ON THE ASYMPTOTIC THEORY OF FIXED-SIZE SEQUENTIAL CONFIDENCE BOUNDS FOR LINEAR REGRESSION PARAMETERS[1]

By Leon J. Gleser

Columbia University

1. Introduction. Chow and Robbins [3] have considered the problem of finding a confidence interval of prescribed width $2d$ and prescribed coverage probability α for the unknown mean μ of a population Ω having fixed distribution function F with unknown, but finite, variance $\sigma^2 > 0$. Since no fixed sample procedure can possibly work, they consider a certain class of sequential procedures and show that the members of this class are asymptotically "consistent" (i.e., cover μ with probability α) and asymptotically "efficient" (i.e., have expected sample size equal to the smallest sample size one could use if σ^2 were known) as d goes to zero. The purpose of this paper is to extend these results to the linear regression problem.

2. The problem. Consider $y_1, y_2, \cdots$ a sequence of independent observations with

$$y_i = \beta x^{(i)} + \epsilon_i, \tag{2.1}$$

β an unknown $1 \times p$ vector, $x^{(i)}$ a known $p \times 1$ column vector, and ϵ_i a random observation obeying a (possibly) unknown distribution function F with finite, but unknown, variance σ^2. We wish to find a region R in p-dimensional Euclidean space such that $P(\beta \varepsilon R) = 1 - \alpha$ and such that the length of the interval cut off on the β_i-axis by R has width $\leqq 2d$, $i = 1, \cdots p$. As has already been noted for $p = 1$, no fixed-sample procedure will meet our requirements; we are thereby led to consider sequential procedures.

To motivate the sequential procedure that we use, consider what classical statistical practice would be if σ^2 were known. Since the least-squares estimate of β has componentwise (by the Gauss-Markov theorem) uniformly minimum variance among all linear unbiased estimates of β, has good asymptotic properties (such as consistency—viz., Eicker [5]), and performs reasonably well against nonlinear unbiased estimates (Anderson [1]), classical practice would be to use the least-squares estimate of β in the construction of our confidence region. It is well-known that the least-squares estimate of β in our problem is

$$\hat{\beta}(n) = Y_n X_n'(X_n X_n')^{-1} \tag{2.2}$$

where $Y_n = (y_1, \cdots, y_n)$, $X_n = (x^{(1)}, \cdots, x^{(n)}) : p \times n$, $p \leqq n$, and where we assume that X_p is of full rank. (This is usually possible to achieve in practice—

Received 5 October 1964.

[1] Research sponsored by National Science Foundation Grant NSF-GP-2074 at Columbia University.

if not, sample until p independent $x^{(i)}$ are found, start with the p corresponding y_i's and save the remainder for future use in the sequential procedure. Such a procedure does not bias the results and is equivalent to starting after a fixed number of observations n_0.)

Since the covariance matrix of the $\hat{\beta}(n)$ is $\sigma^2(X_nX_n')^{-1}$, classical practice would be to construct the confidence region

$$(\hat{\beta}(n) - \beta)(X_nX_n')(\hat{\beta}(n) - \beta)' \leqq d^2$$

which, if F were the cumulative of the normal distribution, would then have probability of coverage equal to $P\{\sigma^2\chi_p^2 \leqq d^2\}$ (and, hopefully, would have this property asymptotically for any F). It is obvious, however, that unless this probability of coverage is equal to α, such a region cannot be of use to us.

