The Annals of Statistics
1985, Vol. 13, No. 2, 663–675

ESTIMATION IN THE GENERAL LINEAR MODEL WHEN THE ACCURACY IS SPECIFIED BEFORE DATA COLLECTION

BY MARK FINSTER

University of Wisconsin

An estimator $\hat{\beta}$ of β is accurate with accuracy A and confidence γ, $0 < \gamma < 1$, if $P(\hat{\beta} - \beta \in A) \geq \gamma$ for all β. Given a sequence $Y_1, Y_2, \cdots$ of independent vector-valued homoscedastic normally-distributed random variables generated via the general linear model $Y_i = X_i\beta + \varepsilon$, the k-dimensional parameter β is accurately estimated using a sequential version of the maximum probability estimator developed by L. Weiss and J. Wolfowitz. The procedure given also generalizes C. Stein's fixed-width confidence sets to several dimensions.

1. Introduction and summary. While observing a sequence $Y_1, Y_2, \cdots$ of independent vector-valued homoscedastic normally-distributed random variables generated via the general linear model

$$Y_i = X_i\beta + \varepsilon_i \tag{1.1}$$

we want to accurately estimate the unknown parameter $\beta \in \mathbb{R}^k$. An estimator $\hat{\beta}$ of β is accurate with accuracy $A \subset \mathbb{R}^k$ and confidence γ, $0 < \gamma < 1$, if $P(\hat{\beta} - \beta \in A) \geq \gamma$ for all β. The accuracy set A need only be a Borel measurable set having an interior point at zero. The set $\hat{\beta} - A$ is a $100\gamma\%$ fixed-accuracy confidence set for β. For such estimation procedures the accuracy of the estimator is specified prior to data collection and, hence, accuracy or error specification is a design feature. Unlike the usual theory of confidence sets, a fixed-accuracy confidence set is determined, with stated confidence, by subtracting the nonrandom preassigned accuracy or error set A from the estimator.

In (1.1) the m-variate real valued vectors $\varepsilon_1, \varepsilon_2, \cdots$ are independent and normally distributed with mean zero and covariance matrix $\sigma^2 \sum_i$ where $\sum_i > 0$ is known but $\sigma > 0$ is unknown. To simplify analysis, and without loss of generality, assume each $\sum_i$ is the identity. For the sake of simplicity, assume also that $X_1, X_2, \cdots$ are known $m \times k$ matrices of minimum rank $r > 0$ and having design measure converging to that of some optimal design. For example, see Karlin and Studden (1966), Kiefer and Studden (1976), Federov (1972), Silvey (1980), and Section 4 of this article. In some special situations (see Example 4 of Section 4) the X_i may be i.i.d. random matrices having a joint distribution independent of the ε_i's. In such cases assume that $r = \min\{n: P(\text{rank}(X_i) = n) > 0\} > 0$ and that β is identifiable with probability one (i.e., $w \in \mathbb{R}^k$, $w \neq 0$

Received February 1984; revised June 1984.
AMS 1980 *subject classifications.* Primary 62L12, 62E20; secondary 60G40.
Key words and phrases. Fixed-accuracy confidence set, sequential methods, nonlinear renewal theory, general linear model, maximum probability estimator.

implies $P(X_i w = 0) = 0$), then the distribution of X_i is continuous on $\{\text{rank}(X_i) < k\}$ and for $n \geq p = [\![k/r]\!]$, the matrix

$$M_n = \sum_{i=1}^{n} X_i' X_i \tag{1.2}$$

is invertible with probability one. Here $[\![x]\!]$ is the smallest integer containing x.

Accurate estimators were probably first developed for i.i.d. observations from the Bernoulli distribution, for which the sample mean accurately estimates the expectation with stated confidence and accuracy, provided the sample size is sufficiently large. Stein (1945, 1949) suggested accurate estimation procedures for estimating the mean, say β, of a one-dimensional normal distribution having unknown variance σ^2. Recognizing that no fixed-sample-size procedure can be accurate, Stein suggested estimating β by the sample mean to accuracy $[-d, d]$, for given $d > 0$, via a two-stage procedure in which the sample size depends on the variance estimate of a pilot sample. However, his "fixed-width" procedures are inefficient when the pilot sample is small relative to N, the minimal sample size required for accurate estimation when the nuisance parameter σ^2 is known, because they don't use the second stage to estimate σ^2. Specifically, for the two-stage procedure the regret in not knowing σ^2, which is the difference between N and the expected sample size (ASN), tends to infinity as $N \to \infty$. See Cox (1952) and Ghosh and Mukhopadhyay (1979). To alleviate this inefficiency Stein (1949) suggested updating the sample size estimate after each observation. See also Anscombe (1952, 1953). Although the bounded regret of this continually updating procedure has been evaluated up to $o(1)$ terms as $N \to \infty$, the procedure is consistent (i.e. the confidence is at least γ) only up to $o(N^{-1})$ terms. Woodroofe (1977, 1982) has given a nonlinear-renewal theoretic development of these second-order asymptotic results. By sampling a predetermined number of observations beyond this stopping rule, Simons (1968) obtained truly consistent fixed-accuracy procedures, that is, procedures having confidence at least γ.

Starr (1966) examined the small and moderate sample size performance of Stein's sequential rule. Chow and Robbins (1965) have showed that this procedure is also asymptotically consistent and asymptotically efficient (i.e. ASN $\sim N$) even when the underlying distribution is unknown. Related ideas are found in Starr and Woodroofe (1968, 1969, 1972), Vardi (1979), and Martinsek (1983).

