APPROXIMATE BIAS CALCULATIONS FOR SEQUENTIALLY DESIGNED EXPERIMENTS

D.S. Coad

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

M.B. Woodroofe

Department of Statistics
University of Michigan
Ann Arbor, MI 48109-1027
U.S.A.

Key words and Phrases: adaptive normal linear model; autoregressive model; fundamental identity of sequential analysis; maximum likelihood estimator; unknown variability; very weak expansion.

ABSTRACT

A linear model is considered in which the design variables may be functions of previous responses and/or auxiliary randomisation. The model is observed t successive times, where t is a stopping time, and interest lies in estimating the parameters of the model. Approximations are derived for the bias and variance of the maximum likelihood estimators of the parameters at time t. The derivations involve differentiating the fundamental identity of sequential analysis. The accuracy of the approximations is assessed by simulation for a multi-armed clinical trial model proposed by Coad (1995), two autoregressive models and the sequential design of Ford and Silvey (1980). Very weak expansions are used to justify the approximations.

1. INTRODUCTION

Consider an adaptive linear model of the form

$$y_k = x_k^T\theta + \sigma\epsilon_k, \quad k = 1, 2, \ldots,$$

where $\sigma > 0$ and $\theta \in \mathbb{R}^p$ are unknown parameters and $\epsilon_1, \epsilon_2, \ldots$ are independent standard normal random variables. Adaptive means that each x_k is a measurable function of previous responses and/or auxiliary randomisation, say

$$x_k = x_k(u_1, \ldots, u_k, y_1, \ldots, y_{k-1}) \in \mathbb{R}^p,$$

where $u_1, u_2, \ldots$ are independent of $\epsilon_1, \epsilon_2, \ldots$. The above model has a number of possible applications. These include control problems, as in Lai and Wei (1982), the adaptive designs of Wu (1985) for estimating non-linear functions, and Siegmund's (1993) test for comparing three treatments.

Suppose that the process is observed t successive times, where t is a stopping time with respect to the filtration induced by $(u_k, x_k, y_k), k = 1, 2, \ldots$. Then the model may be written in the form

$$\mathbf{y}_t = X_t\theta + \sigma\epsilon_t, \tag{1}$$

where $\mathbf{y}_t = (y_1, \ldots, y_t)^T$, $X_t = (x_1, \ldots, x_t)^T$ and $\epsilon_t = (\epsilon_1, \ldots, \epsilon_t)^T$. Let $P_{\sigma,\theta}$ denote the probability measure under which (1) holds. Of the stopping time t, it is required that $X_t^T X_t$ is positive definite with probability one. See also (5) below.

It is well known that the use of a sequential design and/or stopping time does not affect the likelihood function. See, for example, Berger and Walpole (1984). So, for model (1), the maximum likelihood estimators of θ and σ^2 are

$$\hat{\theta}_t = (X_t^T X_t)^{-1} X_t^T \mathbf{y}_t \tag{2}$$

and

$$\hat{\sigma}_t^2 = \frac{\|\mathbf{y}_t - X_t\hat{\theta}_t\|^2}{t}, \tag{3}$$

where $\|.\|$ denotes the Euclidean norm of a vector. However, the sampling distributions of $\hat{\theta}_t$ and $\hat{\sigma}_t^2$ may be affected by the sequential design and stopping time. In particular, $\hat{\theta}_t$ may be a biased estimator of θ.

The purpose of this paper is to derive approximations for the bias and variance of $\hat{\theta}_t$ and $\hat{\sigma}_t^2$. The present paper may be regarded as a continuation of Woodroofe and

Coad (1996) in which approximate confidence regions for θ were derived for the case σ known. The derivations of the bias and variance approximations involve differentiating the fundamental identity of sequential analysis. Similar calculations are detailed by Woodroofe (1990) and Coad (1994) in the context of a one-parameter exponential family. Whitehead (1986) describes a method of calculating adjusted estimates with reduced bias. More recently, Todd, Whitehead and Facey (1996) show how bias-adjusted estimates may be evaluated more accurately.

Expressions for the bias and variance of $\hat{\theta}_t$ and $\hat{\sigma}_t^2$ are obtained in Section 2. Approximations to these are given in Section 3. In Section 4, the approximations are specialised to a multi-armed clinical trial model of Coad (1995), two autoregressive models and the sequential design of Ford and Silvey (1980). The results of a simulation study of these models are reported in Section 5, and the Ford-Silvey example is studied further in Section 7. Theoretical justification of the approximations is provided in Section 6 and an appendix. Possible extensions to the work are indicated in Section 8.

2. FORMULAE FOR THE BIAS AND VARIANCE

As noted above, the likelihood function is unaffected by the adaptive design and optional stopping. So the likelihood function is

$$L_t(\sigma, \theta) = \exp\{-\frac{1}{2\sigma^2}\|\mathbf{y}_t - X_t\theta\|^2 + \frac{1}{2}\|\mathbf{y}_t\|^2 - \frac{t}{2}\log(\sigma^2)\} \tag{4}$$

for all $\sigma > 0$ and $\theta \in \mathbb{R}^p$. The gradient and Hessian of L with respect to θ are

$$\nabla L_t(\sigma, \theta) = \frac{1}{\sigma^2}(X_t^T\mathbf{y}_t - X_t^T X_t\theta)L_t(\sigma, \theta)$$

and

$$\nabla^2 L_t(\sigma, \theta) = \{\frac{1}{\sigma^4}(X_t^T\mathbf{y}_t - X_t^T X_t\theta)(X_t^T\mathbf{y}_t - X_t^T X_t\theta)^T - \frac{1}{\sigma^2}(X_t^T X_t)\}L_t(\sigma, \theta),$$

and the derivative with respect to σ^2 is

$$\frac{\partial}{\partial\sigma^2}L_t(\sigma, \theta) = \frac{1}{2}(\frac{1}{\sigma^4}\|\mathbf{y}_t - X_t\theta\|^2 - \frac{t}{\sigma^2})L_t(\sigma, \theta)$$

for all $\sigma > 0$ and $\theta \in \mathbb{R}^p$.

To proceed further, some conditions are needed. Suppose that

$$E_{\sigma,\theta}(X_t^T X_t) \text{ and } E_{\sigma,\theta}\{(X_t^T X_t)^{-2}\} \text{ are finite and continuous in } \theta \text{ for fixed } \sigma. \quad (5)$$

Let $M_t = (X_t^T X_t)^{-1}$ and $M(\sigma, \theta) = E_{\sigma,\theta}(M_t)$, which is finite and continuous in θ for fixed σ by (5). Then, by the fundamental identity of sequential analysis (e.g. Woodroofe, 1982, Chap.1),

$$E_{\sigma,\theta}(M_{tik}) = \int M_{tik} dP_{\sigma,\theta} = \int M_{tik} L_t(\sigma,\theta) dP_{1,0}$$

for $i, k = 1, 2, \ldots, p$. The condition (5) may be used to justify differentiating under the integral sign. This yields

$$\begin{aligned} \sum_{k=1}^{p} \frac{\partial}{\partial \theta_k} E_{\sigma,\theta}(M_{tik}) &= \frac{1}{\sigma^2} \int \sum_{k=1}^{p} M_{tik} (X_t^T \mathbf{y}_t - X_t^T X_t \theta)_k dP_{\sigma,\theta} \\ &= \frac{1}{\sigma^2} E_{\sigma,\theta}(\hat{\theta}_{ti} - \theta_i) \end{aligned}$$

for $i = 1, 2, \ldots, p$. Hence, we obtain

$$E_{\sigma,\theta}(\hat{\theta}_t) = \theta + \sigma^2 M^{\#}(\sigma,\theta)\mathbf{1}, \quad (6)$$

where

$$M_{ik}^{\#}(\sigma,\theta) = \frac{\partial}{\partial \theta_k} M_{ik}(\sigma,\theta) \quad (7)$$

for $i, k = 1, 2, \ldots, p$, and $\mathbf{1}$ is the unit p-vector.

