Journal of the Indian Statistical Association
Vol. 12, 1974, 39–43

SEQUENTIAL ESTIMATION OF REGRESSION PARAMETERS IN GAUSS-MARKOFF SETUP

by
Nitis Mukhopadhyay
Indian Statistical Institute, Calcutta, India

Abstract

In Gauss-Markoff linear estimation with quadratic loss structure, a sequential point estimator for the regession parameters is suggested The procedure is shown to have asymptotic risk efficiency and bounded regret.

1. Introduction

Motivated by the classical paper of Chow and Robbins (1965), L. J. Gleser investigated the problem of fixed size bounds for regression parameters with Gauss-Markoff setup (1965, 1966). The purpose of this paper is to consider the analogous problem of estimating the regression parameters pointwise.

2. Procedure

Consider a sequence $Z_1, Z_2, \ldots$ of independent and normally distributed random variables (r. v.'s) such that

$$Z_i = x'_{(i)} \underline{\beta} + \epsilon_i \quad (i = 1, 2, \ldots) \tag{2.1}$$

where $\underline{\beta}$ is a $m \times 1$ vector of unknown parameters, $x_{(i)}$ is a $m \times 1$ vector of non-stochastic known constants with ϵ_i distributed

Received : 26th July 1974.

as N $(0, \sigma^2)$, Cov $(\epsilon_i, \epsilon_j) = 0$ for all $i, j \, (i \neq j)$, σ being unknown; (as a convention, for any $p \times q$ matrix A, A' and R (A) mean respectively the transpose and rank of A). We start with a sample size K $(\geqslant m+2)$ making sure that R $(X_K) = m$, where $X'_n = (x_{(1)}, x_{(2)}, \ldots, x_{(n)})$ and $Y'_n = (Z_1, Z_2, \ldots, Z_n)$ for any $n \geq$ K. One is referred to Gleser (1965).

It is well known (See e. g., Rao (1965)) that a least square estimator of β with model (2.1) on the basis of a sample of size n is

$$\underline{\beta}_n = (X'_n X_n)^{-1} X'_n Y_n \tag{2.2}$$

with dispersion matrix

$$V(\underline{\beta}_n) = \sigma^2 \, (X'_n X_n)^{-1} \; . \tag{2.3}$$

Suppose the loss incurred in estmating $\underline{\beta}$ by $\underline{\beta}_n$ from a sample of fixed size n is

$$L_n = n^{-1} (\underline{\beta}_n - \underline{\beta})' (X'_n X_n) (\underline{\beta}_n - \underline{\beta}) + n \tag{2.4}$$

with risk

$$\begin{aligned} v_n(\sigma) &= E_\sigma (L_n) \qquad (2.5) \\ &= E_\sigma \{ n^{-1} \operatorname{tr} (\underline{\beta}_n - \underline{\beta})' (X'_n X_n) (\underline{\beta}_n - \underline{\beta}) \} + n \\ &= n^{-1} \sigma^2 \operatorname{tr} (I_{m \times m}) + n \\ &= m\sigma^2 / n + n \end{aligned}$$

where trA means trace of the matrix A and $I_{m \times m}$ stands for the identity matrix of order $m \times m$. If σ were known, the problem of finding the value of n, say n^0, for which the risk (2.5) is a minimum is perfectly straight forward yielding

$$n^0 = m^{\frac{1}{2}} \sigma \tag{2.6}$$

and minimum risk

$$v(\sigma) = v_{n^0}(\sigma) = 2m^{\frac{1}{2}} \sigma. \tag{2.7}$$

But, in ignorance of σ, no fixed sample size procedure will minimize (2.5) simultaneously for all $0 < \sigma < \infty$. So the possibility of utilising a sample of random size N determined by the following sequential rule $\mathcal{R}$ is considered.

$\mathcal{R}$: The stopping number N is the first positive integer $n \geqslant K$ such that

$$n \geqslant \left[m R^2_{0n}(n-m)^{-1}\right]^{\frac{1}{2}} \tag{2.8}$$

where $R^2_{0n} = Y'_n Y_n - Y'_n X_n \beta_n$, starting sample size being $K\ (\geqslant m+2)$.

The rule $\mathcal{R}$ can be rephrased as

$\mathcal{R}^*$: The stopping number N is the first integer $n \geqslant K$ such that

$$V_n \leqslant l\ (n, \sigma) \tag{2.9}$$

where $V_n = (R_{0n}/\sigma)^2$, $l\ (n, \sigma) = n^2(n-m)/m\sigma^2$.

We now state the following

Lemma For any fixed integer $n\ (\geqslant K)$, $\underline{\beta}_n$ is independent of the vector $(V_K, V_{K+1}, \ldots, V_n)$.