To find a confidence interval of fixed width $2d$ for any one of the β_i, we could (in analogy to [3]) use the interval $\hat{\beta}_i(n) \pm d$. Indeed, for any normed linear combination $a\beta'$, a: $1 \times p$, $aa' = 1$, of the β_i, $i = 1, \cdots p$, we could use the confidence interval $a\hat{\beta}'(n) \pm d$. Let us now ask for a confidence region R_n that would be contained in *all* of these confidence intervals. One such region is

$$R_n = \{z\colon (z - \hat{\beta}(n))(z - \hat{\beta}(n))' \leqq d^2\}, \tag{2.3}$$

since for any a such that $aa' = 1$, any $z \varepsilon R_n$,

$$(a(z - \hat{\beta}(n))')^2 \leqq \max_{aa'=1} (a(z - \hat{\beta}(n))')^2 = (z - \hat{\beta}(n)(z - \hat{\beta}(n))' \leqq d^2.$$

We shall adopt this region for our confidence procedure.

Since in our problem σ^2 is unknown, classical theory would suggest the least-squares estimate

$$\hat{\sigma}^2(n) = Y_n(I_n - X_n'(X_nX_n')^{-1}X_n)Y_n' \tag{2.4}$$

as an estimate of σ^2.

Before presenting our class of sequential procedures $\mathcal{C}$, we digress briefly to consider some asymptotic properties of $\hat{\beta}(n)$ and $\hat{\sigma}^2(n)$ for large n. These properties will be important in our discussion of the asymptotic properties of the class $\mathcal{C}$, and are of interest on their own merits.

3. Asymptotic theory for large n. The asymptotic distribution theory for $\hat{\beta}(n)$ is merely a corollary of the following well-known theorem:

THEOREM 3.1. *If $z_1, z_2, \cdots$ are independent identically distributed random variables, each with mean 0, variance 1, and cumulative distribution function G, and if b_{ni}, $i = 1, \cdots n$, $n = 1, 2, \cdots$, is a fixed array of constants with*

$$\textstyle\sum_{i=1}^n b_{ni}^2 = 1, \qquad n = 1, 2, \cdots,$$

then if

$$\max_{1 \leqq i \leqq n} |b_{ni}| \to 0, \qquad n \to \infty, \tag{3.1}$$

we have

$$\lim \mathcal{L}(\textstyle\sum_{i=1}^n b_{ni}z_i) = \mathfrak{N}(0, 1). \tag{3.2}$$

PROOF. This theorem is an immediate consequence of the "particular case" of Theorem 3 in Gnedenko-Kolmogorov [6], p. 103. Of interest in connection with Theorem 3.1 above is the work of Eicker [4].||

Returning to our problem, let

$$U_n = (X_n X_n')^{-\frac{1}{2}} X_n = ((u_{n \cdot ij})).$$

COROLLARY 3.2. *If*

$$\max_{i,j} |u_{n \cdot ij}| \to 0 \quad as \quad n \to \infty, \tag{3.3}$$

then

$$\lim_{n\to\infty} \mathcal{L}((\hat{\beta}(n) - \beta)(X_n X_n')^{\frac{1}{2}}) = \mathfrak{N}(0, \sigma^2 I_p). \tag{3.4}$$

PROOF. For any $1 \times p$ vector a such that $aa' = 1$ consider

$$t_n(a) = \sigma^{-1} a (X_n X_n')^{\frac{1}{2}} (\hat{\beta}(n) - \beta)' = \sigma^{-1} a U_n (Y_n - EY_n)'.$$

Then $t_n(a)$ is a linear combination of the elements of $Z_n = \sigma^{-1}(Y_n - EY_n)$, the elements of Z_n being independent, identically distributed with means 0, variances 1, and cumulative distribution functions $F(x/\sigma)$. Further $aU_nU_n'a' = 1$ so that we may apply Theorem 3.1 provided that the maximum element of the vector aU_n tends to 0 as $n \to \infty$. This follows from the fact that (letting $U_n = (U_n^{(1)}, \cdots U_n^{(n)})$, $U_n^{(i)}: p \times 1$):

$$|aU_n^{(i)}| = |\textstyle\sum_{i=1}^n a_i u_{n \cdot ij}| \leqq \sum_{i=1}^p |a_i||u_{n \cdot ij}|$$
$$\leqq p^{\frac{1}{2}} \max_{i,j} |u_{n \cdot ij}| \to 0, \quad \text{all } j.$$

