat{\theta}_n$ are the log likelihood function and maximum likelihood estimator, even if n is replace by a stopping time with respect to $\mathcal{A}_k = \sigma\{x_1, \ldots, y_k\}$, $k = 1, 2, \ldots$ See, for example, Berger and Wolpert (1984).

Let $1 \leq m \leq p$ and let A_n and B_n denote $m \times p$ and $p \times p$ matrices, depending measurably on $x_1, \ldots, y_n$, for which

$$A_n A_n' = I_m \quad \text{and} \quad X_n'X_n = B_n B_n', \tag{1}$$

and let

$$Z_n = B_n'(\theta - \hat{\theta}_n) \quad \text{and} \quad W_n = A_n Z_n \tag{2}$$

for $n \geq n_0$. Here Z_n and W_n may be regarded as first approximations to pivotal quantities. If a particular linear functional, say $c'\theta$, is of interest, then $c'(\theta - \hat{\theta}_n)/\sqrt{\{c'(X_n'X_n)^{-1}c\}}$ may be written in the form $A_n Z_n$, where $A_n = (B_n^{-1}c)'/\sqrt{\{c'(X_n'X_n)^{-1}c\}}$ and $A_n A_n' = 1$. There are many ways of factoring $X_n'X_n$ in (1). Some advantages of using a Cholesky decomposition are described at the end of Section 3.

The goal is to find asymptotic expansions for the distribution of W_t for suitable families of stopping times $t = t_a$, $a \geq 1$. It is assumed that the parameter a may be so chosen that $X_t'X_t$ is of order a as $a \to \infty$ in the sense of (11) below, and the expansions take the following form: data dependent vectors $\hat{\mu}_a$ and matrices $\hat{\Gamma}_a$ are found for which

$$W_t^* := \hat{\Gamma}_a^{-1}\Big(W_t - \frac{\hat{\mu}_a}{\sqrt{a}}\Big) \tag{3}$$

are asymptotically standard normal to third order in the very weak sense of Woodroofe (1986). To state the result, let Φ^m denote the standard m-variate normal distribution and write $\Phi^m h = \int_{\Re^m} h d\Phi^m$ for measurable functions $h : \Re^m \to \Re$ for which the integral exists. Then it is shown that

$$\int_\Omega E_\theta[h(W_t^*)]\xi(\theta)d\theta = \Phi^m h + o(\frac{1}{a}) \text{ as } a \to \infty \tag{4}$$

for a large class of measurable $h : \Re^m \to \Re$ and *all* twice continuously differentiable densities ξ with compact support. Woodroofe (1986, 1989) calls expansions of the form (4) very weak expansions and writes $E_\theta[h(W_t^*)] = \Phi^m h + o(1/a)$ (very weakly).

For the application to confidence sets, let $C \subset \Re^m$ be a measurable set and let $\mathcal{C} = \{\theta \in \Re^p : W_t^* \in C\}$ and $\gamma(\theta) = P_\theta[\theta \in \mathcal{C}]$. Then $\mathcal{C}$ may be regarded as a confidence set with confidence level γ. By (4), $\gamma(\theta)$ is approximately $\Phi^m(C)$ in the very weak sense; that is, $\gamma(\theta) = \Phi^m(C) + o(1/a)$ very weakly. Woodroofe (1986, 1989) argues that very weak expansions are strong enough to support a frequentist interpretation of confidence.

The derivation of (4) is outlined in Section 3 with supporting details in Sections 6, 7 and 8. Relation (4) is applied to form simultaneous confidence intervals for contrasts for Siegmund's (1980, 1993) sequential comparison of three treatments in Section 4. The latter includes both adaptive design and optional stopping. Section 2 contains some preliminary material on Stein's Identity, and Section 5 some remarks.

The goals of this paper are similar to those of Woodroofe (1989), but there are important differences in the development. An additional term is computed here *without imposing additional smoothness conditions.* It is mildly surprising that this is possible. There is no analogue of the possibly data dependent matrix A_n in earlier work. Advantages of including this matrix are illustrated in the example. The analysis of the standardized variable W_t^* is entirely different from the earlier work.

2. Stein's Identity

The proof of (4) depends on Stein's (1981) Identity. If $h : \Re^p \to \Re$ is a function of polynomial growth, say $|h(z)| \le C(1 + \|z\|^r)$ for all $z \in \Re^p$ for some $0 < C, r < \infty$, let $h_0 = \Phi^p h,\ h_p = h,\ h_j(y_1, \ldots, y_j) = \int_{\Re^{p-j}} h(y_1, \ldots, y_j, z)\Phi^{p-j}(dz)$, and

$$g_j(y_1, \ldots, y_p) = e^{\frac{1}{2}y_j{}^2} \int_{y_j}^{\infty} [h_j(y_1, \ldots, y_{j-1}, w) - h_{j-1}(y_1, \ldots, y_{j-1})] e^{-\frac{1}{2}w^2} dw$$

for $-\infty < y_1, \ldots, y_p,\ z < \infty$ and $j = 1, \ldots, p$. Here g_j is to be regarded as a function on $\Re^p$, though it depends on $y_1, \ldots, y_p$ only through $y_1, \ldots, y_j$. Let $U_p h = (g_1, \ldots, g_p)'$.

For $r \ge 0$, let $\mathcal{H}_r^p$ denote the class of all measurable functions $h : \Re^p \to \Re$ for which $|h(z)| \le 1 + \|z\|^r$ for all $z \in \Re^p$, and let $\mathcal{H}^p = \cup_{r \ge 0} \cup_{c \ge 0} c\mathcal{H}_r^p$, the class of functions of polynomial growth.

Proposition 1. *U_p is a linear transformation from $\mathcal{H}^p$ into $\mathcal{H}^p$. Moreover, there are constants $c_{p,0}, c_{p,1}, \ldots$ for which $U_p\mathcal{H}_0^p \subseteq c_{p,0}\mathcal{H}_0^p$ and $U_p\mathcal{H}_r^p \subseteq c_{p,r}\mathcal{H}_{r-1}^p$ for all $r = 1, 2, \ldots$*

Proof. This is established in Woodroofe (1992) for the case $p = 1$. The extension from one to several dimensions is not difficult.

The transformation U_p may be iterated. If $h \in \mathcal{H}^p$, let $U_p^2 h = [U_p g_1, \ldots, U_p g_p]$, the $p \times p$ matrix whose jth column is $U_p g_j,\ j = 1, \ldots, p$, where $g_1, \ldots, g_p$ are as above. Then $U_p^2 h$ is an upper triangular matrix. Let

$$V_p h = \frac{U_p^2 h + U_p^2 h'}{2} = \frac{1}{2}\{[U_p g_1, \ldots, U_p g_p] + [U_p g_1, \ldots, U_p g_p]'\}.$$

Then $V_p h$ is a symmetric matrix. For an example, let $h(z) = \|z\|^2 = z_1^2 + \cdots + z_p^2,\ z \in \Re^p$. Then $h_j(y_1, \ldots, y_p) = y_1^2 + \cdots + y_j^2 + (p - j)$ and $g_j(z) = z_j$ for all $j = 1, \ldots, p$. That is, $U_p h(z) = z$, for all $z \in \Re^p$. Similar, simpler calculations then show that $V_p h = I_p$, the $p \times p$ identity. Simple calculations also show that

$$\Phi^p(U_p h) = \int_{\Re^p} z h(z) \Phi^p(dz)$$

and

$$\Phi^p(V_p h) = \frac{1}{2}\int_{\Re^p} (zz' - I_p) h(z) \Phi^p(dz) \tag{5}$$

for all $h \in \mathcal{H}^p$.

If Ω is a convex open subset of $\Re^p$, then a measurable function $f : \Omega \to \Re$ is said to be *almost differentiable on* Ω if there is a measurable function $\nabla f : \Omega \to \Re^p$ for which

$$f(y) - f(x) = \int_0^1 (y - x)' \nabla f[ty + (1 - t)x] dt$$

for a.e. $x \in \Omega$ for each $y \in \Omega$. In this case, ∇f is essentially unique (Lebesgue). Of course, a continuously differentiable function f is almost differentiable with ∇f equal to the gradient. Below ∇f is called the gradient of f, and the components of ∇f are denoted by $\partial f / \partial z_j,\ j = 1, \ldots, p$, even if f is only almost differentiable.

The following properties of almost differentiable functions are needed. If f is a continuous, almost differentiable function, $K \subset \Omega$ is compact, and $\int_K \|\nabla f\|^r dx < \infty$, where $r \geq 1$, then there are infinitely differentiable f_ϵ, $0 < \epsilon \leq \epsilon_0$, for which

$$\lim_{\epsilon \to 0}\Big\{\sup_{x \in K} |f_\epsilon(x) - f(x)| + \int_K \|\nabla f_\epsilon - \nabla f\|^r dx\Big\} = 0. \tag{6}$$

Further, if f and g are continuous, almost differentiable functions for which $\|\nabla f\|$ and $\|\nabla g\|$ are locally integrable (integrable over all compact subsets of Ω), and g has compact support, then

$$\int_\Omega f \nabla g dx = -\int_\Omega \nabla f g dx.$$

In the next proposition, ∇^2 denotes Hessian and $\|\cdot\|$ denotes the trace norm of a matrix, as well as the Euclidean norm in $\Re^p$.

