数理解析研究所講究録
第27巻 1967年 99-115

SOME PROBLEMS IN THE COMPLETE CLASS THEOREMS

by Hirokichi Kudō
Osaka City University and University of California, Berkeley

Since Abraham Wald established the theory of decision functions, many textbooks in Statistics are written in the language of this theory. However very few of them cited the complete class theorem, which I believe is one of the essential parts of his book [21]. The following difficulties laid on the complete class theorem might give the excuse for this situation:

1) In many statistical problems, nontrivial complete classes have not been known up to the present stage.

2) Assumptions of the general class theorems are complicated and restrictive.

3) Methods of construction of nontrivial complete classes in general theory are inaccessible and indirect.

4) There is no general expression of the minimal complete class.

5) The relation between two important complete classes $\bar{B}$ and W has not been investigated, where $\bar{B}$ is the closure, in the sense of 'regular convergence', of the class B of all Bayes solutions, and W is the class of all Bayes solutions in the wide sense.

These difficulties of the complete class should be overcome in a near future for the sake of making complete class theorems useful.

The <u>first difficulty</u> is being resolved by several authors step by step. Several complete (or essentially complete) classes are obtained: e.g. the class of non-randomized estimates in the estimation problem of convex loss function [5, page 294], the class of the convex critical regions in the problem of testing hypothesis when the sample distribution is of exponential type [3], [6], the class of likelihood stopping rules in some sequential analysis [16], [18], [8] and a class of some type of designs in the experimental design problem [7], [22]. Another important application of the concept of complete class is that to the sufficient statistic. This can be stated as follows: A statistic T is sufficient if and only if the class of all decision functions through T is essentially complete in the class of all decision functions for every statistical problem [4], [15], or even for a particular problem [1], [2], [10], [11].

<u>The second difficulty</u>: The most general complete classes were given by Wald in his book [21], and later extended by LeCam [14]. The main part of the restrictions under which Wald proved the completeness of $\overline{B}$ and W is as follows:

W1) the dominatedness of the distribution space Ω by a σ-finite measure λ. The generalized density function $p(x,\theta)$ of the sample distribution with respect to λ is labelled by θ in the parameter space Θ.

W2) the boundedness of the loss function $L(\theta,a)$, where a is a point of action space A.

W3) the subconvexity of the space D of the decision functions available to a statistician, i.e. for any δ_1 and $\delta_2 \in D$ and for any real $\alpha \in [0,1]$, there is a δ_3 such that

$$r(\theta,\delta_3) \leq \alpha\, r(\theta,\delta_1) + (1-\alpha)\, r(\theta,\delta_2) \quad \text{for all } \theta,$$

where the risk function $r(\theta,\delta)$ is defined as

$$r(\theta,\delta) = \int_X \int_A L(\theta,a)\, \delta(da,x)\, p(x,\theta)\, \lambda(dx).$$

W4) the compactness of the action space A.

W5) the compactness of the space D.

(W5 is an implication of W4 and closedness of D.)

The concept of convergence in D is first introduced by Wald. LeCam makes this notion very clear by using the weak topology in a space of bilinear functionals. According to LeCam, a decision function is a bilinear positive normalized functional on spaced $C_0(A)$ and L, where $C_0(A)$ is the linear space of all continuous functions on A with compact support and L is a linear space spanned by all generalized density functions $p(x,\theta)$. The topology in the family $\mathcal{D}$ of all decision functions defined above is defined as a weak topology of the family of bilinear functionals. We shall call the relative topology in D induced from the above topology of $\mathcal{D}$ the regular topology, after the terminology 'regular convergence' used by Wald [21]. The compactness in W5) can be interpreted as that in this regular topology. LeCam [14] proved Wald's theorems without the restriction W2). His restrictions for his complete class theorems are mainly the following five:

L1) the same as W1).

L2) the lower semicontinuity of the loss function $L(\theta,a)$ as a function of a for each fixed value of parameter θ, and its boundedness from bottom.

L3) the same as W3).

L4) A is a δ-compact, locally compact metric space.

L5) the property (W) of D.

