数理解析研究所講究録
第 380 巻 1980 年 74-82

ESSENTIAL COMPLETENESS OF THE CLASS OF MONOTONE DECISION PROCEDURES IN ESTIMATION PROBLEMS

Isao Hashimoto
Kumamoto University

1. INTRODUCTION

For the estimation problem, Karlin and Rubin (1956) proved that the class of monotone decision procedures is essentially complete under the following conditions: Each of the sample space, the parameter space and the decision space is a subset of the real line, the probability density has monotone likelihood ratio and the loss function satisfies some suitable conditions (cf. Karlin and Rubin, 1956, p.293).

In this paper we try to extend the results in Karlin and Rubin (1956) to the following case:
Euclidean k-space is denoted by R^k. The sample space X and the parameter space Θ are subsets of R^m and R^n, respectively. The decision space T is an open convex subset of R^n. Suppose that there are defined a partial ordering $\underset{s}{\leqq}$ in X, a partial ordering $\underset{p}{\leqq}$ in Θ and a partial ordering $\underset{d}{\leqq}$ in T. The probability density $p(x,\theta)$ with respect to a σ-finite measure μ on X has monotone likelihood ratio with respect to the partial orderings $\underset{s}{\leqq}$ and $\underset{p}{\leqq}$; that is, if $x \underset{s}{\leqq} y$ and $\theta \underset{p}{\leqq} \omega$, then it follows

$$p(x,\theta)\ p(y,\omega) - p(x,\omega)\ p(y,\theta) \geqq 0. \qquad (1)$$

The loss function $L(\theta,t)$ is assumed to satisfy the following conditions:

(i) For each θ, $L(\theta,t)$ is continuous and convex as a function of t.

(ii) For any two decisions $t \underset{d}{\leqq} u$, the set $D = \{\theta : L(\theta,t) - L(\theta,u) = 0\}$ is a type Π_1 set: that is, any two points on the set are not greater than or smaller than each other in the sense of $\underset{p}{\leqq}$ and $\{\theta : L(\theta,t) \geqq L(\theta,u)\} = \{\theta : \theta \underset{p}{\geqq} D\}$, where $\theta \underset{p}{\geqq} D$ means that there exists a ω in D such that $\theta \underset{p}{\geqq} \omega$.

For example, the loss function $L(\theta,t) = (\theta - t)'\ M(\theta - t)$ satisfies the conditions (i) and (ii), where $(\theta - t)' = (\theta_1 - t_1, \cdots, \theta_n - t_n)$ and M is a positive definite matrix. Here the partial ordering $(a_1, \cdots, a_n) \leqq (b_1, \cdots, b_n)$ in R^n is defined by $a_i \leqq b_i$ $(i = 1, 2, \cdots, n)$.

2. RESULT

Let X, Θ, T, $\underset{s}{\leqq}$, $\underset{p}{\leqq}$, $\underset{d}{\leqq}$, $p(x,\theta)$ and $L(\theta,t)$ be as for Section 1. Furthermore, we require that $X = \{x : p(x,\theta) > 0\}$ and the inequality in (1) is strict.

Definition 1. An estimator $\hat{\theta}$ of θ is called monotone if $x \underset{s}{\leqq} y$ implies $\hat{\theta}(x) \underset{d}{\leqq} \hat{\theta}(y)$. Denote the class of monotone estimators by M.

An extension of Lemma 1 of Karlin and Rubin (1956) is stated as follows.

Lemma 1. For a prior probability measure ξ and a real valued function h on Θ, let

$$g(x) = \int_{\Theta} p(x,\theta)\ h(\theta)\ d\xi(\theta). \tag{2}$$

If $\theta \in H^+ = \{\theta : h(\theta) \geqq 0\}$ and $\theta \underset{p}{\leqq} \omega$ imply $\omega \in H^+$, then $x \in G^+ = \{x : g(x) \geqq 0\}$ and $x \underset{s}{\leqq} y$ imply $y \in G^+$.

