FIXED SIZE CONFIDENCE ELLIPSOIDS FOR LINEAR REGRESSION PARAMETERS[1]

By Arthur Albert

ARCON Corporation

TABLE OF CONTENTS

1. Introduction. In [2], Chow and Robbins developed an asymptotic theory of fixed width sequential confidence intervals for the mean of a univariate population. In [3], Gleser applied their stopping rule to the problem of constructing fixed size confidence sets for linear regression parameters. Gleser's confidence regions are spheres, centered at the least squares estimate of the regression parameter. In order that his method be valid, a strong assumption must be made concerning the large sample behavior of the least squares estimator's covariance matrix. Specifically, it is assumed in [3], that

$$\lim_{n\to\infty} n\Sigma_n = \Sigma \tag{1.1}$$

exists and is non-singular where Σ_n is the covariance matrix of the l.s.e. The stopping rule depends, in fact, upon the eigenvalues of Σ. The confidence regions proposed here are the more conventional ellipsoidal ones (whose kernels are proportional to Σ_n^{-1}), and while we are at it, we develop the theory to include estimable functions of the regression vector. Having done so, it is a simple task to adapt our results to the task of constructing sequential tests of the general linear hypothesis.

Our methods require that the least squares estimator for the regression parameter and its associated residual error be updated constantly as each datum is taken. To facilitate the attendant computations, we supply algorithms which allow these quantities to be computed iteratively "in real time" (as the data are collected).

In the last section, we exhibit formulae for the exact coverage probabilities in the case of normally distributed residuals.

Received 23 August 1965; revised 11 May 1966.

[1] This research was sponsored in part by the Office of Naval Research Contract No. Nonr 4401(00). Reproduction in whole or in part is permitted for any purpose of the United States Government.

2. Sequential confidence ellipsoids. Let $\{z(n)\}$ be a sequence of real valued random variables of the form

$$z(n) = h_n^t X + v(n) \tag{2.1a}$$

where the $v(n)$ are independent identically distributed random variables having a density with respect to Lebesgue measure with zero mean and common unknown variance σ^2. The row vectors, h_n^t, are the transpose of the known vectors, h_n, and X is an unknown m dimensional vector.

Let H_n be the $n \times m$ matrix whose row vectors are h_j^t $(j = 1, 2, \cdots, n)$, let Z_n be the n dimensional column vector whose components are $z(j)$ $(j = 1, 2, \cdots, n)$ and let $\hat{X}_n$ be the least squares estimator of X, based upon the observation vector Z_n:

$$\hat{X}_n = (H_n^t H_n)^{-1} H_n^t Z_n \qquad (n \geqq m). \tag{2.1b}$$

(We tacitly assume that the vectors $h_1, h_2, \cdots, h_m$, are linearly independent so that $H_n^t H_n = \sum_{j=1}^n h_j h_j^t$ is non-singular for $n \geqq m$.) Finally, denote the residual error associated with $\hat{X}_n$ by E_n:

$$E_n = \|Z_n - H_n \hat{X}_n\|^2. \tag{2.2}$$

We desire a $100\alpha\%$ confidence set of "fixed size" for the parameter

$$\psi = GX \tag{2.3}$$

where G is a $k \times m$ matrix of rank k.

We accomplish this as follows:

(1) Choose a so that

$$\Pr[\chi_k^2 \leqq a] = \alpha \tag{2.4}$$

where χ_k^2 is a central chi-square rv with k degrees of freedom.

(2) Let λ_n be any sequence of positive real numbers such that for some integer $c \geqq 1$,

$$\lambda_{c(n+2)} - \lambda_{c(n+1)} \geqq \lambda_{c(n+1)} - \lambda_{cn}, \tag{2.5}$$

$$\lambda_{n-1} \leqq \lambda_n \to \infty \quad \text{and} \quad \lambda_n/\lambda_{n+1} \to 1. \tag{2.6}$$

(3) Let a_n be any positive sequence such that $a_n \to a$.

(4) Let $N(d)$ be the rv defined by: $N(d)$ is the first value of $n \geqq m + 1$ such that

$$n^{-1} E_n \leqq d^2 \lambda_n / a_n. \tag{2.7}$$

[The assumption that the $v(n)$ are continuous rv's can be dropped if E_n is replaced by $(E_n + 1)$ in the stopping rule. (c.f. Chow & Robbins [2] eq. (7))].

(5) Let

$$R_n(d) = \{\xi: (\xi - G\hat{X}_n)^t [G(H_n^t H_n)^{-1} G^t]^{-1} (\xi - G\hat{X}_n) \leqq d^2 \lambda_n\}. \tag{2.8}$$

($G\hat{X}_n$ is the Gauss-Markov estimate of GX. $R_n(d)$ is an ellipsoid centered at $G\hat{X}_n$, whose kernel is the inverse of $G\hat{X}_n$'s covariance matrix.)

THE PROCEDURE. Take $N(d)$ observations (i.e., sample until (2.7) obtains) and use $R_{N(d)}(d)$ as the confidence region.

The "size" of $R_{N(d)}(d)$ can be interpreted variously as the length of its longest semi-axis or its volume. The size can be controlled by the choice of the sequence $\{\lambda_n\}$ and by d. In the next section we will study the effect of various sequences $\{\lambda_n\}$. For the present, assume that $\{\lambda_n\}$ satisfying (2.5) has been decided upon. We study the properties of the procedure as $d \to 0$. The theorems which follow depend upon four lemmas which, for the sake of continuity of exposition, are stated and proved at the end of this section.

THEOREM 2.1. *If* $\operatorname{tr}(H_n{}^t H_n)^{-1} \to 0$ *as* $n \to \infty$, *then*

$$\text{(a)} \qquad N(d) < \infty \qquad \text{wp } 1;$$

$$(2.9) \quad \text{(b)} \qquad \lim_{d\to 0} N(d) = \infty \qquad \text{wp } 1;$$

$$\text{(c)} \qquad \lim_{d\to 0} \lambda_{N(d)} d^2/a_{N(d)}\sigma^2 = 1 \qquad \text{wp } 1;$$

$$\text{(d)} \qquad \lim_{d\to 0} N(d)/g(a\sigma^2/d^2) = 1 \qquad \text{wp } 1;$$

where $g(t) = \max\{n : \lambda_n \leqq t\}$.

PROOF. Under the present assumptions,

$$E_n/n \to \sigma^2 \qquad \text{wp } 1.$$

(c.f. Gleser's correction note, [3].) There is therefore wp 1. a finite value of n such that $E_n/n \leqq d^2\lambda_n/a_n$, which proves (a). (b) and (c) follow from Lemma 1 of Chow and Robbins, [2], if we set

$$y_n = E_n/\sigma^2(n - m), \qquad f(n) = na\lambda_n/(n - m)a_n$$

and $t = a\sigma^2/d^2$ for then, $N(d)$ is the first value of $n \geqq m + 1$ such that $y_n \leqq f(n)/t$. Since $f(n) \geqq 0, f(n) \to \infty$ and $f(n)/f(n+1) \to 1$, their lemma obviously applies.

To establish (d), we build upon the argument of Chow and Robbins' Lemma 1. By definition of $N(d)$, $E_{N-1}/N - 1 > d^2\lambda_{N-1}/a_{N-1}$ and $E_N/N \leqq d^2\lambda_N/a_N$. Thus

$$(a_N/d^2)(E_N/N) \leqq \lambda_N \leqq (\lambda_N/L_{N-1})(a_{N-1}/d^2)(E_{N-1}/N - 1).$$

Since $g(\cdot)$ is non-decreasing,

$$g(a_N E_N/d^2 N) \leqq g(\lambda_N) \leqq g(\lambda_N a_{N-1} E_{N-1}/(d^2\lambda_{N-1})(N - 1)).$$

Since $\lambda_n{}^* = \lambda_{cn}$ possesses non-negative second differences and approaches ∞ as $n \to \infty$, there is a value of n_0 such that

$$\lambda_{nc+c} - \lambda_{nc} = \lambda_{n+1}^* - \lambda_n{}^* \geqq \lambda_{n_0+1}^* - \lambda_{n_0}^* > 0 \quad \text{if} \quad n \geqq n_0 .$$

Thus, if n is suitably large (say greater than n_1) $\lambda_{n+2c} > \lambda_n$. Since $g(\lambda_N)$ is the *largest* value of n such that $\lambda_n \leqq \lambda_N$, it follows that $g(\lambda_N) \geqq N$ if $N \geqq n_1$. On the other hand, $g(\lambda_N) \leqq N + 2c$ if $N \geqq n_1$ (otherwise $\lambda_{N+2c} \leqq \lambda_N$ which cannot be if $N > n_1$).

Since $N(d) \to \infty$ wp 1 as $d \to 0$,

$$\begin{aligned} &\lim_{d\to 0} [g(a_N E_N/d^2 N) - 2c]/g(a\sigma^2/d^2) \\ (2.10)\qquad &\leqq \lim_{d\to 0} N(d)/g(a\sigma^2/d^2) \\ &\leqq \lim_{d\to 0} [g(\lambda_N a_{N-1} E_{N-1}/(N-1)\, d^2 \lambda_{N-1})]/g(a\sigma^2/d^2) \qquad \text{wp } 1. \end{aligned}$$

Since $(E_N/N\sigma^2)$, (a_N/a) and $(\lambda_N/\lambda_{N-1})$ all approach unity as $d \to 0$, part (d) follows from (2.10) and Lemma 2.1 (c.f. 1609 the end of this section).

In what follows, we make use of the following abbreviations repeatedly:

If B is any matrix,

$$\lambda_{\max}(B) = \text{the largest eigenvalue of } B,$$

$$\lambda_{\min}(B) = \text{the smallest eigenvalue of } B,$$

$$\operatorname{tr}(B) = \text{the trace of } B,$$

$$\|B\| = [\lambda_{\max}(B^t B)]^{\frac{1}{2}}.$$

THEOREM 2.2. *In addition to the hypotheses of Theorem* 2.1, *assume that*

$$(2.11)\qquad \max_{1\leqq i\leqq n} \|h_i\|^2 \operatorname{tr}(H_n{}^t H_n)^{-1} \to 0,$$

$$\limsup_{n\to\infty} [\lambda_{\max}(H_n{}^t H_n)/\lambda_{\min}(H_n{}^t H_n)] < \infty$$

and

$$(2.12)\qquad \lim_{c\to 0} \sup_n |\{\textstyle\sum_{j=1}^{[n(1+c)]} \|h_j\|^2/\sum_{j=1}^{n} \|h_j\|^2\} - 1| = 0.$$

(*Here*, $[x]$ *is the largest integer in* x.) *Then*

$$\lim_{d\to 0} \Pr [GX \,\varepsilon\, R_{N(d)}(d)] = \alpha.$$

COMMENT. Assumptions (2.11) and (2.12) are not as complicated as they appear. The quantity $\lambda_{\max}(H_n{}^t H_n)/\lambda_{\min}(H_n{}^t H_n)$ is called the conditioning number of $H_n{}^t H_n$ and is a measure of how "close" this matrix is to being singular. If this number is large (but finite), $H_n{}^t H_n$ is non-singular, but the solution of the equation $H_n{}^t H_n x = y$ poses numerical difficulties which increase as the conditioning number increases. The second part of (2.11) merely places an upper bound on the sequence of conditioning numbers associated with the matrix sequence $\{H_n{}^t H_n\}$. The first part of (2.11) is equivalent to

$$\max_{1\leqq i\leqq n} \|h_i\|^2/\textstyle\sum_{j=1}^{n} \|h_j\|^2 \to 0$$

when the second part of (2.11) holds. This assumption places a limitation on the rate of growth of the sequence $\{\|h_n\|^2\}$. Finally, (2.12) requires that

$$\lim_{c\to 0} \textstyle\sum_{j=1}^{[n(1+c)]} \|h_j\|^2/\sum_{j=1}^{n} \|h_j\|^2 = 1$$

uniformly in n.