Hall (1981) initiated the quite interesting and more practical triple-sampling approach to "fixed-width" accuracy, in which a pilot sample is used to obtain an initial projection of the sample size needed. After collecting only a fraction of this projected sample size, the estimated sample size is updated and sampling is then completed after this one update. Hall presents second-order asymptotic and Monte Carlo results for triple sampling from the normal distribution, results that compare favorably with the continually updating procedure.

In all these accurate procedures the one-dimensional parameter is estimated to accuracy $[-d, d]$, leading to the "fixed-width" confidence set $\{\beta: |\beta - \hat{\beta}| < d\}$. Gleser (1965) extended the asymptotic consistency and asymptotic efficiency results to several-dimensional linear regression where the fixed-accuracy confidence set again is $\{\beta: |\hat{\beta} - \beta| < d\}$, corresponding to the spherical accuracy set $A = \{\beta: |\beta| < d\}$. Here $|\cdot|$ is Euclidean distance.

However, in both one- and several-dimensional problems, the appropriate

accuracy need not be given by a sphere centered at zero. In practice $\hat{\beta}$ might be a quantity used in place of the actual value of β. Overestimating a coordinate of β by a fixed amount might be more serious than underestimating the coordinate by that same fixed amount, as in dosage determination of a drug toxic at high levels, or as in estimation of the intensity of a radioactive substance. See Lehmann (1959, page 78). In other words, spherical or even ellipsoidal accuracy and confidence sets might not be appropriate. For example, if each coordinate of β is to be simultaneously estimated, the ith coordinate to accuracy A_i (e.g. $A_i = [-d_i, d_i]$ for some $d_i > 0$), then β should be estimated to accuracy ΠA_i. In practice, the accuracy desired for a parameter depends on the given problem and on the particular use of the estimator. The one constraint placed on the accuracy is that it be star shaped with respect to zero. That is, if $\hat{\beta}$ is an accurate estimate of β, then any estimate closer to β must also be accurate.

Without placing traditional constraints on the type of accuracy, accurate procedures are derived, in Section 2, for estimators of β in model (1.1). The associated fixed-accuracy confidence sets generalize Stein's (1945, 1949) fixed-width confidence sets to several dimensions. Relevant properties of the required sample size are also presented. For example, nonlinear renewal theory is used to derive the second-order asymptotic properties of the ASN (such as its regret) as in Woodroofe (1982). Siegmund (1978, 1980) has also used nonlinear renewal theory in a very eloquent way to estimate and to obtain confidence sets after sequential hypothesis testing, but his estimators are not accurate in our sense.

In Section 3, a general theorem is given which can be used to determine the second-order confidence of this procedure under various design strategies. A simple algorithm for computing the relevant parameters is also presented.

In Section 4, several examples of accurate estimation procedures are presented for specific design optimality criteria (e.g., D-optimality and grid search designs) and for specific accuracy sets (e.g., spherical, rectangular, and ellipsoidal accuracy).

It should be noted here that our notion of accurate estimators and their related fixed-accuracy confidence sets are quite different from Lehmann's (1959) uniformly most accurate (perhaps invariant or unbiased) confidence sets. Although motivated by similar statistical needs, our methods yield fixed-accuracy confidence sets for a broader range of problems, and our estimators are conceptually more similar to Weiss and Wolfowitz's (1969, 1970, 1974) and Wolfowitz's (1975) maximum probability estimators. Uniformly most accurate confidence sets are not, in general, fixed-accuracy confidence sets.

Many of the techniques and methods used in this paper depend on nonlinear renewal theory as developed by Woodroofe (1976, 1977, 1978, 1979), Lai and Siegmund (1977, 1979), and Siegmund (1977, 1978, 1980). Woodroofe (1982) is an excellent survey.

2. Accurate estimation and the sampling procedure. After sampling (X_i, Y_i) for $i = 1, 2, \cdots, n$, the estimator $\hat{\beta}_n = \beta_n$, where β_n is the usual least squares and maximum likelihood estimator

$$\beta_n = M_n^{-1} \sum_1^n X_i' Y_i, \tag{2.1}$$

is an accurate estimator of β, having accuracy A and confidence γ, provided

$$P(\hat{\beta}_n - \beta \in A) \geq \gamma. \tag{2.2}$$

However, shifting β_n by a vector $v = v(\sigma^2)$ to

$$\hat{\beta}_n = \beta_n + v \tag{2.3}$$

may increase the probability of accuracy, as is suggested by the work of Weiss and Wolfowitz (1969, 1970, 1974) and Wolfowitz (1975). Their maximum probability estimator is the d that maximizes the integral, with respect to Lebesgue measure, of the likelihood function over the set $d - A$. For the general linear model with σ^2 known, the maximum probability estimator is precisely the estimator (2.3) where v is the vector that maximizes the left-hand side of (2.2).

For the sake of simplicity, suppose v is independent of σ^2 and n. For example, if A is bounded, symmetric, and star-shaped with respect to a vector v, then v maximizes (2.2) independently of σ^2 and n. If the v maximizing (2.2) is not unique, choose a maximizing v closest to zero. For example, if $c'\beta$ ($c \in \mathbb{R}^k$) is to be estimated to accuracy (δ, ε), for $\delta < 0 < \varepsilon$, or, equivalently, if β is to be estimated to accuracy $\{\beta\colon \delta < c'\beta < \varepsilon\}$, the maximum probability estimator is $c'\beta_n + (\varepsilon + \delta)/2$. The existence of such a v is a standing assumption.

