Ann. Inst. Statist. Math.
37 (1985), Part A, 437-442

# MINIMAXITY OF A PRELIMINARY TEST ESTIMATOR FOR THE MEAN OF NORMAL DISTRIBUTION

YASUSHI NAGATA AND TAICHI INABA

(Received Mar. 23, 1984; revised Sept. 5, 1984)

## Summary

The problem is to estimate the mean of the normal distribution under the situation where there is vague information that the mean might be equal to zero. A minimax property of the preliminary test estimator obtained by the use of AIC (Akaike Information Criterion) procedure is proved under a loss function based on the Kullback-Leibler information measure.

## 1. Introduction

Let a $p$-dimensional random variable $X$ 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 $\theta$ in the situation where we have vague information about $\theta$ that it is equal to zero. We are concerned with preliminary test estimation, i.e. $\theta$ is estimated by one of two estimators according as the hypothesis $H_0: \theta=0$ is rejected or accepted. A familiar preliminary test estimator in this case is given by

$$d_*(X)=\begin{cases} 0 & \text{if } \|X\|\leqq c \\ X & \text{if } \|X\|>c\,, \end{cases} \tag{1.1}$$

where $\|\cdot\|$ denotes a Euclidean norm. This type of estimator is inadmissible under various standard loss functions for estimation, because it is not smooth in $X$ and so is neither a Bayes estimator nor its limit. Sclove, Morris and Radhakrishnan [6] showed that it is not minimax for $p\geqq 3$ under the squared error loss function. And for $p=1$ and 2, since $d(X)=X$ is admissible and its risk is constant under the squared error loss function, preliminary test estimators (1.1) cannot be minimax. However, in our opinion, the problem should be formulated

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

as a hybrid problem of model selection and estimation. From this point of view Meeden and Arnold [4] showed the admissibility of (1.1) when $p=1$ under the loss function incorporating a cost of the model complexity. Nagata [5] showed that the estimator (1.1) with $c=\sqrt{2p}$, which is determined by AIC (Akaike Information Criterion) procedure for model selection (see Hirano [1]), is admissible when $p=1$ and is inadmissible when $p\geqq 3$ under another loss function which incorporates model fitting and evaluation of an estimate. This loss function is based on the Kullback-Leibler information measure and was introduced by Inagaki [2].

In this paper we will show that when $p=1$, the estimator (1.1) with $c=\sqrt{2}$ is minimax under Inagaki's loss function. In Section 2 we will describe the loss function, the formulation of the problem and the result. A proof of the result will be given in Section 3.

## 2. Formulation of the problem and result

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

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

Inagaki's [2] loss function has the following form

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

where $k(X)$, $d(X)$ and $\zeta_\gamma(d(X))$ are estimators of the index $\gamma$, the unknown parameter $\theta$ and $\zeta_\gamma(\theta)$, respectively. He introduced the first term, the log-likelihood ratio, as a smooth loss for the model fitting and the second term as a loss incurred by an estimate. It is noticeable that this loss function (2.2) is not always nonnegative, but the first term is decomposed into the sum of the following two parts, $J_0$ and $J_1$, $J_0$ being common to all $\gamma$ and $J_1$ nonnegative:

$$(2.3)\qquad \begin{aligned} J_0(\theta)&=\log\{f(x:\theta)/f^0(x:\theta)\}\\ J_1(k,\theta)&=\log\{f^0(x:\theta)/f_k(x:\zeta_k(\theta))\}\,, \end{aligned}$$

where $f^0(x:\theta)=\sup_r f_r(x:\zeta_r(\theta))$. The second term is, of cource, non-negative.

Now, let $X$ follow a one-dimensional normal distribution $N(\theta, 1)$ with p.d.f. $f(x:\theta)$. We consider two models, $\mathcal{F}_0=\{f(x:0)\}$ and $\mathcal{F}_1=\{f(x:\theta);\ \theta\in(-\infty, \infty)\}$. In this problem Hirano [1] proposed the following preliminary test estimator by using AIC procedure,

$$d_0(X)=\begin{cases} 0 & \text{if } |X|\leqq\sqrt{2} \\ X & \text{if } |X|>\sqrt{2}\,. \end{cases} \tag{2.4}$$

This means that the model $\mathcal{F}_0$ is selected when $|X|\leqq\sqrt{2}$, while $\mathcal{F}_1$ is selected and the unknown parameter $\theta$ is estimated as $X$ when $|X|>\sqrt{2}$. Inagaki's loss function (2.2) becomes

$$L((k, d), \theta, x)=\begin{cases} \{x^2-(x-\theta)^2\}/2 & \text{if } k=0 \\ (d-\theta)^2/2 & \text{if } k=1\,. \end{cases} \tag{2.5}$$

The upper formula in (2.5) denotes a loss accompanied with model fitting and implies $d=0$ in (2.2), whereas the lower formula is derived from the second term of (2.2). Under this formulation of the problem, we obtain the following result.

THEOREM. *If the random variable $X$ follows $N(\theta, 1)$, then the estimator* (2.4) *is minimax under the loss function* (2.5).