In his correction note, [3], Gleser relies on the assumptions that $n^{-1}H_n{}^t H_n$ converges to a positive definite matrix Σ, and that

$$\lim_n \max_{1 \leqq i \leqq n} \|h_i\|^2/n = 0.$$

It is not hard to see that these conditions imply conditions (2.11) and (2.12). In fact, the positive definiteness of Σ implies that its conditioning number is finite and it is clear that the conditioning numbers of $H_n{}^tH_n$ converge to Σ's. To prove the first part of (2.11), we notice that the second part of (2.11) implies that

$$\max_{1 \leqq i \leqq n} \|h_i\|^2 \operatorname{tr}(H_n{}^tH_n)^{-1} \leqq K \max_{1 \leqq i \leqq n} \|h_i\|^2/\sum_{j=1}^n \|h_j\|^2.$$

Under Gleser's assumptions,

$$\begin{aligned}\limsup_n [\max_{1 \leqq i \leqq n} \|h_i\|^2/\textstyle\sum_{j=1}^n \|h_j\|^2] &= \limsup_n [n^{-1} \max_{1 \leqq i \leqq n} \|h_i\|^2/n^{-1} \textstyle\sum_{j=1}^n \|h_j\|^2] \\ &= 0.\end{aligned}$$

To prove that Gleser's assumption implies (2.12), it suffices to show that
(a) $\lim_{c \to 0} \sup_n \sum_{j=1}^{[n(1+c)]} \|h_j\|^2/\sum_{j=1}^n \|h_j\|^2 \leqq 1$
and
(b) $\lim_{c \to 0} \inf_n \sum_{j=1}^{[n(1+c)]} \|h_j\|^2/\sum_{j=1}^n \|h_j\|^2 \geqq 1$:
Given $\epsilon > 0$, choose M so large that

$$(1 - \epsilon) \operatorname{tr}(\Sigma) \leqq n^{-1} \sum_{j=1}^n \|h_j\|^2 \leqq (1 + \epsilon) \operatorname{tr}(\Sigma)$$

and

$$(1 + c) \leqq [n(1 + c) + 1]/n \leqq (1 + \epsilon)(1 + c) \quad \text{for all} \quad |c| \leqq \tfrac{1}{2}$$

whenever $n \geqq M$. Since

$$\sup_n R_n = \max \{\sup_{n \leqq 2M} R_n, \sup_{n > 2M} R_n\},$$

we have

$$\begin{aligned}&\lim_{c \to 0} \sup_n \textstyle\sum_{j=1}^{[(1+c)n]} \|h_j\|^2/\sum_{j=1}^n \|h_j\|^2 \\ &\quad \leqq \lim_{c \to 0} \max \{1, \sup_{n > 2M} ([(1 + c)n]^{-1} \textstyle\sum_{j=1}^{[(1+c)n]} \|h_j\|^2/n^{-1} \sum_{j=1}^n \|h_j\|^2) \\ &\qquad \cdot (((1 + c)n + 1)n^{-1})\} \\ &\quad \leqq \lim_{c \to 0} \max \{1, (1 + \epsilon)^2(1 + c) \operatorname{tr}(\Sigma)/(1 - \epsilon)(\operatorname{tr}(\Sigma))\} \\ &\quad = (1 + \epsilon)^2/(1 - \epsilon).\end{aligned}$$

Since ϵ is arbitrary, (a) follows and (b) is proved in the same way.

Proof of Theorem 2.2. $Y_n = \sigma^{-2}U_n{}^tP_nU_n$ has an asymptotic χ^2 distribution with k df (where $U_n = (H_n{}^tH_n)^{\frac{1}{2}}(\hat{X}_n - X)$ has an asymptotic normal distribution if P_n is the projection onto the k dimensional sub space spanned by the columns of $(H_n{}^tH_n)^{-\frac{1}{2}}G^t$:

$$P_n = (H_n{}^tH_n)^{-\frac{1}{2}}G^t[G(H_n{}^tH_n)^{-1}G^t]^{-1}G(H_n{}^tH_n)^{-\frac{1}{2}}. \tag{2.13}$$

By virtue of Lemma 2.3 (Anscombe's theorem) and Theorem 2.1, $Y_{N(d)}$ converges in law to a χ_k^2 rv as $d \to 0$ if we can establish C2 of Lemma 2.3 for the sequence $\{Y_n\}$. Once this is done, Theorem 2.1c implies

$$\Pr [GX \varepsilon R_N] = \Pr [Y_{N(d)} \leqq d^2\lambda_{N(d)}/\sigma^2] \to \Pr [\chi_k^2 \leqq a] = \alpha \quad \text{as} \quad d \to 0.$$

To establish $C2$, it suffices to show that for any positive ϵ and η, there exist values of ν and c such that

$$(2.14) \qquad \sup_{n \geqq \nu} \Pr [\sup_{(1-c)n \leqq n' \leqq (1+c)n} \|U_n - U_{n'}\| \geqq \epsilon] \leqq \eta.$$

For then, if η and ϵ are given, we can, by virtue of U_n's asymptotic normality,
(1) choose ν and K so large that

$$(2.15) \qquad \Pr [\|U_n\| > K\epsilon^{\frac{1}{2}}] < \eta/6 \quad \text{for all} \quad n \geqq \nu,$$

and
(2) choose c so small that

$$(2.16) \quad \sup_{(1-c)n \leqq n' \leqq (1+c)n} \|P_{n'} - P_n\| \leqq \sigma^2/3K^2 \quad \text{for all } n \quad \text{(c.f. Lemma 2.4)},$$

and

$$(2.17) \quad \sup_{n \geqq \nu} \Pr [\sup_{(1-c)n \leqq n' \leqq (1+c)n} \|U_n - U_{n'}\| \geqq \min (\sigma^2/12K, \sigma/3)\epsilon^{\frac{1}{2}}] \leqq \eta/6.$$

It is true that

$$\begin{aligned} Y_n - Y_{n'} &= \sigma^{-2}[U_n{}^t P_n U_n - U_{n'}^t P_{n'} U_{n'}] \\ &= \sigma^{-2}\{(U_n - U_{n'})^t(P_n - P_{n'})(U_n - U_{n'}) \\ &\quad - (U_{n'} - U_n)^t P_n(U_{n'} - U_n) - U_n{}^t P_n(U_{n'} - U_n) \\ &\quad - U_n{}^t(P_{n'} - P_n)(U_{n'} - U_n) - U_n{}^t(P_{n'} - P_n) U_n \\ &\quad - (U_{n'} - U_n)^t P_{n'} U_n\}. \end{aligned}$$

Since P_n is a projection, $\|P_n\| = 1$, so that

$$\begin{aligned} &\sup_{(1-c)n \leqq n' \leqq n(1+c)} |Y_n - Y_{n'}| \\ &\quad \leqq \sigma^{-2}[\sup_{(1-c)n \leqq n' \leqq (1+c)n} \{3\|U_n - U_{n'}\|^2 + 4\|U_n\| \cdot \|U_{n'} - U_n\| \\ &\qquad + \|P_{n'} - P_n\| \cdot \|U_n\|^2\}]. \end{aligned}$$

$$\begin{aligned} &\Pr [\sup_{(1-c)n \leqq n' \leqq n(1+c)} |Y_n - Y_{n'}| \geqq \epsilon] \\ &\quad \leqq \Pr [\sup_{(1-c)n \leqq n' \leqq n(1+c)} 3\sigma^{-2}\|U_n - U_{n'}\|^2 \geqq \epsilon/3] \\ &\qquad + \Pr [\sup_{(1-c)n \leqq n' \leqq (1+c)n} 4\sigma^{-2}\|U_n\| \cdot \|U_n - U_{n'}\| \geqq \epsilon/3] \\ &\qquad + \Pr [\sup_{(1-c)n \leqq n' \leqq n(1+c)} \sigma^{-2}\|P_{n'} - P_n\| \cdot \|U_n\|^2 \geqq \epsilon/3] \\ &\quad = 1° + 2° + 3°. \end{aligned}$$

By (2.17),

$1° \leqq \eta/6 < \eta/3$ if $n \geqq \nu$,

$2° \leqq \Pr[\|U_n\| \geqq K\epsilon^{\frac{1}{2}}] + \Pr[\sup_{(1-c)n \leqq n' \leqq n(1+c)} 4\|U_n - U_{n'}\| \geqq \sigma^2\epsilon^{\frac{1}{2}}/3K]$

$\leqq \eta/3$ (by (2.15) and (2.17)) if $n \geqq \nu$,

and finally, if $n \geqq \nu$

$$3° \leqq \Pr[(\sigma^2/3K^2)\|U_n\|^2 \geqq \epsilon\sigma^2/3] = \Pr[\|U_n\| \geqq K\epsilon^{\frac{1}{2}}] \leqq \eta/3.$$

Thus

$$\sup_{n \geqq \nu} \Pr[\max_{(1-c)n \leqq n' \leqq (1+c)n} |Y_n - Y_{n'}| > \epsilon] < \eta$$

where establishes C2 of Anscombe's theorem. It remains to prove (2.14): Since

$$\begin{aligned} &\Pr[\max_{(1-c)n \leqq n' \leqq (1+c)n} \|U_n - U_{n'}\| > \epsilon] \\ (2.14a) \qquad &\leqq \Pr[\max_{n \leqq n' \leqq (1+c)n} \|U_n - U_{n'}\| > \epsilon] \\ &\quad + \Pr[\max_{(1-c)n \leqq n' \leqq n} \|U_n - U_{n'}\| > \epsilon], \end{aligned}$$

it suffices to show that for any positive ϵ and η, there exists a (large) value of ν and a (small) value of c such that both terms on the right side of (2.14a) are dominated by $\eta/2$ for all $n \geqq \nu$. We will show this in detail for the first term:

$$U_n = B_n^{\frac{1}{2}} \sum_{j=1}^{n} h_j v(j)$$

where $B_n = (H_n^t H_n)^{-1} = (\sum_{j=1}^{n} h_j h_j^t)^{-1}$. Therefore, if $n' > n$

$$U_{n'} - U_n = B_{n'}^{\frac{1}{2}} \sum_{j=n+1}^{n'} h_j v(j) + B_{n'}^{\frac{1}{2}}(B_n^{-\frac{1}{2}} - B_{n'}^{-\frac{1}{2}}) U_n$$

which in turn implies that

$$\begin{aligned} &\Pr[\max_{n \leqq n' \leqq (1+c)n} \|U_n - U_{n'}\| > \epsilon] \\ (2.18) \qquad &\leqq \Pr[\max_{n \leqq n' \leqq (1+c)n} \|B_{n'}\|^{\frac{1}{2}} \cdot \|\textstyle\sum_{j=n+1}^{n'} h_j v(j)\| \geqq \epsilon/2] \\ &\quad + \Pr[\max_{n \leqq n' \leqq (1+c)n} \|B_{n'}^{\frac{1}{2}}\| \cdot \|B_n^{-\frac{1}{2}} - B_{n'}^{-\frac{1}{2}}\| \cdot \|U_n\| \geqq \epsilon/2]. \end{aligned}$$

Since $\|B_n\| \geqq \|B_{n'}\|$ if $n \leqq n'$, the first term on the right side of (2.18) is dominated by

$$\Pr[\max_{n \leqq n' \leqq n(1+c)} \|\textstyle\sum_{j=n+1}^{n'} h_j v(j)\| \geqq \epsilon/2\|B_n^{\frac{1}{2}}\|]$$

which, in turn, is dominated by

$$(2.19) \qquad 4\sigma^2\epsilon^{-2}\|B_n^{\frac{1}{2}}\|^2 \sum_{j=n+1}^{[n(1+c)]} \|h_j\|^2$$

by virtue of the vector version of Kolmogorov's inequality (Lemma 2.3b). In turn,

$$\begin{aligned} \|B_n^{\frac{1}{2}}\|^2 &= \lambda_{\max}(B_n) \leqq \text{const} \times (\operatorname{tr} B_n^{-1})^{-1} \\ &= \text{const} \times (\textstyle\sum_{j=1}^{n} \|h_j\|^2)^{-1} \qquad \text{(by virtue of (2.11))}, \end{aligned}$$

so that (2.19) and hence the first term on the right side of (2.18) is dominated by

$$\text{const} \times [\textstyle\sum_{j=1}^{[n(1+c)]} \|h_j\|^2 / \sum_{j=1}^{n} \|h_j\|^2 - 1].$$

By virtue of (2.12), if η and ϵ are given, we can therefore choose c so that the first term on the right side of (2.18) is less than $\eta/4$ for all n. Now for the second term on the right side of (2.18):

$$\begin{aligned} \|B_{n'}^{\frac{1}{2}}\| \cdot \|B_n^{-\frac{1}{2}} - B_{n'}^{-\frac{1}{2}}\| &\leqq \|B_n\|^{\frac{1}{2}} \cdot \|(B_n^{-1})^{\frac{1}{2}} - (B_n^{-1} + \textstyle\sum_{j=n+1}^{n'} h_j h_j^t)^{\frac{1}{2}}\| \\ &= \|M_n^{\frac{1}{2}}[I - (I + S_{n,n'})^{\frac{1}{2}}]M_n^{\frac{1}{2}}\| \\ &\leqq \|M_n\|^{\frac{1}{2}} \|(I + S_{n,n'})^{\frac{1}{2}} - I\| \end{aligned}$$

where

$$(2.20) \qquad M_n = (\beta_n B_n^{-1}), \qquad S_{n,n'} = \beta_n M_n^{-\frac{1}{2}} \textstyle\sum_{j=n+1}^{n'} h_j h_j^t M_n^{-\frac{1}{2}}$$

and $\beta_n = \|B_n\|$. By (2.11), $\|\beta_n B_n^{-1}\|$ is uniformly bounded. Since $S_{n,n'}$ is non-negative definite and since we use the $\lambda_{\max}$-norm,

$$\|(I + S_{n,n'})^{\frac{1}{2}} - I\| = (1 + \|S_{n,n'}\|)^{\frac{1}{2}} - 1 \leqq \tfrac{1}{2}\|S_{n,n'}\|.$$

In turn,

$$\|S_{n,n'}\| \leqq \|M_n^{-1}\| \beta_n \textstyle\sum_{j=n+1}^{n'} \|h_j\|^2 = \beta_n \sum_{j=n+1}^{n'} \|h_j\|^2.$$

Since

$$\beta_n \textstyle\sum_{j=n+1}^{n'} \|h_j\|^2 \leqq \text{const} \times [\sum_{j=1}^{n'} \|h_j\|^2 / \sum_{j=1}^{n} \|h_j\|^2 - 1],$$

it follows from (2.12) that

$$\sup_n \max_{n \leqq n' \leqq n(1+c)} \|B_{n'}^{\frac{1}{2}}\| \cdot \|B_n^{-\frac{1}{2}} - B_{n'}^{-\frac{1}{2}}\| \to 0 \quad \text{as} \quad c \to 0.$$

Given η, choose K and ν so large that $\sup_{n \geqq \nu} \Pr[\|U_n\| \geqq K] < \eta/4$. Choose c so small that

$$\sup_n \max_{n \leqq n' \leqq n(1+c)} \|B_{n'}^{\frac{1}{2}}\| \cdot \|B_n^{-\frac{1}{2}} - B_{n'}^{-\frac{1}{2}}\| < \epsilon/2K.$$

Then

$$\begin{aligned} \sup_{n \geqq \nu} \Pr[\max_{n \leqq n' \leqq n(1+c)} \|B_{n'}^{\frac{1}{2}}\| \cdot \|(B_n^{-\frac{1}{2}} - B_{n'}^{-\frac{1}{2}}) U_n\| \geqq \epsilon/2] \\ \leqq \sup_{n \geqq \nu} \Pr[\|U_n\| \geqq K] \leqq \eta/4. \end{aligned}$$

This shows that

$$\sup_{n \geqq \nu} \Pr[\max_{n \leqq n' \leqq n(1+c)} \|U_n - U_{n'}\| > \epsilon] < \eta/2.$$

The other term on the right side of (2.14a) is treated analogously. This completes our proof.

