Ann. Inst. Statist. Math.
39 (1987), Part A, 49-54

MINIMAXITY AND NONMINIMAXITY OF A PRELIMINARY TEST ESTIMATOR FOR THE MULTIVARIATE NORMAL MEAN

TAICHI INABA AND YASUSHI NAGATA

(Received Jan. 17, 1985; revised Dec. 16, 1985)

Summary

The problem is to estimate the mean of a p-dimensional normal distribution in the situation where there is vague information that the mean vector might be equal to zero vector. Minimax property of the preliminary test estimator obtained by the use of AIC (Akaike's Information Criterion) procedure is discussed under a loss function which is based on Kullback-Leibler information measure and evaluates both an error of model selection and that of estimation. When p is even, the minimaxity is shown to hold for small values of p but not for large values.

1. Introduction

Let a p-dimensional random vector X follow a normal distribution $N_p(\theta, I_p)$, where I_p is the p-dimensional identity matrix. We wish to estimate the unknown parameter θ in the situation where there is vague information about θ that it is equal to zero vector. We are concerned with the preliminary test estimation and the following estimator is familiar in this case:

$$d_c(X)=\begin{cases} 0 & \text{if } X'X\leqq c \\ X & \text{otherwise}. \end{cases} \tag{1.1}$$

This type of estimator is inadmissible under various standard loss functions for estimation and is not minimax for the squared error loss function (see Sclove, Morris and Radhakrishnan [7]). However, our opinion is that the problem should be formulated as a hybrid problem of model selection and estimation and that it should be discussed under a loss function which is suitable for the situation. From this point of

Key words and phrases: Preliminary test estimator, minimaxity, AIC, Kullback-Leibler information measure.

view we employ the loss function introduced by Inagaki [2] which is based on Kullback-Leibler information measure and evaluates both an error of model selection and that of estimation. Under this loss function, Nagata [4] showed that the procedure (1.1) with $c=2p$ which is obtained by the use of AIC for model selection (see Hirano [1]) is admissible for $p=1$ and inadmissible for $p\geqq 3$, and Nagata and Inaba [5] showed that it is also minimax when $p=1$.

This paper treats a multivariate version of Nagata and Inaba [5]. At first we conjectured that the procedure (1.1) with $c=2p$ would be minimax for any p. Restricting ourselves only to the case with even p because of the simplicity of calculation, we have seen that it is minimax for $p\leqq 12$, but to our surprise that the procedure is not minimax for $p=14$. In Section 2 we formulate the problem and give the result. Proofs are given in Section 3.

2. Formulation of the problem and a result

We describe the loss function by Inagaki [2] based on Kullback-Leibler information measure. Let X be a random variable with p.d.f. (probability density function) $f(x:\theta)\in \boldsymbol{F}=\{f(x:\theta);\theta\in\Theta\}$, where Θ is a parameter space. Suppose that $\boldsymbol{F}_\gamma=\{f_\gamma(x:\zeta);\zeta\in\Theta_\gamma\}$ is a model for $\boldsymbol{F}$ and Θ_γ a parameter space indexed by γ and that $\zeta_\gamma(\theta)$ is defined by the following equation:

$$\int \log\{f(x:\theta)/f_\gamma(x:\zeta_\gamma(\theta))\}f(x:\theta)dx = \min_{\zeta\in\Theta_\gamma}\int \log\{f(x:\theta)/f_\gamma(x:\zeta)\}f(x:\theta)dx\ . \tag{2.1}$$

Inagaki's loss function has the following form:

$$L((k,d),\theta)=\log\{f(x:\theta)/f_k(x:\zeta_k(\theta))\} + \int \log\{f_k(y:\zeta_k(\theta))/f_k(y:\zeta_k(d))\}f_k(y:\zeta_k(\theta))dy\ , \tag{2.2}$$

where $k=k(X)$, $d=d(X)$ and $\zeta_\gamma(d)=\zeta_\gamma(d(X))$ are estimators of the index γ, the unknown parameter θ and $\zeta_\gamma(\theta)$, respectively. He introduced the first term as a loss for model fitting and the second term as a loss incurred by estimation. For further discussion on this loss function, see the original paper by Inagaki [2] or Nagata and Inaba [5].