Thus by Theorem 3.1

$$\lim_{n\to\infty} \mathcal{L}(t_n(a)) = \mathfrak{N}(0, 1).$$

Since this is true for all a, $aa' = 1$, it follows from well-known theorems in multivariate analysis and large-sample theory that (3.4) holds.||

A sufficient condition for (3.3) to hold is that:

ASSUMPTION 3.1. There exists a $p \times p$ positive definite matrix Σ such that

$$\lim_{n\to\infty} n^{-1}(X_n X_n') = \Sigma. \tag{3.5}$$

ASSUMPTION 3.2. $\lim_{n\to\infty} X_n/n^{\frac{1}{2}} = 0$.

Under these assumptions we can find the asymptotic probability of coverage of the region R_n.

COROLLARY 3.3. *Under Assumptions* 3.1 *and* 3.2,

$$P\{(\hat{\beta}(n) - \beta)(\hat{\beta}(n) - \beta)'/n \leqq d^2\} = P\{T(\lambda_1, \cdots \lambda_p) \leqq d^2/\sigma^2\} \tag{3.6}$$

where $\lambda_1, \cdots \lambda_p$ are the characteristic roots of Σ^{-1} and $T(\lambda_1, \cdots \lambda_p)$ has the distribution of a weighted sum of p independent chi-squared variables with one degree of freedom, the λ_i's being the weights.

PROOF. By Corollary 3.2 we have that $\lim \mathcal{L}((\hat{\beta}(n) - \beta)X_nX_n') = \mathfrak{N}(0, \sigma^2 I_p)$. Further

$(\hat{\beta}(n) - \beta)(\hat{\beta}(n) - \beta)' = (\hat{\beta}(n) - \beta)(X_nX_n)^{\frac{1}{2}}(X_nX_n')^{-1}(X_nX_n')^{\frac{1}{2}}(\hat{\beta}(n) - \beta)'.$

Since $n^{-1}X_nX_n'$ converges to Σ, it follows that

$$\lim \mathcal{L}(n^{-1}(\hat{\beta}(n) - \beta)(\hat{\beta}(n) - \beta)') = \mathcal{L}(w\Sigma^{-1}w')$$

where $\mathcal{L}(w) = \mathfrak{N}(0, \sigma^2 I_p)$. An application of a well-known theorem in multivariate analysis completes the proof.||

Finally we need to consider the question of the consistency of $\hat{\sigma}^2(n)$.

THEOREM 3.4. *If* $\max_{i,j} |u_{n\cdot ij}| \to 0$ *as* $n \to \infty$, *then*

$$\lim \hat{\sigma}^2(n) = \sigma^2, \quad \text{a.s.} \tag{3.7}$$

PROOF. Since for $Z_n = Y_n - EY_n$, $\hat{\sigma}^2(n) = n^{-1}Z_n(I - U_n'U_n)Z_n'$ and since by the given $U_n \to 0$ as $n \to \infty$, then

$$\hat{\sigma}^2(n) = n^{-1}Z_nZ_n'(1 + o(1))$$

which by the strong law of large numbers converges a.s. to σ^2.||

Before closing this section, we would like to note that Theorem 3.1 is really a special case of a more general theorem by Eicker [4]. Generalizations of the present paper in directions suggested by this theorem are possible and will be considered in a subsequent paper.

We are now in a position to discuss the main topic of this paper, namely the class of sequential procedures $\mathcal{C}$. This is taken up in the next section.