The following lemmas were used to establish Theorems 2.1 and 2.2:

LEMMA 2.1. *Let* $\{\lambda_n\}$ *be a sequence of positive scalars such that for some integer* $c \geqq 1$,

(a) $\lambda_{n-1} \leqq \lambda_n \to \infty$,
(b) $\lambda_{c(n+2)} - \lambda_{c(n+1)} \geqq \lambda_{c(n+1)} - \lambda_{cn}$
for all n.

Let $g(A) = \max\{n: \lambda_n \leqq A\}$, $(A \geqq \lambda_1)$. *Then*

$$\lim_{A\to\infty} g(Ay(A))/g(A) = 1 \quad \text{if} \quad \lim_{A\to\infty} y(A) = 1.$$

Proof. Let $\varphi(t)$ be the continuous, piecewise linear function taking the value λ_{cn} when $t = n$ $(n = 1, 2, \cdots)$. By virtue of assumptions (a) and (b), $\varphi(\cdot)$ is convex and approaches ∞ as $t \to \infty$ so that $\varphi(\cdot)$ is strictly increasing beyond some integer t_0. Let $\epsilon > 0$ be given. Choose A_0 so large that

$$(1 - \epsilon) \leqq y(A) \leqq (1 + \epsilon)$$

and

$$g((1 - \epsilon)A) > t_0 \qquad \text{if } A > A_0.$$

Since $\varphi(\cdot)$ is convex, its graph lies above the line of support through $(t, \varphi(t))$ with slope $\dot{\varphi}_R(t)(\dot{\varphi}_R(t)$ is the value of φ's right hand derivative at t.) In particular, if

$$\varphi(t_1) = A(1 - \epsilon) \quad \text{and} \quad \varphi(t_2) = A(1 + \epsilon)$$

then

$$0 \leqq (t_2 - t_1)\dot{\varphi}_R(t_1) \leqq \varphi(t_2) - \varphi(t_1) = (2\epsilon/(1 - \epsilon))\varphi(t_1).$$

Hence,

$$\begin{aligned} 0 \leqq t_2/t_1 - 1 &\leqq (2\epsilon/(1 - \epsilon))[\varphi(t_1)/(\varphi(t_1) - \lambda_c)][(t_1 - 1)/t_1] \\ &\quad \cdot [(\varphi(t_1) - \varphi(1))/(t_1 - 1))/\dot{\varphi}_R(t_1) \\ &\leqq (2\epsilon/(1 - \epsilon))[\varphi(t_1)/(\varphi(t_1) - \lambda_c)][(t_1 - 1)/(t_1)], \end{aligned}$$

since the slope of the secant line connecting the point $(1, \varphi(1))$ with $(t_1, \varphi(t_1))$ is no greater than the slope of the tangent at $(t_1, \varphi(t_1))$.

But $\varphi(t_1) = (1 - \epsilon)A$ so that $\varphi([t_1]) = \lambda_{c[t_1]} \leqq (1 - \epsilon)A = \varphi(t_1)$. Hence $c(t_1 - 1) \leqq c[t_1] \leqq g((1 - \epsilon)A)$. Similarly, $(1 + \epsilon)A = \varphi(t_2) \leqq \varphi([t_2] + 1) = \lambda_{c[t_2]+1}$ so that $ct_2 + 1 \geqq c[t_2] + 1 \geqq g((1 + \epsilon)A)$. Thus, if $A > A_0$,

$$\begin{aligned} &[g((1 + \epsilon)A) - 1]/[g((1 - \epsilon)A) + c] \\ &\qquad \leqq t_2/t_1 \leqq 1 + (2\epsilon/(1 - \epsilon))[\varphi(t_1)/(\varphi(t_1) - \lambda_c)][(t_1 - 1)/t_1]. \end{aligned}$$

Since t_1, $\varphi(t_1)$, $g((1 + \epsilon)A)$ and $g((1 - \epsilon)A)$ all go to ∞ as $A \to \infty$, we see that

$$g((1 + \epsilon)A)/g((1 - \epsilon)A)(1 + o(1)) \leqq 1 + (2\epsilon/(1 - \epsilon))(1 + o(1)) \qquad \text{as} \quad A \to \infty,$$

which implies that

$$g((1 + \epsilon)A)/g((1 - \epsilon)A) \leqq 1 + (2\epsilon/(1 - \epsilon))(1 + o(1)) \quad \text{as} \quad A \to \infty.$$

Since $g(\cdot)$ is non-decreasing,

$$[g((1+\epsilon)A)/g((1-\epsilon)A)]^{-1} \leqq g(Ay(A))/g(A) \leqq g((1+\epsilon)A)/g((1-\epsilon)A) \quad \text{if } A > A_0.$$

Therefore, for any $\epsilon > 0$,

$$\limsup_{A\to\infty} g(Ay(A))/g(A) \leqq 1 + 2\epsilon$$

and

$$\liminf_{A\to\infty} g(Ay(A))/g(A) \geqq (1+2\epsilon)^{-1}.$$

The conclusion follows.

LEMMA 2.2. *If* $\max_{1\leqq i\leqq n} \|h_i\|^2 \operatorname{tr} (H_n{}^tH_n)^{-1} \to 0$, *and if* T_n *is any sequence of* $m \times m$ *orthogonal matrices, then*

(a) $T_nU_n = T_n(H_n{}^tH_n)^{\frac{1}{2}}(\hat{X}_n - X)$ *converges in law to an* m *dimensional vector random variable whose distribution is* $N(0, \sigma^2 I)$.

(b) *If* P_n *is any sequence of projections with* k *dimensional ranges, then*

$$Y_n = \sigma^{-2}U_n{}^tP_nU_n$$

converges in law to a chi-square random variable with k *degrees of freedom.*

PROOF. We prove (a) by a trivial extension of Corollary 3.2 in [3]: Let ξ be any m vector. Then

$$\mathcal{E} \exp i\xi^tT_nU_n = \mathcal{E} \exp i\|\xi\| \textstyle\sum_{i=1}^n b_{ni}v(i)$$

where $v(i) = z(i) - h_i{}^tX$ is the residual associated with the ith observation and

$$b_{ni} = \|\xi\|^{-1}(T_n{}^t\xi)^t(H_n{}^tH_n)^{-\frac{1}{2}}h_i.$$

Since

$$\begin{aligned}\textstyle\sum_{i=1}^n b_{ni}^2 &= \|\xi\|^{-2}(T_n{}^t\xi)^t(H_n{}^tH_n)^{-\frac{1}{2}}(\textstyle\sum_{i=1}^n h_ih_i{}^t)(H_n{}^tH_n)^{-\frac{1}{2}}(T_n{}^t\xi)\\ &= \|T_n{}^t\xi\|^2/\|\xi\|^2 = 1\end{aligned}$$

and

$$\begin{aligned}|b_{ni}|^2 &\leqq \|(H_n{}^tH_n)^{-\frac{1}{2}}\|^2\|h_i\|^2 \leqq \|h_i\|^2\lambda_{\max}(H_n{}^tH_n)^{-1}\\ &\leqq \|h_i\|^2 \operatorname{tr}(H_n{}^tH_n)^{-1},\end{aligned}$$

the hypothesis of the lemma guarantees that $\max_{1\leqq i\leqq n} b_{ni}^2 \to 0$. Theorem 3, p. 103 of [4] applies: $\sigma^{-2}\sum_{i=1}^n b_{ni}v(i) \to_L N(0, 1)$, hence

$$\exp i\xi^tT_nU_n = \exp i\|\xi\| \textstyle\sum_{i=1}^n b_{ni}v(i) \to \exp - (\sigma^2\|\xi\|^2/2)$$

which in turn implies that $T_nU_n \to_L N(0, \sigma^2 I)$.

(b) follows immediately from the fact that there is a sequence of orthogonal matrices, T_n, which reduce each P_n to the *same* diagonal matrix, D, having k

ones and $m - k$ zeros on its diagonal: $T_n P_n T_n^t = D; n = 1, 2, \cdots$. Thus

$$Y_n = \sigma^{-2} U_n^t P_n U_n = \sigma^{-2} (T_n U_n)^t D (T_n U_n).$$

Since $T_n U_n \rightarrow_L N(0, \sigma^2 I)$, Y_n must therefore converge in law to χ_k^2.

LEMMA 2.3. (a) *Let* $\{N(t); t \geqq 0\}$ *be an integer valued stochastic process and suppose* $g(\cdot)$ *is a non-decreasing integer valued function which approaches* $+\infty$ *as* $t \rightarrow \infty$, *having the further property that* $N(t)/g(t) \rightarrow 1$ *in probability. If* $\{Y_n; n = 1, 2, \cdots\}$ *is a stochastic process satisfying condition* C2 *below, and if* $\lim_{n\to\infty} \Pr[Y_n \leqq x] = F(x)$ *at all continuity points of* F, *then*

$$\lim_{t\to\infty} \Pr[Y_{N(t)} \leqq x] = F(x) \quad \text{at all continuity points of } F.$$

C2: *For any positive* ϵ *and* η *there is a (large) value of* ν *and a (small) value of* c *such that*

$$\sup_{n \geqq \nu} \Pr\{\max_{(1-c)n \leqq n' \leqq (1+c)n} |Y_n - Y_{n'}| > \epsilon\} < \eta.$$

(b) *If* $\{V(n); n = 1, 2, \cdots\}$ *is an independent vector process, then*

$$\Pr[\max_{k \leqq n} \|\textstyle\sum_{j=1}^k V_j\| > \epsilon] \leqq \epsilon^{-2} \sum_{j=1}^n \mathcal{E}\|V_j\|^2.$$

PROOF. (a) This is a special case of Anscombe's theorem 1 [1].

(b) This is an easy generalization of Kolmogorov's inequality. The proof is obtained by substituting norms and inner products for absolute values and products in Loève's proof, pg. 235 of [5].

LEMMA 2.4. *Let* $\{h_1, h_2, \cdots\}$ *be a sequence of* m *dimensional column vectors such that the matrix* $\sum_{j=1}^{n_0} h_j h_j^t$ *has rank* m *for some* n_0. *If*

$$\limsup_n \lambda_{\max}(\textstyle\sum_{j=1}^n h_j h_j^t)/\lambda_{\min}(\sum_{j=1}^n h_j h_j^t) < \infty$$

and

$$\lim_{c\to 0} [\sup_n |(\textstyle\sum_{j=1}^{[n(1+c)]} \|h_j\|^2 / \sum_{j=1}^n \|h_j\|^2) - 1|] = 0$$

then

$$\lim_{c\to 0} [\sup_{n > n_0} \max_{(1-c)n \leqq n' \leqq (1+c)n} \|P_n - P_{n'}\|] = 0$$

where

$$P_n = B_n^{\frac{1}{2}} G^t (G B_n G^t)^{-1} G B_n^{\frac{1}{2}},$$
$$B_n = (\textstyle\sum_{j=1}^n h_j h_j^t)^{-1}$$

and G *is any* $k \times m$ *matrix of rank* k.

PROOF. Let $C_n = B_n^{\frac{1}{2}} G^t$ and $E_n = C_n^t C_n$.