Now let p-dimensional random vector $X=(X_1, X_2,\cdots, X_p)'$ follow $N_p(\theta, I_p)$ with p.d.f. $f(x:\theta)$ and $\theta=(\theta_1,\theta_2,\cdots,\theta_p)'$. We consider two models, $\boldsymbol{F}_0=\{f(x:0)\}$ and $\boldsymbol{F}_1=\{f(x:\theta);\theta\in R^p\}$. In this problem Hirano [1] discussed the following preliminary test estimator by using AIC procedure,

$$d_{2p}(X)=\begin{cases} 0 & \text{if } X'X\leqq 2p \\ X & \text{otherwise}. \end{cases} \tag{2.3}$$

And Inagaki's loss function (2.2) becomes

$$L((k,d),\theta)=\begin{cases} x'x/2-(x-\theta)'(x-\theta)/2 & \text{if } k=0, \\ (d-\theta)'(d-\theta) & \text{if } k=1. \end{cases} \tag{2.4}$$

Under this formulation of the problem, we obtain the following result.

THEOREM. *If the random vector X follows $N_p(\theta, I_p)$ with even p, then, under the loss function* (2.4), *the procedure* (2.3) *is minimax for $p\leqq 12$, but not for $p=14$.*

A proof of the theorem will be given in Section 3.

3. Proof of the theorem

We shall follow a multivariate extension of the proof in Nagata and Inaba [5] using the following lemma (see Lehmann [3]).

LEMMA 1. *Suppose that there exists a class $\{\pi_\tau\}$ of distributions such that the Bayes risk $r(\tau, d_\tau)$ of the Bayes solution d_τ of θ with respect to π_τ converges to some constant C as τ tends to infinity. If the risk, $R(\theta, d_0)$, of d_0 satisfies that $R(\theta, d_0)\leqq C$ for all θ, then d_0 is a minimax estimator of θ.*

The Bayes solution of θ with respect to the prior distribution $N_p(0, \tau^2 I_p)$ is given by

$$d_\tau(X)=\begin{cases} 0 & \text{if } X'X\leqq 2p/(2-a(\tau)) \\ a(\tau)X & \text{otherwise}, \end{cases} \tag{3.1}$$

where $a(\tau)=\tau^2/(1+\tau^2)$ (see Nagata [4] for the derivation). The risk function of the procedure (2.3) is

$$R(\theta, d_{2p})=\frac{p}{2}+\frac{1}{2}\int_{x'x\leqq 2p}\{x'x-2(x-\theta)'(x-\theta)\}f(x:\theta)dx\ . \tag{3.2}$$

Noting that the conditional distribution of θ given X is $N_p(a(\tau)X, a(\tau)I_p)$, the Bayes risk of the procedure (3.1) is

$$r(\tau, d_\tau)=\frac{p}{2}a(\tau)+\frac{1}{2}\int_{x'x\leqq 2p/(2-a(\tau))}\{x'x(2a(\tau)-a(\tau)^2)-2pa(\tau)\}f_\tau(x)dx\ , \tag{3.3}$$

where $f_\tau(x)$ is the p.d.f. of $N_p(0, (1+\tau^2)I_p)$, the marginal distribution of X. Therefore $r(\tau, d_\tau)\to p/2$ as $\tau\to\infty$. In order to prove the minimaxity of the procedure (2.3) by using Lemma 1, we must establish that the second term of the right-hand side of (3.2) is nonpositive for all $\theta \in R^p$. On the other hand, noting the form of the loss function (2.4), the risk function of the usual estimator $d^*(X)=X$ is

$$R(\theta, d^*)=\frac{p}{2} \qquad \text{for all } \theta \in R^p, \tag{3.4}$$

which establishes the minimaxity of d^* for all p. Putting

$$g(\theta)=\int_{x'x\leq 2p} \{x'x-2(x-\theta)'(x-\theta)\}f(x:\theta)dx, \tag{3.5}$$

we obtain the following lemma.

LEMMA 2. *A necessary and sufficient condition for the procedure* (2.3) *to be minimax is* $g(\theta)\leqq 0$ *for all* $\theta \in R^p$.

PROOF. The sufficiency of this condition is clearly obtained from Lemma 1 and above discussion. The necessity is as follows. Suppose $g(\theta_0)>0$ for some θ_0. Then $R(\theta_0, d_{2p})>p/2$, which shows from (3.4) that d_{2p} is not minimax.

We need the following lemma at this stage of the proof of the theorem.