COROLLARY. *Under the same condition of theorem, $d(X)=X$ is also minimax but is inadmissible.*

A proof of the theorem will be given in the next section.

## 3. Proof of the theorem

In order to prove the theorem we need the following well known lemma, which is stated without proof. (See Lehmann [3], Theorem 2.2 in page 256.)

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 $d_\tau$ of $\theta$ w.r.t. $\pi_\tau$ converges to some constant $r$ as $\tau$ tends to infinity. If the risk, $R(\theta, d_0)$, of $d_0$ satisfies that $R(\theta, d_0)\leqq r$ for all $\theta$, then $d_0$ is a minimax estimator of $\theta$.*

To prove the theorem we can use $\{N(0, \tau^2),\ \tau\in(0, \infty)\}$ for $\{\pi_\tau\}$. The Bayes solution of $\theta$ w.r.t. this class is given by

(3.1) $$d_\tau(X)=\begin{cases} 0 & \text{if } |X|\leqq\sqrt{2/(2-a(\tau))} \\ a(\tau)X & \text{if } |X|>\sqrt{2/(2-a(\tau))}\ , \end{cases}$$

where $a(\tau)=\tau^2/(1+\tau^2)$. See Nagata [5] for this derivation. Noting that the posterior distribution of $\theta$ given $X$ is $N(a(\tau)X, a(\tau))$ and that the marginal distribution of $X$ is $N(0, 1+\tau^2)$, the risk of $d_0$ is found to be

(3.2) $$R(\theta, d_0)=1/2+(1/2)\int_{-\sqrt{2}}^{\sqrt{2}}\{x^2-2(x-\theta)^2\}f(x:\theta)dx\ ,$$

and the Bayes risk of $d_\tau$ is

(3.3) $$r(\tau, d_\tau)=a(\tau)/2+(1/2)\int_{-\sqrt{2/(2-a(\tau))}}^{\sqrt{2/(2-a(\tau))}}\{x^2(2a(\tau)-a(\tau)^2)-2a(\tau)\}f_\tau(x)dx\ ,$$

where $f_\tau(x)$ is the p.d.f. of $N(0, 1+\tau^2)$. Now we prove the following

LEMMA 2. *$r(\tau, d_\tau)$ converges to 1/2 as $\tau\to\infty$.*

PROOF. Lemma 2 is a simple consequence of the facts that $a(\tau)\to 1$ as $\tau\to\infty$ and that

(3.4) $$|\text{the second term of (3.3)}| \leqq\frac{K}{(1+\tau^2)^{1/2}}\int_{-\sqrt{2}}^{\sqrt{2}}|x^2(2a(\tau)-a(\tau)^2)-2a(\tau)|dx\ ,$$

where $K$ is a positive constant. Note that the right-hand-side of (3.4) clearly converges to zero.

For the proof of the theorem we have only to show that the second term of the right-hand-side of (3.2) is non-positive. Putting

(3.5) $$g(\theta)=\int_{-\sqrt{2}}^{\sqrt{2}}\{x^2-2(x-\theta)^2\}\exp\{-(x-\theta)^2/2\}dx\ ,$$

we shall prove the next lemma.

LEMMA 3. *It holds that for all $\theta\in(-\infty,\infty)$,*

(3.6) $$g(\theta)\leqq 0\ .$$

PROOF. Since we can easily show that $g(-\theta)=g(\theta)$, we may confine ourselves to the case when $\theta\in[0,\infty)$. Furthermore, for $\theta\in(\sqrt{2}+1,\infty)$ the quadratic function of the integrand in (3.5) is always negative in the domain of integration $(-\sqrt{2},\sqrt{2})$. Hence we have only to prove (3.6) for $\theta\in[0,\sqrt{2}+1]$. Now carrying out the integral (3.5) we obtain

$$(3.7)\quad g(\theta)=(\sqrt{2}-3\theta)\exp\{-(\sqrt{2}-\theta)^2/2\}+(\sqrt{2}+3\theta)\exp\{-(\sqrt{2}+\theta)^2/2\} +(\theta^2-1)\int_{-\sqrt{2}-\theta}^{\sqrt{2}-\theta}\exp(-x^2/2)dx\,.$$

We note that $g(0)=-\int_{-\sqrt{2}}^{\sqrt{2}} x^2\exp(-x^2/2)dx<0$ from (3.5) and that $g(1)=\exp(-3/2)\{(\sqrt{2}-3)\exp\sqrt{2}+(\sqrt{2}+3)\exp(-\sqrt{2})\}<0$ from (3.7). We shall separate following two cases.