Proposition 2. *Let $r \geq 0$, and let Ψ be a signed measure of the form $d\Psi = fd\Phi^p$, where f is an almost differentiable function on $\Re^p$, for which*

$$\int_{\Re^p} |f|d\Phi^p + \int_{\Re^p} (1 + \|z\|^r)\|\nabla f(z)\|\Phi^p(dz) < \infty.$$

Then

$$\int_{\Re^p} hd\Psi = \Phi^p h \times \Psi 1 + \int_{\Re^p} (U_p h)' \nabla f d\Phi^p$$

for all $h \in \mathcal{H}_r^p$. If, in addition, f is continuously differentiable, $\partial f/\partial z_j$, $j = 1, \ldots, p$, are almost differentiable, and $\int_{\Re^p} (1 + \|z\|^r)\|\nabla^2 f(z)\|\Phi^p(dz) < \infty$, then

$$\int_{\Re^p} hd\Psi = \Phi^p h \times \Psi 1 + \Phi^p(U_p h)' \int_{\Re^p} \nabla f d\Phi^p + \int_{\Re^p} \mathrm{tr}[(V_p h)\nabla^2 f]d\Phi^p \tag{7}$$

for all $h \in \mathcal{H}_r^p$, where $\Phi^p(U_p h) = [\Phi^p g_1, \ldots, \Phi^p g_p]'$.

Proof. For the first assertion, see Woodroofe (1989, Proposition 1). The second assertion follows easily from the first.

Proposition 2 may be applied to a function of m variables, where $1 \leq m \leq p$, as follows.

Corollary. *Let $1 \leq m \leq p$ and let A be an $m \times p$ matrix for which $AA' = I_m$. If $h \in \mathcal{H}_r^m$ and $h^*(z) = h(Az)$, $z \in \Re^p$, then*

$$\begin{aligned}\int_{\Re^p} h^* d\Psi = \Phi^m h \times \Psi 1 &+ \int_{\Re^p} \Phi^m(U_m h)' A\nabla f(z)\Phi^p(dz) \\ &+ \int_{\Re^p} \mathrm{tr}[(V_m h)(Az)A\nabla^2 f(z)A']\Phi^p(dz).\end{aligned} \tag{8}$$

Proof. Using a singular value decomposition, A may be written as $A = HJK$, where H and K are orthogonal matrices of dimensions $m \times m$ and $p \times p$ and $J = [I_m, 0]$ $(m \times p)$. Thus, it suffices to verify (8) for $A = K$ (and $m = p$) and for $A = J$. For $A = K$, (8) follows from (7) and two transparent changes of variables. For $A = J$, (8) follows from the easily verified relations $\Phi^p h^* = \Phi^m h$, $U_p h^*(z) = J' U_m h(Jz)$ and $V_p h^*(z) = J' V_m h(Jz) J$ and (7).

3. Expansions

The derivation of (4) depends on the following simple observation. If h is measurable, ξ is a density on $\Re^p$, and the expectations exist, then $\int_{\Re^p} E_\theta[h(W_t)] \xi(\theta)d\theta = E_\xi[h(W_t)]$, where E_ξ denotes expectation in the Bayesian model in which θ is replaced by a random variable Θ which has prior density ξ and is independent of $e_1, e_2, \ldots$ and $u_1, u_2 \ldots$ Let t denote any stopping time with respect to $\mathcal{A}_n = \sigma\{x_1, \ldots, y_n\}$ for which $t \geq n_0$ w.p.1 and let E^t_ξ denote conditional expectation given $x_1, \ldots, y_t$. Then

$$\int_{\Re^p} E_\theta[h(W_t)]\xi(\theta)d\theta = E_\xi[h(W_t)] = E_\xi\{E^t_\xi[h(W_t)]\}.$$

The approach is to generate expansions for the posterior expectations and then integrate them.

If Θ has a density ξ, then the posterior densities of Θ and Z_n given $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ are $\xi_n(\theta) \propto \xi(\theta)e^{l_n(\theta)}$ and

$$\zeta_n(z) \propto \xi(\hat{\theta}_n + B_n'^{-1}z)e^{-\frac{1}{2}\|z\|^2}$$

for all $\theta, z \in \Re^p$ and $n \geq n_0$. That is, the posterior distribution of Z_n is of the form considered in Proposition 2, with $f(z) \propto \xi(\hat{\theta}_n + B_n'^{-1}z),\ z \in \Re^p$. Moreover, if ξ is twice continuously differentiable with compact support, then

$$\frac{\nabla f}{f}(Z_n) = B_n^{-1}\frac{\nabla\xi}{\xi}(\Theta) \quad \text{and} \quad \frac{\nabla^2 f}{f}(Z_n) = B_n^{-1}\frac{\nabla^2\xi}{\xi}(\Theta)B_n'^{-1}.$$

So, replacing n by t and appealing to the Corollary to Proposition 2 leads to

$$\begin{aligned} E^t_\xi\,[h(W_t)] = \Phi^m h &+ E^t_\xi\Big\{(\Phi^m U_m h)' A_t B_t^{-1}\frac{\nabla\xi}{\xi}(\Theta)\Big\} \\ &+ E^t_\xi\Big\{\mathrm{tr}[(V_m h)(W_t)A_t B_t^{-1}\frac{\nabla^2\xi}{\xi}(\Theta)B_t'^{-1}A_t']\Big\} \end{aligned} \tag{9}$$

w.p.1 for all $h \in \mathcal{H}^m$ on $\{t \geq n_0\}$.

To proceed further some conditions on the design matrices $X_n,\ n \geq n_0$, are needed. Let λ_n denote the minimum eigenvalue of $X_n'X_n$. It is assumed throughout that there is a $\lambda^0 > 0$ for which

$$\inf_{n \geq n_0} \lambda_n \geq \lambda^0 \quad \text{and} \quad \lim_{n\to\infty} \lambda_n = \infty \tag{10}$$

w.p.1 (P_θ) for a.e. $\theta \in \Re^p$. This condition insures that $\hat{\theta}_n$ is consistent for a.e. $\theta \in \Omega$. See Lemma 3 of Woodroofe (1989). Next let $Q^a = \sqrt{a}A_t B_t^{-1},\ a \geq 1$, and suppose that there are matrices $Q_\theta,\ \theta \in \Re^p$, for which

$$\int_K \|Q_\theta\|^2 d\theta < \infty \quad \text{and} \quad \lim_{a\to\infty}\int_K E_\theta\{\|Q^a - Q_\theta\|^2\}d\theta = 0 \tag{11}$$

for all compact $K \subseteq \Re^p$. Then

$$\begin{aligned} E^t_\xi\,[h(W_t)] = \Phi^m h &+ \frac{1}{\sqrt{a}}E^t_\xi\Big\{(\Phi^m U_m h)' Q_\Theta \frac{\nabla\xi}{\xi}(\Theta)\Big\} \\ &+ \frac{1}{a}E^t_\xi\Big\{\mathrm{tr}[(\Phi^m V_m h)Q_\Theta\frac{\nabla^2\xi}{\xi}(\Theta)Q_\Theta']\Big\} \\ &+ \frac{1}{\sqrt{a}}(\Phi^m U_m h)' I_a + \frac{1}{a}II_a(h), \end{aligned} \tag{12}$$

where

$$I_a = E^t_\xi \left\{ [Q^a - Q_\Theta] \frac{\nabla \xi}{\xi}(\Theta) \right\}$$

and

$$II_a(h) = E^t_\xi \left\{ \mathrm{tr} \left[V_m h(W_t) Q^a \frac{\nabla^2 \xi}{\xi}(\Theta) Q^{a\prime} - (\Phi^m V_m h) Q_\Theta \frac{\nabla^2 \xi}{\xi}(\Theta) Q'_\Theta \right] \right\}.$$

Ignoring the remainder terms for the moment and taking expectations in (12), this suggests the approximation

$$\begin{aligned} E_\xi \left[h(W_t) \right] \approx \Phi^m h &+ \frac{1}{\sqrt{a}} (\Phi^m U_m h)' \int_\Omega Q_\theta \nabla \xi(\theta) d\theta \\ &+ \frac{1}{a} \int_\Omega \mathrm{tr}[(\Phi^m V_m h) Q_\theta \nabla^2 \xi(\theta) Q'_\theta] d\theta. \end{aligned} \tag{13}$$

In some cases the integrals on the right side of (13) may be written in the form of expectations with respect to the prior density. Write $Q_\theta = [q_{i,j}(\theta) : i = 1, \ldots, m,\ j = 1, \ldots, p]$ for $\theta \in \Re^p$. If ξ is twice continuously differentiable with compact support and $q_{i,j}$ are almost differentiable with locally integrable gradients, then

$$\int_{\Re^p} Q_\theta \nabla \xi(\theta) d\theta = - \int_{\Re^p} Q^\#_\theta \mathbf{1} \xi(\theta) d\theta, \tag{14}$$

where $q^\#_{i,j}(\theta) = \partial q_{i,j}(\theta)/\partial \theta_j$ for a.e. $\theta \in \Re^p$, for all $i = 1, \ldots, m\ , j = 1, \ldots, p$, $Q^\#_\theta = [q^\#_{i,j}(\theta) :\ i = 1, \ldots, m,\ j = 1, \ldots, p]$, and $\mathbf{1} = (1, \ldots, 1)'$. If also $q_{i,j}$ are continuously differentiable and their partial derivatives are almost differentiable with locally square integrable gradients, then

$$\int_{\Re^p} Q_\theta \nabla^2 \xi(\theta) Q'_\theta d\theta = \int_{\Re^p} M(\theta) \xi(\theta) d\theta, \tag{15}$$

where

$$m_{i,j}(\theta) = \sum_{k=1}^p \sum_{l=1}^p \frac{\partial^2}{\partial \theta_k \partial \theta_l} [q_{i,k}(\theta) q_{j,l}(\theta)]$$

for $i, j = 1, \ldots, m$.