The property (W) of D is defined by the following

<u>Definition</u>. Let $\mathcal{F}$ be a family of all the non-negative extended functions on a space Θ and $\mathcal{R}$ a subset of $\mathcal{F}$. $\mathcal{R}$ is called <u>half-closed</u> if, for every cluster point $\bar{f}$ of $\mathcal{R}$ with respect to everywhere convergence topology of $\mathcal{F}$, there is an element f of $\mathcal{R}$ such that $f(\theta) \leq \bar{f}(\theta)$ holds for all θ in Θ.

The space D of decision functions is said to have <u>the property</u> (W) if $\mathcal{R} = \{r(\theta,\delta); \delta \in D\}$ is half-closed in $\mathcal{F}$.

The property (W) is a generalization of the compactness in the sense of weak intrinsic convergence in Wald's Book [21].

We shall give a criterion for the property (W).

<u>Theorem</u> 1. Suppose that T is a Hausdorff space and Θ an arbitrary space. Let $f(\theta,t)$ be a nonnegative extended

function defined on $\Theta \times T$, which is lower semicontinuous in t for any fixed $\theta \in \Theta$. Then $\mathcal{R} = \{f(\cdot,t);\ t \in T\}$ is half-closed if, for any $\theta \in \Theta$ and any positive number k less than $\sup_{t\in T} f(\theta,t)$, there exists a relatively compact subset $C_{\theta,k}$ of T such that

$$\inf_{t \notin C_{\theta,k}} f(\theta,t) > k.$$

(The notation $\inf_{t\notin C_{\theta,k}} f(o,t)$ may stand for $\sup_{t\in T} f(\theta,t)$ wher $T = C_{\theta,k}$).

From Theorem 1 we can easily see that the Wald-LeCam theorem holds for the decision problem with a compact action space, e.g. the problems of testing hypothesis and finite-action problems.

However the above theorem is not able to be applied directly to the case of noncompact action space.

Theorem 2. Suppose that W1) and L4) hold and, in addition, the following conditions hold:

1) The loss function $L(\theta,a)$ is a Borel measurable nonnegative real function of a for each fixed θ.

2) $\mathcal{L} = \{L(\cdot,a) : a \in A\}$ is, in the relative sense of pointwise convergence topology of $\mathcal{F}$, a σ-compact, locally compact metrizable space.

3) There is a mapping τ of $\overline{\mathcal{L}}$, the closure of $\mathcal{L}$ in the pointwise convergence topology in $\mathcal{F}$, onto $\mathcal{L}$ such that

a) $\{\ell \in \bar{\mathfrak{L}} : (\tau\ell)(\Theta) \leq \alpha\}$ is a Borel set, in the relative topology of $\bar{\mathfrak{L}}$ in $\mathfrak{F}$, for every fixed $\Theta \in \Theta$ and every real positive α.

b) $(\tau\ell)(\Theta) \leq \ell(\Theta)$ for all $\Theta \in \Theta$ and all $\ell \in \bar{\mathfrak{L}}$.

(b) implies the half-closedness of $\mathfrak{L}$.)

Then $\mathcal{D}$ has the property (W).

One of the applications of this theorem is as follows: Let us consider the problem of interval estimation with the loss function

$$L(\Theta,(\underline{\Theta},\overline{\Theta})) = \alpha\ \mu(|\overline{\Theta}-\underline{\Theta}|) + \beta C(\Theta,\underline{\Theta},\overline{\Theta}), \quad \text{where}$$

$\underline{\Theta} < \overline{\Theta}$, $\alpha > 0$, $\beta > 0$, $\mu(t)$ is a monotone nondecreasing, left continuous nonnegative function of t and $\mu(t) = 0$ if $t < t_0$ for some $t_0 > 0$ and

$$C(\Theta,\underline{\Theta},\overline{\Theta}) = \begin{cases} 0, & \text{if } \underline{\Theta} < \Theta < \overline{\Theta}, \\ 1, & \text{otherwise.} \end{cases}$$

For this problem the class of the interval estimates $[\underline{\Theta},\overline{\Theta}]$ of real Θ has the property (W). The case stated by LeCam [14, Miscellaneous remark (6), p. 80] is also an implication of Theorem 2.