Then

$$\begin{aligned} P_n - P_{n'} &= C_n E_n^{-1} C_n^t - C_{n'} E_{n'}^{-1} C_{n'}^t \\ &= (C_n - C_{n'})[E_n^{-1}(E_{n'} - E_n) E_{n'}^{-1}](C_n - C_{n'})^t \\ &\quad + C_{n'} E_n^{-1} (C_n - C_{n'})^t + C_n [E_n^{-1}(E_{n'} - E_n) E_{n'}^{-1}] C_{n'}^t \\ &\quad + (C_n - C_{n'}) E_{n'}^{-1} C_n^t. \end{aligned} \tag{2.21}$$

Let $b_n = (\sum_{j=1}^n \|h_j\|^2)^{\frac{1}{2}}$. Then if $n \leqq n'$:

$$\|C_n\|^2 \leqq \|G\|^2\|B_n\| \leqq m_1 b_n^{-2},$$

$$\text{(2.22)} \quad \|C_n - C_{n'}\| \leqq \|G\|\cdot\|B_n^{\frac{1}{2}} - B_{n'}^{\frac{1}{2}}\| = \|G\|\cdot\|B_n^{\frac{1}{2}}[I - (I + S_{n,n'}^*)^{-\frac{1}{2}}]B_n^{\frac{1}{2}}\| \leqq \|G\|\cdot\|B_n\|^{\frac{1}{2}}\|I - (I + S_{n,n'}^*)^{-\frac{1}{2}}\|,$$

where $S_{n,n'}^* = (B_n^{\frac{1}{2}}\sum_{j=n+1}^{n'} h_j h_j^t B_n^{\frac{1}{2}})$ is nonnegative definite. Therefore, since we use the $\lambda_{\max}$-norm,

$$\|I - (I + S_{n,n'}^*)^{-\frac{1}{2}}\| = 1 - (1 + \|S_{n,n'}^*\|)^{-\frac{1}{2}} \leqq \|S_{n,n'}^*\|.$$

But

$$\|S_{n,n'}^*\| \leqq \|B_n\| \sum_{j=n+1}^{n'} \|h_j\|^2 \leqq m_2 b_n^{-2}(b_{n'}^2 - b_n^2)$$

so that

$$\text{(2.23)} \quad \|C_n - C_{n'}\| \leqq m_3 b_n^{-3}(b_{n'}^2 - b_n^2).$$

$$\text{(2.24)} \quad \|E_n^{-1}\| = (\lambda_{\min} E_n)^{-1} \leqq m_4(\lambda_{\min} B_n)^{-1} \leqq m_5 \lambda_{\max}(B_n^{-1}) \leqq m_5 b_n^2$$

and

$$\text{(2.25)} \quad \begin{aligned} \|E_n - E_{n'}\| &= \|G(B_n - B_{n'})G^t\| \\ &\leqq \|G\|^2\|B_{n'}\|\cdot\|B_{n'}^{-1}B_n - I\| \\ &\leqq m_6 b_{n'}^{-2}(b_{n'}^2 - b_n^2)b_n^{-2}. \end{aligned}$$

By assumption,

$$\text{(2.26)} \quad \lim_{c\to 0} \sup_n \max_{(1-c)n \leqq n' \leqq n(1+c)} |b_{n'}/b_n - 1| = 0.$$

The conclusion of Lemma 2.4 follows when (2.21)–(2.26) are combined in the obvious way.

3. Particular choices of the stopping sequence $\{\lambda_n\}$. The stopping rule depends upon d, $\{\lambda_n\}$ and $\{a_n\}$. The "size" of the confidence ellipsoid depends upon $\{\lambda_n\}$. The exact confidence level, α, depends upon d, $\{\lambda_n\}$ and $\{a_n\}$, whereas the asymptotic level depends only on $\lim_n a_n$. In this section, we will fix $d > 0$ and the sequence $\{a_n\}$, and examine the effects on the size of the confidence ellipsoid that are exerted by various choices of $\{\lambda_n\}$:

The volume of the ellipsoid $R_n(d)$ (c.f. (2.8)) is equal to a constant times the product of the lengths of its semi-axis. We choose our units so that this constant is unity. Then the volume of $R_n(d)$ is equal to the square root of the product of the eigenvalues of $d^2\lambda_n G(H_n{}^t H_n)^{-1}G^t$:

$$\text{Volume of } R_n(d) = (d^2\lambda_n)^{k/2}\Delta_n^{\frac{1}{2}}$$

where Δ_n is the determinant of $G(H_n{}^t H_n)^{-1}G^t$.

Thus, if we choose

$$\lambda_n = d^{-2}[V^2/\Delta_n]^{1/k},$$

the volume of $R_n(d)$ will be equal to V for every n.

On the other hand, the length of $R_n(d)$'s longest semi-axis is equal to the square root of the largest eigenvalue of $d^2\lambda_n[G(H_n{}^tH_n)^{-1}G^t]$.

Thus, if we choose λ_n = the smallest eigenvalue of $[G(H_n{}^tH_n)^{-1}G^t]^{-1}$, the longest semi-axis of $R_n(d)$ will have length d.

From the computational point of view, neither of these choices of λ_n are de-desirable, since λ_n must be recomputed after every sample. If, however, we allow ourselves to lose a bit of control over the size of $R_n(d)$ (say by only requiring that the longest semi-axis of $R_n(d)$ be no longer than d) then

$$\lambda_n{}^\circ = [\text{tr } G(H_n{}^tH_n)^{-1}G^t]^{-1}$$

will do the trick. In the next section we exhibit a convenient algorithm which relates λ_{n+1}° to $\lambda_n{}^\circ$.

It only remains to state the conditions under which $\lambda_n \uparrow \infty$ and $\lambda_n/\lambda_{n+1} \to 1$. For notational convenience we

DEFINE.

$$\Sigma_n = G(H_n{}^tH_n)^{-1}G^t. \tag{3.1}$$

THEOREM 3.1. *If*

$$\max\nolimits_{1 \leq i \leq n} \|h_i\|^2 \text{ tr } (H_n{}^tH_n)^{-1} \to 0$$

and

$$\limsup\nolimits_n \lambda_{\max}(H_n{}^tH_n)/\lambda_{\min}(H_n{}^tH_n) < \infty$$

then $\lambda_n \uparrow \infty$ *and* $\lambda_n/\lambda_{n+1} \to 1$ *for the following choices of* λ_n :

(3.2) (a) $$\lambda_n = (\det \Sigma_n)^{-1/k},$$

(3.3) (b) $$\lambda_n = \lambda_{\min}(\Sigma_n^{-1}),$$

(3.4) (c) $$\lambda_n = (\text{tr } \Sigma_n)^{-1}.$$

For choice (a), *the volume of* $R_{N(d)}$ *is* d^k. *For choice* (b), *the length of* $R_{N(d)}$*'s longest semi-axis is d. For choice* (c), *the length of* $R_{N(d)}$*'s longest semi-axis is no greater than d.*

COMMENT. In order to apply Theorem 2.2 (i.e., to establish that $\Pr[GX \varepsilon R_{N(d)}] \to \alpha$), it is also necessary that $\{\lambda_n\}$ satisfy the "convexity" condition of (2.5) (that λ_{cn}'s second differences are non-negative for some integer c). We will not pursue this question but rather, carry it as an assumption where needed.

PROOF OF THEOREM 3.1. We begin by proving that all three choices of λ_n tend to ∞. To prove this, it suffices to show that

$$\lambda_{\max}[G(H_n{}^tH_n)^{-1}G^t] \to 0. \tag{3.5}$$

The celebrated Courant-Fischer characterization of eigenvalues tells us that for any unit vector u,

$$\begin{aligned} u^tG(H_n{}^tH_n)^{-1}G^tu &\leqq \lambda_{\max}(H_n{}^tH_n)^{-1}\|G^tu\|^2 \\ &\leqq \lambda_{\max}(H_n{}^tH_n)^{-1}\lambda_{\max}(GG^t) \\ &\leqq \text{tr } (H_n{}^tH_n)^{-1}\lambda_{\max}(GG^t). \end{aligned}$$

By assumption, the right hand side approaches zero. Thus,

$$\lambda_{\max}[G(H_n{}^tH_n)^{-1}G^t] = \sup_{\|u\|=1} u^tG(H_n{}^tH_n)^{-1}G^tu \to 0.$$

To show that $\lambda_n/\lambda_{n+1} \to 1$, we will use the following notation repeatedly:

$$(3.6) \qquad B_n = (H_n{}^tH_n)^{-1}, \qquad n \geqq m.$$

Since

$$(3.7) \qquad H^t_{n+1}H_{n+1} = H_n{}^tH_n + h_{n+1}h^t_{n+1}$$

it is easy to verify that

$$(3.8) \qquad B_{n+1} = B_n - (B_nh_{n+1})(B_nh_{n+1})^t/(1 + h^t_{n+1}B_nh_{n+1}).$$

(Multiply the right side of (3.8) by the right side of (3.7). The identity matrix results after some algebra.) We also will make repeated use of the fact that

$$(3.9) \qquad \|h_{n+1}\|^2\lambda_{\max}(B_n) \to 0.$$

To see this, we notice that $\lambda_{\min}(H_n{}^tH_n) \geqq c\lambda_{\max}(H_n{}^tH_n)$ (by assumption) and $\lambda_{\max}(H_n{}^tH_n) \geqq k^{-1}$ tr $(H_n{}^tH_n)$. Thus,

$$\begin{aligned} \lambda_{\max}(B_n) &= \lambda^{-1}_{\min}(H_n{}^tH_n) \\ &\leqq c'/\text{tr } (H_n{}^tH_n). \end{aligned}$$

Since $H_n{}^tH_n = H^t_{n+1}H_{n+1} - h_{n+1}h^t_{n+1}$ and since tr $h_{n+1}h^t_{n+1} = \|h_{n+1}\|^2$, we have

$$\begin{aligned} \|h_{n+1}\|^2\lambda_{\max}(B_n) &\leqq c'\|h_{n+1}\|^2/[\text{tr } H^t_{n+1}H_{n+1} - \|h_{n+1}\|^2] \\ &\leqq c'\|h_{n+1}\|^2/[\lambda_{\min}B^{-1}_{n+1} - \|h_{n+1}\|^2] \\ &= c'[(\|h_{n+1}\|^2\lambda_{\max}B_{n+1})^{-1} - 1]^{-1} \\ &\leqq c'[(\|h_{n+1}\|^2 \text{ tr } B_{n+1})^{-1} - 1]^{-1}. \end{aligned}$$

Since $\|h_{n+1}\|^2$ tr $B_{n+1} \to 0$ by assumption, (3.9) follows.

Pre- and postmultiplying (3.8) by G and G^t we find that

$$\begin{aligned} (3.10) \qquad GB_{n+1}G^t = (GB_nG^t)^{\frac{1}{2}}[I &- (GB_nG^t)^{-\frac{1}{2}}(GB_nh_{n+1})(GB_nh_{n+1})^t \\ &\cdot (GB_nG^t)^{-\frac{1}{2}}(1 + h^t_{n+1}B_nh_{n+1})^{-1}](GB_nG^t)^{\frac{1}{2}}. \end{aligned}$$

Letting $\lambda_{n+1} = [\det GB_{n+1}G^t]^{-1/k}$, we find that

$$(3.11)\quad (\lambda_n/\lambda_{n+1})^k = \det [I - \Sigma_n^{-\frac{1}{2}}(GB_nh_{n+1})(GB_nh_{n+1})^t\Sigma_n^{-\frac{1}{2}} \cdot(1 + h_{n+1}^tB_nh_{n+1})^{-1}].$$

To show that $\lambda_n/\lambda_{n+1} \to 1$, it suffices to show that the trace of the second term in the square brackets approaches zero:

$$(3.12)\quad \begin{aligned} &\operatorname{tr}(\Sigma_n^{-\frac{1}{2}}(GB_nh_{n+1})(GB_nh_{n+1})^t\Sigma_n^{-\frac{1}{2}}(1 + h_{n+1}^tB_nh_{n+1})^{-1}) \\ &\qquad = h_{n+1}^tB_nG^t\Sigma_n^{-1}GB_nh_{n+1}(1 + h_{n+1}^tB_nh_{n+1})^{-1} \\ &\qquad \leqq \|h_{n+1}\|^2\lambda_{\max}(B_n^2)\lambda_{\max}(G^tG)\lambda_{\max}(\Sigma_n^{-1}). \end{aligned}$$

But

$$\begin{aligned} \lambda_{\max}(\Sigma_n^{-1}) &= 1/\lambda_{\min}(\Sigma_n) \\ &\leqq 1/\lambda_{\min}(GG^t)\lambda_{\min}(B_n) \\ &\leqq 1/c\lambda_{\min}(GG^t)\lambda_{\max}(B_n). \end{aligned}$$

Thus, the right side of (3.12) is less than or equal to a constant times $\|h_{n+1}\|^2\lambda_{\max}(B_n)$ (since GG^t is non-singular, its smallest eigenvalue is positive). The conclusion follows by (3.9).