Then the maximum probability estimator (2.3) is an accurate estimate of β provided the standardized error of the least squares estimator

$$e_n = \sqrt{n}\sigma^{-1}(\beta_n - \beta) \tag{2.4}$$

satisfies

$$\gamma \leq P(\hat{\beta}_n - \beta \in A) = P(\sigma n^{-1/2} e_n \in A - v) = F_n(n/\sigma^2). \tag{2.5}$$

Here

$$F_n(w) = P(w^{-1/2} e_n \in A - v), \quad w > 0 \tag{2.6}$$

is an increasing function and does not depend on β or σ^2 since $\beta_n \sim N(\beta, \sigma^2 M_n^{-1})$. Thus $\hat{\beta}_n$ has accuracy A and confidence γ provided $n\sigma^{-2} \geq a_n$ where

$$a_n = F_n^{-1}(\gamma) \tag{2.7}$$

or, equivalently, provided $n \geq \sigma^2 a_n$. Although σ^2 is unknown, it can be estimated by its usual (MVUE) estimator

$$\hat{\sigma}_n^2 = (nm - k)^{-1} \textstyle\sum_1^n \mid Y_i - X_i\beta_n \mid^2, \tag{2.8}$$

suggesting termination of sampling when $n \geq \hat{\sigma}_n^2 a_n$. However, due to the randomness of $\hat{\sigma}_n^2$ ($\hat{\sigma}_n^2$ acts roughly as the average of a random walk, cf. Finster, 1983), this rule tends to stop sampling early. To protect against early termination, the variance estimate is inflated by the factor

$$R_n = 1 + r_n/n + o(n^{-1})$$

for design-dependent parameters $r_n \geq 0$ belonging to a compact set. Often $r_n \equiv r \geq 0$. R_n will be discussed further in Section 3 and Section 4. Thus, after taking

a pilot sample of size $\eta \geq p$, our procedure uses a sample of total size

$$t = \inf\{n \geq \eta: n \geq R_n \hat{\sigma}_n^2 a_n\} \tag{2.9}$$

to obtain the accurate estimator $\hat{\beta}_t$.

Later we will show that, in most practical design strategies,

$$a_n = a(1 + \alpha_n/n) + o(n^{-1}) \tag{2.10}$$

where the α_n are design-dependent parameters belonging to a compact set. See (3.20) for a determination of α_n using a linear function. The parameter a is a measure of the precision desired. This precision parameter measures the "size" of and the "confidence" in the accuracy set according to (2.6), increasing to infinity as either the confidence γ increases to one, or as the accuracy set A decreases to the empty set (decrease here being defined with respect to the partial order of set inclusion). Setting $N = a\sigma^2$ and letting

$$\ell_n = R_n a_n / a = 1 + \Delta_n/n + o(n^{-1}) \tag{2.11}$$

where $\Delta_n = r_n + \alpha_n$ is constrained to be positive, (2.9) can be written

$$t = \inf\{n \geq \eta: n(\sigma^2/\hat{\sigma}_n^2)/\ell_n \geq N\}. \tag{2.12}$$

Here N is approximately the integer representing the minimal sample size necessary for accurate estimation of β with confidence γ. Note that $N \to \infty$ as the precision $a \to \infty$ or as $\sigma \to \infty$.

The sampling procedure t terminates with probability one and has a distribution that is independent of β and that depends on the unknown parameter σ^2 only through N. Given $X_1, X_2, \cdots, X_t$, $\hat{\beta}_t \sim N(\beta + v, \sigma^2 M_t^{-1})$. Furthermore, t is both pointwise and momentwise asymptotically efficient in that $t/N \to 1$ with probability one, and, for all $q > 0$, $N^{-q}E(t^q) \to 1$ as $N \to \infty$. These and the second-order results stated in Theorem 1 below can be derived from (2.12) as in Section 3 of Finster (1983). The derivation will not be repeated here.

THEOREM 1. *As* $N \to \infty$

(i) $(t - N)/\sqrt{N} \to_{\mathcal{L}} N(0, 2/m)$, *and*
(ii) *the probability of early termination is*

$$P(t < N/2) \sim P(t = \eta) \sim C_\eta N^{-b}$$

where $b = (\eta m - k)/2$ *and* $C_\eta = b^b(\eta/\ell_\eta)^b\Gamma(b + 1)$.

If $\eta > \max[p, (2 + k)/m]$ *then*

(iii) $E(t) = N + \frac{1}{2} - (\nu + 1)/m + \Delta + o(1)$, *where* $\Delta = E(\Delta_t)$, *and*
(iv) $N^{-1}E(t - N)^2 \to 2/m$.

Here $\nu = \nu(m) = \sum_{n=1}^{\infty} n^{-1}E\{[\chi_{nm}^2 - 2mn]^+\}$ where χ_{nm}^2 represents a variable having that distribution. See Table 1 ($\alpha = 10{,}000$) of Finster (1983) for specific values of ν. The function $\nu(m)$ is decreasing and is always less than .7. In many practical situations (see Section 4) $\alpha_n \equiv 0$, $r_n \equiv r$, and $E(\Delta_t) = r$ so that the regret

in not knowing σ^2 is $r + \frac{1}{2} - (\nu + 1)/m + o(1)$. In Section 3 a constraint on r_n is determined to ensure second-order confidence of at least γ.

When using sampling procedure (2.12) it is necessary to update and hence compute M_n^{-1}, β_n, and $\hat{\sigma}_n^2$ for each $n \geq \eta$. If $m = 1$ a three-step algorithm that avoids inverting M_n follows.

(1) $M_n^{-1} = M_{n-1}^{-1} - C_n C_n'/\xi_n$ where $C_n = M_{n-1}^{-1} X_n'$ and $\xi_n = 1 + X_n C_n$. Here prime denotes transpose.