The class treated in Theorem 2 is the whole space $\mathcal{D}$. However there are many important problems concerning the class D of decision functions which are not the whole $\mathcal{D}$ but just a subset of $\mathcal{D}$, e.g. the class of unbiased estimates, the class of level α tests, or the class of internal estimates of confidence coefficient $1 - \alpha$. For closed subsets of $\mathcal{D}$ we have another criterion.

Theorem 3. Suppose that

1) A is a σ-compact, locally compact metric space,

2) the loss function $L(\Theta,a)$ is a lower semicontinuous function of a for every fixed Θ,

3) for any positive integer n and for any Θ in Θ, there exists a compact $C_{n,\Theta} \subset A$ such that

$$n \leq \inf_{a \notin C_{n,\Theta}} L(\Theta,a).$$

Then every closed subset D of $\mathcal{D}$ has the property (W).

From this theorem we can easily verify that the Wald-LeCam theorem holds for every closed convex subset D of the class $\mathcal{D}$ of all estimates when the loss is quadratic. And moreover Theorem 3 has an application to the sequential problem of decision procedure. The assumption for the cost function $c(\lambda_1,\cdots\lambda_n,\Theta)$ at the experimental stages $\lambda_1,\cdots,\lambda_n$ when Θ is true value of the parameter should satisfy $c(\lambda_1,\cdots,\lambda_n,\Theta) \geq C_n(\Theta)$ and $\lim_{n\to\infty} C_n(\Theta) = \infty$ [14, Assumption 5][21, Assumption 3.5.(iii)], and that in each stage the statistician is allowed to choose his experiments out of a finite number of possibilities. (Both Wald and LeCam considered the case where the cost depends also on the sample x observed, but we shall restrict ourselves to the case of the cost independent of x.) Let A be a set of all combinations $(\lambda_1,\cdots,\lambda_n)$ of a finite number of experiments. Since only a finite number of experiments is available for the statistician up to every stage where the terminal decision is made, A is a countable set and so it can be regarded as a discrete topological space. When the cost function $c(\lambda_1,\cdots,\lambda_n\theta)$ is considered as a loss function $L(\theta,a)$ with $a = (\lambda_1,\cdots,\lambda_n)$, the condition assumed by Wald and LeCam is completely the same as 3) of Theorem 3. Thus the sequential decision problem is covered by the general one-stage decision problem as far as the complete class theorems are concerned.

The following theorem is useful but not difficult to prove.

Theorem 4. If a class D of decision procedures has the property (W) and C is an essentially complete class in D, then C has also the property (W).

One of the applications of this theorem is as follows: If the loss function $L(\theta,a)$ is convex in a, and $\lim_{a\to\pm\infty} L(\theta,a) = \infty$, then the class of all non-randomized estimates is an essentially complete class [5, page 294]. Since the class $\mathcal{D}$ of the estimates of θ has the property W (Theorem 3), C has the property (W).. In addition, C is subconvex with respect to the loss function $L(\theta,a)$. Therefore the Wald-LeCam theorem holds for this class C.

As we have seen, the property (W) is an essential condition for the completeness of $\overline{B}$ and W. Are there any other good criteria for this property?

The third difficulty: Roughly speaking, Wald [21, Th. 3.20] showed that the class B itself is essentially complete under the additional assumption that

W6) the parameter space Θ is a compact topological space.

W7) $p(x,\theta)$ and $L(\theta,a)$ are measurable on respective product spaces $X \times \Theta$ and $\Theta \times A$.

W8) the class $\{L(\cdot,a) ; a \in A\}$ of functions of Θ is equicontinuous.

Noticing that any Bayes procedure can be gotten by minimizing posterior loss, it can be understood that B is a nontrivial complete class easily found. However W6) and W7) are not satisfied in the usual statistical problem, in which we have to take $\overline{B}$ or W as a complete class. In building up $\overline{B}$, there is usually a problem in taking a limit in the sense of regular convergence. The problem in getting W is more serious. W is defined as the class of δ_0's such that

$$\inf_{\xi} \{r(\xi,\delta_0) - \inf_{\delta \in D} r(\xi,\delta)\} = 0,$$

where ξ's are prior distributions on Θ. According to LeCam [14, Th. 3], $\delta_0 \in W$ if and only if δ_0 is not improvable uniformly; i.e., for any $\varepsilon > 0$ and any $\delta \in D$ there exists at least one element $\theta \in \Theta$ such that

$$r(\theta,\delta_0) - \varepsilon < r(\theta,\delta),$$

provided that the conditions L1) - L5) hold.