Suppose next that $E_{\sigma,\theta}(\|M_t\|^4)$ is finite and continuous in θ for fixed σ, where $\|.\|$ denotes the trace norm of a matrix. Then, again by the fundamental identity,

$$E_{\sigma,\theta}(M_{tik} M_{tj\ell}) = \int M_{tik} M_{tj\ell} dP_{\sigma,\theta} = \int M_{tik} M_{tj\ell} L_t(\sigma,\theta) dP_{1,0}.$$

Differentiating the integral yields

$$\begin{aligned} \sum_{k=1}^{p} \sum_{\ell=1}^{p} \frac{\partial^2}{\partial \theta_k \partial \theta_\ell} E_{\sigma,\theta}(M_{tik} M_{tj\ell}) &= \frac{1}{\sigma^4} \int \sum_{k=1}^{p} \sum_{\ell=1}^{p} \{ M_{tik} M_{tj\ell} (X_t^T \mathbf{y}_t - X_t^T X_t \theta)_k \\ &\times \ (X_t^T \mathbf{y}_t - X_t^T X_t \theta)_\ell \} dP_{\sigma,\theta} \\ &- \ \frac{1}{\sigma^2} \int \sum_{k=1}^{p} \sum_{\ell=1}^{p} M_{tik} M_{tj\ell} (X_t^T X_t)_{k\ell} dP_{\sigma,\theta} \end{aligned}$$

for $i, j = 1, 2, \ldots, p$, which leads to the relation

$$E_{\sigma,\theta}\{(\hat{\theta}_t - \theta)(\hat{\theta}_t - \theta)^T\} = \sigma^2 M(\sigma, \theta) + \sigma^4 M^{\#\#}(\sigma, \theta), \tag{8}$$

where

$$M^{\#\#}_{i,j}(\sigma, \theta) = \sum_{k=1}^{p} \sum_{\ell=1}^{p} \frac{\partial^2}{\partial\theta_k \partial\theta_\ell} E_{\sigma,\theta}(M_{tik} M_{tj\ell}) \tag{9}$$

for $i, j = 1, 2, \ldots, p$. Notice that the expressions (6) and (8) simplify when $X_t^T X_t$ is a diagonal matrix, which will be the case for Example 1 in Section 4.

We can also obtain a simple expression for the bias of $\hat{\sigma}_t^2$. By the fundamental identity,

$$E_{\sigma,\theta}(\frac{1}{t}) = \int \frac{1}{t} dP_{\sigma,\theta} = \int \frac{1}{t} L_t(\sigma, \theta) dP_{1,0}.$$

If $E_{\sigma,\theta}(t)$ is finite and continuous in σ for fixed θ, then we may differentiate under the integral sign, and

$$\begin{aligned} \frac{\partial}{\partial\sigma^2} E_{\sigma,\theta}(\frac{1}{t}) &= \frac{1}{2\sigma^4} \int \frac{1}{t} (\|\mathbf{y}_t - X_t\theta\|^2 - t\sigma^2) dP_{\sigma,\theta} \\ &= \frac{1}{2\sigma^4} \int \frac{1}{t} (\|\mathbf{y}_t - X_t\hat{\theta}_t\|^2 + \|X_t\hat{\theta}_t - X_t\theta\|^2 - t\sigma^2) dP_{\sigma,\theta}. \end{aligned} \tag{10}$$

Hence, by (3), we have the relation

$$E_{\sigma,\theta}(\hat{\sigma}_t^2) = \sigma^2 + 2\sigma^4 \frac{\partial}{\partial\sigma^2} E_{\sigma,\theta}(\frac{1}{t}) - E_{\sigma,\theta}(\frac{\|X_t\hat{\theta}_t - X_t\theta\|^2}{t}). \tag{11}$$

Proceeding further, it follows easily from (10) that

$$\frac{\partial}{\partial\sigma^2}\{2\sigma^4 \frac{\partial}{\partial\sigma^2} E_{\sigma,\theta}(\frac{1}{t^2})\} = \frac{1}{2\sigma^4} \int \frac{1}{t^2} \{(\|\mathbf{y}_t - X_t\hat{\theta}_t\|^2 + \|X_t\hat{\theta}_t - X_t\theta\|^2 - t\sigma^2)^2 - 2t\sigma^4\} dP_{\sigma,\theta}.$$

Hence, again by (3), we have the relation

$$\begin{aligned} E_{\sigma,\theta}\{(\hat{\sigma}_t^2 - \sigma^2)^2\} &= 2\sigma^4 E_{\sigma,\theta}(\frac{1}{t}) + 2\sigma^4 \frac{\partial}{\partial\sigma^2}\{2\sigma^4 \frac{\partial}{\partial\sigma^2} E_{\sigma,\theta}(\frac{1}{t^2})\} \\ &- E_{\sigma,\theta}(\frac{\|X_t\hat{\theta}_t - X_t\theta\|^4}{t^2}) - 2E_{\sigma,\theta}\{(\hat{\sigma}_t^2 - \sigma^2)\frac{\|X_t\hat{\theta}_t - X_t\theta\|^2}{t}\}. \end{aligned} \tag{12}$$

3. EXPANSIONS

In many problems, asymptotic expansions for biases and variances may be obtained, at least for some values of the parameters. Let Ω denote a convex open set of $\mathbb{R}^p$ or, more generally, a set of the form (23) below.

For the asymptotic expansions, the stopping time t is supposed to depend on a design parameter $a \geq 1$ in such a manner that

$$\lim_{a \to \infty} \frac{1}{a}(X_t^T X_t) = \mathcal{I}(\sigma, \theta) \tag{13}$$

in $P_{\sigma,\theta}$-probability for almost every σ and $\theta \in \Omega$ under Lebesgue measure. Asymptotic expansions are sought for the bias and matrix of second moments

$$b_a(\sigma, \theta) = E_{\sigma,\theta}(\hat{\theta}_t - \theta)$$

and

$$\Sigma_a(\sigma, \theta) = E_{\sigma,\theta}\{(\hat{\theta}_t - \theta)(\hat{\theta}_t - \theta)^T\}.$$

Suppose for the moment that $\mathcal{I}(\sigma, \theta)$ is positive definite for all $\sigma > 0$ and $\theta \in \Omega$, and let $\eta(\sigma, \theta) = \mathcal{I}(\sigma, \theta)^{-1}$. Further, let $\eta_a(\sigma, \theta) = E_{\sigma,\theta}\{a(X_t^T X_t)^{-1}\}$. Then (6) may be written as $b_a(\sigma, \theta) = \sigma^2 \eta_a^{\#}(\sigma, \theta)\mathbf{1}/a$, where $\eta_a^{\#}(\sigma, \theta)$ is obtained from $\eta_a(\sigma, \theta)$ as in (7), which suggests the relation

$$b_a(\sigma, \theta) \simeq \frac{\sigma^2}{a}\eta^{\#}(\sigma, \theta)\mathbf{1}. \tag{14}$$

Similarly, (8) suggests the relation

$$\Sigma(\sigma, \theta) \simeq \frac{\sigma^2}{a}\eta_a(\sigma, \theta) + \frac{\sigma^4}{a^2}\eta^{\#\#}(\sigma, \theta), \tag{15}$$

where $\eta^{\#\#}(\sigma, \theta)$ is obtained from $\eta(\sigma, \theta)$ as in (9). Of course, (15) would be nicer if η_a could be replaced by η, or by $\eta + \eta_1/a$, where η_1 is easily computable. This is possible in specific cases. However, we do not know how to do this in general without imposing quite strong conditions on t and the sequence of designs.

Asymptotic expansions are also sought for the bias and mean square error of $\hat{\sigma}_t^2$,

$$\mu_a(\sigma^2, \theta) = E_{\sigma,\theta}(\hat{\sigma}_t^2 - \sigma^2)$$

and

$$v_a(\sigma^2, \theta) = E_{\sigma,\theta}\{(\hat{\sigma}_t^2 - \sigma^2)^2\}.$$

Suppose that

$$\lim_{a\to\infty} \frac{a}{t} = \kappa(\sigma^2, \theta) \tag{16}$$

in $P_{\sigma,\theta}$-probability for almost every σ and $\theta \in \Omega$ under Lebesgue measure. Further, let $\kappa_a(\sigma^2, \theta) = E_{\sigma,\theta}(a/t)$. Then (11) suggests the relation

$$\mu_a(\sigma^2, \theta) \simeq \frac{2\sigma^4}{a}\kappa'(\sigma^2, \theta) - \frac{p\sigma^2}{a}\kappa(\sigma^2, \theta), \tag{17}$$

where $'$ denotes differentiation with respect to σ^2. Of course, the usual estimator of σ^2 for a fixed-sample design is $\tilde{\sigma}_t^2 = t\hat{\sigma}_t^2/(t-p)$. It follows from (16) and relation (17) that

$$E_{\sigma,\theta}(\tilde{\sigma}_t^2) \simeq \sigma^2 + \frac{2\sigma^4}{a}\kappa'(\sigma^2, \theta).$$

In a similar manner to (17), (12) suggests the relation

$$\begin{aligned} v(\sigma^2, \theta) \quad \simeq \quad & \frac{2\sigma^4}{a}\kappa_a(\sigma^2, \theta) + \frac{2\sigma^4}{a^2}\kappa(\sigma^2, \theta)^2\{4\sigma^4\frac{\kappa''}{\kappa}(\sigma^2, \theta) \\ & +4\sigma^4\frac{\kappa'}{\kappa}(\sigma^2, \theta)^2 - 4(p-2)\sigma^2\frac{\kappa'}{\kappa}(\sigma^2, \theta) + \frac{1}{2}p(p-2).\} \end{aligned} \tag{18}$$

Conditions under which (14), (15), (17) and (18) are valid and proofs are given in Section 6.