Letting $h(w) = w_i,\ w \in \Re^m,\ i = 1, \ldots, m$, in (13) and (14) leads to the very weak approximation,

$$E_\theta(W_t) \approx -\frac{1}{\sqrt{a}} Q^\#_\theta \mathbf{1} = \frac{1}{\sqrt{a}} \mu(\theta), \text{ say}, \tag{16}$$

very weakly. Here $\mu(\theta)$ may be estimated. If μ is bounded and continuous a.e., then $\hat{\mu}_a = \mu(\hat{\theta}_t),\ a \geq 1$, are suitable estimators. More general situations are considered in Proposition 3. Letting $\hat{\mu}_a$ denote suitable estimators, approximations like those described above lead to

$$E_\theta\Big\{[W_t - \frac{\hat{\mu}_a}{\sqrt{a}}][W_t - \frac{\hat{\mu}_a}{\sqrt{a}}]'\Big\} \approx I_m + \frac{1}{a}\Delta(\theta), \tag{17}$$

very weakly, where

$$\Delta_{i,j}(\theta) = \sum_{k=1}^{p}\sum_{l=1}^{p}\Big(\frac{\partial q_{i,k}}{\partial \theta_l}\Big)\Big(\frac{\partial q_{j,l}}{\partial \theta_k}\Big)$$

for $i, j = 1, \ldots, m$ and $\theta \in \Re^m$. As above, the matrix $\Delta(\theta)$ may be estimated, by $\hat{\Delta}_a = \Delta(\hat{\theta}_t)$ if $\Delta(\theta)$ is bounded and continuous a.e., and by other estimators in more general situations. Let

$$\hat{\Gamma}_a = I_m + \frac{\hat{\Delta}_a}{2a}. \tag{18}$$

The main result asserts that (4) holds with these choices of $\hat{\mu}_a$ and $\hat{\Gamma}_a$.

Relations (16) and (17) are intended to provide motivation for the choice of $\hat{\Gamma}_a$ in (18). They will not be explicitly proved. It may seem clear that (4) can be deduced from (13), if the functions μ and Γ are sufficiently smooth. The proof of (4) does not require smoothness of these functions, however. This is an important point, since optimal designs may not lead to smooth functions. The example in the next section illustrates this point.

There are advantages to using a Cholesky decomposition in (1). If B_n is lower (or upper) triangular in (2) and $X_t'X_t/a \to L_\theta > 0$ w.p.1 (P_θ) for a.e. $\theta \in \Re^p$, then $B_t/\sqrt{a} \to B_\theta$ w.p.1 for a.e. $\theta \in \Re^p$, where $L_\theta = B_\theta B_\theta'$. If, in addition,

$$\int_K \|L_\theta^{-1}\| d\theta < \infty \quad \text{and} \quad \lim_{a\to\infty}\int_K E_\theta\|a(X_t'X_t)^{-1} - L_\theta^{-1}\|d\theta = 0$$

for a given compact set $K \subset \Re^p$, then (11) holds for the same K with $m = p$ and $A = I_p$. Then (11) also holds for any $m \le p$ and any convergent sequence A_n for which $A_nA_n' = I_m$ for all n.

4. Comparing Treatments

The problem of comparing three treatments is considered in this section. It is assumed that three treatments produce normally distributed responses with unknown means θ_1, θ_2, and θ_3 and unit variances. Siegmund (1980, 1993) proposed sequential tests for the hypothesis $H : \theta_1 = \theta_2 = \theta_3$. Here interest centers on simultaneous confidence intervals for contrasts following these tests. Of course the problem may be formulated in terms of three samples or a linear model, and elements of both formulations are used below. Let $y_{i,j}$ denote the ith observation on the jth treatment, so that the $y_{i,j}$ are independent and $y_{i,j}$ is normally distributed with mean θ_j and unit variance; alternatively, each $y_{i,j}$ may be written in the form $\theta' x + e$, where $\theta = (\theta_1, \theta_2, \theta_3)'$, x is chosen from $(1,0,0)', (0,1,0)'$, or $(0,0,1)'$, and e has a standard normal distribution.

A Sequential Test. Siegmund's (1980) sequential test for this problem depends on three design parameters, an initial sample size n_0, a boundary parameter $a \geq 1$, and a truncation parameter $\epsilon > 0$. Let $N = \lfloor a/\epsilon^2 \rfloor$, the greatest integer which is less than or equal to a/ϵ^2. Next, let Π denote the projection operator on the orthogonal complement, $\mathcal{C} \subset \Re^3$ say, of the linear subspace $\{\alpha \mathbf{1}; \alpha \in \Re\} \subseteq \Re^3$. Below, $\mathcal{C}$ is called the contrast space. Triples $(y_{k,1}, y_{k,2}, y_{k,3})'$, $k = 1, 2, \ldots$, are observed until time

$$s = \inf\{n : n \geq n_0 \text{ and } \|\Pi \mathbf{S}_n\| > \sqrt{an}\} \wedge N,$$

where $\mathbf{S}_n = \sum_{i=1}^{n} (y_{i,1}, y_{i,2}, y_{i,3})'$ and $\wedge$ denotes minimum. Thus $X_s' X_s = sI_3$ is a diagonal matrix in this example, and $\lim_{a\to\infty} a/s = \|\Pi\theta\|^2 \vee \epsilon^2 = \sigma^2(\theta)$, say, w.p.1 (P_θ) for all $\theta \in \Re^3$ by standard arguments, where $\vee$ denotes maximum. Let $B_s = \sqrt{s} I_3$ and let A denote a 2×3 matrix whose rows form an orthonormal basis for $\mathcal{C}$. If $A_s = A$, then it is easily seen that (11) holds with $Q_\theta = \sigma(\theta) A$ for all $\theta \in \Re^3$, again using standard arguments (maximal inequalities). It follows easily that $\mu(\theta) = -A \nabla \sigma(\theta)$ and $\Delta(\theta) = \mu(\theta)\mu(\theta)'$ for a.e. θ.

By (4), $W_s^* = \hat{\Gamma}_a^{-1}(W_s - a^{-\frac{1}{2}} \hat{\mu}_a)$ is approximately standard bivariate normal to order $o(1/a)$. Simulations which illustrate the accuracy of this approximation are presented in Table 1 for the case in which

$$A = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} \end{pmatrix}, \tag{19}$$

$a = 12.25$, $n_0 = 10$, and $N = 50$, and selected values of θ. Clearly the mean of W_s^* is approximately zero to within the accuracy of the simulations for the θ's considered. The standard deviations appear to be too large by about 1 or 2 percent. Let $\hat{\nu}_a = A' \hat{\mu}_a / \sqrt{as}$. Then simultaneous confidence intervals for all contrasts may be obtained from the relations, $\sup_{0 \neq c \in \mathcal{C}} |c'(\Theta - \hat{\theta}_s - \hat{\nu}_a)| / \|\hat{\Gamma}_a A c\| = \|W_s^*\| / \sqrt{s}$ and

$$P_\theta[\|W_s^*\| \leq \kappa] = 1 - e^{-\frac{1}{2}\kappa^2} + o(\frac{1}{a}) \tag{20}$$

(very weakly) as $a \to \infty$. From Table 1, the latter approximation appears to be very good.