LEMMA 3. *Suppose the random vector* X *follows* $N_p(0, D)$ *with p.d.f.* $f(x: 0, D)$. *Let* S *be the set of* x *such that* $(x+a)'D^{-1}(x+a)\geqq c$ *for a nonnegative constant* c. *Then*

$$\int_S xf(x: 0, D)dx=a[\Pr\{\chi^2(p;\delta)\leqq c\}-\Pr\{\chi^2(p+2;\delta)\leqq c\}] \tag{3.6}$$

and

$$\begin{aligned}\int_S xx'f(x: 0, D)dx=D[1-&\Pr\{\chi^2(p+2;\delta)\leqq c\}]\\ -aa'[&\Pr\{\chi^2(p;\delta)\leqq c\}-2\Pr\{\chi^2(p+2;\delta)\leqq c\}\\ &+\Pr\{\chi^2(p+4;\delta)\leqq c\}],\end{aligned} \tag{3.7}$$

where a *is a p-dimensional constant vector,* $\delta:=a'D^{-1}a$ *and* $\chi^2(k;\delta)$ *is a random variable of a noncentral* χ^2*-distribution with degrees of freedom* k *and noncentral parameter* δ.

These equalities are described in Sen [8] without proof. They can be proved by induction with respect to p. Now, $g(\theta)$ can be rewritten as follows by using (3.6) and (3.7):

$$\begin{aligned}g(\theta)=-p\Pr\{\chi^2(p+2;\theta'\theta)\leqq 2p\}+\theta'\theta[&-2\Pr\{\chi^2(p;\theta'\theta)\leqq 2p\}\\ &+4\Pr\{\chi^2(p+2;\theta'\theta)\leqq 2p\}-\Pr\{\chi^2(p+4;\theta'\theta)\leqq 2p\}]\end{aligned} \tag{3.8}$$

$$=e^{-r}\sum_{j=0}^{\infty}\frac{r^j}{j!}\Big[-p\beta\Big(\frac{p}{2}+j\Big)+2r\Big\{-2\beta\Big(\frac{p}{2}+j-1\Big)$$
$$+4\beta\Big(\frac{p}{2}+j\Big)-\beta\Big(\frac{p}{2}+j+1\Big)\Big\}\Big],$$

where $\beta(k)=\int_0^p e^{-x}x^k dx/\Gamma(k+1)$, $\Gamma(k+1)=\int_0^\infty e^{-x}x^k dx$ and $r=\theta'\theta/2$. We shall use the following lemma in order to modify (3.8).

LEMMA 4. (I) *It follows* (i) $\beta(k)-\beta(k-1)=-e^{-p}p^k/k!$, (ii) $\beta(k)=1-e^{-p}\sum_{i=0}^{k}p^i/i!$ *and* (iii) $\beta(k-1)/\beta(k)=1+\left[\sum_{m=1}^{\infty}\prod_{i=1}^{m}p/(k+i)\right]^{-1}$.
(II) *If* $c>1$ *and* $i\geqq cp-1$, *then* $\beta(i-1)/\beta(i)>c$.
(III) *Putting* $\eta(j):=\{2(j+1)-p\}\{3\beta(p/2+j)-2\beta(p/2+j-1)\}/(j+1)!$, (i) *if* $j\geqq p-1$, *then* $\eta(j)<0$, (ii) *if* $j=p-2$ *and* $p\geqq 4$, *then* $\eta(j)<0$ *and* (iii) *if* $j\leqq p/2-1$, *then* $\eta(j)\leqq 0$.
(IV) *Putting* $\varepsilon(j):=\eta(j)p^{j+1}$, *if* $j\geqq 3p/2+1$, *then* $|\varepsilon(j)/\varepsilon(j-1)|<2/3$.
(V) *It follows* $\sum_{j=3p/2+1}^{\infty}\varepsilon(j)>2\varepsilon(3p/2)$.

PROOF. Parts (I) and (IV) are obvious. Part (II) can be proved by considering the function $\beta(i-1)-c\beta(i)$. Part (III)-(i) is established by considering $c=3/2$ in (II), whereas (ii) and (iii) are obtained by considering $p/(k+i)\leqq 2/3$ for $i\geqq 6$ and $\prod_{i=1}^{m}p/(k+i)\geqq 0$ for $m\geqq 3$, respectively, in (I)-(iii). Part (V) follows from (III) and (IV).