4. EXAMPLES

Example 1. Multi-armed clinical trial model

Suppose that we wish to compare three treatments which produce normally distributed responses with unknown means θ_1, θ_2 and θ_3 and a common unknown variance σ^2. Let y_{ij} denote the ith observation on the jth treatment for $i = 1, 2, \ldots$ and $j = 1, 2, 3$. Then y_{ij} may be written in the form $\theta^T x + \sigma\epsilon$, where $\theta = (\theta_1, \theta_2, \theta_3)^T$, x is chosen from $(1,0,0)^T$, $(0,1,0)^T$, or $(0,0,1)^T$, and ϵ has a standard normal distribution. Thus, model (1) holds with $p = 3$.

Coad (1995) proposed several sequential allocation rules for the above problem. We consider one of these here, the triangular test. This test depends on two design parameters, $a \geq 1$ and $d > 0$. Let

$$\nu_n = 1 + \frac{b_1}{\sqrt{n}} + \frac{b_2}{n},$$

where $b_1, b_2 \geq 0$. Then, if σ is estimated, triples $(y_{k1}, y_{k2}, y_{k3})^T$, $k = 1, 2, \ldots$, are observed until time

$$s = \inf\{n : n \geq 2 \text{ and } \frac{\max S_n^{(i)} - \min S_n^{(j)}}{\tilde{\sigma}_n} \geq a\nu_n - dn\},$$

where i and j range over the three treatments, $S_n^{(i)} = \sum y_{ki}$ for $i = 1, 2, 3$ and $\tilde{\sigma}_n^2$ is the usual pooled estimate of σ^2. The quantity ν_n is an adjustment to the stopping boundary to allow for the extra variability due to the estimation of σ^2. Denote the ordered θ's by $\theta_{(1)} \leq \theta_{(2)} \leq \theta_{(3)}$. Then

$$\lim_{a\to\infty} \frac{a}{s} = \frac{\theta_{(3)} - \theta_{(1)}}{\sigma} + d = d_1(\sigma, \theta),$$

say, in $P_{\sigma,\theta}$-probability for all $\sigma > 0$ and $\theta \in \mathbb{R}^3$.

Let the data-dependent indices $I = I^a$, $J = J^a$ and $K = K^a$ be determined with probability one by $\hat{\theta}_{s,I} < \hat{\theta}_{s,J} < \hat{\theta}_{s,K}$, where $\hat{\theta}_s$ denotes the maximum likelihood estimator of θ at time s. After time s, treatment I is eliminated and observations are taken on treatments J and K until time

$$t = \inf\{n : n \geq s \text{ and } \frac{|S_n^{(K)} - S_n^{(J)}|}{\tilde{\sigma}_n} \geq a\nu_n - dn\}.$$

Sampling is terminated at time t, and

$$\lim_{a\to\infty} \frac{a}{t} = \frac{\theta_{(3)} - \theta_{(2)}}{\sigma} + d = d_2(\sigma, \theta),$$

say, in $P_{\sigma,\theta}$-probability for all $\sigma > 0$ and $\theta \in \mathbb{R}^3$. Now, it is clear that $X_t^T X_t$ is a diagonal matrix. Define $i = i_{\sigma,\theta}$, $j = j_{\sigma,\theta}$ and $k = k_{\sigma,\theta}$ 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

$$\eta(\sigma, \theta) = \lim_{a\to\infty} a(X_t^T X_t)^{-1} = \text{diag}\{\eta_1(\sigma, \theta), \eta_2(\sigma, \theta), \eta_3(\sigma, \theta)\}$$

in $P_{\sigma,\theta}$-probability for almost every $\sigma > 0$ and $\theta \in \mathbb{R}^3$, where $\eta_i = d_1$ and $\eta_j = \eta_k = d_2$.

The set Ω in this example consists of θ for which $\theta_{(1)} < \theta_{(2)} < \theta_{(3)}$. This is the union of open convex sets as in (23). However, to determine the approximations given by (14) and (15), it is only necessary to consider separate cases, according to the ordering of the treatment means at time s. For each case, different expressions are obtained for the $\eta^{\#}$ and $\eta^{\#\#}$ matrices.

Now consider the bias and variance of $\hat{\sigma}_t^2$. First note that $\hat{\sigma}_t^2$ is based on $s + 2t$ observations. Thus,

$$\kappa(\sigma^2, \theta) = \lim_{a\to\infty} \frac{a}{s+2t} = \frac{d_1(\sigma,\theta)d_2(\sigma,\theta)}{d_2(\sigma,\theta) + 2d_1(\sigma,\theta)}$$

in $P_{\sigma,\theta}$-probability for all $\sigma > 0$ and $\theta \in \mathbb{R}^3$. So an approximation to the bias of $\hat{\sigma}_t^2$ may be determined from (17). Further, by using (18), it follows easily that

$$\begin{aligned}\mathrm{var}_{\sigma,\theta}(\hat{\sigma}_t^2) \simeq \frac{2\sigma^4}{a}\kappa_a(\sigma^2,\theta) &+ \frac{2\sigma^4}{a^2}\kappa(\sigma^2,\theta)^2\{4\sigma^4\frac{\kappa''}{\kappa}(\sigma^2,\theta) \\ &+ 2\sigma^4\frac{\kappa'}{\kappa}(\sigma^2,\theta)^2 + 2\sigma^2\frac{\kappa'}{\kappa}(\sigma^2,\theta) - 3\}.\end{aligned}$$

Example 2. First-order autoregression

Consider the autoregressive model

$$y_k = \theta y_{k-1} + \sigma\epsilon_k, \qquad k = 1, 2, \ldots,$$

where $y_0 = 0$, $\sigma > 0$ and $|\theta| < 1$. This is of the form (1) with $p = 1$, $x_1 = 0$ and $x_k = y_{k-1}$ for $k = 2, 3, \ldots$. The maximum likelihood estimator of θ is

$$\hat{\theta}_n = \frac{\sum y_k y_{k-1}}{\sum y_{k-1}^2}.$$

Further,

$$\eta(\sigma,\theta) = \lim_{n\to\infty} n(X_n^T X_n)^{-1} = \frac{1-\theta^2}{\sigma^2}$$

in $P_{\sigma,\theta}$-probability for all $\sigma > 0$ and $|\theta| < 1$. So, from (14), we have that

$$E_{\sigma,\theta}(\hat{\theta}_n) \simeq \theta - \frac{2\theta}{n}$$

and, from (15), a straightforward calculation leads to

$$\mathrm{var}_{\sigma,\theta}(\hat{\theta}_n) \simeq \frac{1-\theta^2}{n} + \frac{10\theta^2 - 1}{n^2}.$$

Example 3. Second-order autoregression

Consider the autoregressive model

$$y_k = \theta_1 y_{k-1} + \theta_2 y_{k-2} + \sigma\epsilon_k, \qquad k = 1, 2, \ldots,$$

where $y_{-1} = y_0 = 0$, $\sigma > 0$ and $\theta = (\theta_1, \theta_2)^T \in \Omega$, a triangular region defined by $\theta_1 + \theta_2 < 1$, $\theta_1 - \theta_2 > -1$ and $|\theta_2| < 1$. See, for example, Brockwell and Davis (1991, Chap.8). Then model (1) holds with $p = 2$, $x_1 = 0$, $x_2 = (y_1, 0)^T$ and $x_k = (y_{k-1}, y_{k-2})^T$ for $k = 3, 4, \ldots$, so that

$$X_n^T X_n = \begin{pmatrix} \sum_{k=2}^n y_{k-1}^2 & \sum_{k=3}^n y_{k-1}y_{k-2} \\ \sum_{k=3}^n y_{k-1}y_{k-2} & \sum_{k=3}^n y_{k-2}^2 \end{pmatrix}.$$

Thus, it follows that

$$\eta(\sigma, \theta) = \lim_{n\to\infty} n(X_n^T X_n)^{-1} = \frac{1}{\sigma^2}\begin{pmatrix} 1-\theta_2^2 & -\theta_1(1+\theta_2) \\ -\theta_1(1+\theta_2) & 1-\theta_2^2 \end{pmatrix}$$

in $P_{\sigma,\theta}$-probability for all $\sigma > 0$ and $\theta \in \Omega$. So, from (14), we have that

$$E_{\sigma,\theta}(\hat{\theta}_n) \simeq \theta - \frac{1}{n}\begin{pmatrix} \theta_1 \\ 3\theta_2 + 1 \end{pmatrix}.$$