Let us now examine the case where

$$\lambda_n = \lambda_{\min}(\Sigma_n^{-1}) = 1/\lambda_{\max}(\Sigma_n).$$

If A and C are non-negative definite and $\|x\| = 1$, then $x^tAx \geqq x^tAx - x^tCx \geqq x^tAx - \lambda_{\max}(C)$. Taking supremums over all such x's,

$$(3.13)\quad \lambda_{\max}(A) \geqq \lambda_{\max}(A - C) \geqq \lambda_{\max}(A) - \lambda_{\max}(C).$$

Since

$$(3.14)\quad \begin{aligned} \lambda_n/\lambda_{n+1} &= \lambda_n\lambda_{\max}(\Sigma_{n+1}) \\ &= \lambda_n\lambda_{\max}(\Sigma_n - (GB_nh_{n+1})(GB_nh_{n+1})^t/(1 + h_{n+1}^tB_nh_{n+1}) \end{aligned}$$

and since $\lambda_n\lambda_{\max}(\Sigma_n) = 1$, it therefore suffices to show that

$$(3.15)\quad \begin{aligned} &\lambda_n\lambda_{\max}((GB_nh_{n+1})(GB_nh_{n+1})^t/(1 + h_{n+1}^tB_nh_{n+1})) \\ &\qquad \equiv \lambda_nh_{n+1}^tB_nG^tGB_nh_{n+1}/(1 + h_{n+1}^tB_nh_{n+1}) = o(1) \quad \text{as} \quad n \to \infty. \end{aligned}$$

(We use the fact that $\lambda_{\max}(xx^t) = x^tx$ if x is a vector.)
But

$$h_{n+1}^tB_nG^tGB_nh_{n+1}(1 + h_{n+1}^tB_nh_{n+1})^{-1} \leqq \|h_{n+1}\|^2\lambda_{\max}(B_n)\operatorname{tr}(B_n)\lambda_{\max}(G^tG)$$

and

$$\lambda_n = [\lambda_{\max}(\Sigma_n)]^{-1} \leqq [\lambda_{\min}(\Sigma_n)]^{-1} \leqq [\lambda_{\min}(GG^t)]^{-1}[\lambda_{\min}(B_n)]^{-1}.$$

By assumption

$$[\lambda_{\max}(B_n)/\lambda_{\min}(B_n)]\|h_{n+1}\|^2 \operatorname{tr} B_n \to 0 \quad \text{as} \quad n \to \infty$$

and this proves (3.15).

Next suppose $\lambda_n = (\operatorname{tr} \Sigma_n)^{-1}$. Then

$$\begin{aligned}\lambda_{n+1}^{-1} &= \operatorname{tr} GB_{n+1}G^t \\ &= \operatorname{tr} GB_nG^t - \operatorname{tr} (GB_nh_{n+1})(GB_nh_{n+1})^t(1 + h_{n+1}^tB_nh_{n+1})^{-1} \\ &= \lambda_n^{-1} + O(\|h_{n+1}\|^2\lambda_{\max}(B_n^2)).\end{aligned}$$

Since

$$\lambda_n^{-1} \geqq \lambda_{\max}(\Sigma_n) \geqq \lambda_{\min}(GG^t)\lambda_{\min}(B_n) \geqq c\lambda_{\max}(B_n),$$

it follows that

$$\lambda_n/\lambda_{n+1} = 1 + O(\|h_{n+1}\|^2\lambda_{\max}(B_n)).$$

Again, apply (3.9).

To show that the three choices of λ_n are monotone: By (3.8),

$$\begin{aligned}\det \Sigma_{n+1} = \det GB_{n+1}G^t &= \det [GB_nG^t - (GB_nh_{n+1})(GB_nh_{n+1})^t(1 + h_{n+1}^tB_nh_{n+1})^{-1}] \\ &= [\det GB_nG^t] \det [I - (GB_nG^t)^{-1}\eta\eta^t]\end{aligned}$$

where $\eta = GB_nh_{n+1}/(1 + h_{n+1}^tB_nh_{n+1})^{\frac{1}{2}}$. The eigenvalues of $I - (GB_nG^t)^{-1}\eta\eta^t$ are unity (with multiplicity $k - 1$) and $1 - \eta^t(GB_nG^t)^{-1}\eta$ (with multiplicity 1). Thus

$$\det \Sigma_{n+1} = [1 - \eta^t(GB_nG^t)^{-1}\eta] \det \Sigma_n \leqq \det \Sigma_n$$

(since $(GB_nG^t)^{-1}$ is non-negative definite). This proves the monotonicity of (a).

Since $\Sigma_{n+1} = GB_{n+1}G^t = \Sigma_n - \eta\eta^t$, $\lambda_{\max}\Sigma_{n+1} \leqq \lambda_{\max}\Sigma_n$. The monotonicity of (b) follows.

Since $\operatorname{tr} \Sigma_{n+1} = \operatorname{tr} \Sigma_n - \eta^t\eta \leqq \operatorname{tr} \Sigma_n$, the monotonicity of (c) follows.

4. Computational considerations. The stopping rule associated with the sequential confidence set $R_{N(d)}(d)$ requires that the least squares residual, E_n, be computed after each datum is collected, and compared with the threshold

$$nd^2\lambda_n/a_n,$$

which itself must either be computed in "real time," or else stored in memory for every conceivable value of n.

When sampling terminates, the confidence ellipsoid, centered at $G\hat{X}_N$ with kernel $[d^2\lambda_NG(H_N^tH_N)^{-1}G^t]^{-1}$ must be constructed.

In the next theorem, we exhibit convenient recursions which relate the $n + 1$st value of these quantities to the nth, for a particular choice of the norming sequence $\{\lambda_n\}$:

THEOREM 4.1. *Let*

$$B_n = (H_n^t H_n)^{-1}, \tag{4.1}$$

$$C_n = [G(H_n^t H_n)^{-1} G^t]^{-1}, \tag{4.2}$$

$$\hat{X}_n = B_n H_n^t Z_n, \tag{4.3}$$

$$E_n = \|Z_n - H_n \hat{X}_n\|^2, \tag{4.4}$$

and

$$\lambda_n = \{\operatorname{tr} G(H_n^t H_n)^{-1} G^t\}^{-1}. \tag{4.5}$$

We assume that the rows of G are linearly independent and that $\{h_1, h_2, \cdots, h_m\}$ are a linearly independent set. (We are maintaining the previously developed notation, so that H_n is $n \times m$, with rows $h_1^t, \cdots, h_n^t$ and G is $k \times m$.

Then the following recursions hold for $n \geqq m$:

$$B_{n+1} = B_n - (B_n h_{n+1})(B_n h_{n+1})^t/(1 + h_{n+1}^t B_n h_{n+1}), \tag{4.6}$$

$$\hat{X}_{n+1} = \hat{X}_n + [B_n h_{n+1}/(1 + h_{n+1}^t B_n h_{n+1})][z(n+1) - h_{n+1}^t \hat{X}_n], \tag{4.7}$$

$$E_{n+1} = E_n + [z(n+1) - h_{n+1}^t \hat{X}_n]^2/[1 + h_{n+1}^t B_n h_{n+1}], \tag{4.8}$$

$$\lambda_{n+1}^{-1} = \lambda_n^{-1} - h_{n+1}^t B_n G^t G B_n h_{n+1}/(1 + h_{n+1}^t B_n h_{n+1}) \tag{4.9}$$

and

$$C_{n+1} = C_n + (C_n G B_n h_{n+1})(C_n G B_n h_{n+1})^t \cdot (1 + h_{n+1}^t B_n h_{n+1} - h_{n+1}^t B_n G^t C_n G B_n h_{n+1})^{-1}, \tag{4.10}$$

with initial conditions

$$\begin{aligned} B_m &= (H_m^t H_m)^{-1}, \\ \hat{X}_m &= B_m H_m^t Z_m, \\ E_m &= 0, \\ \lambda_m^{-1} &= \operatorname{tr} G(H_m^t H_m)^{-1} G^t, \end{aligned} \tag{4.11}$$

and

$$C_m = (G B_m G^t)^{-1}.$$

PROOF. The B_n recursion has been derived in Section 3. Since $\hat{X}_n = B_n H_n^t Z_n$ for $n \geqq m$ and since

$$H_{n+1}^t Z_{n+1} = H_n^t Z_n + z(n+1) h_{n+1},$$

it follows that

$$\hat{X}_{n+1} = B_{n+1}[H_n^t Z_n + z(n+1) h_{n+1}]$$

and (4.7) follows after applying the B_n recursion.

$$E_{n+1} = \|Z_{n+1} - H_{n+1}\hat{X}_{n+1}\|^2$$
$$= \|Z_n - H_n\hat{X}_{n+1}\|^2 + [z(n+1) - h_{n+1}^t\hat{X}_{n+1}]^2$$

and (4.8) follows by applying (4.7), after a little algebra.

$$\lambda_{n+1}^{-1} = \operatorname{tr} GB_{n+1}G^t = \operatorname{tr} GB_nG^t - \operatorname{tr}(g_{n+1}g_{n+1}^t)$$

where

$$(4.12) \qquad g_{n+1} = GB_nh_{n+1}/(1 + h_{n+1}^tB_nh_{n+1})^{\frac{1}{2}}.$$

Since $\operatorname{tr}(gg^t) = g^tg$, (4.9) follows.

It is true in general that for non-singular symmetric matrices C, the matrix $C^{-1} - gg^t$ is non-singular if and only if $g^tCg \neq 1$.

In this case

$$(4.13) \qquad (C^{-1} - gg^t)^{-1} = C + (Cg)(Cg)^t/(1 - g^tCg).$$

(To see the if and only if part, notice that $(C^{-1} - gg^t)\chi = 0$ if and only if $\chi/\|\chi\| = Cg/\|Cg\|$ and $g^tCg = 1$. To verify (4.13), multiply the right side of (4.13) by $(C^{-1} - gg^t)$. The identity matrix obtains.) Since $C_n^{-1} = G(H_n^tH_n)^{-1}G^t$ is non-singular for all $n \geqq m$ and since $C_{n+1}^{-1} = C_n^{-1} - g_{n+1}g_{n+1}^t$ where g_{n+1} is given by (4.12), (4.10) follows from (4.13).

5. Spherical confidence regions. There is something appealing about the geometric simplicity of spherical confidence regions. If one is able to satisfy stronger assumptions, the method described in Theorem 2.1 (which lead to an ellipsoidal region) can be modified to this end. Besides requiring stronger assumptions, spherical confidence regions are slightly less efficient (in an asymptotic sense to be made precise shortly) than their ellipsoidal brothers.

In this section, the quantity

$$(5.1) \qquad a^*(\mu_1, \mu_2, \cdots, \mu_k)$$

is the upper 100α percentile of the random variable $\sum_{j=1}^k \mu_jY_j^2$, where the Y_j are independent $N(0, 1)$:

$$(5.2) \qquad \Pr[\sum_{j=1}^k \mu_jY_j^2 \leqq a^*(\mu_1, \cdots, \mu_k)] = \alpha.$$

The following theorem generalizes Gleser's main result [3], and compares the "efficiency" of spherical and ellipsoidal confidence regions:

THEOREM 5.1. *Suppose there exists a sequence of real numbers, $\{\lambda_n\}$, satisfying (2.5) such that*

$$(5.3) \qquad \lambda_nG(H_n^tH_n)^{-1}G^t \to \Sigma \quad (\textit{where } \Sigma \textit{ is non-singular with positive eigenvalues } \mu_1, \mu_2, \cdots, \mu_k)$$

and suppose that the conditions of Theorem 2.2 are met. Let

$$(5.4) \qquad a_n^* \to a^*(\mu_1, \cdots, \mu_k)$$

and let $N(d)$ be the first value of n such that

$$n^{-1}E_n \leqq d^2\lambda_n/a_n^*, \tag{5.5}$$

where E_n is the residual least squares error. Let $S_n(d)$ be the sphere of radius d, centered at $G\hat{X}_n$:

$$S_n(d) = \{\xi: \|\xi - G\hat{X}_n\|^2 \leqq d^2\}, \tag{5.6}$$

and let $R_n(d)$ be the ellipsoid:

$$R_n(d) = \{\xi: (\xi - G\hat{X}_n)^t[\lambda_n G(H_n{}^tH_n)^{-1}G^t]^{-1}(\xi - G\hat{X}_n) \leqq a_n d^2/a_n^*\} \tag{5.7}$$

where a_n is any sequence such that

$$a_n \to a^*(1, 1, \cdots, 1). \tag{5.8}$$

The following conclusions hold:

(5.9) (a) $\alpha = \lim_{d\to 0} \Pr[GX \varepsilon S_{N(d)}(d)] = \lim_{d\to 0} \Pr[GX \varepsilon R_{N(d)}(d)]$,

(5.10) (b) $\lim_{d\to 0}$ *shortest semi-axis of* $R_{N(d)}(d)$/*radius of* $S_{N(d)}(d) \leqq 1$ wp 1,

(5.11) (c) $\lim_{d\to 0}$ *longest semi-axis of* $R_{N(d)}(d)$/*radius of* $S_{N(d)}(d) \geqq 1$ wp 1,

(5.12) (d) $\lim_{d\to 0}$ *Volume of* $R_{N(d)}(d)$/*Volume of* $S_{N(d)}(d) \leqq 1$ wp 1.

Inequalities (b), (c) and (d) are all strict unless Σ is a multiple of the identity matrix. (Gleser [3] proved the first half of part (a) for the special norming sequence $\lambda_n = n$.)