Proof. Define

$$h_1(\theta) = \begin{cases} h(\theta), & \theta \in H^+ \\ 0, & \theta \in H^- \end{cases} \tag{3}$$

and $h_2(\theta) = h_1(\theta) - h(\theta)$, where $H^- = \{\theta : h(\theta) < 0\}$. Clearly, $h_i(\theta) \geqq 0$ for i = 1, 2. Let

$$g_i(x) = \int_{\Theta} p(x,\theta)\ h_i(\theta)\ d\xi(\theta). \tag{4}$$

Then $g_i(x) \geqq 0$ for i = 1, 2.

For $x \underset{s}{\leqq} y$ we have

$$\begin{aligned} & g_1(y)\ g_2(x) - g_2(y)\ g_1(x) \\ &= \int_{H^+} \int_{H^-} \{p(y,\omega)p(x,\theta) - p(y,\theta)p(x,\omega)\}\ h_1(\omega)h_2(\theta)d\xi(\theta)d\xi(\omega) \\ &\geqq 0. \end{aligned} \tag{5}$$

If $y \notin G^+$, then $g(y) = g_1(y) - g_2(y) < 0$. On the other hand, from the assumption that $x \in G^+$, it follows that $g(x) = g_1(x) - g_2(x) \geqq 0$. Hence $g_2(y)\ g_1(x) > g_1(y)\ g_2(x)$. This contradicts (5).

The following lemma is an extension of Theorem 11 of Karlin and Rubin (1956).

Lemma 2. If $L(\theta,t)$ satisfies (i) and (ii) of Section 1, then the Bayes estimator with respect to a prior probability measure ξ is a monotone procedure.

Proof. We remark from the assumption (ii) that $L^+ = \{\theta : L(\theta,t) \geqq L(\theta,u)\}$ $(t \underset{d}{\leqq} u)$ has the property that $\theta \in L^+$ and $\theta \underset{p}{\leqq} \omega$ imply $\omega \in L^+$.

For t in T, put

$$\rho_t(x) = \int_\Theta L(\theta,y)\ p(x,\theta)\ d\xi(\theta). \tag{6}$$

If, for a given x, min $\rho_t(x)$ is attained at u, then for $t \underset{d}{\leqq} u$ and $x \underset{s}{\leqq} y$ we have

$$\rho_u(y) < \rho_t(y) \tag{7}$$

by the above remark and Lemma 1.

Thus, for $x \underset{s}{\leqq} y$, the minimum of $\rho_t(y)$ is attained in the set t with $u \underset{d}{\leqq} t$. This fact shows that the Bayes estimator must be monotone.

Denote the class of Bayes estimators by B. Then it follows from Lemma 2 that $B \subset M$.

Definition 2. $\lim_{n\to\infty} \hat{\theta}_n = \hat{\theta}$ in the regular sense if $\lim_{n\to\infty} \int \hat{\theta}_n(x)\ f(x)\ d\mu(x) = \int \hat{\theta}(x)\ f(x)\ d\mu(x)$ for any μ-integrable function f.

We remark that if $\lim_{n\to\infty} \hat{\theta}_n = \hat{\theta}$ in the regular sense then $\lim_{n\to\infty} \hat{\theta}_n = \hat{\theta}$ a.e. Furthermore, it is well known that the closure $\bar{B}$ of B in the regular sense is essentially complete.

Lemma 3. $\bar{M} = M$.

Proof. Let $\hat{\theta} \in \bar{M}$. Then there exists a sequence $\{\hat{\theta}_n\}$ such that $\hat{\theta}_n \in M$ and $\lim_{n\to\infty} \hat{\theta}_n = \theta$ in the regular sense. Let $x \underset{s}{\leqq} y$. From $\hat{\theta}_n \in M$, $\hat{\theta}_n(x) \underset{t}{\leqq} \hat{\theta}_n(y)$. From the above remark, $\hat{\theta}(x) \underset{t}{\leqq} \hat{\theta}(y)$. This shows that $\hat{\theta} \in M$. This completes the proof.