By using (15), straightforward calculations also show that

$$\begin{aligned} \mathrm{cov}_{\sigma,\theta}(\hat{\theta}_n) \;\; &\simeq \;\; \frac{1}{n}\begin{pmatrix} 1-\theta_2^2 & -\theta_1(1+\theta_2) \\ -\theta_1(1+\theta_2) & 1-\theta_2^2 \end{pmatrix} \\ &\quad + \frac{1}{n^2}\begin{pmatrix} \theta_1^2 - 2(1-3\theta_2)(1+\theta_2) & \theta_1(5+7\theta_2) \\ \theta_1(5+7\theta_2) & 11\theta_2^2 + 2\theta_2 - 5 \end{pmatrix}. \end{aligned}$$

Example 4. Ford-Silvey example

Ford and Silvey (1980) considered the model

$$y_k = \theta_1 x_k + \theta_2 x_k^2 + \epsilon_k, \qquad k = 1, 2, \ldots,$$

where $|x_k| \leq 1$ for all k and $\theta \in \mathbb{R}^2$. Interest lay in estimating the non-linear function $g(\theta) = -\theta_1/(2\theta_2)$, the value at which the regression function attains its maximum. They used the following sequential design. First take an observation each at $x = \pm 1$. For any $n > 2$, let y_n^+ denote the sum of all observations at $x = +1$ and let y_n^- denote the corresponding sum at $x = -1$. Then choose $x_{n+1} = +1$ or -1 according as $|y_n^+|$ is less than or greater than $|y_n^-|$. Clearly, the above model is of the form (1) with $p = 2$ and $\mathbf{x}_k = (x_k, x_k^2)^T$ for $k = 1, 2, \ldots$, so that

$$X_n^T X_n = \begin{pmatrix} \sum x_k^2 & \sum x_k^3 \\ \sum x_k^3 & \sum x_k^4 \end{pmatrix} = \begin{pmatrix} n & n - 2s_n \\ n - 2s_n & n \end{pmatrix},$$

where s_n is the number of observations taken at $x = -1$. Ford and Silvey (1980) showed that

$$\lim_{n\to\infty} \frac{1}{n}(X_n^T X_n) = \begin{pmatrix} 1 & k(\theta) \\ k(\theta) & 1 \end{pmatrix}$$

in P_θ-probability for all $\theta \in \mathbb{R}^2$, where $k(\theta) = 2g$ or $(2g)^{-1}$ according as $|g| \leq \frac{1}{2}$ or $|g| \geq \frac{1}{2}$. Note that the limiting matrix above is singular if $|g| = \frac{1}{2}$, or, equivalently, $|\theta_1| = |\theta_2|$.

The set Ω consists of θ for which $|\theta_1| \neq |\theta_2|$ here, and to determine approximations to the bias and covariance matrix of $\hat{\theta}_n$, we need to consider two separate cases, $|g| < \frac{1}{2}$ and $|g| > \frac{1}{2}$. To simplify notation, let $h(\theta) = \max \theta_i^2 - \min \theta_j^2$. Then a simple calculation shows that (14) yields

$$E_\theta(\hat{\theta}_n) \simeq \theta + \frac{1}{nh(\theta)} \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}.$$

By using (15), a lengthy but straightforward calculation leads to the approximation

$$\mathrm{cov}_\theta(\hat{\theta}_n) \simeq \frac{1}{nh(\theta)} \begin{pmatrix} \max \theta_i^2 & \theta_1\theta_2 \\ \theta_1\theta_2 & \max \theta_i^2 \end{pmatrix} + \frac{1}{n^2 h(\theta)^2} \begin{pmatrix} m_1(\theta) & \theta_1\theta_2 \\ \theta_1\theta_2 & m_2(\theta) \end{pmatrix},$$

where $m_1(\theta) = \theta_1^2 + 2\theta_2^2$ or $-\theta_1^2$ and $m_2(\theta) = -\theta_2^2$ or $2\theta_1^2 + \theta_2^2$ according as $|g| < \frac{1}{2}$ or $|g| > \frac{1}{2}$.

5. SIMULATION RESULTS

5.1. General

In order to assess the accuracy of the approximations presented in Section 4, a simulation study based on 10,000 replications was conducted, for selected values of the design parameters. In each case, the value of σ was 1.0. The results are reported separately for the multi-armed clinical trial model, the autoregressive models and the Ford-Silvey example.

5.2. Multi-armed clinical trial model

We first consider the sequential procedure in Example 1 in Section 4. Monte Carlo results are reported in detail for two sets of design parameters, $a = 15.0$ and $d = 0.5$, and $a = 30.0$ and $d = 0.5$. These were chosen to agree with Coad (1995) and correspond to truncation at $n = 30$ and $n = 60$, respectively. For adjustment ν_n, we have taken $(b_1, b_2) = (0, 0)$ in all cases.

Table I gives approximate and Monte Carlo values for the bias and covariance matrix of $\hat{\theta}_t$. By comparing columns 4 and 5, we see that the approximation to the bias is quite good, especially when there is a strict ordering in the treatment means or when there is a single inferior treatment. Columns 6 to 11 indicate that the approximation to the covariance matrix of $\hat{\theta}_t$ is most accurate when there is a single better treatment. The Monte Carlo values are slightly underestimated in the other cases.

TABLE I ABOUT HERE

Approximate and Monte Carlo values for the bias and variance of $\hat{\sigma}_t^2$ are presented in Table II. The fourth and fifth columns show that the approximation to the bias is very accurate for the first two configurations of means, and tends to underestimate the true value in the other cases. The values for the variance of $\hat{\sigma}_t^2$ in columns 6 and 7 exhibit a similar behaviour.

TABLE II ABOUT HERE

5.3. Autoregressive models

We now consider Examples 2 and 3. In both cases, Monte Carlo results are reported in detail when $n = 25$ and $n = 50$. Table III gives approximate and Monte Carlo values for the bias and variance of $\hat{\theta}_n$ for Example 2. In almost all cases, the approximations are remarkably accurate. Indeed, when $n = 50$, the discrepancy is at most 0.002.

TABLE III ABOUT HERE

Approximate and Monte Carlo values for the bias and covariance matrix of $\hat{\theta}_n$ for Example 3 are presented in Table IV. It is clear that the approximations are not quite as accurate as those in Table 3, especially when θ is near the boundary of the set Ω. In particular, the approximations tend to slightly underestimate the true values.

TABLE IV ABOUT HERE

5.4. Ford-Silvey example

Monte Carlo results for Example 4 are reported in detail when $n = 25$ and $n = 50$. These are summarised in Table V. Note that we have taken $|\theta_1| < |\theta_2|$, and that the choice of values for θ follows Ford and Silvey (1980). Similar results have been obtained when $|\theta_1| > |\theta_2|$.

TABLE V ABOUT HERE

A comparison of columns 3 and 4 in Table V indicates that the approximation to the bias of $\hat{\theta}_n$ tends to slightly overestimate the true value. Comparing columns 5 and 6 with columns 7 and 8, the accuracy of the approximation to the covariance matrix of $\hat{\theta}_n$ is impressive, especially when $n = 25$. The discrepancy is no more than 0.002.

6. VERY WEAK JUSTIFICATION

If C is a $q \times p$ matrix of constants, then clearly $Cb_a(\sigma, \theta) = E_{\sigma,\theta}\{C(\hat{\theta}_t - \theta)\}$. In a more general formulation, the matrix C may be allowed to depend on a and the data so that $C_a = C_a(y_1, y_2, \ldots, y_t)$. Thus, let

$$c_a(\sigma, \theta) = aE_{\sigma,\theta}\{C_a(\hat{\theta}_t - \theta)\}, \tag{19}$$

assumed to exist for all $\sigma > 0$ and $\theta \in \Omega$. Of course, (19) includes the normalised bias $ab_a(\sigma, \theta)$ as a special case with $C_a = I_p$. The more general formulation may be useful if there is interest in estimating a non-linear function, say $g(\theta)$. Then a Taylor series approximation includes a term of the form (19) with $C_a = \nabla g(\hat{\theta}_t)$.

The first step is to find a limiting function $c(\sigma, \theta)$ for which $c_a \to c$ in the very weak sense of Woodroofe (1989); that is,

$$\lim_{a\to\infty} \int_\Omega \{c_a(\sigma, \theta) - c(\sigma, \theta)\}\xi(\theta)d\theta = 0 \tag{20}$$

for all continuously differentiable, compactly supported densities ξ on Ω. Relation (20) holds under fairly modest assumptions. For a fixed family $C_a = C_a(y_1, y_2, \ldots, y_t)$, $a \geq 1$, of matrices, suppose that there are matrices $N(\sigma, \theta)$, $\sigma > 0$, $\theta \in \Omega$, for which

$$\int_K \|N(\sigma, \theta)\|^\alpha d\theta < \infty \tag{21}$$

and

$$\lim_{a\to\infty} \int_K \|aC_a(X_t^T X_t)^{-1} - N(\sigma, \theta)\|^\alpha d\theta = 0 \tag{22}$$

for all compact $K \subset \Omega$ and appropriate $\alpha > 0$. If $C_a = I_p$ for all a, then (21) imposes a strong condition on $\mathcal{I}(\sigma, \theta)$ in (13). In other cases, (21) may hold even if $\mathcal{I}(\sigma, \theta)$ is singular for some σ and θ.