PROOF. (a) By Theorem 2.1,

$$\lim_{d\to 0} d^2\lambda_{N(d)}/\sigma^2 a^*_{N(d)} = 1 \quad \text{wp } 1. \tag{5.13}$$

By Lemma 2.2

$$\begin{aligned}\sigma^{-2}U_n{}^tP_nU_n \\ &= \sigma^{-2}(GX - G\hat{X}_n)^t[G(H_n{}^tH_n)^{-1}G^t]^{-1}(GX - G\hat{X}_n) \to_L \textstyle\sum_{j=1}^k Y_j^2,\end{aligned} \tag{5.14}$$

and

$$\begin{aligned}\sigma^{-2}U_n{}^tQ_nU_n &= (\lambda_n/\sigma^2)(GX - G\hat{X}_n)^t(GX - G\hat{X}_n) \\ &= \sigma^{-2}(GX - G\hat{X}_n)^t\Sigma_n^{-\frac{1}{2}}(\lambda_n\Sigma_n)\Sigma_n^{-\frac{1}{2}}(GX - G\hat{X}_n) \\ &\to_L \sigma^{-2}Y^t\Sigma Y = \textstyle\sum_{j=1}^k \mu_j Y_j^2\end{aligned} \tag{5.15}$$

where

$$P_n = B_n^{\frac{1}{2}}G[GB_nG^t]^{-1}GB_n^{\frac{1}{2}}, \tag{5.16}$$

$$Q_n = \lambda_n B_n^{\frac{1}{2}}G^tGB_n^{\frac{1}{2}}, \tag{5.17}$$

$$\Sigma_n = G(H_n{}^tH_n)^{-1}G^t \tag{5.18}$$

$$\Sigma = \lim_{n\to\infty} \lambda_m\Sigma_n, \tag{5.19}$$

$\mu_1, \mu_2, \cdots, \mu_k$ are the eigenvalues of Σ, and the Y_j are independent, $N(0, 1)$.

By Theorem 2.2 and (5.13),

$$\Pr[GX \varepsilon R_N] = \Pr[\sigma^{-2}U_N{}^tP_NU_N \leqq a_N\, d^2\lambda_N/a_N{}^*\sigma^2] \to \Pr[\chi_k{}^2 \leqq a] = \alpha$$

as $d \to 0$.

Using arguments akin to those of Lemma (2.4), one can show that

$$\lim_{c\to 0} \sup_n \max_{(1-c)n \leqq n' \leqq (1+c)n} \|Q_n - Q_{n'}\| = 0 \quad (5.20)$$

under the present assumptions. The arguments of Theorem 2.2 can then be repeated verbatim for the rv's $U_n{}^tQ_nU_n$. We conclude therefore, that

$$\sigma^{-2}U^t_{N(d)}Q_{N(d)}U_{N(d)} \to_L \textstyle\sum_{j=1}^k \mu_jY_j^2 \quad \text{as} \quad d \to 0$$

so that

$$\begin{aligned} \Pr[GX \varepsilon S_N] &= \Pr[\|GX - G\hat{X}_N\|^2 \leqq d^2] \\ &= \Pr[\sigma^{-2}U_N{}^tQ_NU_N \leqq \lambda_N\, d^2/\sigma^2] \\ &\to \Pr[\textstyle\sum_{j=1}^k \mu_jY_j^2 \leqq a^*] = \alpha. \end{aligned} \quad (5.21)$$

This proves part (a).

(b) Without loss of generality, suppose that

$$\mu_1 \geqq \mu_2 \geqq \cdots \geqq \mu_k > 0. \quad (5.22)$$

For notational conveneince, let $a = a^*(1, 1, \cdots, 1)$ and $a^* = a^*(\mu_1, \mu_2, \cdots, \mu_k)$. By (5.2)

$$\begin{aligned} \alpha = \Pr[\textstyle\sum_{j=1}^k Y_j^2 \leqq a] &= \Pr[\textstyle\sum_{j=1}^k (\mu_j/\mu_1)Y_j^2 \leqq a^*/\mu_1] \\ &= \Pr[\textstyle\sum_{j=1}^k (\mu_j/\mu_k)Y_j^2 \leqq a^*/\mu_k]. \end{aligned} \quad (5.23)$$

Since $\mu_j/\mu_k \geqq 1$ and $\mu_j/\mu_1 \leqq 1$, we have from (5.23)

$$\alpha \leqq \Pr[\textstyle\sum_{j=1}^k Y_j^2 \leqq a^*/\mu_k] \quad (5.24)$$

and

$$\alpha \geqq \Pr[\textstyle\sum_{j=1}^k Y_j^2 \leqq a^*/\mu_1] \quad (5.25)$$

with strict inequalities holding unless $\mu_1 = \mu_k$.

Thus,

$$a^*/a\mu_1 \leqq 1 \quad \text{and} \quad a^*/a\mu_k \geqq 1 \quad (5.26)$$

with strict inequalities holding unless $\mu_1 = \mu_k$.

Let the (random) eigenvalues of $\lambda_N\Sigma_N$ be denoted by $\mu_1(d) \geqq \mu_2(d) \geqq \cdots \geqq \mu_k(d)$. Since $\lambda_n\Sigma_n \to \Sigma$, it follows that $\lambda_N\Sigma_N \to \Sigma$ wp 1, as $d \to 0$ so that

$$\lim_{d\to 0} \mu_j(d) = \mu_j \quad \text{wp } 1. \quad (5.27)$$

The length of $R_N(d)$'s largest and smallest semi-axes are respectively

$$[\mu_1(d)a_N/a_N{}^*]^{\frac{1}{2}}\, d \quad (5.28)$$

and

$$[\mu_k(d)a_N/a_N^*]^{\frac{1}{2}}\, d. \tag{5.29}$$

The radius of $S_N(d)$ is d, so parts (b) and (c) follow from (5.27) and (5.26) (since $a_N/a_N^* \to a/a^*$ wp 1), both inequalities being strict unless $\mu_1 = \mu_k$ (i.e. unless Σ is a multiple of I).

(c) We base our conclusion on the following lemma which we will prove later:

LEMMA 5.1. *Let* $\{e(t);\, t \geqq 0\}$, $\{Y(t);\, t \geqq 0\}$ *and* $\{\Sigma(t);\, t \geqq 0\}$ *be stochastic processes on the non-negative reals. For each* t, $e(t)$ *is real valued,* $Y(t)$ *is a* k*-vector and* $\Sigma(t)$ *is a* $k \times k$ *positive definite matrix. We assume that there is a constant,* e, *a non-negative function,* $\varphi(t)$, *a constant positive definite matrix* Σ^* *and a spherically symmetric* k*-dimensional normal* rv, Y, *such that*

(i) $\lim_{t\to\infty} e(t) = e$ wp 1,

(ii) $Y(t) \to_L Y$,

(iii) $\lim_{t\to\infty} \Sigma(t)/\varphi(t) = \Sigma^*$ wp 1 *and*

(iv) $\Pr[\|\Sigma^{*\frac{1}{2}}Y\|^2 \leqq 1] = \Pr[\|Y\|^2 \leqq e]$.

Then,

$$\lim_{t\to\infty} \gamma[\|\xi\|^2 \leqq \varphi(t)]/\gamma[\|\Sigma^{-\frac{1}{2}}(t)\xi\|^2 \leqq e(t)] \geqq 1$$

with strict inequality holding unless Σ^* *is a multiple of* I, *where* $\gamma[\cdot]$ *is the* k*-dimensional Lebesgue measure of the set in the square brackets.*

To apply the lemma to the case at hand, let

$$t = 1/d, \quad \Sigma(t) = \Sigma_{N(d)}(\Sigma_n = G(H_n{}^tH_n)^{-1}G^t), \quad Y(t) = \Sigma_N^{-\frac{1}{2}}(GX - G\hat{X}_{N(d)}),$$
$$e(t) = a_{N(d)}\lambda_{N(d)}\, d^2/a^*_{N(d)}, \qquad e = a(1, 1, \cdots 1)\sigma^2$$

and $\varphi(t) = 1/t^2$: By (5.14), (5.5) and (5.3),

$$\lim_{t\to\infty} \Sigma(t)/\varphi(t) = \lim_{d\to 0} \lambda_{N(d)}\Sigma_{N(d)}/\lambda_{N(d)}d^2$$
$$= \Sigma/a^*\sigma^2 \qquad \text{wp } 1,$$

so (iii) holds with $\Sigma^* = \Sigma/a^*\sigma^2$.

To establish (iv),

$$\Pr[\|\Sigma^{*\frac{1}{2}}Y\|^2 \leqq 1] - \Pr[\|Y\|^2 \leqq e] = \Pr[\sigma^{-2}Y^t\Sigma Y \leqq a^*] - \Pr[\sigma^{-2}Y^tY \leqq a] = 0.$$

(i) and (ii) follow from earlier results. Thus,

$$\lim_{d\to 0} \gamma[\|\xi\|^2 \leqq d^2]/\gamma[\xi^t\Sigma_{N(d)}^{-1}\xi \leqq a_N\, d^2\lambda_N/a_N^*] \geqq 1 \qquad \text{wp } 1$$

with strict inequality holding unless Σ is a multiple of I.

Since Lebesgue measure is translation invariant,

$$\lim_{d\to 0} \gamma[S_N(d)]/\gamma[R_N(d)] \geqq 1 \qquad \text{wp } 1$$

with strict inequality etc., etc.

PROOF OF LEMMA. It suffices to show that

$$\gamma[\|\xi\|^2 \leqq 1] \geqq \gamma[\xi^t\Sigma^{*-1}\xi \leqq e] \tag{5.30}$$

with strict inequality holding unless Σ^* is a multiple of I: For then,

$$\begin{aligned}\gamma[\|\xi\|^2 \leqq \varphi(t)]/\gamma[\|\Sigma^{-\frac{1}{2}}(t)\xi\|^2 \leqq e(t)]\\ &= \{\gamma[\|\xi\|^2 \leqq \varphi(t)]/\gamma[\xi^t(\varphi(t)\Sigma^*)^{-1}\xi \leqq e]\}\\ &\quad\cdot\{\gamma[\xi^t(\varphi(t)\Sigma^*)^{-1}\xi \leqq e]/\gamma[\xi^t\Sigma^{-1}(t)\xi \leqq e(t)]\}\\ &= \{\gamma[\|\xi\|^2 \leqq \varphi(t)]/\gamma[\xi^t(\varphi(t)\Sigma^*)^{-1}\xi \leqq e]\}\\ &\quad\cdot\{\gamma[\xi^t\Sigma^{*-1}\xi \leqq e]/\gamma[\xi^t(\Sigma(t)/\varphi(t))^{-1}\xi \leqq e(t)]\}\\ &= \{\gamma[\|\xi\|^2 \leqq 1]/\gamma[\xi^t\Sigma^{*-1}\xi \leqq e]\}\\ &\quad\cdot\{\gamma[\xi^t\Sigma^{*-1}\xi \leqq e]/\gamma[\xi^t(\Sigma(t)/\varphi(t))^{-1})\xi \leqq e(t)]\}.\end{aligned}$$

The second factor tends to unity wp 1 and the first is greater than unity by (5.30) unless Σ^* is a multiple of I.
Therefore

$$\lim_{t\to\infty} \gamma[\|\xi\|^2 \leqq \varphi(t)]/\gamma[\|\Sigma^{-\frac{1}{2}}(t)\xi\|^2 \leqq e(t)] \geqq 1.$$

To prove (5.30), let $Z = \Sigma^{*\frac{1}{2}}Y$.
Then Z is normally distributed with a density whose surfaces of constant density are the ellipsoidal surfaces $\xi^t\Sigma^{*-1}\xi = c$. Let $R(e)$ be the ellipsoid $[\xi^t\Sigma^{*-1}\xi \leqq e]$ and let A be any Lebesgue measurable set whose volume is the same as $R(e)$.

(5.31) $\Pr[Z \,\varepsilon\, R(e)] - \Pr[Z \,\varepsilon\, A] = \Pr[Z \,\varepsilon\, R(e) - A] - \Pr[Z \,\varepsilon\, A - R(e)]$.

Since $\gamma(A) = \gamma(A - R) + \gamma(AR) = \gamma(R) = \gamma(R - A) + \gamma(RA)$, we see that $\gamma(A - R) = \gamma(R - A)$. Since the density function of Z is uniformly larger over $R(e) - A$ than over $A - R(e)$, (5.31) is positive unless $\gamma(A - R) + \gamma(R - A) = 0$. Since the density of Z is everywhere positive, the last implies that if $\Pr[Z \,\varepsilon\, R(e)] = \Pr[Z \,\varepsilon\, A]$, then $\gamma(A) \geqq \gamma(R(e))$ with equality holding if and only if $\gamma(A - R) + \gamma(R - A) = 0$. (5.30) is the special case where $A = [\|\xi\|^2 \leqq 1]$. For, by (iv),

$$\Pr[\|Z\|^2 \leqq 1] = \Pr[Z^t\Sigma^{*-1}Z \leqq e]$$

or equivalently,

$$\Pr[Z \,\varepsilon\, A] = \Pr[Z \,\varepsilon\, R(e)].$$

Since $\gamma(A - R) + \gamma(R - A) = 0$ if and only if $A = R(e)$, the conclusion follows.