(2) $\beta_n = \beta_{n-1} + \varepsilon_n M_n^{-1} X_n'$ where $\varepsilon_n = Y_n - X_n \beta_{n-1}$ is the prediction error in predicting Y_n from $Y_1, Y_2, \cdots, Y_{n-1}$.

(3) $(n - k)\hat{\sigma}_n^2 = (n - k - 1)\hat{\sigma}_{n-1}^2 + \varepsilon_n^2/\xi_n$.

If $m > 1$ the above algorithm can be used to update M_n^{-1} one row at a time. See Finster (1983) and Brown, Durbin, and Evans (1975).

3. Confidence in the sampling procedure. To calculate the confidence note that (2.5), (2.12), and the independence of $\{\hat{\sigma}_i^2, i \leq n\}$ and $\hat{\beta}_n$ imply

$$\begin{aligned} P(\hat{\beta}_t - \beta \in A) &= \textstyle\sum_{n=\eta}^{\infty} P(\hat{\beta}_n - \beta \in A, t = n) = \sum_{n=\eta}^{\infty} F_n(n/\sigma^2) P(t = n) \\ &= E[F_t(t/\sigma^2)] = E[F_t(at/N)]. \end{aligned} \tag{3.1}$$

Since e_n, given in (2.4), has distribution $N(0, nM_n^{-1})$, one might hope to calculate asymptotics for the coverage probability provided M_n/n converges elementwise to some matrix $\mathscr{I}$ where $\mathscr{I}' = \mathscr{I} > 0$, so that, for all w, $F_n(w)$ converges to

$$F(w) = P(w^{-1/2} X \in A - v), \quad X \sim N(0, \mathscr{I}). \tag{3.2}$$

In fact, the asymptotic consistency (i.e., convergence of the coverage probability to γ as $N \to \infty$) follows easily from (3.1), (3.2), and the asymptotic properties of the sample size. However, to derive the second-order asymptotics, additional regularity conditions are needed.

Let $C^2(0, \infty)$ symbolize all nonnegative twice continuously differentiable functions defined on $(0, \infty)$ with range in $(0, 1]$. If $F \in C^2(0, \infty)$ let $\dot{F}$ and $\ddot{F}$ denote first and second derivatives respectively.

THEOREM 2. *Let $0 < \gamma < 1$, let $\{a_n\}$ be a sequence of constants converging to a, and suppose $F_n \in C^2(0, \infty)$, $n \geq 1$, satisfy the following assumptions.*

(i) *There exists a function $F \in C^2(0, \infty)$ such that $\ddot{F}_n(w) \to \ddot{F}(w)$ uniformly on a neighborhood about a.*

(ii) *$\dot{F}(a) \neq 0$ and $|\dot{F}_n(a) - \dot{F}(a)| = o(n^{-1/2})$.*

(iii) *$F_n(a_n) = \gamma$ for all n and $F_n(a) - F(a) = O(1/n)$.*

(iv) *There exists $L > 0$, $\varepsilon > 0$, and $\alpha \geq 0$ such that, for all n, $|\ddot{F}_n(w)| \leq L w^{-\alpha}$ if $0 < w < a + \varepsilon$.*

Then, for $\eta > \max[(2\alpha + k + 2)/m, p]$, t *defined by* (2.12), *and* $\bar{r} = E(r_t)$,

$$P(\hat{\beta}_t - \beta \in A) = \gamma + aN^{-1}\{\dot{F}(a)[\tfrac{1}{2} - (\nu + 1)m^{-1} + \bar{r}] + a\ddot{F}(a)/m\} + o(N^{-1}).$$

Note that Theorem 2 implies the inflation factor $R_n = 1 + r_n/n + o(n^{-1})$ must satisfy

$$\bar{r} \geq (\nu + 1)/m - \tfrac{1}{2} - a\ddot{F}(a)/m\dot{F}(a) \tag{3.3}$$

if the probability of coverage is to be at least γ up to $o(N^{-1})$ terms. Note also that for $m \geq 2(\nu + 1) - 2a\ddot{F}(a)/\dot{F}(a)$, any positive r_n suffices.

PROOF. A Taylor's expansion of $a_n = F_n^{-1}(\gamma)$ about $F_n(a)$ gives

$$a_n = a - [\dot{F}_n(a)]^{-1}[F_n(a) - F(a)] + o(F_n(a) - F(a)). \tag{3.4}$$

Thus by assumptions (i), (ii), and (iii) there exists α_n, $n \geq 1$ such that (2.10) holds. A Taylor's expansion of F_t about a_t yields

$$F_t(at/N) = \gamma + GHN^{-1} + DH^2N^{-2} \tag{3.5}$$

where $G = a_t\dot{F}_t(a_t)$, $H + N = at/a_t$, and $D = \tfrac{1}{2}a_t^2\ddot{F}_t(w_t)$ for some w_t between a_t and at/N. Let I_N be the indicator function of $\{t > N/2\}$ and let $I_N' = 1 - I_N$ indicate the complimentary event. Assumptions (i), (ii), (2.10), a first-order Taylor's expansion of $\dot{F}_n$ about a, and the triangle inequality show

$$\sup_{n \geq N/2} | a_n\dot{F}_n(a_n) - a\dot{F}(a) | = o(N^{-1/2}). \tag{3.6}$$

Hence $| G |$ is bounded and by Theorem 1 (ii) there exists a constant $K > 0$ so that

$$E(GHI_N') \leq KNP(t < N/2) = o(1) \tag{3.7}$$

for $\eta > (2 + k)/m$. Theorem 1(i) implies H^2N^{-1} converges in distribution to $(2/m)\chi_1^2$ and (3.6) implies

$$[G - a\dot{F}(a)]^2NI_N = o(1)$$

uniformly in N. Hence by Schwartz's inequality and Theorem 1

$$| E([G - a\dot{F}(a)]HI_N) |^2 \leq E([G - a\dot{F}(a)]^2NI_N)E(H^2N^{-1}) = o(1).$$