In Theorems 1, 2 and 3 below, $\Omega \subseteq \mathbb{R}^p$ denotes a countable union of convex open sets, say

$$\Omega = \bigcup_{i=1}^{\infty} \Omega_i, \tag{23}$$

where $\Omega_1, \Omega_2, \ldots$ are convex open sets, and approximations are asserted to hold only on Ω.

Theorem 1. Suppose that (21), (22) and (23) hold with $\alpha = 1$ and that the entries $n_{ij}(\sigma, \theta)$ of $N(\sigma, \theta)$ are continuously differentiable on Ω. Let $N^\#(\sigma, \theta)$ be obtained from $N(\sigma,\theta)$ as in (7) and suppose that $\|N^{\#}(\sigma,\theta)\|$ is locally integrable on Ω. Then (20) holds for all continuously differentiable, compactly supported densities ξ on Ω with

$$c(\sigma,\theta) = N^{\#}(\sigma,\theta)\mathbf{1},$$

where $\mathbf{1}$ is the unit p-vector.

Proof. Since σ is fixed, there is no loss of generality in taking $\sigma = 1$. Further, $c(\theta)$ and $N(\theta)$ are written for $c(1,\theta)$ and $N(1,\theta)$ throughout the proof. Thus, for a given ξ, consider a Bayesian model in which θ is replaced by a random vector Θ and model (1) holds conditionally given $\Theta = \theta$ for every $\theta \in \Omega$. Expectation in the Bayesian model is denoted by E_ξ. Then the integral on the left-hand side of (20) is

$$\int_\Omega \{c_a(\theta) - c(\theta)\}\xi(\theta)d\theta = E_\xi\{aC_a(\hat{\theta}_t - \Theta) - c(\Theta)\}. \tag{24}$$

Let E_ξ^t denote conditional expectation given $x_1,\ldots,y_t$. Then the expectation on the right-hand side of (24) may be simplified by using the two relations

$$E_\xi^t\{aC_a(\Theta - \hat{\theta}_t)\} = aC_a(X_t^T X_t)^{-1}E_\xi^t\{\frac{\nabla\xi}{\xi}(\Theta)\} \tag{25}$$

and

$$\int_\Omega N^{\#}(\theta)\mathbf{1}\xi(\theta)d\theta = -\int_\Omega N(\theta)\nabla\xi(\theta)d\theta = -E_\xi\{N(\Theta)\frac{\nabla\xi}{\xi}(\Theta)\}. \tag{26}$$

The first of these relations is a direct consequence of Lemma 3 of Woodroofe (1989); the second follows from a straightforward integration by parts, as in Corollary 1 of Woodroofe (1989). Combining (24), (25) and (26), we have that

$$\begin{aligned} |\int_\Omega \{c_a(\theta) - c(\theta)\}\xi(\theta)d\theta| &= |E_\xi[\{aC_a(X_t^T X_t)^{-1} - N(\Theta)\}\frac{\nabla\xi}{\xi}(\Theta)]| \\ &\le \int_\Omega E_\theta\{\|aC_a(X_t^T X_t)^{-1} - N(\theta)\|\|\nabla\xi(\theta)\|\}d\theta \\ &\to 0 \end{aligned} \tag{27}$$

as $a \to \infty$, by (22). $\square$

There is an analogous result for the second moments. Suppose that the entries in $N(\sigma, \theta)$ are twice differentiable on Ω and let

$$\Sigma_a(\sigma, \theta) = E_{\sigma,\theta}\{C_a(\hat{\theta}_t - \theta)(\hat{\theta}_t - \theta)^T C_a^T\}.$$

Further, let $N_a(\sigma, \theta) = E_{\sigma,\theta}\{aC_a(X_t^T X_t)^{-1} C_a^T\}$ and let $N^{\#\#}(\sigma, \theta)$ be obtained from $N(\sigma, \theta)$ as in (9).

Theorem 2. Suppose that (21), (22) and (23) hold with $\alpha = 2$ and that the entries of $N(\sigma, \theta)$ are twice continuously differentiable. Then

$$\Sigma_a(\sigma, \theta) = \frac{\sigma^2}{a} N_a(\sigma, \theta) + \frac{\sigma^4}{a^2} N^{\#\#}(\sigma, \theta) + o(\frac{1}{a^2})$$

as $a \to \infty$ in the very weak sense; that is,

$$\int_\Omega [\Sigma_a(\sigma, \theta) - \{\frac{\sigma^2}{a} N_a(\sigma, \theta) + \frac{\sigma^4}{a^2} N^{\#\#}(\sigma, \theta)\}]\xi(\theta) d\theta = o(\frac{1}{a^2}) \tag{28}$$

as $a \to \infty$ for all twice continuously differentiable, compactly supported densities ξ on Ω.

Proof. As in the proof of Theorem 1, we take $\sigma = 1$. For a given ξ, the integral on the left-hand side of (27) is

$$E_\xi\{C_a(\hat{\theta}_t - \Theta)(\hat{\theta}_t - \Theta)^T C_a^T - C_a M_t C_a^T - \frac{1}{a^2} N^{\#\#}(\Theta)\}.$$

But, as a direct consequence of (9) in Woodroofe and Coad (1996), we have that

$$E_\xi^t\{C_a(\hat{\theta}_t - \Theta)(\hat{\theta}_t - \Theta)^T C_a^T - C_a M_t C_a^T\} = E_\xi^t\{C_a M_t \frac{\nabla^2 \xi}{\xi}(\Theta) M_t^T C_a^T\}.$$

Also, a straightforward integration by parts, as in (15) of Woodroofe and Coad (1996), yields

$$E_\xi\{N^{\#\#}(\Theta)\} = E_\xi\{N(\Theta)\frac{\nabla^2 \xi}{\xi}(\Theta) N(\Theta)^T\}.$$

The result (28) then follows as in (27). □

For the bias of $\hat{\sigma}_t^2$, suppose that a/t has a limit $\kappa(\sigma^2,\theta)$ for which

$$\lim_{a\to\infty}\int_K E_{\sigma,\theta}\{|\frac{a}{t}-\kappa(\sigma^2,\theta)|\}d\theta d\sigma^2 = 0 \tag{29}$$

for all compact subsets $K \subseteq (0,\infty)\times\Omega$. Sufficient conditions for (29) are that (16) holds, and that for every compact $K\subset(0,\infty)\times\Omega$ there exists an $\epsilon>0$ for which $P_{\sigma,\theta}(t\le\epsilon a)=o(1/a)$ uniformly on K, as $a\to\infty$. Suppose also that κ is absolutely continuous in σ^2 for fixed θ, let

$$\begin{aligned}\tilde{c}_a(\sigma,\theta) &= aE_{\sigma,\theta}(\hat{\sigma}_t^2-\sigma^2)\\ &= aE_{\sigma,\theta}(\frac{\|\mathbf{y}_t-X_t\theta\|^2}{t}-\sigma^2)-aE_{\sigma,\theta}(\frac{\|X_t\hat{\theta}_t-X_t\theta\|^2}{t})\\ &= \tilde{c}_{a,1}(\sigma,\theta)-\tilde{c}_{a,2}(\sigma,\theta),\end{aligned}$$

say, and let

$$\tilde{c}(\sigma,\theta)=2\sigma^4\kappa'(\sigma^2,\theta)-p\sigma^2\kappa(\sigma^2,\theta)$$

for $a,\sigma>0$ and $\theta\in\Omega$. It is shown that $\tilde{c}_a(\sigma,\theta)$ converges to $\tilde{c}(\sigma,\theta)$ in the very weak sense.

Theorem 3. Suppose that (21), (22), (23) and (29) hold, that κ is absolutely continuous in σ^2 for fixed θ, and that $\kappa'(\sigma^2,\theta)$ is locally integrable over $(\sigma^2,\theta)\in(0,\infty)\times\Omega$. Then

$$\lim_{a\to\infty}\int_0^\infty\int_\Omega\{\tilde{c}_a(\sigma,\theta)-\tilde{c}(\sigma,\theta)\}\xi(\sigma^2,\theta)d\theta d\sigma^2=0$$

for all continuously differentiable densities ξ with compact support in $(0,\infty)\times\Omega$.