6. Application to hypothesis testing. The method of constructing fixed size confidence sets can be immediately adapted to the problem of constructing sequential tests of the general linear hypothesis. The essential idea is quite simple and was pointed out to me by H. Robbins (private communication): Suppose one wishes to test the hypothesis

$$H_0 : GX = \xi_0$$

against the alternative

$$H_1 : \|GX - \xi_0\| > 2r.$$

Suppose that the fixed size confidence region $R_N(d)$ is constructed according to the methods of Sections 2 and 3 in such a way that $R_N(d)$ is contained inside a sphere of radius r. Consider the test which accepts H_0 if and only if R_N covers ξ_0. The size of this test is equal to

$$\begin{aligned} \Pr[\text{reject } H_0 \mid GX = \xi_0] &= \Pr[\xi_0 \not\varepsilon R_N(d) \mid GX = \xi_0] \\ &= \Pr[R_N(d) \text{ fails to cover } GX \mid GX = \xi_0] \\ &= 1 - \Pr[GX \varepsilon R_N(d)]. \end{aligned} \tag{6.1}$$

On the other hand, if $\|\xi_1 - \xi_0\| > 2r$, $R_N(d)$ cannot simultaneously cover ξ_1 and ξ_0 and so,

$$\begin{aligned} \Pr[\text{reject } H_0 \mid GX = \xi_1] \\ \geqq \Pr[R_N(d) \text{ covers } \xi_1 \mid GX = \xi_1] = \Pr[GX \varepsilon R_N(d)]. \end{aligned} \tag{6.2}$$

Thus, whatever the exact confidence level of the set (call it $\alpha(d; X) = \Pr[GX \varepsilon R_N(d)]$), the aforementioned test has size equal to $1 - \alpha(d; X)$ over the null hypothesis and power of at least $\alpha(d; X)$ for all alternatives at a distance $2r$ from the null hypothesis. If the distribution of the residuals, $v(n)$, does not depend upon X, then $\alpha(d; X)$ does not depend on X so $\alpha(d; X) \to \alpha$ *uniformly* in X as $d \to 0$.

7. Exact confidence levels for the normal case. If the residuals are normally distributed, it is possible to derive useful formulae for the exact coverage probabilities associated with the regions $R_{N(d)}(d)$ which were described in Section 3.

The main result of this section is

THEOREM 7.1. *Let $z(n)$ be a sequence of real valued random variables of the form*

$$z(n) = h_n^t X + v(n) \qquad n = 1, 2, \cdots,$$

where the $v(n)$ are independent, $N(0, \sigma^2)$, the h_n are known m-dimensional column vectors and X is the unknown regression vector. Let $\hat{X}_n$ be the least squares estimate of X based upon the first n observations and let E_n be the associated residual error.

Let G be any $k \times m$ matrix of rank k, let λ_n be any sequence of non-decreasing positive numbers and let H_n be the $n \times m$ matrix whose rows are $h_1^t, h_2^t, \cdots, h_n^t$. We assume H_n has rank m for all $n \geqq m$.

Let N be the first integer greater than m such that

$$E_n \leqq d^2 \lambda_n n / a_n$$

where a_n is any sequence of positive numbers, and let

$$R_n(d) = \{\xi : (\xi - G\hat{X}_n)^t [G(H_n^t H_n)^{-1} G^t]^{-1} \cdot (\xi - G\hat{X}_n) \leqq d^2 \lambda_n\} \quad (n = m + 1, m + 2, \cdots).$$

Then the following conclusions hold:

(a) $$\alpha = \Pr[GX \varepsilon R_N(d)]$$

depends upon d and σ through their ratio $\rho = d^2/\sigma^2$.

(b) $\alpha(\rho) = \Pr[GX \varepsilon R_N(d)]$

$$= \Pr[\chi_k^2 \leqq \lambda_{m+1}\rho] + \textstyle\sum_{n=m+1}^{\infty} \Pr[\lambda_n\rho \leqq \chi_k^2 \leqq \lambda_{n+1}\rho]Q_n(\rho) \quad (7.1)$$

where χ_k^2 is a chi-square rv *with k* df,

$$Q_{m+n}(\rho) = F_n(0), \qquad n = 1, 2, \cdots, \quad (7.2)$$

$$F_1(t) = 1 - (2/\pi)^{\frac{1}{2}} \int_0^{(\rho\beta_1 - t)^{\frac{1}{2}}} e^{-1/2y^2}\, dy, \qquad t \leqq 0,$$

$$F_{n+1}(t) = F_n(0) - (2/\pi)^{\frac{1}{2}} \cdot \int_0^{(\rho\beta_{n+1}-t)^{\frac{1}{2}}} [F_n(0) - F_n(u^2 - \rho\beta_{n+1} + t)]e^{-1/2y^2}\, du, \quad t \leqq 0; \quad (7.3)$$

$$\beta_1 = (m+1)\lambda_{m+1}/a_{m+1}$$

and

$$\beta_n = (m+n)\lambda_{m+n}/a_{m+n} - (m+n-1)\lambda_{m+n-1}/a_{m+n-1}, \qquad n > 1. \quad (7.4)$$

(c) *Let*

$$\alpha_M(\rho) = \Pr[\chi_k^2 \leqq \lambda_{m+1}\rho] + \textstyle\sum_{j=m+1}^{m+M} \Pr[\lambda_j\rho \leqq \chi_k^2 \leqq \lambda_{j+1}\rho]Q_j(\rho). \quad (7.5)$$

For every $M \geqq 1$,

$$\alpha_M(\rho) \leqq \alpha(\rho) \leqq \alpha_M(\rho) + \textstyle\sum_{j=m+M+1}^{\infty} Q_j(\rho). \quad (7.6)$$

(d) *Suppose $0 < \gamma < \frac{1}{2}$. Let K be any integer such that*

(i)

$$\delta = \delta(\gamma, \rho, K) = \inf_{j \geqq K+1} \{1 - [(2/\pi)((1 - 2\gamma)/\gamma^2\rho^3\beta_j^3)]^{\frac{1}{2}} - [(\gamma\rho\beta_j)^{-1}(\log(1 - 2\gamma)^{-\frac{1}{2}})]\} > 0 \quad (7.7)$$

and

(ii) $\inf_{j \geqq K} (\beta_{j+1} - \beta_j) \geqq 0$.

Then for all $M \geqq K$,

$$\alpha_M(\rho) \leqq \alpha(\rho) \leqq \alpha_M(\rho) + [A(\gamma, \rho, K)/\beta_M] \exp - [\gamma\rho\delta \textstyle\sum_{j=1}^{M-1} \beta_j] \quad (7.8)$$

where

$$A(\gamma, \rho, K) = (\gamma\rho\delta)^{-1} \exp -\gamma\rho[\textstyle\sum_{j=1}^{K} \beta_j\{(1 - \delta) - [(2/\pi)((1 - 2\gamma)/\gamma^2\rho^3\beta_j^3)]^{\frac{1}{2}} - [(\gamma\rho\beta_j)^{-1} \log(1 - 2\gamma)^{-\frac{1}{2}}]\}]. \quad (7.9)$$

Parts (c) and (d) furnish very useful approximations to the coverage probability $\alpha(\rho)$. Typically, the sequences, λ_n, which we discussed in Section 3 go to

∞ like n (or faster) so that β_n behaves like n and $\sum_{j=1}^n \beta_j$ behaves like n^2. For moderately large values of ρ, α_M (which in any event furnishes a lower bound for α), is going to be close to α when M is small, and (7.8) furnishes a useable estimate for the M-term truncation error.

Computations are in progress for various choices of the λ_n and a_n sequences and will be reported upon in the near future.

We prove Theorem 7.1 via two lemmas:

LEMMA 7.1. *Assume normal residuals.*

(a) *If* $j \leqq n$, *the random variables*

$$[\hat{X}_{n+m} - X] \quad \text{and} \quad [z(m + j) - h_{m+j}^t \hat{X}_{m+j-1}]$$

are independent.

(b) *For* $j = 1, 2, \cdots$ *the random variables*

$$Y_j = [z(m + j) - h_{m+j}^t \hat{X}_{m+j-1}]/\sigma(1 + h_{m+j}^t B_{m+j-1} h_{m+j})^{\frac{1}{2}}$$

(*where* $B_n = (H_n{}^t H_n)^{-1}$) *are independent,* $N(0, 1)$.

(c) *For every* $n \geqq j$,

$$\sigma^{-2} E_{m+j} = \sum_{i=1}^j Y_i^2$$

is independent of $\hat{X}_{m+n}$.

PROOF. (a) By (4.1), (4.7) and (2.1a),

$$\begin{aligned} \mathcal{E}[z(j) - h_j{}^t \hat{X}_{j-1}][\hat{X}_j - X]^t = \mathcal{E}[v(j) + h_j{}^t (X - \hat{X}_{j-1})] \\ \cdot[(I - B_{j-1} h_j h_j{}^t / (1 + h_j{}^t B_{j-1} h_j))(\hat{X}_{j-1} - X) \\ + B_{j-1} h_j v(j) / (1 + h_j{}^t B_{j-1} h_j)]^t \end{aligned} \tag{7.10}$$

provided $j \geqq m + 1$.

Since $\mathcal{E}(X - \hat{X}_{j-1})(X - \hat{X}_{j-1})^t = \text{Cov}\, \hat{X}_{j-1} = \sigma^2 (H_{j-1}^t H_{j-1})^{-1} = \sigma^2 B_{j-1}$, the right side of (7.10) vanishes.

Since

$$\begin{aligned} \hat{X}_{n+1} - X = [I - B_n h_{n+1} h_{n+1}^t / (1 + h_{n+1}^t B_n h_{n+1})](\hat{X}_n - X) \\ + B_n h_{n+1} v(n + 1) / (1 + h_{n+1}^t B_n h_{n+1}) \end{aligned}$$

when $n \geqq m$ (c.f. (2.1a) and (4.7)) and since $v(n + 1)$ is independent of $z(j) - h_j{}^t \hat{X}_{j-1}$ when $n \geqq j$, the independence of $(z(j) - h_j{}^t \hat{X}_{j-1})$ and $(\hat{X}_n - X)$ implies the independence of $(z(j) - h_j{}^t \hat{X}_{j-1})$ and $(\hat{X}_{n+1} - X)$. Part (a) is thus proved by induction.

(b) The mean of Y_j is clearly zero and its variance is one. By part (a) $\hat{X}_{m+n-1}$ is independent of Y_j if $n > j$.

(c) By part (a), $\hat{X}_{m+n}$ is independent of $Y_1, Y_2, \cdots, Y_j$ and hence, is independent of $\sum_{i=1}^j Y_i^2$. By (4.8) the last is equal to $\sigma^{-2} E_{m+j}$.

LEMMA 7.2. *Let* $\{\beta_n\}$ *be a sequence of positive numbers and suppose* $\inf_{n \geqq K} \beta_{n+1} - \beta_n \geqq 0$. *Then*

$$\sum_{n=M+1}^{\infty} \exp -\left(\sum_{j=1}^n \beta_j\right) \leqq \beta_M{}^{-1} \exp -\left(\sum_{j=1}^{M-1} \beta_j\right) \quad \textit{if} \quad M \geqq K. \tag{7.11}$$

PROOF. Let $f(t)$ be the continuous piecewise linear function defined over $(0, \infty)$ such that $f(0) = 0, f(n) = \beta_n$ $(n = 1, 2, \cdots)$ and $f'(t) = \beta_n - \beta_{n-1}$ $(n - 1 < t < n)$. Then f is non-negative so that

$$F(t) = \int_0^t f(u)\, du$$

is non-negative and non-decreasing. When $t \geqq K$, the graph of $f(\cdot)$ lies below the graph of the staircase function which takes on the value(s) β_{n+1} in the interval(s) $[n, n + 1]$. Thus,

$$F(n) - F(K) \leqq \sum_{j=K+1}^{n} \beta_j \quad \text{if} \quad n > K.$$

On the other hand, if $t \geqq K$, the graph of $f(\cdot)$ lies entirely below the graph of the staircase function taking the value(s) β_n in the interval(s) $[n, n + 1)$. Thus,

$$F(n + 1) - F(K) \geqq \sum_{j=K}^{n} \beta_j \quad \text{if} \quad n \geqq K.$$

Thus for all $n > K$,

$$F(n) - F(K) \leqq \sum_{j=K+1}^{n} \beta_j \leqq F(n + 1) - F(K) - \beta_K, \tag{7.12}$$

so that

$$\sum_{n=M+1}^{\infty} \exp - \sum_{j=1}^{n} \beta_j \leqq [\exp (F(K) - \sum_{j=1}^{K} \beta_j)] \sum_{n=M+1}^{\infty} \exp -F(n) \quad \text{if} \quad M \geqq K. \tag{7.13}$$

For all positive values of t, the graph of $\exp -F(\cdot)$ lies above the graph of the staircase function taking the value(s) $\exp -F(n + 1)$ in the interval(s) $[n, n + 1)$ so that

$$\sum_{n=M+1}^{\infty} \exp -F(n) = \sum_{n=M}^{\infty} \exp -F(n + 1) \leqq \int_M^{\infty} [\exp -F(t)]\, dt. \tag{7.14}$$

Letting $u = F(t)$, $du = f(t)\, dt \geqq f(M)\, dt$ if $t \geqq M \geqq K$ so that

$$\int_M^{\infty} [\exp -F(t)]\, dt \leqq [f(M)]^{-1} \int_{F(M)}^{\infty} e^{-u}\, du = e^{-F(M)}/\beta_M. \tag{7.15}$$

The right side of (7.12) implies

$$e^{-F(M)}/\beta_M \leqq \beta_M^{-1} \exp -(F(K) + \sum_{j=K}^{M-1} \beta_j) \quad \text{if} \quad M \geqq K. \tag{7.16}$$