For the sake of avoiding such problems, a device is proposed by J. Sacks [17]. If a decision procedure δ minimizes the value of 'posterior' loss with respect to a σ-finite 'prior' measure ξ when it remains finite, then we call δ as a generalized Bayes solution. Sacks showed that the class B^* of the generalized Bayes solutions contains the class $\bar{B}$ and hence it is a complete class under the assumption that the sample distribution is of exponential type and the loss is a convex function. Another approach in this direction is [12]. We will consider a sequence $\{\xi_n\}$ of prior distributions on Θ and a sequence $\{A_n\}$ of positive numbers so that the sequence $\{A_n \xi_n\}$ tends to a linear functional on the linear space spanned by the risk functions (as functions of Θ), i.e.

$$\lim_{n \to \infty} A_n \int r(\Theta,\delta)\, \xi_n(d\Theta) = F(r(\cdot,\delta)) \quad \text{for all} \quad \delta \in D.$$

We shall restrict ourselves to the problems of testing a simple hypothesis $\Theta = \Theta_0$ against the alternative $\Theta \neq \Theta_0$ in one-dimensional parameter space. In this case it is doubtful if the behavior of the risk function in neighborhoods of the infinity of Θ has a great importance. From this point of view, we define a concept of locally complete class. A class C of test functions is called locally complete if any compact contraction of the parameter space Θ makes the class C a complete class.

Suppose that

1) Θ is an open subset of Euclidean R^k, and Θ_0 is an inner point of Θ,

2) the generalized density function $p(x,\Theta)$ is measurable on $X \times \Theta$ and continuous as a function of Θ

for any x,

3) the derivatives of the third order of $p(x,\theta)$ with respect to θ exists and is continuous in θ for almost all $x \in X$,

4) for any compact subset $C \subset \Theta$ and any positive integer $n_1,\cdots,n_k$, $\Sigma_{i=1}^{k} n_i \leq n$, there exists a function $In_1,\cdots,n_k(x)$ integrable on X such that

$$\left| \frac{\partial^{n_1+\cdots+n_k}}{\partial x_1^{n_1}\cdots\partial x_k^{n_k}} p(x,\theta) \right| \leq In_1,\cdots,n_k(x)$$

for almost all $x \in X$ and for any $\theta \in C$, and let

$$(\#) \quad L(f) = -\sum_{i,j=1}^{k} p_{ij}\left(\frac{\partial^2 f(\theta)}{\partial\theta_i\partial\theta_j}\right)_{\theta=\theta_0} - \lim_{\varepsilon\to 0}\left[\sum_{i=1}^{k} p_i(\varepsilon)\left(\frac{\partial f(\theta)}{\partial\theta_i}\right)_{\theta=\theta_0} + p(\varepsilon)f(\theta_0) + \int_{C-S_\varepsilon} f_0(\theta)\eta(d\theta)\right]$$

be a linear functional defined on the space of all functions $f(\theta)$ being twice differentiable continuously, where (p_{ij}) is a positive semidefinite matrix, $p_i(\varepsilon)$ is nonnegative and finite on $C - S_\varepsilon^o$ and S_ε^o is the interior of the sphere at θ^o with radius ε.

Theorem 4 [12]. Under the conditions 1) - 4), the class W^* of the test functions $\phi(x)$, satisfying

$$\phi(x) = \begin{cases} 1, & \text{if } L(p(x,\cdot)) < 0, \\ 0, & \text{if } L(p(x,\cdot)) > 0, \end{cases}$$

for some linear functional of form $(\#)$, contains the closure $\overline{B}$ of the class B of the Bayes solutions relative to prior distributions assigning probability 1 to a compact subset of $\mathcal{D}$. Hence $\mathcal{D}^*$ is a locally complete class.