Proof. As above, consider a Bayesian model in which there are random variables W and Θ with joint density ξ and model (1) holds conditionally given $W=\sigma^2$ and $\Theta=\theta$ for every $\sigma>0$ and $\theta\in\Omega$. Probability and expectation in the Bayesian model are denoted by P_ξ and E_ξ, and conditional expectation given the data by E_ξ^t. The two terms $\tilde{c}_{a,1}(\sigma,\theta)$ and $\tilde{c}_{a,2}(\sigma,\theta)$ are considered separately. For the first, let

$\zeta(\sigma^2,\theta)=\sigma^4\xi(\sigma^2,\theta)$. Then a simple integration by parts shows that

$$\begin{aligned}\int_0^\infty\int_\Omega \tilde{c}_{a,1}(\sigma,\theta)\xi(\sigma^2,\theta)d\theta d\sigma^2 &= aE_\xi(\frac{\|\mathbf{y}_t-X_t\Theta\|^2}{t}-W)\\ &= 2E_\xi\{\frac{a}{t}W^2L_t'(W,\Theta)\}=-2E_\xi\{\frac{a}{t}\frac{\zeta'}{\xi}(W,\Theta)\}.\end{aligned}$$

Using (29) and another integration by parts, it is easily seen that the last line converges to

$$\begin{aligned}-2E_\xi\{\kappa(W,\Theta)\frac{\zeta'}{\xi}(W,\Theta)\} &= -2\int_0^\infty\int_\Omega \kappa(\sigma^2,\theta)\zeta'(\sigma^2,\theta)d\theta d\sigma^2\\ &= 2\int_0^\infty\int_\Omega \sigma^4\kappa'(\sigma^2,\theta)\xi(\sigma^2,\theta)d\theta d\sigma^2,\end{aligned}$$

as $a\to\infty$. So,

$$\lim_{a\to\infty}\int_0^\infty\int_\Omega\{\tilde{c}_{a,1}(\sigma,\theta)-2\sigma^4\kappa'(\sigma^2,\theta)\}\xi(\sigma^2,\theta)d\theta d\sigma^2=0.$$

Next,

$$\begin{aligned}\int_0^\infty\int_\Omega\{\tilde{c}_{a,2}(\sigma,\theta)-p\sigma^2\kappa(\sigma^2,\theta)\}\xi(\sigma^2,\theta)d\theta d\sigma^2 &= E_\xi\{\frac{a}{t}(\|X_t\hat{\theta}_t-X_t\Theta\|^2-pW)\}\\ &+ pE_\xi[W\{\frac{a}{t}-\kappa(W,\Theta)\}].\end{aligned}$$

The second term on the right-hand side approaches zero as $a\to\infty$ by (29), and the first approaches zero by a simple application of Lemma 4 of Woodroofe (1989). □

7. MORE ON THE FORD-SILVEY EXAMPLE

For completeness, verification of the conditions (21), (22), (23) and (29) should be included for each of the four examples. The sets (23) and matrices N were identified in the examples, and (21) is clear. That leaves (22) and (29), which is an issue only in Example 1. Quite similar verifications have been described by Woodroofe and Coad (1996) in the context of Example 1 and will not be repeated here. The verification of (22) for Examples 2 and 3 is straightforward and this too will be omitted. Verification of (22) for Example 4, the Ford-Silvey example, is more interesting and is described next.

Recall that, in the Ford-Silvey example,

$$y_k = \theta_1 x_k + \theta_2 x_k^2 + \epsilon_k, \quad k = 1, 2, \ldots,$$

where θ_1 and θ_2 are unknown parameters, $x_k = \pm 1$, $k \geq 1$, are design variables, and $\epsilon_1, \epsilon_2, \ldots$ are independent standard normal random variables. The sampling design is as follows: to begin $x_1 = -1$ and $x_2 = +1$; thereafter, $x_{k+1} = +1$ if $|y_k^+| < |y_k^-|$ and $x_{k+1} = -1$ otherwise, where

$$y_n^{\pm} = \sum_{k=1}^{n} (\frac{1 \pm x_k}{2}) y_k.$$

Let

$$\Delta_n^{\pm} = \sum_{k=1}^{n} (\frac{1 \pm x_k}{2}) \quad \text{and} \quad R_n^{\pm} = \sum_{k=1}^{n} (\frac{1 \pm x_k}{2}) \epsilon_k.$$

Then

$$y_n^{\pm} = (\theta_2 \pm \theta_1) \Delta_n^{\pm} + R_n^{\pm}.$$

The verification of (22) depends on the following lemma. Since σ is taken to be one throughout, P_θ is written for $P_{\sigma,\theta}$.

Lemma 1. For all $\alpha, b > 0$ and all $\theta = (\theta_1, \theta_2)^T \in \mathbb{R}^2$,

$$P_\theta(|R_k^{\pm}| \geq \frac{1}{2} \alpha \Delta_k^{\pm} + b, \exists k \geq 1) \leq 2e^{-\alpha b}.$$

Proof. It is easily seen that for any $\alpha \in \mathbb{R}$,

$$M_k^{\pm} = \exp(\alpha R_k^{\pm} - \frac{1}{2} \alpha^2 \Delta_k^{\pm}), \quad k \geq 1,$$

is a martingale for which $E_\theta(M_n^{\pm}) = 1$ for all $n = 1, 2, \ldots$. So,

$$\begin{aligned} P_\theta(R_k^{\pm} \geq \frac{1}{2} \alpha \Delta_k^{\pm} + b, \exists k \leq n) &= P_\theta(M_k^{\pm} \geq e^{\alpha b}, \exists k \leq n) \\ &\leq e^{-\alpha b} E_\theta(M_n^{\pm}) = e^{-\alpha b} \end{aligned}$$

for all $n = 1, 2, \ldots$. The lemma follows easily. □

Recall that the set Ω in (23) consists of (θ_1, θ_2) for which $|\theta_1 + \theta_2| > 0$ and $|\theta_1 - \theta_2| > 0$. For definiteness, consider the subset Ω^{++}, say, where $\nu^+ := \theta_1 + \theta_2 > 0$ and $\nu^- := \theta_2 - \theta_1 > 0$. Further, let $\alpha = \min(\nu^+, \nu^-)$ and let B be the event

$$B = \{|R_k^+| \le \frac{1}{2}\alpha\Delta_k^+ + b, \forall k \ge 1, \text{ and } |R_k^-| \le \frac{1}{2}\alpha\Delta_k^- + b, \forall k \ge 1\}. \tag{30}$$

Then $P_\theta(B^c) \le 4e^{-\alpha b}$ for all $\theta \in \mathbb{R}^2$, by Lemma 1. Observe that, if B occurs, then $|y_k^\pm - \nu^\pm \Delta_k^\pm| \le \frac{1}{2}\alpha\Delta_k^\pm + b$ for all $k \ge 1$. So, if B occurs and

$$\Delta_k^+ < \frac{\nu^- k - 4b}{\nu^+ + \nu^-},$$

then

$$\begin{aligned} |y_k^+| - |y_k^-| &\le (\nu^+ + \frac{1}{2}\alpha)\Delta_k^+ + b - \{(\nu^- - \frac{1}{2}\alpha)\Delta_k^- - b\} \\ &\le (\nu^+ + \nu^-)\Delta_k^+ + 2b - \frac{1}{2}\nu^- k < 0 \end{aligned}$$

and, therefore, $x_{k+1} = 1$. It follows easily from this observation, Lemma 2 below, and symmetry that

$$\Delta_k^\pm \ge \frac{\nu^\mp k - 4b}{\nu^+ + \nu^-} - 1 \tag{31}$$

for all $k \ge 1$ on B.

For the verification of (22) with $C_a = I_2$, let ξ be a compactly supported, but not necessarily smooth, density on Ω^{++}. Suppose that the sample size is n and let $b = \log^2(n)$ in (30). Then B implies that (31) holds for all $k \ge 1$ and, in particular, for $k = n$. It follows easily that $\|n(X_n^T X_n)^{-1}\|^2 1_B$, $n \ge 1$, where 1_B is the indicator of the event B, are bounded and, therefore, uniformly integrable. Let η be the infimum of α over the compact support of ξ. Then

$$\int_{B^c} \|n(X_n^T X_n)^{-1}\|^2 dP_\xi \le n^2 P_\xi(B^c) \le 4n^2 e^{-\eta b} \to 0$$

as $n \to \infty$.

It remains to verify (31). This relation follows by applying Lemma 2 below to the sequence

$$Z_k = \Delta_k^+ - \frac{\nu^- k - 4b}{\nu^+ + \nu^-}, \quad k \ge 1.$$

Lemma 2. Let z_k, $k \geq 1$, be a sequence of real numbers for which $z_1 > 0$, $|z_k - z_{k-1}| \leq 1$ for all $k \geq 2$, and $z_{k+1} \geq z_k$ whenever $z_k < 0$. Then $z_k \geq -1$ for all $k \geq 1$.

