ADMISSIBILITY OF THE MAXIMUM LIKELIHOOD ESTIMATOR IN THE REGRESSION OF TWO PREDICTANDS ON ONE PREDICTOR

STANLEY L. SCLOVE

(Received March 25, 1969)

Summary

Stein [2] has shown that the maximum likelihood estimator (MLE) of the regression coefficients is admissible in unvariate regression with one predictor or with two predictors and known means. In a similar way it is shown in the present note that the MLE is admissible when there are two predictands and one predictor and the means are known.

1. Introduction and statement of theorem

Let $Z_1, \cdots, Z_N$ be independent identically distributed random variables, where, for $k=1, \cdots, N$, $Z_k'=(Y_k', X_k)$, Y_k is a vector of two components, X_k is a scalar, and Z_k is normally distributed with mean

$$\mu=\begin{pmatrix}\mu_Y\\ \mu_X\end{pmatrix}$$

and covariance matrix

$$\Sigma=\begin{pmatrix}\Sigma_Y & \Sigma_{YX}\\ \Sigma_{XY} & \Sigma_X\end{pmatrix}.$$

The problem is to estimate the parameters of the *regression function*

$$\rho(x_k)=E[Y_k \mid X_k=x]=\alpha+\gamma x ,$$

where γ is a column vector of two components given by

$$\gamma=\Sigma_{YX}/\Sigma_X ,$$

$$\alpha=\mu_Y-\gamma\mu_X .$$

The 2×2 conditional covariance matrix of Y_k given X_k is

$$\Sigma_{Y\cdot X}=\Sigma_{YY}-\Sigma_{YX}\Sigma_{XY}/\Sigma_X .$$

(We assume $\Sigma_{Y\cdot X}$ is nonsingular; this holds if Σ is nonsingular.) Thus the conditional distribution of Y_k, given $X_k=x$, is $\mathfrak{N}(\alpha+\gamma x, \Sigma_{Y\cdot X})$.

When the mean μ is known (admittedly an artificial situation), the problem is reduced to estimating γ only. The result to be proved is the

THEOREM. *If $N\geqq 5$, the MLE is admissible for the problem of estimating γ when the mean μ is known and the loss function is*

$$L(\Sigma, \hat{\gamma})=\operatorname{tr} \Sigma_{Y\cdot X}^{-1}(\hat{\gamma}-\gamma)\Sigma_X(\hat{\gamma}-\gamma)' .$$

The choice of this type of loss function is explained in [2].

Define the statistics $\bar{Y}=\sum_{k=1}^{N} Y_k/N$ and $\bar{X}=\sum_{k=1}^{N} X_k/N$, $T=\sum_{k=1}^{N}(Y_k-\mu_Y)\cdot(Y_k-\mu_Y)'$, $U=\sum_{k=1}^{N}(X_k-\mu_X)(Y_k-\mu_Y)'$, and $V=\sum_{k=1}^{N}(X_k-\mu_X)^2$. Then the MLE's are: for γ,

$$g=U'/V ,$$

for Σ,

$$S=\frac{1}{N}\begin{pmatrix} T & U' \\ U & V \end{pmatrix},$$

and for $\Sigma_{Y\cdot X}$, $S_{Y\cdot X}/N$, where

$$S_{Y\cdot X}=T-U'U/V=T-Vgg' .$$

2. Proof of the theorem

The proof is accomplished by applying a general result due to Stein (see Lemma 2 of [2]), which we restate here:

LEMMA. *Let $\mathcal{B}$ be the σ-algebra of all Borel subsets of the two-dimensional real coordinate space $\mathcal{X}$ and $\mathcal{G}$ a σ-algebra of subsets of a set $\mathcal{Y}$. Let μ be Lebesgue measure on $\mathcal{B}$ and ν a probability measure on $\mathcal{G}$. Let f be a non-negative-valued $\mathcal{B}\mathcal{G}$ measurable function on $\mathcal{X}\times\mathcal{Y}$ such that*

$$\int f(x, y)dx=1 , \qquad \textit{for all } y , \tag{1}$$

$$\int x_i f(x, y)dx=0 , \qquad i=1, 2 , \quad \textit{for all } y , \tag{2}$$

and

$$\int d\nu(y)\left[\int (x'x)^{1+\varepsilon} f(x, y)dx\right]^2<\infty , \tag{3}$$

for some $\varepsilon>0$, where $dx=d\mu(x)$. Then, if we observe (X, Y) distributed *so that, for some unknown ξ, $(X-\xi, Y)$ has probability density $f(x, y)$ with respect to $\mu\ \nu$, the estimator X is an admissible estimator of ξ with loss function*

$$(\hat{\xi}-\xi)'A(\hat{\xi}-\xi)\,, \tag{4}$$

for any constant positive semidefinite matrix A.