The classes B* and $\mathcal{D}^*$ are obtained only by the same procedure as the class of Bayes solutions, and any limiting procedure in the regular topology is not needed. If we have another method for getting nontrivial complete classes without limiting procedure, it will be a great advantage.

The fourth difficulty: The existence of the minimal complete class of decision procedures is implied by LeCam's assumptions L 1) - L 5) [14, p. 77], and so is the completeness of the class of all admissible procedures. A general criterion for the admissibility of a decision procedure was given by C. Stein [19]. Another criterion was orally given by K. Takeuchi [20] in 1961.

Theorem 5 (Takeuchi). Suppose that the parameter space Θ is a subset of Euclidean space and the risk function $r(\theta,\delta)$ is continuous on Θ for any fixed δ. Then δ is admissible if and only if there exists a sequence $\{\xi_n\}$ of prior distributions, and a sequence $\{A_n\}$ of positive numbers such that

1) $r(\xi_n,\delta) < \infty$ for any $\delta \in D$,

2) for any measurable subset $N \subset \Theta$ for which

$$\liminf_{n\to\infty} A_n \xi_n(N) = 0,$$

where $\Theta - N$ is everywhere dense in Θ,

3) $\liminf_{n\to\infty} A_n \{r(\xi_n,\delta_0) - r(\xi_n,\delta)\} < \infty$ for any $\delta \in D$.

The last condition 3) of Theorem 5 seems to have some connection with the class W^* which is defined in Theorem 4, but it has not yet been clear. This theorem is an abstraction of the methods used by Karlin [9] for the establishment of admissibility.

Though there are several criterions of the admissibility and smaller classes than $\bar{B}$ and W, it is rather difficult to find examples of applications of these results.

The fifth difficulty: As is well known, Wald proved first the completeness of $\bar{B}$ and W, and then LeCam proved that of the intersection $\bar{B} \cap W$ under respective conditions. However, we do not know whether LeCam's assertion about $\bar{B} \cap W$ adds anything new to the completeness of $\bar{B}$ and W. This is the point of the fifth difficulty. To see this, we shall look at the problem of testing hypothesis, in which the null hypothesis is "$\Theta \in \omega$" and the alternative is "$\theta \notin \omega$". If there is at least one boundary point of ω in Θ at which the average $E_\Theta[\varphi]$ of every test function φ is continuous in θ, then every test function is a Bayes solution in the wide sense, i.e., $W = D$, and so $\bar{B} \cap W = \bar{B}$. Thus W is not useful in such a case. Besides, there is a statistical problem in which $\bar{B} = D$.

Theorem 6 [Kusama, 13]. $\bar{B} = D$ if

1) there is a disjoint measurable partition $\{E_i\}$ of the sample space χ and a sequence $\{\theta_i\}$ of elements of Θ such that $P_{\theta_i}(E_i) = 1$ for all i,

2) for every $\theta \in \Theta$ there is at least one element

a_Θ of the action space A such that

$$L(\Theta, a_\Theta) = \inf_{a \in A} L(\Theta, a).$$

Theorem 6 covers the case of the quadratic loss estimation of the location parameter of the uniform distribution. Kusama [13] obtained another interesting result.

<u>Theorem</u> 7 (Kusama). Suppose that

1) Θ is a δ-compact, locally compact metric space,

2) the class $\{p(x,\theta)L(\theta,\cdot)\ ;\ \Theta \in \Theta\}$ of functions of $a \in A$ is equicontinuous on A for almost all $x \in \chi$,

3) there exists a sequence $\{F_n\}$ of compact subsets of Θ such that $\bigcup F_n = \Theta$ and

$$\lim_{n \to \infty} \sup_{\Theta \in F_n} p(x,\Theta)\, L(\Theta,a) = 0 \quad \text{for all } a \text{ and almost all } x.$$

4) $r(\theta,\delta)$ is continuous on Θ.