Using $\eta(j)$ defined in Lemma 4 (III) and rearranging, we can write $g(\theta)$ in (3.8) in the following form:

$$g(\theta)=e^{-r}\Big[-p\beta\Big(\frac{p}{2}\Big)+\sum_{j=0}^{\infty}\eta(j)r^{j+1}\Big]. \tag{3.9}$$

Now for $p=2$ or 4, we can see that $g(\theta)\leqq 0$ for all θ, since $\eta(j)\leqq 0$ for all j from Lemma 4 (III). Therefore the procedure (2.3) is minimax for $p=2$ or 4. For $p=6$, 8, 10 or 12 the procedure (2.3) can be proved to be also minimax by calculating $\eta(j)$'s for $j=0, 1, \cdots, 3p/2$ and showing that

$$g(\theta)<e^{-r}\Big[\sum_{j=0}^{3p/2}\eta(j)r^{j+1}\Big]<0. \tag{3.10}$$

This line of the argument is similar to the case for $p=14$ which is discussed below, but in the latter case it is proved that there exists some $r_0=\theta_0'\theta_0/2$ such that $g(\theta_0)>0$ and that therefore the procedure (2.3) is not minimax. In order to prove the fact for $p=14$ described above, we take $r_0=14$. Making use of Lemma 4 (V), we obtain from (3.9)

$$(3.11)\quad g(\theta_0)=e^{-14}\left[-14\beta(7)+\sum_{j=0}^{\infty}\varepsilon(j)\right]>e^{-14}\left[-14\beta(7)+\sum_{j=0}^{21}\varepsilon(j)+2\varepsilon(21)\right].$$

Now we define the following notations

$$(3.12)\quad A(k):=\sum_{i=0}^{k}\frac{14^i}{i!}\,,\qquad B(k):=e^{14}-A(k)\,(=e^{14}\beta(k))\,,$$

$$C(j):=3B(j+7)-2B(j+6)\qquad\text{and}$$

$$D(j):=\left\{\frac{2(j+1)-14}{(j+1)!}\right\}14^{j-7}C(j)\,(=e^{14}14^{-8}\varepsilon(j))$$

$$(k=0,1,\cdots,28\text{ and }j=0,1,\cdots,21)\,.$$

Using above notations, the extreme right-hand side of (3.11) becomes

$$(3.13)\quad g(\theta_0)>e^{-28}14^8\left[-14^{-7}B(7)+\sum_{j=0}^{21}D(j)+2D(21)\right].$$

Here, we obtain that $-14^{-7}B(7)\geqq -0.02$, $\sum_{j=0}^{21}D(j)\geqq 5.46$ and $2D(21)\geqq -0.27$, exactly. Hence, the bracket of the right-hand side of (3.13) is positive. Therefore we conclude that the procedure (2.3) is not minimax for $p=14$, and the Theorem has been proved.

We conjecture that the procedure (2.3) is not minimax for $p\geqq 16$ (and p even), either. Another multivariate extension of the result in Nagata and Inaba [5] can be seen in Nagata [6].

OSAKA UNIVERSITY*
OSAKA UNIVERSITY

REFERENCES

[1] Hirano, K. (1977). Estimation procedures based on preliminary test, shrinkage technique and information criterion, *Ann. Inst. Statist. Math.*, **29**, 21-34.

[2] Inagaki, N. (1977). Two errors in statistical model fitting, *Ann. Inst. Statist. Math.*, **29**, 131-152.

[3] Lehmann, E. L. (1983). *Theory of Point Estimation*, Wiley, New York.

[4] Nagata, Y. (1983). Admissibility of some preliminary test estimators for the mean of normal distribution, *Ann. Inst. Statist. Math.*, **35**, 365-373.

[5] Nagata, Y. and Inaba, T. (1984). Minimaxity of a preliminary test estimator for the mean of normal distribution, to appear in *Ann. Inst. Statist. Math.*

[6] Nagata, Y. (1984). Minimaxity and inadmissibility of the model selection and estimation procedure for the mean of multivariate normal distribution, to appear in *J. Japan Statist. Soc.*

[7] Sclove, S. L., Morris, C. and Radhakrishnan, R. (1972). Non-optimality of preliminary test estimators for the mean of a multivariate normal distribution, *Ann. Math. Statist.*, **43**, 1481-1490.

[8] Sen, P. K. (1979). Asymptotic properties of maximum likelihood estimators based on conditional specification, *Ann. Statist.*, 7, 1019-1033.

---

* Now at Nagoya University.