Theorem. Let X, Θ, T, $\underset{s}{\leqq}$, $\underset{p}{\leqq}$, $\underset{d}{\leqq}$, $p(x,\theta)$ and $L(\theta,t)$ be as for Section 1. Furthermore, we require that $X = \{x : p(x,\theta) > 0\}$ and the inequality in (1) is strict. Then the class of monotone estimators is essentially complete.

Proof. From Lemma 2, $B \subset M$. From Lemma 3, $\bar{M} = M$. Hence, $\bar{B} \subset \bar{M} = M$. This completes the proof of Theorem.

3. EXAMPLES

Example 1. (Stein.) Consider the problem of estimating the mean of a p-variate normal distribution ($p \geqq 3$) when the covariance matrix is known. Let X be a random variable distributed by $N_p(\theta, I)$, where I is a $p \times p$ unit matrix. Stein (1956) has proved that the estimator $\hat{\theta}(x) = x$ of $\theta = (\theta_1, \cdots, \theta_p)'$ is inadmissible for the loss function $L(\hat{\theta},\theta) = (\hat{\theta} - \theta)'(\hat{\theta} - \theta)$ and it is strictly dominated by the estimator $\theta^*(x) = \left(1 - \frac{p-2}{x'x}\right)x$. It is clear that $\hat{\theta}(x)$ is monotone with respect to the coordinatewise ordering. But $\theta^*(x)$ is not monotone and hence inadmissible. In fact, for $x = (-\sqrt{p-2}, 0, \cdots, 0)'$ and $y = \left(\frac{\sqrt{p-2}}{2}, 0, \cdots, 0\right)'$ we have that

$\theta^*(x) = (0, \cdots, 0)'$ and $\theta^*(y) = \left(\frac{3\sqrt{p-2}}{2}, 0, \cdots, 0\right)'$.

It is also known that the positive part of Stein's estimator $\theta^{**}(x) = \left(1 - \frac{p-2}{x'x}\right)^+ x$ dominates the estimator $\theta^*(x)$, where $a^+ = \max(0,a)$. Here we shall show that $\theta^{**}(x)$ is monotone. In fact, for $x = (x_1, 0, \cdots, 0)'$,

$$\theta^{**}(x) = \begin{cases} \left(x_1 - \frac{p-2}{x_1}, 0, \cdots, 0\right) & \text{if } |x_1| \geqq \sqrt{p-2} \\ (0, 0, \cdots, 0) & \text{if } |x_1| < \sqrt{p-2}. \end{cases}$$

and the function defined by

$$f(x_1) = \begin{cases} x_1 - \frac{p-2}{x_1} & \text{if } |x_1| \geqq \sqrt{p-2} \\ 0 & \text{if } |x_1| < \sqrt{p-2} \end{cases}$$

is monotone non-decreasing in x_1. Hence $\theta^{**}(x)$ is monotone for x such that all the components except x_1 are 0. Futher, for $x \underset{s}{\leqq} y$; that is, $x_i \leqq y_i$ $(i = 1, 2, \cdots, p)$, we have from the above facts that

$$\theta^{**}\left((x_1,x_2,x_3,\cdots,x_p)'\right) \leqq \theta^{**}\left((y_i,x_2,x_3,\cdots,x_p)'\right)$$
$$\leqq \theta^{**}\left((y_1,y_2,x_3,\cdots,x_p)'\right) \leqq \cdots \leqq \theta^{**}\left((y_1,y_2,\cdots,y_p)'\right).$$

This means that $\theta^{**}(x)$ is monotone.