*Case* 1 *when* $\theta\in[0,1)$. We shall show that $g_1(\theta)=g(\theta)/(1-\theta^2)\leqq 0$ for $\theta\in[0,1)$. We have

$$(3.8)\quad g_1(\theta)=\frac{\sqrt{2}-3\theta}{1-\theta^2}\exp\{-(\sqrt{2}-\theta)^2/2\}+\frac{\sqrt{2}+3\theta}{1-\theta^2}\exp\{-(\sqrt{2}+\theta)^2/2\} -\int_{-\sqrt{2}-\theta}^{\sqrt{2}-\theta}\exp(-x^2/2)dx\,.$$

Differentiating (3.8) and simplifying, we obtain,

$$(3.9)\quad g_1'(\theta)=A(\theta)\{(-\sqrt{2}-2\theta+2\sqrt{2}\,\theta^2-\theta^3) +(-\sqrt{2}+2\theta+2\sqrt{2}\,\theta^2+\theta^3)\exp(-2\sqrt{2}\,\theta)\}\,,$$

where $A(\theta)=2\theta\exp\{-(\sqrt{2}-\theta)^2/2\}/(1-\theta^2)^2$ $(>0)$. Clearly for $\theta\in[0,1/\sqrt{2}]$, $g_1'(\theta)<0$. Since for $\theta\in[1/\sqrt{2},1)$ the expression in the second parentheses in the bracket of (3.9) is positive and $\exp(-2\sqrt{2}\,\theta)<\exp(-2)<1/5$, it follows

$$(3.10)\quad g_1'(\theta)<(A(\theta)/5)g_2(\theta)\,,$$

where $g_2(\theta)=-6\sqrt{2}-8\theta+12\sqrt{2}\,\theta^2-4\theta^3$. Examining the behavior of $g_2(\theta)$ by differentiating, we can see that it increases in $[1/\sqrt{2},1)$. Since $g_2(1)=6\sqrt{2}-12<0$, $g_2(\theta)$ is negative in $[1/\sqrt{2},1)$. Hence from (3.10), $g_1'(\theta)<0$ for $\theta\in[1/\sqrt{2},1)$. Now as $g_1'(\theta)<0$ for $\theta\in[0,1)$, $g_1(\theta)$ is decreasing. Therefore noting $g_1(0)=g(0)<0$, we conclude that $g_1(\theta)\leqq 0$ for $\theta\in[0,1)$.

*Case* 2 *when* $\theta\in(1,\sqrt{2}+1]$. We shall show that $g_3(\theta)=g(\theta)/(\theta^2-1)\leqq 0$ for $\theta\in(1,\sqrt{2}+1]$. Since $g_3(\theta)=-g_1(\theta)$, similarly to (3.9) we obtain,

$$(3.11)\quad g_3'(\theta)=A(\theta)\{(\sqrt{2}+2\theta-2\sqrt{2}\,\theta^2+\theta^3) +(\sqrt{2}-2\theta-2\sqrt{2}\,\theta^2-\theta^3)\exp(-2\sqrt{2}\,\theta)\}\,.$$

Since for $\theta\in(1,\sqrt{2}+1]$ the second parentheses in the bracket of (3.11) is negative and $\exp(-2\sqrt{2}\,\theta)<\exp(-2\sqrt{2})<1/16$, it follows

$$(3.12)\quad g_3'(\theta)>(A(\theta)/16)g_4(\theta)\,,$$

where $g_4(\theta)=17\sqrt{2}+30\theta-34\sqrt{2}\,\theta^2+15\theta^3$. Examining the behavior of

$g_4(\theta)$, we can see that it has a local maximum at $\alpha=(34\sqrt{2}-\sqrt{962}/(45$ and a local minimum at $\beta=(34\sqrt{2}+\sqrt{962})/45$. Since $\alpha<1<\beta<\sqrt{2}+1$ and $g_4(\beta)>0$, $g_4(\theta)$ is positive for $\theta\in(1, \sqrt{2}+1]$. Hence from (3.12), $g_3'(\theta)>0$ for $\theta\in(1, \sqrt{2}+1]$, which implies that $g_3(\theta)$ is increasing. Therefore noting $g_3(\sqrt{2}+1)=g(\sqrt{2}+1)/(2+2\sqrt{2})<0$ ($g(\sqrt{2}+1)$ is negative recalling that the quadratic function of the integrand in (3.5) is always negative in the domain of integration $(-\sqrt{2}, \sqrt{2})$.), we conclude that $g_3(\theta)<0$ for $\theta\in(1, \sqrt{2}+1]$. Q.E.D.

Our theorem immediately follows from Lemmas 1, 2 and 3. For the multivariate case ($p\geqq 2$), the theorem would hold if Lemma 3 could be generalized to the multivariate version of (3.5), i.e. $g(\theta)=\int_{||x||^2<2p}\{x'x-2(x-\theta)'(x-\theta)\}\exp\{-(x-\theta)'(x-\theta)/2\}dx$ with $(p\times 1)$-vectors, $x$ and $\theta$, which remains still as an open problem.

Corollary is obtained clearly by the facts that the risk of $d(X)=X$ is 1/2 for all $\theta$ and that it is equal to or greater than the risk of $d_0$, (3.2), from Lemma 3.

## Acknowledgement

The authors wish to express their hearty thanks to Professor M. Okamoto and Dr. N. Inagaki for their kind advice and encouragement.

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] Meeden, G. and Arnold, B. C. (1979). The admissibility of a preliminary test estimator when the loss incorporates a complexity cost, *J. Amer. Statist. Ass.*, **74**, 872-874.

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

[6] 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.