4. Asymptotic properties of the class $\mathcal{C}$. Given d and α and for a fixed sequence of x-vectors, $x^{(1)}, x^{(2)}, \cdots$ arranged so that X_p is non-singular and so that Assumptions 3.1 and 3.2 are satisfied, let $\{a_n\}$ be any sequence of constants converging to the number a^* satisfying

$$P\{T(\lambda_1, \cdots \lambda_p) \leqq a^*\} = \alpha. \tag{4.1}$$

Then this sequence $\{a_n\}$ determines a member of the class $\mathcal{C}$ of sequential procedures as follows:

(I) We start by taking $n_0 \geqq p$ observations $y_1, \cdots y_{n_0}$. We then sample one extra observation at a time, stopping according to the stopping variable N defined by

$$N = \text{smallest } k \geqq n_0 \quad \text{such that} \quad k^{-1}(\hat{\sigma}^2(k) + k^{-1}) \leqq d^2/a_k. \tag{4.2}$$

(II) When sampling is stopped at $N = n$, construct the region R_n described in (2.3).

Then the procedures in the class $\mathcal{C}$ are asymptotically "consistent" and "efficient" as $d \to 0$. That is

THEOREM 4.1. *Under the assumption that* $0 < \sigma^2 < \infty$,

$$\lim_{d\to0} (d^2N)/(a^*\sigma^2) = 1 \quad \text{a.s.}, \tag{4.3}$$

$$\lim_{d\to0} P(\beta \in R_N) = \alpha, \tag{4.4}$$

and

$$\lim_{d\to 0} (d^2 EN)/(a^*\sigma^2) = 1. \tag{4.5}$$

REMARKS. 1. As in [3], the adding of n^{-1} to $\hat{\sigma}^2(n)$ in (4.1) is unnecessary if F is continuous.

2. As in [3], N could be defined as the smallest odd, even, etc. integer $\geqq n_0$ such that (4.1) holds and the above result would go through.

PROOF OF THE THEOREM. Since $n^{-1}\hat{\sigma}^2(n)$ converges to σ^2 a.s. and since $\hat{\sigma}^2(n) + n^{-1}$ is a.s. positive, then Lemma 1 of [3] implies (4.3). Further Lemma 3 and the discussion following in [3] apply to $\hat{\sigma}^2(n) + n^{-1}$ as well, and thus (4.5) follows. It only remains to prove (4.4).

Since

$$P\{\beta \,\varepsilon\, R_N\} = P\{(\hat{\beta}(N) - \beta)(\hat{\beta}(N) - \beta)' \leqq d^2\}$$
$$= P\{N(\hat{\beta}(N) - \beta)(\hat{\beta}(N) - \beta)'/\sigma^2 \leqq Nd^2/\sigma^2\},$$

since $Nd^2/\sigma^2 \to a^*$ a.s., and since by Corollary (3.3)

$$\lim_{n\to\infty} P\{n(\beta(n) - \beta)(\beta(n) - \beta)'/\sigma^2 \leqq a^*\} = \alpha,$$

it follows as a trivial extension (to the distribution of $T(\lambda_1, \cdots \lambda_p)$) of a result of Anscombe [2] that (4.4) holds.||

Very little is known about the properties of any member of the class $\mathcal{C}$ for moderate values of σ^2/d^2. Some work done on this problem by N. Starr will soon be available.

REFERENCES

[1] ANDERSON, T. W. (1962). Least squares and best unbiased estimates. *Ann. Math. Statist.* **33** 266–272.

[2] ANSCOMBE, F. J. (1952). Large sample theory of sequential estimation. *Proc. Camb. Phil. Soc.* **48** 600–607.

[3] 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.

[4] EICKER, F. (1963). Central limit theorems for families of sequences of random variables. *Ann. Math. Statist.* **34** 439–447.

[5] EICKER, F. (1963). Asymptotic normality and consistency of the least squares estimators for families of linear regression. *Ann. Math. Statist.* **34** 447–457.

[6] GNEDENKO, B. V. AND KOLMOGOROV, A. N. (1954). *Limit Distribution for Sums of Independent Random Variables*, (translated from the Russian by K. L. Chung). Addison-Wesley, Cambridge.