Proof For any integer p in $[K, n]$,

$$R^2_{0p} = Y'_p\left[I - X_p(X'_pX_p)^{-1}X'_p\right]Y_p,\ I = (\delta_{i,j}),\ 1 \leqslant i, j \leqslant p \tag{2.10}$$

$$= Y_p\left(\sum_{i=1}^{m} \xi_i \underline{\xi}_i\right) Y_p$$

$$= \sum_{i=1}^{m} (\underline{\xi}'_i\ Y_p)^2$$

where ξ'_i are othonormal eigenvectors of the idempotent matrix $\left[I - X_p\ (X'_p\ X_p)^{-1}\ X'_p\right]$ associated eigenvalues being thereby all unity $(i = 1, 2, \ldots, m)$. Use the symbol 0 for the null vector, irrespective of dimension. Then we can write,

$$R^2_{0p} = \sum_{i=1}^{m} (\underline{\rho}'_i\ Y_n)^2 \tag{2.11}$$

where $\underline{\rho}'_i = (\underline{\xi}'_i \vdots 0')$ is a $1 \times n$ vector. Let $\begin{pmatrix} X_p \\ U_{n-p} \end{pmatrix}$ be the corresponding partition of X_n. From (2.2), $\underline{\beta}_n = B\, Y_n$ where $B = (X'_n X_n)^{-1} X'_n$. A sufficient condition for $B\, Y_n$ and $\rho'_i Y_n$ to be distributed independently is $B\, \rho_i = 0$ (see Rao (1965)). Now for verifying this sufficient condition (using the notations of Rao (1966)), note that $\underline{\xi}_i \in \mathcal{M}[I - X'_p (X'_p X_p)^{-1} X'_p]$ implying $\xi_i \in \Theta[X_p (X'_p X_p)^{-1} X'_p] = \Theta(X_p)$, since $(X'_p X_p)^{-1} X'_p$ is a generalised inverse of X_p. This gives $X'_p \underline{\xi}_i = 0$ implying $X'_n \underline{\rho}_i = 0$. Hence $B\, \underline{\rho}_i = 0$, and it completes the proof of the lemma.

Using this lemma, one can say that the event $[N = n]$ and L_n are independent for all $n \geq K$, and one gets

$$\bar{\nu}(\sigma) = E(L_N)$$
$$= m\sigma^2 E(N^{-1}) + E(N), \quad 0 < \sigma < \infty, \tag{2.12}$$

and

$$\eta(\sigma) = \bar{\nu}(\sigma)/\nu(\sigma)$$
$$= \tfrac{1}{2}\left[n^0 E(N^{-1}) + E(N/n^0)\right] \tag{2.13}$$

Also,

$$\omega(\sigma) = \bar{\nu}(\sigma) - \nu(\sigma)$$
$$= \left[(n^0)^2 E(N^{-1}) - n^0\right] + \left[E(N) - n^0\right]. \tag{2.14}$$

Regarding efficiencies of our procedure $\mathcal{R}$ in (2.3), we have the following theorems.

Theorem 1 $\lim_{\sigma \to \infty} \eta(\sigma) = 1$.

Theorem 2 $\lim_{\sigma \to \infty} \omega(\sigma) = O(1)$.

Modifying the proof of theorem 2 in Mukhopadyay (1973a) or theorem

3 in Starr (1966) one can prove theorem 1. One can get a proof of theorem 2 by modifying the proof of Lemma 4.1 in Mukhopadhyay (1973 b). However, one can refer to Starr & Woodroofe (1969) also. Here main thing to be noted is that V_n is distributed as χ^2 with $(n-m)$ degrees of freedom.

Acknowledgment

It is a pleasure to thank Dr. Malay Ghosh for his guidance during this investigation.

References

1. Chow, Y. S. and Robbins, H. (1965). On the asymptotic theory of fixed width sequential confidence intervals for the mean. *Ann. Math. Statist.* 36, 447–462.
2. Gleser, L. J. (1965). On the asymptotic theory of fixed size sequential confidence bounds for linear regression parameters. *Ann. Math. Statist.* 36. 463–467.
3. Gleser L. J. (1966). Correction to- On the asymptotic theory of fixed size sequential confidence bounds for linear regression parameters. *Ann. Math. Statist.* 37, 1053–1055.
4. Mukhopadhyay, N. (1973a). Sequential estimation of location parameter in exponential distributions. *Ind. Stat. Inst. Tech. Report Math-Stat.* 18/73.
5. Mukhopadhyay, N. (1973 b). Sequential estimation of the difference of two means : the normal case. *Ind. Stat. Inst. Tech. Report Math-Stat.* 22/73.
6. Rao, C. R. (1965). *Linear Statistical Inference and Its Applications.* New York John Wiley and Sons.
7. Starr, N. (1966). On the asymptotic efficiency of a sequential procedure for estimating the mean. *Ann. Math. Statist.* 37, 1173–1185.
8. Starr, N. and Woodroofe, M. B. (1969). Remarks on sequential point estimation. *Proc. Nat. Acad. Sci. USA*, 63, 285–288.