Proof. If $z_k < 0$, then there is a unique $m \geq 1$ for which $z_{m-1} \geq 0$ and $z_j < 0$ for $j = m, \ldots, k$, in which case $z_m \geq -1$ and

$$z_k = z_m + \sum_{j=m+1}^{k} (z_j - z_{j-1}) \geq z_m \geq -1,$$

as required. □

8. DISCUSSION

In this work, we have obtained approximations for the bias and variance of the maximum likelihood estimators of the parameters for model (1). The simulations in Section 5 indicate that the approximations are quite accurate for the four examples described in Section 4. Very weak expansions were used in Section 6 to justify the approximations. One open problem is to obtain fixed-parameter expansions for the bias and variance as in, for example, Aras and Woodroofe (1993). These may require additional conditions. There are also several possible extensions to the examples.

In Example 1, we considered one of the sequential procedures proposed by Coad (1995). It would be interesting to determine the approximations for his other two procedures and to compare them with respect to their estimation properties. The comparison of more than three treatments is also a possibility. A natural extension to Examples 2 and 3 is to consider a general autoregressive model of order p, $p \geq 1$. It should not be difficult to determine an approximation to the bias of $\hat{\theta}_n$. However, the covariance matrix would present a problem, due to the difficulty in computing η_n. Whether the methods in this paper can be applied to other stationary time series models is an open problem.

Perhaps the most interesting of the examples is Example 4. Open problems here include computing approximations for the bias and variance of $g(\hat{\theta}_n)$ and finding a corrected confidence interval for $g(\theta)$. An extension to other sequential designs in which interest lies in estimating a non-linear function would be useful, such as those of Wu (1985). The case of a singular limiting matrix $\mathcal{I}(\sigma, \theta)$ deserves special attention.

APPENDIX: DIFFERENTIATION UNDER THE INTEGRAL SIGN

Let $(\mathcal{X}, \mathcal{B}, \Lambda)$ denote a sigma-finite measure space, let Ω be an open subset of $\mathbb{R}^p$, and let f_θ, $\theta \in \Omega$, be a family of probability densities with respect to the measure Λ. Recall that the family is said to be regular if the square roots $s_\theta = \sqrt{f_\theta}$, $\theta \in \Omega$, are continuously differentiable in mean square. That is, the family is regular if and only if there are measurable functions $\dot{s}_\theta : \mathcal{X} \mapsto \mathbb{R}^p$ for which

$$\int_{\mathcal{X}} \|\dot{s}_\theta\|^2 d\Lambda < \infty, \quad \forall \theta \in \Omega, \tag{A.1}$$

$$\lim_{\omega \to \theta} \int_{\mathcal{X}} \|\dot{s}_\omega - \dot{s}_\theta\|^2 d\Lambda = 0, \quad \forall \theta \in \Omega, \tag{A.2}$$

and

$$\lim_{h \to 0} \int_{\mathcal{X}} \left|\frac{s_{\theta+h\omega} - s_\theta - h\omega^T \dot{s}_\theta}{h}\right|^2 d\Lambda = 0, \quad \forall \theta \in \Omega, \quad \omega \in \mathbb{R}^p. \tag{A.3}$$

If $s_\theta(x)$ are continuously differentiable in θ for almost every x under the measure Λ with gradients denoted by $\nabla s_\theta(x)$, and if $\int_{\mathcal{X}} \|\nabla s_\theta\|^2 d\Lambda$ is finite and continuous in θ, then (A.1), (A.2) and (A.3) hold with $\dot{s}_\theta = \nabla s_\theta$, $\theta \in \Omega$. See the proof of Proposition 1 of Bickel *et al.* (1993, p.13).

The following result is known from Ibragimov and Khasminski (1981, Chap.1).

Proposition. Suppose that f_θ, $\theta \in \Omega$, is regular. If $g : \mathcal{X} \mapsto \mathbb{R}$ is a measurable function for which $\int_{\mathcal{X}} g^2 f_\theta d\Lambda$ is locally bounded near $\theta_0 \in \Omega$, then

$$\bar{g}(\theta) := \int_{\mathcal{X}} g f_\theta d\Lambda$$

is continuously differentiable near θ_0 and

$$\nabla g(\theta) = 2 \int_{\mathcal{X}} g s_\theta \dot{s}_\theta d\Lambda.$$

This result may be applied to the model of Section 1 with $\Lambda = P_{1,0}$,

$$f_{\sigma,\theta}(\mathbf{y}) = \exp\{-\frac{1}{2\sigma^2}\|\mathbf{y} - X_t\theta\|^2 + \frac{1}{2}\|\mathbf{y}\|^2 - \frac{t}{2}\log(\sigma^2)\}$$

and

$$s_{\sigma,\theta}(\mathbf{y}) = \exp\{-\frac{1}{4\sigma^2}\|\mathbf{y} - X_t\theta\|^2 + \frac{1}{4}\|\mathbf{y}\|^2 - \frac{t}{4}\log(\sigma^2)\}$$

for $\sigma > 0$ and $\theta \in \Omega$. For fixed $\sigma > 0$, $s_{\sigma,\theta}(\mathbf{y})$ is continuously differentiable in $\theta \in \Omega$, and

$$\nabla s_\theta(\mathbf{y}) = \frac{1}{2\sigma^2}(X_t^T\mathbf{y} - X_t^T X_t\theta)s_\theta$$

for all $\theta \in \Omega$. That (A.1), (A.2) and (A.3) are satisfied then follows easily since

$$\begin{aligned}\int_{\mathcal{X}} \|\nabla s_\theta\|^2 d\Lambda &= \frac{1}{4\sigma^4}\int_{\mathcal{X}} \|X_t^T\mathbf{y} - X_t^T X_t\theta\|^2 dP_{\sigma,\theta} \\ &= \frac{1}{4\sigma^2}E_{\sigma,\theta}(\|X_t^T X_t\|), \qquad \text{(A.4)}\end{aligned}$$

by the Optional Stopping and Martingale Convergence Theorems. See, for example, Williams (1991, Chaps.10 and 11). The right-hand side of (A.4) is finite and continuous by (5).

Relation (6) follows immediately from the Proposition, under the conditions that $E_{\sigma,\theta}(\|M_t\|^2)$ be finite and continuous in θ for fixed σ. Similarly, if $E_{\sigma,\theta}(\|M_t\|^4)$ is finite and continuous, then

$$\sum_{\ell=1}^{p} \frac{\partial}{\partial\theta_\ell}E_{\sigma,\theta}(M_{tik}M_{tj\ell}) = E_{\sigma,\theta}\{(M_t(\hat{\theta}_t - \theta))_k\} \qquad \text{(A.5)}$$

for all $k = 1, 2, \ldots, p$. Further, if $E_{\sigma,\theta}(\|\hat{\theta}_t - \theta\|^4)$ is locally bounded in θ, then (8) follows by differentiating (A.5) with respect to θ_k and summing over k.

For fixed θ,

$$\frac{\partial}{\partial\sigma^2}s_{\sigma,\theta}(\mathbf{y}) = \frac{1}{4\sigma^4}(\|\mathbf{y} - X_t\theta\|^2 - t\sigma^2)s_{\sigma^2} = \frac{1}{4\sigma^2}\{\sum_{k=1}^{t}(\epsilon_k^2 - 1\}s_{\sigma^2}.$$

So, if $E_{\sigma,\theta}(t)$ is finite and continuous in σ for fixed θ, then

$$\int_{\mathcal{X}} |\frac{\partial}{\partial\sigma^2}s_{\sigma^2}(\mathbf{y})|^2 d\Lambda = (\frac{1}{4\sigma^2})^2 E_{\sigma,\theta}[\{\sum_{k=1}^{t}(\epsilon_k^2 - 1)\}^2] = \frac{1}{8\sigma^4}E_{\sigma,\theta}(t),$$

by Wald's lemma. It follows that the family $f_{\sigma,\theta}$, $\sigma > 0$, is regular in σ for fixed θ, and (11) follows easily from the Proposition. Further,

$$2\sigma^4 \frac{\partial}{\partial \sigma^2} E_{\sigma,\theta}(\frac{1}{t^2}) = E_{\sigma,\theta}\{\frac{1}{t}(\hat{\sigma}_t^2 - \sigma^2 + \frac{\|X_t\hat{\theta}_t - X_t\theta\|^2}{t})\}, \quad \text{(A.6)}$$

and (12) follows by differentiating (A.6) with respect to σ^2.

ACKNOWLEDGEMENTS

Part of this work was carried out while the first author was in receipt of a Fulbright Scholarship Grant. Both authors' research for this paper was supported by the U.S. National Science Foundation.

BIBLIOGRAPHY

Aras, G. and Woodroofe, M. (1993). Asymptotic expansions for the moments of a randomly stopped average. *Ann. Statist.* **21**, 503-519.