This and Theorem 1(iii) imply

$$\begin{aligned} E(GHI_N) &= a\dot{F}(a)E(HI_N) + o(1) \\ &= a\dot{F}(a)[E(r_t) + \tfrac{1}{2} - (\nu + 1)/m] + o(1). \end{aligned} \tag{3.8}$$

Assumptions (i) and (ii) imply D converges dominantly on $\{t > N/2\}$ and hence by Theorem 1(ii) and (iv)

$$E(DH^2N^{-1}I_N) = a^2\ddot{F}(a)/m + o(1). \tag{3.9}$$

Furthermore, there exists a constant C so that $H^2 < CN^2$ on $\{t < N/2\}$. Since $w_t \geq \min\{atN^{-1}, a_t\}$, assumption (iv) and Theorem 1(ii) imply that for sufficiently

large N

$$
\begin{aligned}
E(DH^2N^{-1}I'_N) &\le CNE(\mid D \mid I'_N) \\
&\le CN^{\alpha+1}E(\max\{a^{-\alpha}t^{-\alpha}, a_t^{-\alpha}N^{-\alpha}\}I'_N) \\
&\le Ca^{-\alpha}N^{\alpha+1}P(t < N/2) = o(1)
\end{aligned} \tag{3.10}
$$

provided $\eta > (2\alpha + k + 2)/m$. Combining (3.1), (3.5), (3.7), (3.8), (3.9) and (3.10) completes the proof.

To verify the assumptions of Theorem 2, the designs $X_1, X_2, \cdots$ must be examined further. The design structure can be specified by the design information matrix per observation per unit variance, say $\mathscr{I}_n = n^{-1}M_n$, for each n. See Federov (1972) or Silvey (1980). In most situations the information $\mathscr{I}_n$ satisfies

$$\mathscr{I}_n = n^{-1}M_n = \mathscr{I} + n^{-1}K_n + o(n^{-1}) \tag{3.11}$$

for some $\mathscr{I} = \mathscr{I}' > 0$ and for matrices K_n, $n \ge 1$, belonging to some compact set. Let $\mathscr{I}_0 = \mathscr{I}$. Then F_n defined by (2.6) can be written

$$F_n(w) = w^{k/2}g_{0,n}(w), \quad n = 0, 1, 2, \cdots \tag{3.12}$$

where $g_{i,n}(w) = g_i(\mathscr{I}_n, \mathscr{I}_n, w)$ with

$$g_i(B, C, w) = \int_{A-v} 2^{-i}(z'Bz)^i(2\pi)^{-k/2} \mid C \mid^{1/2} \exp\left\{-\frac{1}{2}(wz'Cz)\right\} dz, \tag{3.13}$$

$$i = 0, 1, \cdots$$

defined for symmetric $k \times k$ matrices B and $C > 0$. Here $\mid C \mid$ denotes the determinant of C and dz is Lebesgue measure on $\mathbb{R}^k$. Note that differentiation with respect to w gives

$$\dot{g}_{i,n} = -g_{i+1,n} \tag{3.14}$$

which with (3.12) implies

$$w^{1-k/2}\dot{F}_n(w) = \tfrac{1}{2}kg_{0,n} - wg_{1,n} \tag{3.15}$$

and

$$w^{2-k/2}\ddot{F}_n(w) = (k/2)(k/2 - 1)g_{0,n} - kwg_{1,n} + w^2g_{2,n}. \tag{3.16}$$

Thus, verification of the assumptions in Theorem 2 involves examination of the $g_{i,n}$. Suppose the design structure satisfies (3.11). Then, for fixed i, $g_{i,n}$ converges to $g_{i,0}$ uniformly on compact sets of w, and assumption (i) is satisfied with $F = F_0$ defined by (3.12) or (3.2). Similarly, assumption (iv) follows from (3.16) with $\alpha = \frac{3}{2}$ if $k = 1$, $\alpha = \frac{1}{2}$ if $k = 3$, and $\alpha = 0$ for other k. To verify assumption (iii) consider

$$g_{0,n}(a) - g_{0,0}(a) = \mid \mathscr{I}_n \mid^{-1/2}\{\mid \mathscr{I}_n \mid^{1/2} - \mid \mathscr{I} \mid^{1/2}\}g_{0,n}(a) + ag/n \tag{3.17}$$

where

$$g = na^{-1}(2\pi)^{-k/2} |\mathscr{I}|^{1/2} \int_{A-v} \exp\{-1/2(az' \mathscr{I}_n z)\} - \exp\{-1/2(az'\mathscr{I} z)\}\, dz.$$

Considering the inner product of two matrices as the trace of their matrix product, denoted tr, a Taylor's expansion of $|\mathscr{I}_n|^{1/2}$ about $|\mathscr{I}|$ gives

$$|\mathscr{I}_n|^{1/2} - |\mathscr{I}|^{1/2} = (2n)^{-1} |\mathscr{I}|^{1/2} \mathrm{tr}(\mathscr{I}^{-1} K_n) + o(n^{-1}). \tag{3.18}$$

A Taylor's expansion of the integrand of g indicates

$$g = -g_1(K_n, \mathscr{I}, a) + o(1). \tag{3.19}$$

Thus (3.17), (3.18), (3.19) and (3.12) show assumption (iii) is satisfied, and, together with (3.4) and (3.15), indicate

$$\alpha_n = -h(K_n)/h(\mathscr{I}) \tag{3.20}$$

in (2.10), where h is the linear function

$$h(K_n) = \gamma\ \mathrm{tr}(I^{-1}K_n) - 2a^{(1/2)k+1} g_1(K_n, \mathscr{I}, a). \tag{3.21}$$

Verification of assumption (ii) is similar and the details are omitted.