Table 1. A sequential test

θ_1	.5	.5	.25	0
θ_2	0	$-.25$	.25	0
θ_3	$-.5$	$-.25$	$-.5$	0
$P_\theta(s < N)$	.958	.870	.873	.017
$E_\theta(s)$	22.7	28.3	28.2	49.6
$E_\theta(W_{s,1})$	$-.120$	$-.204$	$-.003$	$-.004$
$E_\theta(W_{s,2})$	$-.203$	$-.121$	$-.233$	$-.006$
$\sqrt{E_\theta}(W_{s,1}^2)$	1.04	1.04	1.08	1.02
$\sqrt{E_\theta}(W_{s,2}^2)$	1.02	1.08	1.04	1.03
$E_\theta(W_{s,1}W_{s,2})$	$-.07$	$-.02$	.00	.02
$E_\theta(W_{s,1}^*)$	$-.007$	$-.011$	$-.003$	$-.005$
$E_\theta(W_{s,2}^*)$	$-.008$	$-.008$	$-.008$	$-.006$
$\sqrt{E_\theta}(W_{s,1}^{*2})$	.99	.98	1.01	1.01
$\sqrt{E_\theta}(W_{s,2}^{*2})$	1.00	1.01	.98	1.02
$E_\theta(W_{s,1}^*W_{s,2}^*)$	$-.01$	$-.01$	.00	.01
$P_\theta(\Vert W_s^*\Vert \le \kappa_1)$	.901	.900	.901	.893
$P_\theta(\Vert W_s^*\Vert \le \kappa_2)$	.952	.951	.949	.943
$P_\theta(\Vert W_s^*\Vert \le \kappa_3)$	.991	.991	.991	.980

Entries are Monte Carlo estimates based on 10,000 replications; $a = 12.25$, $n_0 = 10$, $N = 50$, $\kappa_1 = 2.146$, $\kappa_2 = 2.448$ and $\kappa_3 = 3.035$.

A Sequential Allocation Rule. Siegmund (1993) added a second stage to the procedure described above. Let the data dependent indices $I = I^a$, $J = J^a$, and $K = K^a$ be determined (w.p.1) by $\hat{\theta}_{s,I} < \hat{\theta}_{s,J} < \hat{\theta}_{s,K}$, where $\hat{\theta}_s = \mathbf{S}_s/s$. After time s, treatment I is dropped and observations are taken on treatments J and K until time

$$t = \inf\Big\{n : n \ge s \text{ and } |\sum_{i=1}^{n}(y_{i,K} - y_{i,J})| > \frac{1}{c}\sqrt{an}\Big\} \wedge N,$$

where $c \ge 2/\sqrt{3}$. Sampling is terminated at time t.

To determine the proper corrections, it is necessary to compute the limit of $X_t'X_t/a$. Denote the ordered θ's by $\theta_{(1)} \le \theta_{(2)} \le \theta_{(3)}$. Then

$$\lim_{a\to\infty} \frac{a}{t} = c^2[\theta_{(3)} - \theta_{(2)}]^2 \wedge \Vert\Pi\theta\Vert^2 \vee \epsilon^2 = \tau^2(\theta), \text{ say},$$

w.p.1 (P_θ) for all $\theta \in \Re^3$. As above, $X_t'X_t$ is a diagonal matrix. Let $B_t = B_t'$ denote the unique diagonal square root of $X_t'X_t$ with non negative entries. Then the diagonal entries of $X_t'X_t = B_t^2$ are $b_{t,II}^2 = s$ and $b_{t,JJ}^2 = b_{t,KK}^2 = t$. Define $i = i_\theta$, $j = j_\theta$, and $k = k_\theta$ by $\theta_{(1)} = \theta_i$, $\theta_{(2)} = \theta_j$, $\theta_{(3)} = \theta_k$, $i < j$ if $\theta_{(1)} = \theta_{(2)}$, and $j < k$ if $\theta_{(2)} = \theta_{(3)}$. Then

$$\lim_{a\to\infty} a(X_t'X_t)^{-1} = D_\theta^2 = \text{diag}[d_1^2(\theta), d_2^2(\theta), d_3^2(\theta)]$$

w.p.1 (P_θ) for a.e. $\theta \in \Re^3$, where $d_i^2 = \sigma^2$, and $d_j^2 = d_k^2 = \tau^2$. The matrix D depends continuously on θ. This is not immediately clear, since the indices i, j, and k are discontinuous. It is clear that D_θ is continuous at every $\theta^o \in \Re^3$ for which $\theta^o_{(1)} < \theta^o_{(2)} \leq \theta^o_{(3)}$, since i is constant on some neighborhood of such a point, and $d_j = d_k$ everywhere. That leaves the case $\theta^o_{(1)} = \theta^o_{(2)} < \theta^o_{(3)}$. In this case, $\|\Pi\theta^o\|^2 = 2[\theta^o_{(3)} - \theta^o_{(2)}]^2/3$, so that $c^2[\theta^o_{(3)} - \theta^o_{(2)}] \geq 2\|\Pi\theta^o\|^2$ and, therefore, $\tau^2 = \sigma^2$ on some neighborhood of θ^o.

Let M be a 2×3 matrix whose rows from a basis for the contrast space $\mathcal{C}$. Further, let L_t be a 2×2 lower triangular matrix for which $L_t L_t' = M(X_t'X_t)^{-1}M'$, let $A_t = L_t^{-1}MB_t^{-1}$, and let $W_t = A_t Z_t = L_t^{-1}M(\Theta - \hat{\theta}_t)$. Then, $A_t A_t' = I_2$ as required in (1), and it is easily seen that (11) is satisfied with $Q_\theta = \mathcal{L}_\theta^{-1} M D_\theta^2$ for all $\theta \in R^3$, where $\mathcal{L}_\theta$ denotes the limit of $\sqrt{a}L_t$. Define μ and Δ by (16) and (17), and let $\hat{\mu}_a = \mu(\hat{\theta}_t)$ and $\hat{\Gamma}_a = I_2 + \Delta(\hat{\theta}_t)/(2a)$. Then (4) holds for all symmetric (sign invariant) functions $h \in \mathcal{H}_2$ by a simple application of Theorem 2 below. The accuracy of this approximation is illustrated in Table 2 for the case in which M is given by (19) and selected values of θ. Let $\hat{\nu}_a = M'L_t\hat{\mu}_a/\sqrt{at}$. Then $\sup_{0\neq c \in \mathcal{C}} |c'(\Theta - \hat{\theta}_t - \hat{\nu}_a)|/\|\hat{\Gamma}_a L_t M c\| = \|W_t^*\|$, as above, and approximate simultaneous confidence intervals for all contrasts may be determined from (20). From Table 2, the accuracy of this approximation appears to be good too.

Table 2. A sequential test with sequential allocation

θ_1	.5	.5	.25	0
θ_2	0	$-.25$	.25	0
θ_3	$-.5$	$-.25$	$-.5$	0
$E_\theta(t)$	31.9	30.2	46.9	49.8
$E_\theta(W_{t,1})$	$-.232$	$-.211$	.012	.012
$E_\theta(W_{t,2})$	$-.163$	$-.110$	$-.176$	.002
$\sqrt{E_\theta}(W_{t,1}^2)$	1.04	1.05	1.14	1.02
$\sqrt{E_\theta}(W_{t,2}^2)$	1.00	1.06	1.02	1.03
$E_\theta(W_{t,1}W_{t,2})$	.05	$-.02$	$-.01$	.00
$E_\theta(W_{t,1}^*)$	.016	.015	.009	.011
$E_\theta(W_{t,2}^*)$	$-.116$	.022	$-.170$	.002
$\sqrt{E_\theta}(W_{t,1}^{*2})$	.91	.97	1.03	1.01
$\sqrt{E_\theta}(W_{t,2}^{*2})$	.99	1.08	1.02	1.03
$E_\theta(W_{t,1}^*W_{t,2}^*)$	$-.04$	$-.16$	$-.01$	.00
$P_\theta(\|W_t^*\| \leq \kappa_1)$	.914	.905	.901	.891
$P_\theta(\|W_t^*\| \leq \kappa_2)$	.959	.952	.954	.941
$P_\theta(\|W_t^*\| \leq \kappa_3)$	.992	.989	.991	.983

Entries are Monte Carlo estimates based on 10,000 replications; $a = 12.25$, $c = 1.198$, $n_0 = 10$, $N = 50$, $\kappa_1 = 2.146$, $\kappa_2 = 2.448$ and $\kappa_3 = 3.035$.

The values of a, c, n_0, and N used in Tables 1 and 2 were chosen to agree with Siegmund (1993). Other simulations were conducted with $a = 15$, 20, $n_0 =$ 1, 5, and $N = 75$ with similar results. The accuracy of the approximations did not deteriorate when n_0 was decreased from 10 to 5.

An alternative to Siegmund's (1993) procedure has been proposed by Betensky (1995) who changed s and t to $s = \inf\{n \geq n_0 : \|\Pi \mathbf{S}_n\| > a\} \wedge N$ and $t = \inf\{n \geq s : |\sum_{i=1}^{n}(y_{i,K} - y_{i,J})| > a/c\} \wedge N$, where $c \geq 2/\sqrt{3}$. Corrected confidence sets for her procedure are similar to those for Siegmund's, and simulations indicate that the approximations are slightly better. They differ in the functional forms of σ and τ.

5. Remarks and Open Questions

The questions addressed in Section 4 are motivated by phase III clinical trials in which new treatments are tested on human subjects. It is straightforward to extend the analysis of the first stage of Siegmund's procedure to the case of several treatments. The second stage presents more difficulty. It is not even clear what form the second stage should take, or whether there should be multiple stages. A question not addressed in Section 4 is that of finding corrected confidence levels for a contrast, like $\theta_K - \theta_J$, in which the indices may be random variables. Such contrasts present technical difficulties in that the resulting Q_θ matrix may be discontinuous in θ.