<u>Example 2.</u> (Stein.) Let $y_1, y_2, \cdots, y_n$ be independently and identically distributed random variables having a p-variate normal distribution $N_p(0,\Sigma)$ with an unknown covariance matrix Σ $(n > p)$. Since a complete sufficient statistic is $x = \sum_{i=1}^{n} y_i y_i'$, we consider only estimators based on x for the problem of estimating Σ. Both the parameter space $\Theta = \{\Sigma\}$ and the sample space $X = \{x\}$ are the set of all $p \times p$ symmetric positive definite matrices. Partial orderings $\underset{p}{\leqq}$ and $\underset{s}{\leqq}$ in Θ and X are defined as follows :

$\Sigma_1 \underset{p}{\leqq} \Sigma_2$ if and only if $\Sigma_2 - \Sigma_1$ is semi-positive definite, and $x_1 \underset{s}{\leqq} x_2$ if and only if $x_2 - x_1$ is semi-positive definite. The distribution of x is the $\frac{p(p+1)}{2}$ - dimensional Wishart distribution and has monotone likelihood ratio with respect to $\underset{s}{\leqq}$ and $\underset{p}{\leqq}$ since $\Sigma_1 \underset{p}{\leqq} \Sigma_2$ implies $\Sigma_2^{-1} \underset{p}{\leqq} \Sigma_1^{-1}$.

Let the decision space $T = \{\hat{\Sigma}\}$ be the set of all $p \times p$ symmetric positive definite matrices. A partial ordering in T is defined as follows : $\hat{\Sigma}_1 \underset{d}{\leqq} \hat{\Sigma}_2$ if and only if $\hat{\Sigma}_2 - \hat{\Sigma}_1$ is semi-positive definite.

It is assumed that a loss function is $L(\Sigma,\hat{\Sigma}) = tr\ \Sigma^{-1}\hat{\Sigma} - log\ det\ \Sigma^{-1}\hat{\Sigma} - p$. We shall show that $L(\Sigma,\hat{\Sigma})$ satisfies the conditions (i) and (ii). Since $\{a\Sigma_1 + (1-a)\Sigma_2\}^{-1} \leqq a\Sigma_1^{-1} + (1-a)\Sigma_2^{-1}$, $L(\Sigma,\hat{\Sigma})$ is convex in Σ for each $\hat{\Sigma}$. For any $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$ with $\hat{\Sigma}_1 \underset{p}{<} \hat{\Sigma}_2$, let $D = \{\Sigma\ ;\ L(\Sigma,\hat{\Sigma}_1) = L(\Sigma,\hat{\Sigma}_2)\}$. Then we have from a simple calculation that $D = \{\Sigma : tr\ \Sigma^{-1}(\hat{\Sigma}_2 - \hat{\Sigma}_1) = log\ det\ \hat{\Sigma}_2 - log\ det\ \hat{\Sigma}_1\}$. Suppose that Σ_1 and Σ_2 are in D. Then $tr\ (\Sigma_1^{-1} - \Sigma_2^{-1})(\hat{\Sigma}_2 - \hat{\Sigma}_1) = 0$. If $\Sigma_1 \underset{p}{<} \Sigma_2$ or $\Sigma_1 \underset{p}{>} \Sigma_2$ had held, we have that $tr\ (\Sigma_1^{-1} - \Sigma_2^{-1})(\hat{\Sigma}_2 - \hat{\Sigma}_1) > 0$ or $tr\ (\Sigma_1^{-1} - \Sigma_2^{-1})(\hat{\Sigma}_2 - \hat{\Sigma}_1) < 0$. This contradicts that $tr\ (\Sigma_1^{-1} - \Sigma_2^{-1})(\hat{\Sigma}_2 - \hat{\Sigma}_1) = 0$. Hence any two elements in D are not greater then or smaller than each other. Next, it is shown that $\{\Sigma : L(\Sigma,\hat{\Sigma}_1) \geqq L(\Sigma,\hat{\Sigma}_2)\} = \{\Sigma : \Sigma \underset{p}{\geqq} D\}$. Suppose that $L(\Sigma_1,\hat{\Sigma}_1) \geqq L(\Sigma_1,\hat{\Sigma}_2)$ and $L(\Sigma_2,\hat{\Sigma}_1) = L(\Sigma_2,\hat{\Sigma}_2)$. Clearly, $L(\Sigma_1,\hat{\Sigma}_1) - L(\Sigma_1,\hat{\Sigma}_2) \geqq L(\Sigma_2,\hat{\Sigma}_1) - L(\Sigma_2,\hat{\Sigma}_2)$. Then $tr\ (\Sigma_1^{-1} - \Sigma_2^{-1})(\hat{\Sigma}_1 - \hat{\Sigma}_2) \geqq 0$. Since the last inequality holds for any $\hat{\Sigma}_1 \underset{d}{<} \hat{\Sigma}_2$, we have $\Sigma_1^{-1} \underset{p}{\leqq} \Sigma_2^{-1}$ and hence $\Sigma_2 \underset{p}{\leqq} \Sigma_1$. This shows that $\{\Sigma : L(\Sigma,\hat{\Sigma}_1) \geqq L(\Sigma,\hat{\Sigma}_2)\} \subset \{\Sigma : \Sigma \underset{p}{\geqq} D\}$. Conversely, suppose that $L(\Sigma_2,\hat{\Sigma}_1) = L(\Sigma_2,\hat{\Sigma}_2)$ and $\Sigma_2 \underset{p}{\leqq} \Sigma_1$. Then $tr\ \Sigma_2^{-1}(\hat{\Sigma}_2 - \hat{\Sigma}_1)$ $= \mathit{log\ det}\ \hat{\Sigma}_2 - \mathit{log\ det}\ \hat{\Sigma}_1$ and