PROOF OF THE THEOREM. We remark that (2) states $E[X-\xi \mid Y]=0$ and (3) states

$$E\{E^2[((X-\xi)'(X-\xi))^{1+\varepsilon} \mid Y]\} < \infty\,. \tag{5}$$

Since μ_Y and μ_X are known, the sufficient statistic is the triplet $(g, V, S_{Y\cdot X})$. To apply the lemma, take $X=g$, $\xi=\gamma$, and $Y=(V, S_{Y\cdot X})$; take $A=\Sigma_X\Sigma_{Y\cdot X}^{-1}$ so that (4) becomes $L(\Sigma, \hat{\gamma})$.

Given V, g is $\mathcal{N}(\gamma, V^{-1}\Sigma_{Y\cdot X})$; the statistic V is $\Sigma_X\chi_N^2$, $S_{Y\cdot X}$ has the Wishart distribution $\mathcal{W}(\Sigma_{Y\cdot X}, N-1)$ ([1], p. 158) and is independent of g and V. The density $f(x, y)$ is the conditional density of $X-\xi=g-\gamma$ given $(V, S_{Y\cdot X})=y$. Since $S_{Y\cdot X}$ is independent of g and V, this is the conditional density of $g-\gamma$ given V, which is $\mathcal{N}(0, V^{-1}\Sigma_{Y\cdot X})$. First suppose that $\Sigma_{Y\cdot X}$ and Σ_X are known. Then $f(x, y)$ and ν are specified and the conditions (1) and (2) are met.

It remains to show that condition (5) is satisfied. Define $U^*=U-V\gamma'$. Then we have $(X-\xi)'=g'-\gamma'=V^{-1}U-\gamma'=V^{-1}U^*$, and $(X-\xi)'(X-\xi)=U^*U^{*\prime}/V^2$.

The finiteness of the left-hand side of (5) does not depend upon the values of $\Sigma_{Y\cdot X}$ and Σ_X; we assume $\Sigma_{Y\cdot X}=I$ and $\Sigma_X=1$. Then the conditional distribution of $U^{*\prime}$ given V is $\mathcal{N}(0, VI)$ and that of $U^{*\prime}/V^{1/2}$ is $\mathcal{N}(0, I)$, so that the conditional distribution of $W=U^*U^{*\prime}/V$ is χ_2^2. Since this conditional distribution does not depend upon V, it is also the unconditional distribution. The statistic $U^*U^{*\prime}/V^2=W/V$. We have, for any fixed $\varepsilon>0$,

$$\begin{aligned} E^2[(W/V)^{1+\varepsilon} \mid V] &= E^2(W^{1+\varepsilon} \mid V)/V^{2+2\varepsilon} \\ &= E^2(W^{1+\varepsilon})/V^{2+2\varepsilon} \\ &= E^2[(\chi_2^2)^{1+\varepsilon}]/V^{2+2\varepsilon} \\ &\leqq MV^{-2-2\varepsilon} \end{aligned}$$

where $0<E^2[(\chi_2^2)^{1+\varepsilon}]<M<\infty$. Thus

$$\begin{aligned} E\{E^2[(U^*U^{*\prime}/V^2)^{1+\varepsilon} \mid V]\} &\leqq ME(V^{-2-2\varepsilon}) \\ &= ME[(\chi_N^2)^{-2-2\varepsilon}] \\ &= M\int_0^\infty x^{N/2-1-2-2\varepsilon}e^{-x/2}dx\,, \end{aligned}$$

which is finite for $N/2-1-2-2\varepsilon>-1$, that is, for $N\geqq 5$, if $\varepsilon<1/4$. This proves that the MLE is admissible if Σ_X and $\Sigma_{Y\cdot X}$ are known; consequently, it is also admissible when they are unknown. (An estimator admissible over each set of a partition of the parameter space is admissible.)

Strictly speaking, we have demonstrated only the admissibility of the MLE in the class of estimators which depend on the sufficient statistic, for we have taken $(g, (V, S))$ as the basic observation. However, by the convexity of the loss function this class is complete, so the MLE is admissible in the class of all estimators.

CARNEGIE-MELLON UNIVERSITY

REFERENCES

[1] Anderson, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*, John Wiley and Sons, Inc., New York.

[2] Stein, C. (1960). Multiple regression, *Contributions to Probability and Statistics—Essays in Honor of Harold Hotelling*, Stanford University Press, Stanford, California, 424–443.