The case of unknown variability presents another question. If the model is changed to $y_k = x_k'\theta + \sigma e_k$, $k = 1, 2, \ldots$, where $\sigma > 0$ is unknown, then Z_t may be changed to $\hat{Z}_t = Z_t/\hat{\sigma}_t$, where $\hat{\sigma}_t$ denotes an estimator of σ–for example, $\hat{\sigma}_t^2 = \|\boldsymbol{y}_t - X_t\hat{\theta}_t\|^2/(t-p)$. For this case and $A_n = I_p$, considerations like those presented in Section 7 suggest that $\mu(\theta)$ and $\Delta(\theta)$ should be replaced by $\sigma\mu(\theta)$ and $\sigma^2\Delta(\theta) + a(v - b\sigma^2)I_p/\sigma^4$, where b and v denote the bias and variance of $\hat{\sigma}_t^2$ (which must be of order $1/a$). The authors hope to present the details of this extension elsewhere and to relate it to the work of Coad (1995).

Relation (13) may be useful in obtaining higher order approximations to the integrated risk of sequential designs, like Wu's (1985) adaptive design for estimating non-linear functions, with respect to a large class of prior densities. In principle, such approximations may lead to refinements of designs which are optimal to first order.

6. Proof of (13)

Let $\Omega \subseteq \Re^p$ be open and let Ξ_Ω denote the class of twice continuously differentiable densities ξ with compact support in Ω. The inclusion of Ω in the model allows the expansions to fail on a subset of $\Re^p$ (the complement of Ω).

Recall that $t = t_a$, $a \geq 1$, denote stopping times.

Theorem 1. *Let $\Omega \subseteq \Re^p$ be open; suppose that* (11) *holds for all compact $K \subset \Omega$; and define I_a and II_a by* (12). *If $\xi \in \Xi_\Omega$, then*

$$\lim_{a\to\infty} E_\xi\{\|I_a\|\} = 0, \tag{21}$$

and

$$\lim_{a\to\infty} E_\xi\{\text{essup}_{h\in\mathcal{H}_2^m}|II_a(h)|\} = 0. \tag{22}$$

If also,

$$\lim_{a\to\infty} \sqrt{a}\int_K \|E_\theta\{Q^a - Q_\theta\}\|\, d\theta = 0, \text{ for all compact } K \subset \Omega, \tag{23}$$

then

$$\lim_{a\to\infty} \sqrt{a}E_\xi(I_a) = 0. \tag{24}$$

Proof. In the proof, ξ is written for $\xi(\Theta)$, $\nabla\xi$ for $\nabla\xi(\Theta)$, etc. Relations (21) and (24) are easy, since

$$E_\xi\{\|I_a\|\} \le E_\xi\Big\{\|Q^a - Q_\theta\|\,\|\frac{\nabla\xi}{\xi}\|\Big\} = \int_\Omega E_\theta\{\|Q^a - Q_\theta\|\}\|\nabla\xi\|d\theta \to 0,$$

as $a \to \infty$, by the assumption (11), since ξ has compact support; and if (23) holds, then

$$\sqrt{a}E_\xi\{I_a\} = \sqrt{a}\int_\Omega E_\theta\{Q^a - Q_\theta\}\nabla\xi(\theta)d\theta \to 0,$$

as $a \to \infty$, again since ξ has compact support.

For (22), write

$$II_a(h) = II_{1,a}(h) + II_{2,a}(h) + II_{3,a}(h),$$

where

$$II_{1,a}(h) = E_\xi^t\Big\{\text{tr}\Big[(V_m h(W_t) - \Phi^m V_m h)Q^a[\frac{\nabla^2\xi}{\xi} - E_\xi^t(\frac{\nabla^2\xi}{\xi})]Q^{a\prime}\Big]\Big\},$$
$$II_{2,a}(h) = E_\xi^t\Big\{\text{tr}\Big[(V_m h(W_t) - \Phi^m V_m h)Q^a E_\xi^t(\frac{\nabla^2\xi}{\xi})Q^{a\prime}\Big]\Big\},$$

and

$$II_{3,a}(h) = E_\xi^t\Big\{\text{tr}\Big[(\Phi^m V_m h)(Q^a\frac{\nabla^2\xi}{\xi}Q^{a\prime} - Q_\Theta\frac{\nabla^2\xi}{\xi}Q_\Theta')\Big]\Big\}.$$

By Proposition 1, there is a constant C for which $\|V_m h(w)\| \le C$ for all $w \in \Re^m$ and all $h \in \mathcal{H}_2^m$. Let

$$M_t = \text{essup}_{h\in\mathcal{H}_2^m}\|E_\xi^t\{\text{tr}[V_m h(W_t) - \Phi^m V_m h]\}\|.$$

Then $M_t \le 2C$ w.p.1, and $M_t \to 0$ in P_ξ-probability, by Lemma 1 of Woodroofe (1989). Now

$$|II_{1,a}(h)| \le 2CE_\xi^t\Big\{\|Q^a[\frac{\nabla^2\xi}{\xi} - E_\xi^t(\frac{\nabla^2\xi}{\xi})]Q^{a\prime}\|\Big\} = II_{1,a}^*, \text{ say},$$
$$|II_{2,a}(h)| \le M_t\|Q^a E_\xi^t(\frac{\nabla^2\xi}{\xi})Q^{a\prime}\| = II_{2,a}^*, \text{ say},$$

and

$$|II_{3,a}(h)| \le CE_\xi^t\Big\{\|Q^a\frac{\nabla^2\xi}{\xi}Q^{a\prime} - Q_\Theta\frac{\nabla^2\xi}{\xi}Q_\Theta'\|\Big\} = II_{3,a}^*, \text{ say},$$

for all $h \in \mathcal{H}_2^m$, and it suffices to show that $E_\xi(II_{i,a}^*) \to 0$ as $a \to \infty$ for $i = 1, 2, 3$. For $i = 3$, this is clear, since

$$\begin{aligned} E_\xi(II_{3,a}^*) &\le CE_\xi\Big\{\|Q^a - Q_\Theta\|\,\|\frac{\nabla^2\xi}{\xi}\|(\|Q^a\| + \|Q_\Theta\|)\Big\} \\ &\le C\sqrt{E}_\xi\Big\{\|Q^a - Q_\Theta\|^2\|\frac{\nabla^2\xi}{\xi}\|\Big\}\sqrt{E}_\xi\Big\{(\|Q^a\| + \|Q_\Theta\|)^2\|\frac{\nabla^2\xi}{\xi}\|\Big\} \end{aligned}$$

and

$$E_\xi\Big\{\|Q^a - Q_\Theta\|^2\|\frac{\nabla^2\xi}{\xi}\|\Big\} \le \int_\Omega E_\theta\{\|Q^a - Q_\theta\|^2\}\|\nabla^2\xi\|d\theta \to 0 \qquad (25)$$

as $a \to \infty$ by (11). Moreover, it follows from (25) that

$$\Big\|Q^a\frac{\nabla^2\xi}{\xi}Q^{a\prime}\Big\|,\ a \ge 1, \quad \text{and} \quad \Big\|Q^a E_\xi^t(\frac{\nabla^2\xi}{\xi})Q^{a\prime}\Big\|,\ a \ge 1,$$

are uniformly integrable with respect to P_ξ, since the first sequence converges in $L^1(P_\xi)$, and the second is bounded by the conditional expectation of the first. That $E_\xi(II_{1,a}^* + II_{2,a}^*) \to 0$ as $a \to \infty$ then follows directly, since $M_t \to 0$ in $P_\xi-$probability and

$$\lim_{n\to\infty} E_\xi^n\Big(\frac{\nabla^2\xi}{\xi}\Big) = \frac{\nabla^2\xi}{\xi} \text{ w.p.1 } (P_\xi)$$

by the Martingale Convergence Theorem.

In the Corollary, $\mathcal{H}_{r,0}^m$ denotes the set of $h \in \mathcal{H}_r^m$ for which $\Phi^m U_m h = 0$. This class includes all sign invariant $h \in \mathcal{H}_r^m$.

Corollary. *Let $\xi \in \Xi_\Omega$. If* (11) *holds for all compact $K \subset \Omega$, then* (13) *holds uniformly with respect to $h \in \mathcal{H}_{2,0}^m$ with "$\approx$" replaced by "$= +o(1/a)$;" and if* (24) *holds, then* (13) *holds uniformly with respect to $h \in \mathcal{H}_2^m$.*

Proof. The corollary is clear from (12) and Theorem 1.

The uniformity asserted in the theorem is much stronger than that of the Corollary.

7. Proof of (4)

The following lemma is needed in the proof of (4).