$$\begin{aligned} & L(\Sigma_1,\hat{\Sigma}_1) - L(\Sigma_1,\hat{\Sigma}_2) \\ &= tr\ \Sigma_1^{-1}(\hat{\Sigma}_1 - \hat{\Sigma}_2) + \log\det\hat{\Sigma}_2 - \log\det\hat{\Sigma}_1 \\ &= tr\ \Sigma_1^{-1}(\hat{\Sigma}_1 - \hat{\Sigma}_2) + tr\ \Sigma_2^{-1}(\hat{\Sigma}_2 - \hat{\Sigma}_1) \\ &= tr\ (\Sigma_2^{-1} - \Sigma_1^{-1})(\hat{\Sigma}_2 - \hat{\Sigma}_1) \geqq 0. \end{aligned}$$

This shows that $\{\Sigma : \Sigma \underset{p}{\geqq} D\} \subset \{\Sigma : L(\Sigma,\hat{\Sigma}_1) \geqq L(\Sigma,\hat{\Sigma}_2)\}$. From the above facts it is seen that $L(\Sigma,\hat{\Sigma}) = tr\ \Sigma^{-1}\hat{\Sigma} - \log\det\Sigma^{-1}\hat{\Sigma} - p$ satisfies the conditions (i) and (ii). Accordingly, it follows from Theorem that the class of monotone estimators based on x constitutes an essentially complete class. This example is suggested by G. Ishii.

ACKNOWLEDGMENT

The author is grateful to Professor Goro Ishii for helpful suggestions and discussions.

REFERENCES

Karlin, S. and Rubin, H. (1956). The theory of decision procedures for distributions with monotone likelihood ratio. Ann. Math. Statist., 27, 272-299.

Blackwell, D. and Girshick, M. A. (1954). Theory of Games and Statistical Decisions, John Wiley & Sons.

Lehmann, E. L. (1952). Testing multiparameter hypotheses. Ann. Math. Statist., 23, 541-552.

Takeuchi, K. (1965). Theory of Statistical estimation (II). Sugaku, 16, 139-149. (in Japanese)

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multi-variate normal distribution. Proc. Third Berkeley Symp. Math. Statist. Prob., 1, 197-206.

James, W. and Stein, C. (1961). Estimation with quodratic loss. Proc. Fourth Berkeley Symp. Math. Statist. Prob., 1, 361-379.

Ishii, G. (1976). Test of hypotheses for distributions with monotone likelihood ratio: case of vector valued parameter. Proc. Third Japan-USSR Symp. on Prob. Theory, 223-231.