Jour. Ind. Soc. Ag. Statistics
Vol. XL, No. 2 (1988), pp. 111-115

ON A MODIFIED SEQUENTIAL PROCEDURE TO CONSTRUCT CONFIDENCE REGIONS FOR REGRESSION PARAMETERS

AJIT CHATURVEDI
Lucknow University, Lucknow
(Received : February, 1987)

SUMMARY

A "modified" sequential procedure is developed to construct an ellipsoidal confidence region for regression parameters in a linear model. The proposed procedure requires fewer sampling operations than that required by the purely sequential procedure and is strongly competitive with it. The moderate sample-size performance is studied to demonstrate practical applicability of the procedure. The basic ideas are borrowed from Hall [3].

Keywords: Linear model; Regression parameters; Ellipsoidal confidence region; Sampling stages.

Introduction and Preliminaries

Let us consider the linear model

$$\underset{\sim}{Y}_n = X_n \underset{\sim}{\beta} + \underset{\sim}{\varepsilon}_n \tag{1.1}$$

where $\underset{\sim}{Y}_n$ is an observed $n \times 1$ random vector, X_n is a known $n \times p$ matrix of rank p, $\underset{\sim}{\beta}$ is $p \times 1$ vector of unknown parameters, and ε_n is the disturbance term following $N_n(\underset{\sim}{0}, \sigma^2 I_n)$ distribution.

For specified d, α $(d > 0, 0 < \alpha < 1)$, suppose one wishes to construct an ellipsoidal confidence region R_n for β, such that the maximum diameter of $R_n \leqslant 2d$ and $P(\underset{\sim}{\beta} \in R_n) \geqslant \alpha$. Following Srivastava [5], it is defined by

$$R_n = [Z : n^{-1} (\underset{\sim}{Z} - \hat{\underset{\sim}{\beta}}_n)' (X_n' X_n) (\underset{\sim}{Z} - \hat{\underset{\sim}{\beta}}_n) \leqslant d^2]$$

where $\hat{\underset{\sim}{\beta}}_n = (X_n'X_n)^{-1} X_n' \underset{\sim}{Y}_n$. It is easy to verify that

$$P(\underset{\sim}{\beta} \varepsilon R_n) = F(nd^2/\sigma^2) \tag{1.1}$$

where $F(.)$ stands for the c.d.f. of a $\chi^2_{(p)}$ r.v. Let "a" by any constant such that

$$F(a) = \alpha \tag{1.2}$$

It is obvious from (1.1) and (1.2) that to achieve $P(\underset{\sim}{\beta} \varepsilon R_n) \geqslant \alpha$, the sample size needed is the smallest positive integer $n \geqslant n_0$, where $n_0 = a\sigma^2/d^2$. In the absence of any knowledge about σ, we adopt the following sequential procedure.

Let, for $n \geqslant m$ $(\geqslant p + 1)$, $\hat{\sigma}^2_n = (n - p)^{-1} \underset{\sim}{Y}_n' [I_n - X_n (X_n'X_n)^{-1} X_n'] \underset{\sim}{Y}_n$. Then, the stopping time $N \equiv N(d)$ is defined by

$$N = \inf [n \geqslant m : n \geqslant a \hat{\sigma}^2_n/d^2] \tag{1.3}$$

Recently, Mukhopadhyay and Abid [4] obtained second-order approximations for the sequential procedure (1.3) and proved the following theorem.

THEOREM 1. *For the rule* (1.3), *we have as* $d \to 0$,

$$E(N) = n_0 + \nu - 2 - p + 0(1); \quad \text{if } m \geqslant p + 3 \tag{1.4}$$

$$P(\underset{\sim}{\beta} \varepsilon R_N) = \alpha + (d/\sigma)^2 [\nu - 3 - \{(p + a)/2\}] f(a) + 0(\alpha^2) \tag{1.5}$$

if (i) $m \geqslant p + 3$ *for* $p = 2$ *or* $p \geqslant 4$ *and (ii)* $m \geqslant 7$ *for* $p = 3$. *Here* ν *is specified and* $f(.)$ *denotes the p.d.f. of a* $\chi_{(p)}$ r.v.

In the next section, following Hall [3], we shall develop a "modified" sequential procedure, which requires fewer number of sampling operations than that required by the purely sequential procedure (1.3) and is strongly competitive with it.

2. A Modified Sequential Procedure

Let $0 < \rho < 1$ and $K(\geqslant 0)$ be specified constants. Take m $(\geqslant p + 3)$ as the initial sample size and start sampling sequentially with stopping rule N_1, defined by

$$N_1 = \inf\,[n_1 \geqslant m : n_1 \geqslant \rho\,(a/d^2)\,\hat{\sigma}^2_{n_1}]$$

Then jump ahead by taking N_2 observations, given by

$$N_2 = [\,(a/d^2)\ \hat{\sigma}\ N_1^2 + K] + 1 \qquad (2.1)$$

where K is to be evaluated. Define $M = \max\ (N_1,\ N_2)$ and construct R_M for $\underset{\sim}{\beta}$.

Now we state the following theorem.