Lemma 1. *Let* $\nu \in \Re^m$ *and let* Γ *be a non singular* $m \times m$ *matrix for which* $\|\nu\| \le 1$ *and* $\|\Gamma - I_m\| \le 1$. *For* $h \in \mathcal{H}_2^m$, *let* $h^*(x) = h[\Gamma^{-1}(x-\nu)]$ *for all* $x \in \Re^m$. *Then there is a constant* C, *independent of* h, ν, *and* Γ, *for which*

$$\Phi^m h^* - \Phi^m h = -(\Phi^m U_m h)'\nu + \mathrm{tr}\{(\Phi^m V_m h)[\nu\nu' - 2(\Gamma - I_m)]\} + III(h;\Gamma,\nu), \quad (26)$$

$$\Phi^m U_m h^* - \Phi^m U_m h = -2(\Phi^m V_m h)\nu + IV(h;\Gamma,\nu),$$

and

$$\|\Phi^m V_m h^* - \Phi^m V_m h\| \le C[\|\nu\| + \|\Gamma - I_m\|],$$

where $|III(h;\Gamma,\nu)| \le C[\|\nu\|^3 + \|\Gamma - I_m\|^{\frac{3}{2}}]$ *and* $|IV(h;\Gamma,\nu)| \le C[\|\nu\|^2 + \|\Gamma - I_m\|]$.

Proof. The proof of (26) depends on the observations that

$$\Phi^m h^* = \int_{\Re^m} h(z)|\det(\Gamma)|\phi^m(\Gamma z + \nu)dz,$$

where ϕ^m denotes the standard m-variate normal density, and that $|\det(\Gamma)|$ $\phi^m(\Gamma z + \nu)$ form an exponential family of densities with natural parameters $\Gamma'\Gamma$ and $\Gamma'\nu$. It follows that $\Phi^m h^*$ is differentiable, and (26) then follows from a straightforward Taylor series expansion for any C that is an upper bound for the partial derivatives of order three. That C may be chosen independently of h then follows from basic analytic properties of exponential families. See, for example, Brown (1986, pp. 34-36). The proof of the remainder of the lemma is similar.

In the proof of (4), the estimators $\hat{\mu}_a$ and $\hat{\Delta}_a$ must be so chosen that

$$\|\hat{\mu}_a\| \le \sqrt{a}, \qquad \|\hat{\Delta}_a\| \le a, \quad (27)$$

$$\lim_{a\to\infty}\Big[E_\xi\{\|\hat{\mu}_a - \mu(\Theta)\|^2\} + E_\xi\{\|\hat{\Delta}_a - \Delta(\Theta)\|\}\Big] = 0 \quad (28)$$

and

$$\lim_{a\to\infty}\sqrt{a}E_\xi\{\hat{\mu}_a - \mu(\Theta)\} = 0 \quad (29)$$

for all $\xi \in \Xi_\Omega$. If the partial derivatives of $q_{i,j}$ are bounded and continuous, then such estimators may be constructed by letting $\hat{\mu}_a = \mu_a(\hat{\theta}_t)$ and $\hat{\Delta}_a = \Delta(\hat{\theta}_t)$. That such estimators exist more generally is shown in Proposition 3 below, provided that

$$\int_K (\|\mu(\theta)\|^2 + \|\Delta(\theta)\|)d\theta < \infty \quad (30)$$

for all compact $K \subset \Omega$.

Theorem 2. *Let $\Omega \subseteq \Re^p$ be a convex open set and let $\xi \in \Xi_\Omega$. Suppose that* (11) *holds for all compact $K \subset \Omega$ and that the entries of Q_θ are almost differentiable on Ω with locally square integrable gradients. Let $\hat{\mu}_a$ and $\hat{\Gamma}_a = I_m + \hat{\Delta}_a/(2a)$ be estimators for which* (27) *and* (28) *hold. Then*

$$E_\xi\{h(W_t^*)\} = \Phi^m h + o(\frac{1}{a}) \tag{31}$$

uniformly with respect to $h \in \mathcal{H}_{2,0}^m$ as $a \to \infty$. If, in addition, (23) *and* (29) *hold, then* (31) *holds uniformly with respect to $h \in \mathcal{H}_2^m$.*

Proof. Fix $\xi \in \Xi$ throughout the proof, as above, and write ξ for $\xi(\Theta)$, etc. Given $h \in \mathcal{H}_2^m$, let $h_a(w) = h[\hat{\Gamma}_a^{-1}(w - a^{-\frac{1}{2}}\hat{\mu}_a)]$ for $w \in \Re^m$, so that $h(W_t^*) = h_a(W_t)$ and, therefore, $E_\xi\{h(W_t^*)\} = E_\xi\{E_\xi^t[h_a(W_t)]\}$. Using (12) and Lemma 1,

$$\begin{aligned} E_\xi^t[h_a(W_t)] - \Phi^m h_a &= \frac{1}{\sqrt{a}}(\Phi^m U_m h_a)' E_\xi^t\Big\{Q^a \frac{\nabla \xi}{\xi}\Big\} \\ &\quad + \frac{1}{a} E_\xi^t\Big\{\mathrm{tr}[(\Phi^m V_m h_a) Q_\Theta \frac{\nabla^2 \xi}{\xi} Q_\Theta']\Big\} + \frac{1}{a} I\!I_a(h_a), \\ \Phi^m h_a - \Phi^m h &= -\frac{1}{\sqrt{a}}(\Phi^m U_m h)'\hat{\mu}_a + \frac{1}{a}\,\mathrm{tr}\{(\Phi^m V_m h)[\hat{\mu}_a\hat{\mu}_a' - \hat{\Delta}_a]\} + \frac{1}{a} I\!I\!I_a(h), \end{aligned}$$

and

$$\Phi^m U_m h_a - \Phi^m U_m h = -\frac{2}{\sqrt{a}}(\Phi^m V_m h)\hat{\mu}_a + I\!V\Big(h; \hat{\Gamma}_a, \frac{\hat{\mu}_a}{\sqrt{a}}\Big),$$

where $I\!I_a(h)$ is as in (12) and $I\!I\!I_a(h) = aI\!I\!I(h; \hat{\Gamma}_a, a^{-\frac{1}{2}}\hat{\mu}_a)$ with $I\!I\!I$ as in Lemma 1. So,

$$\begin{aligned} E_\xi^t[h_a(W_t)] - \Phi^m h &= E_\xi^t[h_a(W_t)] - \Phi^m h_a + \Phi^m h_a - \Phi^m h \\ &= \frac{1}{\sqrt{a}}(\Phi^m U_m h)'\Big[E_\xi^t(Q^a \frac{\nabla \xi}{\xi}) - \hat{\mu}_a\Big] + \frac{1}{a} E_\xi^t\{tr[(\Phi^m V_m h) M_a]\} \\ &\quad + \frac{1}{a}[I\!I_a(h_a) + I\!I\!I_a(h) + I\!V_a(h)], \end{aligned}$$

where

$$M_a = Q_\Theta\Big(\frac{\nabla^2 \xi}{\xi}\Big)Q_\Theta' + \hat{\mu}_a\hat{\mu}_a' - \hat{\Delta}_a - 2E_\xi^t\Big\{Q^a(\frac{\nabla \xi}{\xi})\Big\}\hat{\mu}_a',$$

and

$$\begin{aligned} I\!V_a(h) &= \sqrt{a} I\!V(h; \hat{\Gamma}_a, a^{-\frac{1}{2}}\hat{\mu}_a)' E_\xi^t\Big\{Q^a \frac{\nabla \xi}{\xi}\Big\} \\ &\quad + E_\xi^t\Big\{\mathrm{tr}[(\Phi^m V_m h_a - \Phi^m V_m h) Q_\Theta (\frac{\nabla^2 \xi}{\xi}) Q_\Theta']\Big\} \end{aligned}$$

with IV as in Lemma 1.

The remainder terms, $II_a(h_a)$, $III_a(h)$, and $IV_a(h)$, are negligible. It was shown in the proof of Theorem 1 that $\lim_{a\to\infty} E_\xi\{\mathrm{essup}_{h\in\mathcal{H}_2^m}|II_a(h)|\} = 0$, and it follows that $E_\xi\{\mathrm{essup}_{h\in\mathcal{H}_2^m}|II_a(h_a)|\} \to 0$ as $a \to \infty$. For III_a, it is clear from Lemma 1, (27) and (28) that there is a constant C for which

$$E_\xi\{\mathrm{essup}_{h\in\mathcal{H}_2^m}|III_a(h)|\} \le CE_\xi\Big\{\frac{\|\hat{\mu}_a\|^3 + \|\hat{\Delta}_a\|^{\frac{3}{2}}}{\sqrt{a}}\Big\} \to 0,$$

as $a \to \infty$, since the integrand converges to zero in P_ξ-probability and is bounded by $C[\|\hat{\mu}_a\|^2 + \|\hat{\Delta}_a\|]$ which converges in the mean and is, therefore, uniformly integrable. Similarly, there is a constant C for which

$$\begin{aligned}
&\mathrm{essup}_{h\in\mathcal{H}_2^m}\|IV_a(h)\| \\
&\le C\Big[\frac{\|\hat{\mu}_a\|^2+\|\hat{\Delta}_a\|}{\sqrt{a}}\Big]\|Q^a\|E_\xi^t\Big(\frac{\|\nabla\xi\|}{\xi}\Big)+C\|\Phi^m V_m h_a-\Phi^m V_m h\|E_\xi^t\Big[\|Q_\theta\Big(\frac{\nabla^2\xi}{\xi}\Big)Q_\theta'\|\Big] \\
&= IV_{1,a} + IV_{2,a},\ \text{say.}
\end{aligned}$$

Clearly, $IV_{2,a} \to 0$ in P_ξ-probability, and $IV_{2,a}$ is uniformly integrable, so that $E_\xi(IV_{2,a}) \to 0$ as $a \to \infty$. Similarly, $IV_{1,a} \to 0$ in P_ξ-probability, and $IV_{1,a}$ is uniformly integrable, since $E_\xi(IV_{1,a}^2)$ remains bounded. So, $E_\xi(IV_{1,a}) \to 0$ and, therefore,

$$\begin{aligned}
E_\xi\{h(W_a^*)\} - \Phi^m h = {}& \frac{1}{\sqrt{a}}(\Phi^m U_m h)E_\xi\Big\{Q^a(\frac{\nabla\xi}{\xi}) - \hat{\mu}_a\Big\} \\
&+\frac{1}{a}E_\xi\Big\{\mathrm{tr}[(\Phi^m V_m h)M_a]\Big\} + o(\frac{1}{a}),
\end{aligned} \tag{32}$$

uniformly with respect to $h \in \mathcal{H}_2^m$ as $a \to \infty$.