Note that $h(K_n)$ is very easy to calculate for changing K_n since it is linear. Hence (3.20) gives a simple method for calculating α_n and thus, by (2.10), for determining a_n directly from the design information $\mathscr{I}_n$ satisfying (3.11).

Summarizing the above, we have the following.

COROLLARY. *If* $\mathscr{I}_n = \mathscr{I} + n^{-1}K_n + o(n^{-1})$ *for matrices* K_n *belonging to a compact set, and if* $\alpha_n = -h(K_n)/h(\mathscr{I})$, *then the results of Theorem* 2 *are valid for* $a_n = a(1 + n^{-1}\alpha_n) + o(n^{-1})$, $F_n(w) = w^{k/2} g_{0,n}(w)$, $F \equiv F_0$, *and for* $\eta > \max[y, p]$ *where* $y = 5 + k$ *if* $m = 1$, $y = 1 + k/3$ *if* $m = 3$, *and* $y = (2 + k)/m$ *for other* m.

4. Illustrative examples. In this section we discuss various regression examples having specific design strategies and given accuracy sets.

EXAMPLE 1. Suppose $c \in \mathbb{R}^k$ and $c'\beta$ is to be estimated to accuracy (δ, ε), for $\delta < 0 < \varepsilon$. The maximum probability estimator of $c'\beta$ is $c'\hat{\beta}_n = c'\beta_n + (\varepsilon + \delta)/2$ where β_n is the MLE of β. Thus, if Φ denotes the standard normal cumulative distribution function, $F_n(n/\sigma^2) = P(\delta < c'\hat{\beta}_n - c'\beta < \varepsilon) = 2\Phi(\sigma^{-1}n^{1/2}s\xi_n^{1/2}) - 1$ where $s = (\varepsilon - \delta)/2[c'\mathscr{I}^{-1}c]^{1/2}$ and $\xi_n = c'\mathscr{I}^{-1}c/c'\mathscr{I}_n^{-1}c$. Hence $F_n(w) = F(w\xi_n)$ and $F(w) = 2\Phi(w^{1/2}s) - 1$, so that $a = z^2/s^2$ and $a_n\xi_n = a$ where $z = \Phi^{-1}((1 + \gamma)/2)$. The procedure of Section 2 takes a pilot sample of size $\eta \geq p$ and then stops sampling as soon as $n \geq \hat{\sigma}^2 a_n R_n$ where $R_n = 1 + n^{-1}r_n$. The maximum probability estimator is then used. Theorem 1 applies when $\eta > 3/m$ and Theorem 2 applies when $\eta \geq 6/m$ and r_n satisfies (3.3).

Note that if the design information satisfies (3.11) then

$$\begin{aligned} n(\mathscr{I}_n^{-1} - \mathscr{I}^{-1}) &= \mathscr{I}^{-1}(n\mathscr{I} - n\mathscr{I}_n)\mathscr{I}^{-1} + (\mathscr{I}_n^{-1} - \mathscr{I}^{-1})(n\mathscr{I} - n\mathscr{I}_n)\mathscr{I}^{-1} \\ &= -\mathscr{I}^{-1}K_n\mathscr{I}^{-1} + o(1) \end{aligned}$$

so that the α_n, which determine the a_n by (2.10), can be given by

$$\alpha_n = -c'\mathscr{I}^{-1}K_n\mathscr{I}^{-1}c/c'\mathscr{I}^{-1}c. \tag{4.1}$$

When calculating a_n from $\mathscr{I}_n$ for many n, (4.1) is much easier to use than the formula for ξ_n above which involves inverting $\mathscr{I}_n$.

For example, in one variable polynomial regression of degree $d = k - 1$, suppose $\beta = (\beta_0, \beta_1, \cdots, \beta_d)'$ determines the coefficients of the dth degree polynomial $\sum_0^d \beta_i x^i$, $x \in [-1, 1]$. More specifically, the jth row of X_i is given by $(1, x_{ij}, x_{ij}^2, \cdots, x_{ij}^d)$ for $x_{ij} \in [-1, 1]$.

Suppose further that the leading coefficient β_d is to be accurately estimated. If n is fixed, and if nm is a multiple of $2d$, then an optimal design places observations at the zeroes of $(1 - x^2)\dot{T}_d(x)$ where $T_d(x)$ is the dth Chebyshev polynomial of the first kind. More specifically, if $\{-1 = \pi_0 < \pi_1 < \cdots < \pi_d = 1\}$ are the roots, $(1/2d)$th of the observations are placed at each of $x = 1$ and $x = -1$, and $(1/d)$th of the observations are taken from the remaining design points $\{\pi_i\colon i = 1, \cdots, d - 1\}$. See Karlin and Studden (1966) and Federov (1972) or Silvey (1980). Suppose the observations are taken $2d$ at a time (i.e. $m = 2d$) according to the above proportions. That is, suppose the second column of each X_i consists of the elements $\Pi = \{\pi_0, \pi_1, \pi_1, \cdots, \pi_{d-1}, \pi_{d-1}, \pi_d\}$, where each π_j, for $j = 1, \cdots, d - 1$, is listed twice but $+1$ and -1 are listed only once. Then the sampling procedure always stops at an optimal design, $\mathscr{I} \equiv \mathscr{I}_n$, $\alpha_n = 0$, and the (i, j)th coordinate of $\mathscr{I}$ is $(2d)^{-1}\,[2\sum_{w=1}^{d-1} \pi_w^{i+j-2} + 1 + (-1)^{i+j}]$. Furthermore Theorem 1 and Theorem 2 apply with second-order confidence at least γ if $a_n = a = z^2/s^2$, if $r_n = r > 0$, if $\eta > \max[p, 6/m]$, and, according to (3.3), if either

$$r \geq (2\nu + z^2 + 3)/2m - \tfrac{1}{2} \quad \text{or} \quad m \geq 2\nu + z^2 + 3. \tag{4.2}$$