Combining (7.13)–(7.16),

$$\sum_{n=M+1}^{\infty} \exp -(\sum_{j=1}^{n} \beta_j) \leqq \beta_M^{-1} \exp - (\sum_{j=1}^{M-1} \beta_j) \quad \text{if} \quad M \geqq K.$$

PROOF OF THEOREM 7.1.

$$\begin{aligned} \alpha &= \Pr [GX \,\varepsilon\, R_N] \\ &= \sum_{n=m+1}^{\infty} \Pr [GX \,\varepsilon\, R_N \mid N = n] \Pr [N = n]. \end{aligned} \tag{7.17}$$

Since the event $[N = n]$ is in the Borel Field induced by the rv's $E_1, E_2, \cdots, E_n$ and since the event $[GX \,\varepsilon\, R_n(d)]$ is in the Borel Field induced by $G\hat{X}_n$, which by Lemma 7.1c, is independent of the first Borel Field, we have

$$\alpha = \sum_{n=m+1}^{\infty} \Pr [GX \,\varepsilon\, R_n(d)] \Pr [N = n]. \tag{7.18}$$

But

$$\text{(7.19)} \qquad \Pr[GX \varepsilon R_n(d)] = \Pr[\chi_k^2 \leqq \lambda_n \rho].$$

The event $[N = m + n]$ occurs if and only if

$$\min_{m+1 \leqq j < m+n} [E_j - j\lambda_j d^2/a_j] > 0$$

and

$$E_{m+n} - (m + n)\lambda_{m+n} d^2/a_{m+n} \leqq 0.$$

Since $\sigma^{-2}E_{m+j} = \sum_{i=1}^{j} Y_i^2$ (by Lemma 7.1c) we see that $[N = n + m]$ occurs if and only if

$$\min_{1 \leqq j < n} \{\textstyle\sum_{i=1}^{j} Y_i^2 - [(m + j)\lambda_{m+j}\rho/(a_{m+j})]\} = \min_{1 \leqq j < n} [\textstyle\sum_{i=1}^{j} (Y_i^2 - \rho\beta_i)] > 0$$

and

$$\text{(7.20)} \qquad S_n \equiv \textstyle\sum_{i=1}^{n} [Y_i^2 - \rho\beta_i] \leqq 0$$

where $\{Y_j\,; j = 1, 2, \cdots\}$ is a sequence of independent $N(0, 1)$ rv's. Thus, $\Pr[N = m + n]$ is strictly a function of $\rho = d^2/\sigma^2$. Combining this result with (7.18) and (7.19), we obtain (a) of Theorem 7.1.

To prove part (b), we notice that

$$\text{(7.21)} \qquad \Pr[N = n] = \Pr[N > n - 1] - \Pr[N > n].$$

Combining (7.21) and (7.18), then summing by parts we see that

$$\text{(7.22)} \quad \alpha(\rho) = \Pr[\chi_k^2 \leqq \lambda_{m+1}\rho] + \textstyle\sum_{n=m+1}^{\infty} \Pr[\lambda_n\rho \leqq \chi_k^2 < \lambda_{n+1}\rho]Q_n(\rho)$$

where $Q_n(\rho) = \Pr[N > n]$. This establishes (7.1).

The event $[N > n + m]$ occurs if and only if

$$\min_{1 \leqq j \leqq n} \{\textstyle\sum_{i=1}^{j} (Y_i^2 - \rho\beta_i)\} > 0$$

or equivalently, if and only if

$$\max_{1 \leqq j \leqq n} \{\textstyle\sum_{i=1}^{j} (\rho\beta_i - Y_i^2\} < 0.$$

Let us now define a random variable T_n, which is equal to S_n until S_n makes its first zero upcrossing. Thereafter, this rv will stay "frozen" at the S process' first non-negative value:

DEFINE. $\quad T_1 = \rho\beta_1 - Y_1^2$

$$\text{(7.23)} \qquad \begin{aligned} T_{n+1} &= T_n && \text{if } T_n \geqq 0 \\ &= T_n + (\rho\beta_{n+1} - Y_{n+1}^2) && \text{if } T_n < 0. \end{aligned}$$

$\{T_n\,; n \geqq 1\}$ is a Markov process. The transition probabilities are

$$\text{(7.24)} \qquad \begin{aligned} &\Pr[T_{n+1} \leqq t \mid T_n = s] \\ &= 1 && \text{if } 0 < s \leqq t \\ &= 0 && \text{if } 0 < s \text{ and } t < s \\ &= \Pr[\rho\beta_{n+1} - Y_{n+1}^2 \leqq t - s] && \text{if } s < 0. \end{aligned}$$

We study the T_n process because

$$\Pr[N > n + m] = \Pr[T_n < 0]. \tag{7.25}$$

If we denote the distribution of T_n by $F_n(t)(=\Pr[T_n \leqq t])$ we see that for $t \leqq 0$,

$$F_{n+1}(t) = F_n(0) - \int_{-\infty}^{0} \Pr[\rho\beta_{n+1} - Y_{n+1}^2 > t - s]\, dF_n(s) \tag{7.26}$$

where $F_1(t) = \Pr[Y_1^2 > \rho\beta_1 - t]$. Since

$$\Pr[\rho\beta_{n+1} - Y_{n+1}^2 > t - s] = (2/\pi)^{\frac{1}{2}} \int_0^{[\min(\rho\beta_{n+1}-t+s,0)]^{\frac{1}{2}}} e^{-1/2y^2}\, dy$$

we see that for $t \leqq 0$,

$$F_1(t) = 1 - (2/\pi)^{\frac{1}{2}} \int_0^{(\rho\beta_1-t)^{\frac{1}{2}}} e^{-1/2y^2}\, dy,$$

$$F_{n+1}(t) = F_n(0) - \int_{-(\rho\beta_{n+1}-t)}^{0} (2/\pi)^{\frac{1}{2}} \int_0^{(\rho\beta_{n+1}-t+s)^{\frac{1}{2}}} e^{-1/2y^2}\, dy\, dF_n(s).$$

Make the change of variable: $u = (\rho\beta_{n+1} - t + s)^{\frac{1}{2}}$ after integrating by parts (F_n is absolutely continuous as can be proven by induction), and equation (7.2) results. By (7.25)

$$Q_{n+m}(\rho) = F_n(0),$$

thereby establishing (7.3). To prove (c), we use (7.1): The M-term truncation, $\alpha_M(\rho)$ defined by (7.5), must be smaller than $\alpha(\rho)$. Since the coefficient of $Q_n(\rho)$ is no greater than unity, the remainder term is bounded by $\sum_{n=m+M+1}^{\infty} Q_n(\rho)$, thereby establishing (7.6).

To prove part (d), we point out that one can prove by induction that for $n \geqq 1, t \leqq 0$ and $0 < \gamma < \frac{1}{2}$, $F_n(t) \leqq A_n \exp \gamma t$, where

$$A_n = \exp -\gamma\rho[\textstyle\sum_{j=1}^{n} \beta_j\{1 - [(2/\pi)((1 - 2\gamma)/\gamma^2\rho^3\beta_j^3)]^{\frac{1}{2}} - [(\gamma\rho\beta_j)^{-1} \log (1 - 2\gamma)^{-\frac{1}{2}}]\}]. \tag{7.27}$$

In fact, by (7.2),

$$\begin{aligned} F_1(t) &= 1 - (2/\pi)^{\frac{1}{2}} \int_0^{(\rho\beta_1-t)^{\frac{1}{2}}} e^{-1/2y^2}\, dy \\ &= (2/\pi)^{\frac{1}{2}} \int_{\rho\beta_1-t}^{\infty} e^{-1/2u}\, du/2u^{\frac{1}{2}} \\ &\leqq (2/\pi)^{\frac{1}{2}}[2(\rho\beta_1)^{-\frac{1}{2}}] \int_{\rho\beta_1-t}^{\infty} e^{-1/2u}\, du \leqq [(2/\pi)(\rho\beta_1)^{-1}]^{\frac{1}{2}} e^{-\gamma\rho\beta_1} e^{\gamma t} \end{aligned}$$

if $0 < \gamma < \frac{1}{2}$ and $t \leqq 0$. Therefore,

$$\begin{aligned} F_1(t) &\leqq \{[(2/\pi)(\rho\beta_1)^{-1}]^{\frac{1}{2}} + (1 - 2\gamma)^{-\frac{1}{2}}\} e^{-\gamma\rho\beta_1} e^{\gamma t} \\ &= \{1 + [(2/\pi)((1 - 2\gamma)/\rho\beta_1)]^{\frac{1}{2}}\}(1 - 2\gamma)^{-\frac{1}{2}} e^{-\gamma\rho\beta_1} e^{\gamma t}. \end{aligned}$$

Since $1 + x \leqq e^x$, we conclude that $F_1(t) \leqq A_1 e^{\gamma t}$. Continuing by induction, suppose that $F_n(t) \leqq A_n e^{\gamma t}$ for $t \leqq 0$. By (7.2),

$$\begin{aligned} F_{n+1}(t) \leqq A_n[1 &- (2/\pi)^{\frac{1}{2}} \int_0^{(\rho\beta_{n+1}-t)^{\frac{1}{2}}} e^{-1/2y^2}\, dy \\ &+ (2/\pi)^{\frac{1}{2}} \int_0^{\infty} \exp(-1/2y^2 + \gamma y^2 - \gamma\rho\beta_{n+1} + \gamma t)\, dy] \end{aligned}$$

$$
\begin{aligned}
&< (A_n \exp -\gamma\rho\beta_{n+1})/(1 - 2\gamma)^{\frac{1}{2}}\{1 + [(2/\pi)((1 - 2\gamma)/\rho\beta_{n+1})]^{\frac{1}{2}}\} \exp \gamma t \\
&\leqq A_n[\exp\{-\gamma\rho\beta_{n+1} + \log(1 - 2\gamma)^{-\frac{1}{2}} \\
&\quad + [(2/\pi)((1 - 2\gamma)/\rho\beta_{n+1})]^{\frac{1}{2}}\}] \exp \gamma t \\
&= A_{n+1} \exp \gamma t.
\end{aligned}
$$

In particular,

$$Q_{M+n}(\rho) = F_n(0) \leqq A_n . \tag{7.28}$$

If K is chosen so large that

$$
\begin{aligned}
\delta &= \delta(\gamma, \rho, K) \\
&= \inf_{j \geqq K+1} \{1 - [(2/\pi)((1 - 2\gamma)/\gamma^2\rho^3\beta_j^3)]^{\frac{1}{2}} - [(\gamma\rho\beta_j)^{-1} \log(1 - 2\gamma)^{-\frac{1}{2}}]\} > 0,
\end{aligned}
$$

and if $n \geqq K + 1$, we have, by (7.27),

$$
\begin{aligned}
A_n &\leqq [\exp -\gamma\rho \textstyle\sum_{j=1}^{K} \beta_j\{(1 - \delta) - [(2/\pi)((1 - 2\gamma)/\gamma^2\rho^3\beta_j^3)]^{\frac{1}{2}} \\
&\quad - [(\gamma\rho\beta_j)^{-1} \log(1 - 2\gamma)^{-\frac{1}{2}}]\}][\exp -\gamma\rho\delta \textstyle\sum_{j=1}^{n} \beta_j] \\
&= A^*(\gamma, \rho, K) \exp -\gamma\rho\delta \textstyle\sum_{j=1}^{n} \beta_j .
\end{aligned} \tag{7.29}
$$

Thus, if $M \geqq K$ is chosen so large that

$$\inf_{j \geqq M} (\beta_{j+1} - \beta_j) \geqq 0,$$

we have by Lemma 7.2, (7.28) and (7.29) that

$$
\begin{aligned}
\textstyle\sum_{j=m+M+1}^{\infty} Q_j(\rho) &\leqq \textstyle\sum_{j=M+1}^{\infty} A_j \\
&\leqq A^*(\gamma, \rho, K) \textstyle\sum_{n=M+1}^{\infty} [\exp -\gamma\rho\delta \textstyle\sum_{j=1}^{n} \beta_j] \\
&\leqq (A^*/\gamma\rho\delta\beta_M) \exp -\gamma\rho\delta \textstyle\sum_{j=1}^{M-1} \beta_j \\
&= (A(\gamma, \rho, K)/\beta_M) \exp -\gamma\rho\delta \textstyle\sum_{j=1}^{M-1} \beta_j ,
\end{aligned}
$$

where $A(\gamma, \rho, K)$ is defined by (7.9).

Acknowledgment. I wish to thank the referee for his efforts in connection with the revision of this paper. Several large and small errors which had escaped my scrutiny were patched up as a result of his attentiveness.

REFERENCES

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

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

[3] GLESER, LEON J. (1965). On the asymptotic theory of fixed-size sequential confidence bounds for linear regression parameters. *Ann. Math. Statist.* **36** 463–467. (See correction note. *Ann. Math. Statist.* **37** (1966) 1053–1055.)

[4] GNEDENKO, B. V. AND KOLMOGOROV, A. N. (1954). *Limit Distributions for Sums of Independent Random Variables*. (Translated by K. L. Chung) Addison-Wesley, Cambridge.

[5] LOÈVE, MICHEL (1955). *Probability Theory*. Van Nostrand, New York.