Next consider $E_\xi^t\{\mathrm{tr}[(\Phi^m V_m h)M_a]\}$. Let

$$M(\theta) := Q_\theta\frac{\nabla^2\xi}{\xi}Q_\theta' + Q_\theta^{\#}\mathbf{1}\mathbf{1}'Q_\theta^{\#\prime} + 2Q_\theta(\frac{\nabla\xi}{\xi})(Q_\theta^{\#}\mathbf{1})' - \Delta(\theta) \tag{33}$$

for $\theta \in \Omega$. Then $E_\xi[\|M_a - M(\Theta)\|] \to 0$ as $a \to \infty$ by (27), (28), and the Martingale Convergence Theorem. So,

$$\lim_{a\to\infty} E_\xi\Big\{E_\xi^t\{\mathrm{tr}[(\Phi^m V_m h)M_a]\}\Big\} = \mathrm{tr}\Big[(\Phi^m V_m h)\int_\Omega M(\theta)\xi(\theta)d\theta\Big],$$

uniformly with respect to $h \in \mathcal{H}_2^m$. The term on the right, $\mathrm{tr}[(\Phi^m V_m h)\int_\Omega M\xi d\theta]$, is zero by an integration by parts. See Lemma 2 below.

This establishes the first assertion of the theorem, since $\Phi^m U_m h = 0$ for all $h \in \mathcal{H}_{2,0}^m$.

For the second assertion, it suffices to show that the first expectation on the right side of (32) is $o(1/\sqrt{a})$ as $a \to \infty$. This is clear, however. For

$$\sqrt{a}E_\xi\Big\{Q^a(\frac{\nabla\xi}{\xi}) - \hat{\mu}_a\Big\} = \sqrt{a}E_\xi\Big\{(Q^a - Q_\Theta)(\frac{\nabla\xi}{\xi})\Big\} + \sqrt{a}E_\xi\Big\{Q_\Theta(\frac{\nabla\xi}{\xi}) + \mu(\Theta)\Big\} + \sqrt{a}E_\xi\{\mu(\Theta) - \hat{\mu}_a\}.$$

The first and third terms approach zero as $a \to \infty$, by (23) and (29), and the middle term is zero by an integration by parts. This completes the proof, except for the proof of Lemma 2.

Lemma 2. *Let $\Omega \subseteq \Re^p$ be a convex open set and $\xi \in \Xi_\Omega$. Let Q be an $m \times p$ matrix which has almost differentiable entries with locally square integrable gradients and define M by* (33)*. Then $\int_\Omega M\xi d\theta$ is a skew symmetric matrix.*

Proof. Suppose first that the entries of Q are twice continuously differentiable. Then the (i,j)th entry in $\int_\Omega M\xi d\theta$ is

$$\sum_{r=1}^p\sum_{s=1}^p\int_\Omega\Big\{(q_{ir}q_{js})\frac{\partial^2\xi}{\partial\theta_r\partial\theta_s} + 2q_{ir}(\frac{\partial q_{js}}{\partial\theta_s})(\frac{\partial\xi}{\partial\theta_r}) + (\frac{\partial q_{ir}}{\partial\theta_r})(\frac{\partial q_{js}}{\partial\theta_s})\xi - (\frac{\partial q_{ir}}{\partial\theta_s})(\frac{\partial q_{js}}{\partial\theta_r})\xi\Big\}d\theta.$$

If the terms involving the partial derivatives of ξ are integrated by parts, then the latter expression becomes

$$\sum_{r=1}^p\sum_{s=1}^p\int_\Omega\Big\{(\frac{\partial^2}{\partial\theta_r\partial\theta_s}q_{ir})q_{js} - q_{ir}(\frac{\partial^2}{\partial\theta_r\partial\theta_s}q_{js})\Big\}\xi d\theta,$$

which is the (i,j)th entry of a skew symmetric matrix. This establishes the lemma when the entries of Q are twice continuously differentiable. The general case then follows from a simple approximation argument using (6).

8. Construction of Estimators

In Theorem 2 the estimators $\hat{\mu}_a$ and $\hat{\Delta}_a$ were required to satisfy conditions (27), (28), and (29). The existence of such estimators is shown in this section. Throughout this section, $\Omega \subseteq \Re^p$ denotes an open set, condition (30) is assumed, and (11) is required to hold with $m = p$ and $A_n = I_m$ for all $n \geq n_0$. Recall that λ_n denotes the minimal eigenvalue of $X_n'X_n$ and that there are n_0 and $\lambda^o > 0$ for which $\lambda_n \geq \lambda^o$ w.p.1 (P_θ) for all $\theta \in \Re^p$ and all $n \geq n_0$. By (11), $a/\lambda_t,\ a \geq 1$, are uniformly integrable.

Three lemmas are needed.

Lemma 3. *If $\xi \in \Xi_\Omega$, then $P_\xi[\|\hat{\theta}_t - \Theta\| \geq \epsilon] = o(1/a)$ for all $\epsilon > 0$ and $E_\xi[\|\hat{\theta}_t - \Theta\|^2] = O(1/a)$, as $a \to \infty$.*

Proof. This follows easily from (11), Theorem 1, and the observation that $\|Z_t\|^2 = (\hat{\theta}_t - \Theta)'(X_t'X_t)(\hat{\theta}_t - \Theta) \geq \lambda_t\|\hat{\theta}_t - \Theta\|^2$. Let $h_t(z)$ be the indicator of $|z| \geq \sqrt{\lambda_t}\epsilon$. Then

$$\begin{aligned} P_\xi[\|\hat{\theta}_t - \Theta\| \geq \epsilon] &\leq E_\xi[h_t(Z_t)] \\ &= E_\xi\Big\{\Phi^p h_t + \frac{1}{a}\operatorname{tr}[(\Phi^p V_p h_t)Q^a(\frac{\nabla^2\xi}{\xi})Q^{a\prime}]\Big\} \\ &\leq \frac{1}{a}E_\xi\Big\{(\frac{a}{\lambda_t})\lambda_t\Phi^p h_t + \|\Phi^p V_p h_t\|\,\|Q^a(\frac{\nabla^2\xi}{\xi})Q^{a\prime}\|\Big\}. \end{aligned}$$

The terms $\lambda_t\Phi^p h_t$ and $\|\Phi^p V_p h_t\|$ converge to zero boundedly in P_ξ-probability, and the last line is uniformly integrable by (11) and the proof of Theorem 1. The first assertion of the lemma follows. The proof of the second is similar. In fact,

$$E_\xi[\|\hat{\theta}_t - \Theta\|^2] \leq E_\xi[\lambda_t^{-1}\|Z_t\|^2] = E_\xi\Big[\lambda_t^{-1}\{p + \frac{1}{a}\operatorname{tr}[Q^a(\frac{\nabla^2\xi}{\xi})Q^{a\prime}]\}\Big],$$

which is of order $1/a$ by (11).

Lemma 4. *Let d be a positive integer, and let $g : \Omega \to \Re^d$ be a locally square integrable function. Then there is a bounded, twice continuously differentiable, positive density π on Ω for which $\int_\Omega \|g\|^2\pi d\theta < \infty$ and $P_\pi[\|\hat{\theta}_t - \Theta\| \geq \epsilon] = o(1/a)$, as $a \to \infty$ for all $\epsilon > 0$.*

Proof. There are compact K_n, $n \geq 1$, for which $K_n \subset K_{n+1}^o$ for all $n \geq 1$, and $\Omega = \cup_{n=1}^\infty K_n^\circ$, where K_n° denotes the interior of K_n; and there are twice continuously differentiable functions π_n, $n \geq 1$, for which $\mathbf{1}_{K_n} \leq \pi_n \leq \mathbf{1}_{K_{n+1}}$, for all $n \geq 1$. Clearly, each π_n is integrable. Write P_{π_n} for the mixture measures $\int_\Omega P_\theta\pi_n(\theta)d\theta$, $n \geq 1$, even though these are not normalized to be probability measures. Then $P_{\pi_n}[\|\hat{\theta}_t - \Theta\| \geq \epsilon] = o(1/a)$, as $a \to \infty$ for all $\epsilon > 0$ and all $n \geq 1$ by Lemma 3. So, there are $1 \leq b_1 < b_2 < \cdots$ for which

$$P_{\pi_n}\Big[\|\hat{\theta}_t - \Theta\| \geq \frac{1}{n}\Big] \leq \frac{1}{an}$$

for all $a \geq b_n$ and $n \geq 1$. Let $\alpha_n = 2n^2\int_\Omega(b_n + \|g\|^2)\pi_n d\theta$ for all $n \geq 1$ and let $\pi = c^{-1}\sum_{n=1}^\infty \pi_n/\alpha_n$, where c is a normalizing constant and $c < 1$. Then π has the desired properties.