The same approach leads to similar results when accurately estimating any other coordinate of β, say β_ρ, in which case the design points might be taken from the roots of $(1 - x^2)\dot{T}_d(x)$ if $d - \rho$ is even or $(1 - x^2)\dot{T}_{d-1}(x)$ if $d - \rho$ is odd.

EXAMPLE 2. More generally, again consider polynomial regression as described in Example 1, and suppose β is to be estimated to some fixed accuracy. For example, if each coefficient is to be estimated to accuracy $[-d, d]$, then β must be estimated to accuracy $[-d, d]^k$. A rational discrete design measure, say μ, like the optimal designs mentioned above, distributes its mass on (i.e. takes observations from) its finite support set, assigning a rational number to each mass. Suppose κ is the smallest positive integer for which κ times each mass is an integer. Let $\Pi = \{\pi_i\colon i = 1, 2, \cdots, \kappa\}$ be a listing of the support points of μ, with each support point listed as often as κ times its mass, and set $q(n) = nm \bmod \kappa$, $n = 1, 2, \cdots$. Then the information matrix $\mathscr{I}$ corresponding to the rational discrete design measure μ has (i, j)th entry

$$\mathscr{I}_{ij} = \int x^{i+j-2}\, d\mu(x) = \kappa^{-1} \sum_{w=1}^{\kappa} \pi_w^{i+j-2}. \tag{4.3}$$

Whenever a rational discrete design measure is implemented and κ divides m, the number of observations taken at one time, then $\mathscr{I}_n = \mathscr{I}$, $\alpha_n = 0$ by (3.20),

$a_n = a$ by (2.10), and Theorem 1 and the Corollary to Theorem 2 apply with second-order confidence at least γ if $r_n = r > 0$ satisfies (3.3).

For arbitrary m, not necessarily a multiple of κ, a_n can also be determined from Π. Suppose that, at the $(n + 1)$st stage, observations are taken at the design points $\{\pi_{q(w)}: w = nm + 1, \cdots, (n + 1)m\}$. In other words, this set determines the matrix X_{n+1}, the ith entry of the second row being $\pi_{q(nm+i)}$. Then the (i, j)th entry of K_n, defined in (3.11), is given by

$$-[q(n) + 1]I_{ij} + \sum_{w=1}^{q(n)} \pi_w^{i+j-2}. \tag{4.4}$$

One can now calculate α_n linearly from (3.20) and a_n from (2.10), and the conclusions of Theorem 1 and Theorem 2 are valid for $r_n > 0$ satisfying (3.3).

Another example of a rational discrete design, practical because it allows for continuous model checking, takes observations along the lattice determined by the $2\ell + 1$ points $\Pi = \{\pm j\ell^{-1}: j = 0, 1, 2, \cdots, \ell\}$, the positive integer ℓ being a design parameter. Here $\Pi = \{\pi_w: w = 1, \cdots, 2\ell + 1\}$, where $\pi_w = w/\ell - 1$, determines $\mathscr{I}_n$ and K_n according to (4.3) and (4.4), and α_n by (3.20).

Note also that a D-optimal design, which places an equal number of observations at the set of $d + 1$ zeros of $(1 - x^2)\dot{P}_d(x)$ where P_d is the dth Legendre polynomial, has a rational discrete design measure, as do other designs with supports defined by roots of orthogonal polynomials.

EXAMPLE 3. Gleser (1965) first investigated the asymptotic properties of a spherical fixed-accuracy confidence set $\{\beta: |\beta_t - \beta| < d\}$, for $d > 0$. Here the corresponding accuracy set is $A = \{\beta: |\beta| < d\}$. This confidence set is contained in all the fixed-accuracy confidence sets for $c'\beta$ of Example 1 ($c \in \mathbb{R}^k$), with $(\delta, \varepsilon) = (-d, d)$ and $|c| = 1$. However, unlike Example 1, this sampling procedure uses a stopping rule t that is independent of c. Hence, for arbitrary c satisfying $|c| = 1$, $c'\beta_t$ will estimate $c'\beta$ to accuracy $(-d, d)$. Theorem 1 and the Corollary to Theorem 2 then extend Gleser's results by giving the second-order asymptotics corresponding to spherical accuracy.

EXAMPLE 4. The classical ellipsoidal confidence set

$$C_n = \{\beta: (\beta_n - \beta)'\mathscr{I}_n(\beta_n - \beta) < d\}$$

generates an approximate fixed-accuracy confidence set C_t by following the development of Section 2 with $F_n = F \sim d^{-1}\chi_k^2$, $a_n = a = F^{-1}(\gamma)$, and with t defined by (2.9). The assumptions of Theorem 2 then hold for $\alpha = 3/2$ if $k = 1$, for $\alpha = 1/2$ if $k = 3$, and for $\alpha = 0$ otherwise. Here $r_n = r > 0$ must exceed $(2m)^{-1}(4 + 2\nu + ad - k) = 1/2$ for second-order coverage probability of at least γ. The advantage of these ellipsoidal confidence sets is that the a_n's do not depend on the matrices X_i, $i = 1, 2, \cdots$, and hence Theorem 1 and Theorem 2 hold for arbitrary and possibly random matrices X_i. However, unless $\mathscr{I}_n = \mathscr{I}$ $\forall n$, the ellipsoidal confidence set C_t has approximate fixed accuracy only in that the accuracy set converges to a fixed ellipsoid. Note that the ellipsoid C_t is also "fixed" in the sense that each axis length is bounded by d.