Berger, J.O. and Walpole, R.L. (1984). *The Likelihood Principle.* Hayward: California: Institute of Mathematical Statistics

Bickel, P., Klassen, C., Ritov, Y. and Wellner, J. (1993). *Efficient and Adaptive Estimation for Semi-Parametric Models.* John Hopkins.

Brockwell, P.J. and Davis, R.A. (1991). *Time Series: Theory and Methods.* New York: Springer.

Coad, D.S. (1994). Estimation following sequential tests involving data-dependent treatment allocation. *Statist. Sinica* **4**, 693-700.

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

Ford, I. and Silvey, S.D. (1980). A sequentially constructed design for estimating a nonlinear parametric function. *Biometrika* **67**, 381-388.

Ibragimov, I.A. and Khasminski, R.Z. (1981). *Statistical Estimation: Asymptotic Theory.* New York: Springer.

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.

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

Todd, S., Whitehead, J. and Facey, K.M. (1996). Point and interval estimation following a sequential clinical trial. *Biometrika* **83**, 453-461.

Whitehead, J. (1986). On the bias of maximum likelihood estimation following a sequential test. *Biometrika* **73**, 573-581.

Williams, D. (1991). *Probability with Martingales.* Cambridge University Press.

Woodroofe, M. (1982). *Nonlinear Renewal Theory in Sequential Analysis.* Philadelphia: Society for Industrial and Applied Mathematics.

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

Woodroofe, M. (1990). On stopping times and stochastic monotonicity. *Sequential Anal.* **9**, 335-342.

Woodroofe, M. and Coad, D.S. (1996). Corrected confidence sets for sequentially designed experiments. *Statist. Sinica* **6** (To appear).

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

TABLE I

Approximate and simulated values for the bias and covariance matrix of $\hat{\theta}_t$ for the multi-armed clinical trial model

θ_1	θ_2	θ_3	Bias Approx	Bias MC	Covariance matrix Approx			Covariance matrix MC		
			(a) $a = 15.0$ and $d = 0.5$							
-0.5	0.0	0.5	-0.059	-0.073	0.102	0.001	0.000	0.121	0.000	-0.001
			-0.059	-0.067	0.001	0.076	-0.004	0.000	0.092	-0.011
			0.066	0.060	0.000	-0.004	0.073	-0.001	-0.011	0.084
-0.25	-0.25	0.5	-0.033	-0.069	0.088	0.002	0.000	0.104	0.002	-0.007
			-0.033	-0.077	0.002	0.088	-0.004	0.002	0.106	-0.005
			0.066	0.064	0.000	-0.004	0.090	-0.007	-0.005	0.090
-0.5	0.25	0.25	-0.064	-0.074	0.088	0.000	0.000	0.114	-0.001	0.001
			0.000	-0.005	0.000	0.037	0.000	-0.001	0.088	-0.027
			0.001	-0.002	0.000	0.000	0.037	0.001	-0.027	0.087
0.0	0.0	0.0	0.000	-0.034	0.035	0.000	0.000	0.089	-0.014	-0.013
			0.000	-0.030	0.000	0.035	0.000	-0.014	0.092	-0.013
			0.000	-0.032	0.000	0.000	0.035	-0.013	-0.013	0.086
			(b) $a = 30.0$ and $d = 0.5$							
-0.5	0.0	0.5	-0.032	-0.034	0.051	0.000	0.000	0.055	0.000	-0.001
			-0.033	-0.031	0.000	0.035	-0.001	0.000	0.035	-0.001
			0.033	0.028	0.000	-0.001	0.036	-0.001	-0.001	0.037
-0.25	-0.25	0.5	-0.016	-0.032	0.043	0.000	-0.001	0.046	0.000	-0.001
			-0.016	-0.037	0.000	0.043	-0.001	0.000	0.046	-0.001
			0.033	0.027	-0.001	-0.001	0.043	-0.001	-0.001	0.040
-0.5	0.25	0.25	-0.033	-0.038	0.043	0.000	0.000	0.050	0.000	-0.001
			0.000	0.001	0.000	0.017	0.000	0.000	0.031	-0.007
			0.000	-0.002	0.000	0.000	0.017	-0.001	-0.007	0.031
0.0	0.0	0.0	0.000	-0.016	0.017	0.000	0.000	0.031	-0.004	-0.004
			0.000	-0.015	0.000	0.017	0.000	-0.004	0.032	-0.003
			0.000	-0.012	0.000	0.000	0.017	-0.004	-0.003	0.031

TABLE II

Approximate and simulated values for the bias and variance of $\hat{\sigma}_t^2$ for the multi-armed clinical trial model

			Bias		Variance	
θ_1	θ_2	θ_3	Approx	MC	Approx	MC
			(a) $a = 15.0$ and $d = 0.5$			
-0.5	0.0	0.5	-0.089	-0.102	0.048	0.057
-0.25	-0.25	0.5	-0.100	-0.107	0.053	0.061
-0.5	0.25	0.25	-0.043	-0.087	0.027	0.046
0.0	0.0	0.0	-0.033	-0.075	0.021	0.038
			(b) $a = 30.0$ and $d = 0.5$			
-0.5	0.0	0.5	-0.044	-0.045	0.024	0.026
-0.25	-0.25	0.5	-0.050	-0.050	0.027	0.028
-0.5	0.25	0.25	-0.022	-0.035	0.014	0.019
0.0	0.0	0.0	-0.017	-0.029	0.011	0.015

TABLE III

Approximate and simulated values for the bias and variance of $\hat{\theta}_n$ for a first-order autoregression

	$n = 25$				$n = 50$			
	Bias		Variance		Bias		Variance	
θ	Approx	MC	Approx	MC	Approx	MC	Approx	MC
0.0	0.000	-0.001	0.039	0.039	0.000	-0.002	0.020	0.020
0.2	-0.016	-0.017	0.038	0.039	-0.008	-0.009	0.019	0.019
0.4	-0.032	-0.032	0.035	0.036	-0.016	-0.017	0.017	0.017
0.6	-0.048	-0.047	0.030	0.031	-0.024	-0.024	0.014	0.014
0.8	-0.064	-0.061	0.023	0.023	-0.032	-0.032	0.009	0.009
0.9	-0.072	-0.067	0.019	0.018	-0.036	-0.035	0.007	0.007

TABLE IV

Approximate and simulated values for the bias and covariance matrix of $\hat{\theta}_n$ for a second-order autoregression

θ_1	θ_2	Bias Approx	Bias MC	Covariance matrix Approx		Covariance matrix MC	
			(a) $n = 25$				
0.0	0.0	0.000	-0.002	0.037	0.000	0.044	0.001
		-0.040	-0.041	0.000	0.032	0.001	0.042
0.0	0.5	0.000	-0.003	0.032	0.000	0.041	0.001
		-0.100	-0.096	0.000	0.028	0.001	0.039
1.5	-0.6	-0.060	-0.061	0.026	-0.022	0.033	-0.029
		0.032	0.029	-0.022	0.022	-0.029	0.031
1.9	-0.9	-0.076	-0.098	0.012	-0.012	0.026	-0.027
		0.068	0.090	-0.012	0.011	-0.027	0.029
			(b) $n = 50$				
0.0	0.0	0.000	-0.002	0.019	0.000	0.021	0.000
		-0.020	-0.021	0.000	0.018	0.000	0.020
0.0	0.5	0.000	-0.001	0.016	0.000	0.018	0.000
		-0.050	-0.049	0.000	0.014	0.000	0.017
1.5	-0.6	-0.030	-0.031	0.013	-0.012	0.015	-0.013
		0.016	0.016	-0.012	0.012	-0.013	0.014
1.9	-0.9	-0.038	0.051	0.005	-0.005	0.009	-0.009
		0.034	0.047	-0.005	0.005	-0.009	0.009

TABLE V

Approximate and simulated values for the bias and covariance matrix of $\hat{\theta}_n$ for the Ford-Silvey example

		Bias		Covariance matrix			
θ_1	θ_2	Approx	MC	Approx		MC	
			(a) $n = 25$				
1.0	1.5	0.032	0.025	0.078	0.050	0.078	0.050
		0.048	0.039	0.050	0.070	0.050	0.072
1.0	2.0	0.013	0.009	0.055	0.027	0.056	0.026
		0.027	0.022	0.027	0.053	0.026	0.053
1.0	4.0	0.003	0.004	0.043	0.011	0.043	0.010
		0.011	0.009	0.011	0.043	0.010	0.043
			(b) $n = 50$				
1.0	1.5	0.016	0.012	0.037	0.024	0.038	0.025
		0.024	0.020	0.024	0.035	0.025	0.037
1.0	2.0	0.007	0.005	0.027	0.013	0.028	0.013
		0.013	0.011	0.013	0.026	0.013	0.026
1.0	4.0	0.001	0.003	0.021	0.005	0.022	0.006
		0.005	0.005	0.005	0.021	0.006	0.021