THEOREM 2. $\lim\limits_{d\to 0} N_1 = \lim\limits_{d\to 0} N_2 = \infty$ a.s. (2.2)

$$\lim_{d\to 0}\ (M/n_0) = 1 \text{ a.s.} \qquad (2.3)$$

$$\lim_{d\to 0}\ E(M/n_0) = 1 \qquad (2.4)$$

$$(2\rho n_0)^{-1/2}\,(N_1 - \rho n_0) \overset{\mathscr{X}}{\to} N(0,\,1) \text{ as } d \to 0 \qquad (2.5)$$

Proof. The proofs of (2.2) and (2.3) follow from the definitions of N_1, N_2 and M. For the proofs of (2.4) and (2.5) we refer to Hall [3] and Ghosh and Mukhopadhyay [2], respectively.

The main results of this section are stated in the next theorem.

THEOREM 3. *For all* $m \geqslant p + 3$, *as* $d \to 0$,

$$E(M) = n_0 + K - p - 2 - \rho^{-1} + 0\,(1) \qquad (2.6)$$

$$\text{Var}\ (M) = 2\rho^{-1}\, n_0 + 0\,(d^{-2}) \qquad (2.7)$$

$$E/M - EM/^r = 0\,(d^{-r}),\ \text{ for } r > 0 \qquad (2.8)$$

and, for all $K \geqslant \max\,\{p + 2 + \rho^{-1}\,(5 + 2a - 2p),\,0\}$

$$P(\underset{\sim}{\beta} \in R_M) \geqslant \alpha + 0\,(d^2) \qquad (2.9)$$

Proof. The proofs of (2.6)-(2.8) can be obtained exactly along the lines of that of Theorem 2 in Hall [3] after necessary modifications at various places.

Using a Taylor series expansion, we obtain

$$P(\underset{\sim}{\beta} \varepsilon R_M) - \alpha = E[F(Ma/n_0) - F(a)]$$
$$= E[(Ma/n_0 - a) F'(a) + (1/2)(Ma/n_0 - a)^2 F''(a)] + r(d)$$

where the remainder $r(d) = 0 \; (d^6 E\diagdown M - EM\diagdown^3) = 0 \, (d^2)$, on using (2.8). Now, using (2.6) and (2.7), we obtain after some algebra

$$P(\underset{\sim}{\beta} \varepsilon R_M) - \alpha = (a/n_0)(K - p - 2 - \rho^{-1}) f(a) + (a^2/2n_0^2) \{ - 1/2$$
$$+ (p/2 - 1) a^{-1}\} \{\text{Var}(M)\} f(a) + 0 \, (d^2)$$

$= (a/n_0) [K - \{p + 2 + \rho^{-1} (5 + 2a - 2p)\}] f(a) + 0 \, (d^2)$ and the theorem follows.

3. The Moderate Sample Performance

Table 1 presents the results of Monte-Carlo experiment with pseudo-random bivariate normal deviates. We fix $p = 2$, $\sigma = 1$, $m = 8$, $\alpha = .95$. For 6 values of d, we conducted 500 trials for $\rho = .25, .5, .75$. For each set of 500 values of M, we computed the mean $\bar{M}$, as well as, the coverage probability P that the confidence region covers the origin. The procedure behaves quite satisfactorily. As expected, $\bar{M}$ approaches to the optimum sample size n_0 as ρ increases.

TABLE 1 : RESULTS OF 500 MONTE-CARLO TRIALS WITH $p = 2, \sigma = 1, m = 8, \alpha = .95$

		$\rho = .25, K = 9$		$\rho = .5, K = 7$		$\rho = .75, K = 6$	
d	n_0	$\bar{M}$	P	$\bar{M}$	P	$\bar{M}$	P
.015	458	467.2	.958	465.7	.952	459.6	.959
.025	165	178.3	.959	175.2	.985	163.7	.963
.05	42	45.7	.948	44.9	.951	44.5	.968
.075	19	22.7	.954	21.3	.954	20.9	.964
.095	12	15.8	.951	15.2	.956	13.8	.955
.100	11	16.3	.950	14.9	.954	12.6	.953

ACKNOWLEDGEMENT

I am thankful to the referee for his valuable comments. I am also grateful to Mr. S. A. Akbar, Arkansas State University, for his computational assistance.

REFERENCES

[1] Chaturvedi, A. (1987) : Sequential point estimation of regression parameters in a linear model, *Ann. Inst. Statist. Math*. A39 : 55-67.

[2] Ghosh, M. and Mukhopadhyay, N. (1979) : Sequential point estimation of the mean when the distribution is unspecified, *Commun. Statist.-Theor. Meth*. A8(7) : 637-652.

[3] Hall, P. (1983) : Sequential estimation saving sampling operations, *Jour. Roy. Statist. Soc*. B45 : 219-223.

[4] Mukhopadhyay, N. and Abid, A. D. (1986) : On fixed-size confidence regions for the regression parameters, *Metron* XLIV : 297-306.

[5] Srivastava, M. S. (1967) : On fixed-width confidence bounds for regression parameters and mean vector, *Jour. Roy. Statist. Soc*. B28 : 132-140.