Acknowledgment. Thanks to the referees for a very careful reading.

REFERENCES

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

ANSCOMBE, F. J. (1953). Sequential estimation. *J. Roy. Statist. Soc. Ser. B* **15** 1–21.

BROWN, R. L., DURBIN, J., and EVANS, M. (1975). Techniques for testing the constancy of regression relationships over time. *J. Roy. Statist. Soc. Ser. B* **37** 149–192.

CHOW, Y. S. and ROBBINS, H. (1965). On the asymptotic theory of fixed-width sequential confidence intervals for the mean. *Ann. Math. Statist.* **36** 457–462.

COX, D. R. (1952). Estimation by double sampling. *Biometrika* **39** 217–227.

FEDEROV, V. V. (1972). *Theory of Optimal Experiments*. Academic, New York.

FINSTER, M. (1983). A frequentistic approach to sequential estimation in the general linear model. *J. Amer. Statist. Assoc.* **57** 33–45.

GHOSH, M. and MUKHOPADHYAY, N. (1981). Consistency and asymptotic efficiency of two-stage and sequential estimation procedures. *Sankhya A* **43** 220–227.

GLESER, L. J. (1965). On the asymptotic theory of fixed-size sequential confidence bounds for linear regression parameters. *Ann. Math. Statist.* **40** 935–941.

HALL, P. (1981). Asymptotic theory of triple sampling for sequential estimation of the mean. *Ann. Statist.* **9** 1229–1238.

KARLIN, S. and STUDDEN, W. J. (1966). Optimal experimental designs. *Ann. Math. Statist.* **37** 783–815.

KIEFER, J. and STUDDEN, W. J. (1976). Optimal designs for large degree polynomial regression. *Ann. Statist.* **4** 1113–1123.

LAI, T. L. and SIEGMUND, D. (1977). A non-linear renewal theory with applications to sequential analysis I. *Ann. Statist.* **5** 946–954.

LAI, T. L. and SIEGMUND, D. (1979). A non-linear renewal theory with applications to sequential analysis II. *Ann. Statist.* **7** 60–76.

LEHMANN, E. L. (1959). *Testing Statistical Hypothesis*. Wiley, New York.

MARTINSEK, A. T. (1983). Second order approximation to the risk of a sequential procedure. *Ann. Statist.* **11** 827–836.

SIEGMUND, D. (1977). Repeated significance tests for a normal mean. *Biometrika* **64** 177–189.

SIEGMUND, D. (1978). Estimation following sequential testing. *Biometrika* **65** 341–349.

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

SIMONS, G. (1968). On the cost of not knowing the variance when making a fixed width confidence interval for the mean. *Ann. Math. Statist.* **39** 1946–1952.

SILVEY, S. D. (1980). *Optimal Designs*. Chapman, London.

STARR, N. (1966). The performance of a sequential procedure for the fixed-width interval estimation of the mean. *Ann. Math. Statist.* **37** 36–50.

STARR, N. and WOODROOFE, M. (1968). Remarks on a stopping time. *Proc. Nat. Acad. Sci. USA* **61** 1215–1218.

STARR, N. and WOODROOFE, M. (1969). Remarks on sequential point estimation. *Proc. Nat. Acad. Sci. USA* **63** 285–288.

STARR, N. and WOODROOFE, M. (1972). Further remarks on sequential point estimation: the exponential case. *Ann. Math. Statist.* **43** 1147–1154.

STEIN, C. (1945). A two-sample test for a linear hypothesis whose power is independent of the variance. *Ann. Math. Statist.* **16** 243–252.

STEIN, C. (1949). Some problems in sequential estimation (abstract). *Econometrica* **17** 77–78.

VARDI, Y. (1979). Asymptotic optimal sequential estimation: the Poisson case. *Ann. Statist.* **7** 1040–1051.

WEISS, L. and WOLFOWITZ, J. (1969). Maximum probability estimators. *Ann. Inst. Statist. Math.* **19** 193–206.

WEISS, L. and WOLFOWITZ, J. (1970). Maximum probability estimators. and asymptotic sufficiency. *Ann. Inst. Statist. Math.* **22** 225–244.

WEISS, L. and WOLFOWITZ, J. (1974). *Maximum Probability Estimators and Related Topics*. Springer-Verlag, New York.

WOLFOWITZ, J. (1975). Maximum probability estimators in the classical case and in the "almost smooth" case. *Theor. Probab. Appl.* **20** 363–371.

WOODROOFE, M. (1976). A renewal theorem for curved boundaries and moments of first passage times. *Ann. Probab.* **4** 67–80.

WOODROOFE, M. (1977). Second order approximations for sequential point and interval estimation. *Ann. Statist.* **5** 984–995.

WOODROOFE, M. (1978). Large deviations of the likelihood ratio statistic with applications to sequential testing. *Ann. Statist.* **6** 72–84.

WOODROOFE, M. (1979). Repeated likelihood ratio tests. *Biometrika* **66** 453–463.

WOODROOFE, M. (1982). *Nonlinear Renewal Theory in Sequential Analysis*. Soc. Indust. Appl. Math., Philadelphia.

DEPARTMENT OF STATISTICS
UNIVERSITY OF WISCONSIN
MADISON, WISCONSIN 53706