Let $\{\xi_n\}$ be a sequence of prior distributions for which $\lim_{n\to\infty} \xi_n(C) = 0$ for any compact subset C of Θ and let $W_{\{\xi_n\}}$ be a class of decision functions δ^* satisfying

$$\inf_{1 \le n < \infty} \{r(\xi_n, \delta^*) - \inf_{\delta \in \mathcal{D}} r(\xi_n, \delta)\} = 0.$$

Then if there is a decision procedure δ^* in $W_{\{\xi_n\}}$ such that $\delta^*(N_x : x) = 1$ holds for every x and some subset N_x of A, then

$$\overline{W_{\{\xi_n\}}} = \mathcal{D}.$$

and hence we have

$$\overline{W} = \mathcal{D}.$$

Theorem 7 shows how much the problem about W is embarrassing. Is the class W too large or is the regular topology too weak? In the case above stated, the class W is useless. In what case does $\overline{W} \cap B$ work effectively? And, moreover, is there any case in which W is effectively used for finding a complete class?

本論文は1965年1月，著者がNew York大学Courant Instituteにおいて行つた講演の内容をまとめたもので，ここで扱われた主題がこのシンポジウムにおいても論じられた．

REFERENCES

[1] Bahadur, R.R.: Sufficiency and statistical decision functions, Ann. Math. Statist. 25 (1954), pp. 423-462.

[2] ―――――― : A characterization of sufficiency, Ann. Math. Statist. 26 (1955), pp. 286-293.

[3] Birnbaum, A.: Characterizations of complete classes of tests of some multiparametric hypothesis, with applications to likelihood ratio tests, Ann. Math. Statist. 26 (1955), pp. 21-36.

[4] Blackwell, D.: Equivalent comparisons of experiments, Ann. Math. Statist. 24 (1953), pp. 265-272.

[5] Blackwell, D. and M.A. Girshick: Theory of Games and Statistical Decisions, 1954, Wiley.

[6] Burkholder, D.L.: Effect on the minimal complete class of tests of changes in the testing problem, Ann. Math. Statist. 31 (1960), pp. 325-331.

[7] Ehrenheld, S.: Complete class theorems in experimental design, Proc. Third Berkeley Symp. Math. Statist. and Prob. I (1956), pp. 57-67

[8] DeGroot, M.H.: The essential completeness of the class of generalized sequential probability ratio tests, Ann. Math. Statist. 32 (1961), pp. 602-605.

[9] Karlin, S.: Admissibility for estimation with quadratic loss, Ann. Math. Statist. 29 (1958), 406-454.

[10] Kudo, H.: Dependent experiments and sufficient statistics, Nat. Sci. Rep. Ochanomizu Univ., 4 (1954), pp. 151-163.

[11] ――――: On sufficiency and completeness, Sugaku, 8 (1957), pp. 129-138 (in Japanese).

[12] ――――: Locally complete class of tests, Bulletin of Intern. Statist. Instit. 38 (1961), Part IV, pp. 173-180.

[13] Kusama, T.: On classes of Bayes solutions, submitted to the Ann. Instit. Math. Statist.

[14] LeCam, L.: An extension of Wald's theory of statistical decision functions, Ann. Math. Statist. 26 (1955), pp. 69-81.

[15] ———— : Sufficiency and approximate sufficiency, Ann. Math. Statist. 35 (1964), pp. 1419-1455.

[16] Lehmann, E.L. and C. Stein: Completeness in the sequential case, Ann. Math. Statist. 21 (1950), pp.376-385.

[17] Sacks, J.: Generalized Bayes solutions in estimation problems, Ann. Math. Statist. 34 (1963), pp. 751-768.

[18] Sobel, M.: An essentially complete class of decision functions for certain standard sequential problems, Ann. Math. Statist. 24 (1953), pp. 319-337.

[19] Stein, C.: A necessary and sufficient condition for admissibility, Ann. Math. Statist. 26 (1955), pp. 518-522.

[20] Takeuchi, T: Recent development of estimation problem, Oral talk at the Annual Meeting of Math. Soc. of Japan.

[21] Wald, A.: Statistical Decision Functions, 1950, Wiley.

[22] Zacks, S.: On a complete class of linear unbiased estimators for randomized factorial experiments, Ann. Math. Statist. 34 (1963), pp. 769-779.