That $\|g\|^2\pi$ is integrable is clear. The proof of the second assertion uses the relation

$$P_\pi[\|\hat{\theta}_t - \Theta\| \geq \epsilon] = \Big[\sum_{j=1}^{m_0} + \sum_{j=m_0+1}^{m_1} + \sum_{j=m_1+1}^{\infty}\Big]\frac{1}{c\alpha_j}P_{\pi_j}\{\|\hat{\theta}_t - \Theta\| \geq \epsilon\}, \qquad (34)$$

where $m_0 < m_1$ are integers depending on ϵ. Given $\epsilon > 0$, let m_0 be the least positive integer which exceeds $4/(c\epsilon)$. Then there is an $A > 0$ for which

$$\sum_{j=1}^{m_0} \frac{1}{c\alpha_j} P_{\pi_j}[\|\hat{\theta}_t - \Theta\| \geq \epsilon] \leq \frac{\epsilon}{4a} \tag{35}$$

for all $a \geq A$, since the left side of (35) is $o(1/a)$ as $a \to \infty$, and there is no loss of generality in requiring $A > b_{m_0+1}$. If $a > A$, then there is an integer $m_1 > m_0$ for which $b_{m_1} \leq a < b_{m_1+1}$. Then, since $1/j \leq c\epsilon/4$ for $j \geq m_0$,

$$\sum_{j=m_0+1}^{m_1} \frac{1}{c\alpha_j} P_{\pi_j}\{\|\hat{\theta}_t - \Theta\| \geq \epsilon\} \leq \sum_{j=m_0+1}^{m_1} \frac{1}{\alpha_j} \frac{\epsilon}{4a} \leq \frac{\epsilon}{4a}$$

and

$$\sum_{j=m_1+1}^{\infty} \frac{1}{c\alpha_j} P_{\pi_j}\Big\{\|\hat{\theta}_t - \Theta\| \geq \frac{1}{j}\Big\} \leq \frac{1}{cb_{m_1+1}} \sum_{j=m_1+1}^{\infty} \frac{1}{2j^2} \leq \frac{\epsilon}{2a},$$

by the choice of b_n and α_n. That is, the left side of (34) is at most ϵ/a for all sufficiently large a, as asserted.

In the proofs of Lemma 5 and Proposition 3, it is necessary to compare two measures P_ξ and P_π, where ξ and π are densities on Ω. Clearly, if ξ/π is bounded above, say $\xi/\pi \leq c$, then $P_\xi \leq cP_\pi$.

Lemma 5. *Let π be a positive, twice continuously differentiable density on Ω for which* (35) *holds. If r is continuously differentiable with compact support in Ω, then $E_\pi[|r(\Theta) - r(\hat{\theta}_t)|^2] = O(1/a)$, as $a \to \infty$.*

Proof. Let J denote the compact support of r; let K be a compact set for which $J \subset K^\circ \subset K \subset \Omega$; and let ϵ be the distance from J to K', the complement of K. Then there is a constant C for which

$$E_\pi[|r(\Theta) - r(\hat{\theta}_t)|^2] \leq CE_\pi[\|\Theta - \hat{\theta}_t\|^2 \mathbf{1}_K(\Theta)] + CP_\pi[\|\Theta - \hat{\theta}_t\| \geq \epsilon].$$

The second term on the right is of smaller order of magnitude than $1/a$ by Lemma 4. For the first, there is a positive, twice continuously differentiable function ρ with compact support in Ω for which $\mathbf{1}_K \leq \rho \leq 1$. Let ξ be the density for which $\xi \propto \pi\rho$. Then, since $\pi\mathbf{1}_K \leq \pi\rho \leq \xi$,

$$E_\pi[\|\hat{\theta}_t - \Theta\|^2 \mathbf{1}_K(\Theta)] \leq E_\xi[\|\hat{\theta}_t - \Theta\|^2] = O(\frac{1}{a}),$$

by Lemma 3.

Proposition 3. *If g is as in Lemma 4, then there are estimators $\hat{g}_n$, $n \geq 1$, for which*

$$\lim_{a\to\infty} \{E_\xi[\|\hat{g}_t - g(\Theta)\|^2] + \sqrt{a}\|E_\xi[\hat{g}_t - g(\Theta)]\|\} = 0 \tag{36}$$

for all $\xi \in \Xi_\Omega$.

Proof. Fix $\xi \in \Xi_\Omega$ throughout the proof. Construct π as in Lemma 4, and let $\hat{g}_n = E_\pi^n[g(\Theta)]$ for all $n \geq 1$. Then $\hat{g}_n \to g(\Theta)$ w.p.1 (P_π) and $\sup_{n\geq 1} E_\pi[\|\hat{g}_n\|^2] \leq 4E_\pi[\|g(\Theta)\|^2] < \infty$ by the Martingale Convergence Theorem and Doob's Inequality. It then follows from the Dominated Convergence Theorem that $\lim_{a\to\infty} E_\pi[\|\hat{g}_t - g(\Theta)\|^2] = 0$ and, therefore, $\lim_{a\to\infty} E_\xi[\|\hat{g}_t - g(\Theta)\|^2] = 0$, since ξ/π is bounded above. Next, let $\tilde{g}_n = E_\xi^n[g(\Theta)]$, $n \geq 1$. Then $E_\xi[\hat{g}_t - g(\Theta)] = E_\xi[\hat{g}_t - \tilde{g}_t]$ for all $a \geq 1$. So, it suffices to show that $\lim_{a\to\infty} \sqrt{a}E_\xi[\|\hat{g}_t - \tilde{g}_t\|] = 0$. Let $r(\theta) = \xi(\theta)/\pi(\theta)$ for all $\theta \in \Omega$. Then $\tilde{g}_n = E_\pi^n[g(\Theta)r(\Theta)]/E_\pi^n[r(\Theta)]$, and

$$\tilde{g}_n - \hat{g}_n = \frac{\mathrm{Cov}_\pi^n[g(\Theta), r(\Theta)]}{E_\pi^n[r(\Theta)]},$$

where Cov_π^n denotes the $d \times 1$ vector of conditional covariances of the components of g with r. It follows that

$$E_\xi[\|\tilde{g}_t - \hat{g}_t\|] = E_\pi\{\|\tilde{g}_t - \hat{g}_t\| E_\pi^t[r(\Theta)]\} = E_\pi\{\|\mathrm{Cov}_\pi^t[g(\Theta), r(\Theta)]\|\};$$

and

$$\begin{aligned} E_\pi[\|\mathrm{Cov}_\pi^t[g(\Theta), r(\Theta)]\|] &\leq \sqrt{E}_\pi[\|g(\Theta) - \hat{g}_t\|^2] \times \sqrt{E}_\pi[|r(\Theta) - r(\hat{\theta}_t)|^2] \\ &= o(1) \times O(\frac{1}{\sqrt{a}}) = o(\frac{1}{\sqrt{a}}), \end{aligned}$$

as $a \to \infty$ by the first part of the Proposition and Lemma 5.

Acknowledgements

The research of both authors was supported by the National Science Foundation, and that of the second author by a Fulbright Scholarship Grant.

References

Berger, J. O. and Wolpert, R. L. (1984). *The Likelihood Principle.* IMS, Hayward, CA.

Betensky, R. A. (1995). An O'Brien-Fleming sequential test for comparing three treatments. Submitted to *Ann. Statist.*.

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

Coad, D. S. (1995). Sequential allocation rules for multi-armed clinical trials. To appear in *J. Statist. Comput. Simulation* **52**, 239-251.

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

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.

Robbins, H. and Siegmund, D. (1974). Sequential tests involving two populations. *J. Amer. Statist. Assoc.* **69**, 132-139.

Siegmund, D. O. (1980). Sequential χ^2 and F tests and the related confidence intervals. *Biometrika* **67**, 389-402.

Siegmund, D. (1993). A sequential clinical trial for comparing three treatments. *Ann. Statist.* **21**, 464-483.

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

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

Woodroofe, M. (1989). Very weak expansions for sequentially designed experiments: Linear models. *Ann. Statist.* **17**, 1087-1102.

Woodroofe, M. (1992). Integrable expansions for posterior distributions for one-parameter exponential families. *Statist. Sinica* **2**, 91-111.

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

Department of Statistics, University of Michigan, Mason Hall, Ann Arbor, MI 48109, U.S.A.

School of Mathematical Sciences, University of Sussex, Falmer Brighton BN1 9QH, U.K.

(Received May 1995; accepted March 1996)