ON PARTIAL PRIOR INFORMATION AND THE PROPERTY OF PARAMETRIC SUFFICIENCY

HIROKICHI KUDŌ
OSAKA CITY UNIVERSITY and
UNIVERSITY OF CALIFORNIA, BERKELEY

1. Summary

The problem of statistical decisions when there is a partial lack of prior information is considered, and a definition of the optimality of a statistical procedure in such a case is given. This optimality is a generalization of both the minimax property and the Bayes property, in the sense that the former property yields optimality in the case of a complete lack of prior information, whereas the latter coincides with the optimality in the case of complete prior information. A characterization of the sufficiency of a sub-σ-field $\mathcal{B}$ of a σ-field $\mathcal{A}$ of the parameter space is developed from this point of view. The sufficiency of $\mathcal{B}$ is defined as the property that a prior distribution on $\mathcal{B}$ induces the same optimal procedure as a prior distribution on $\mathcal{A}$. In the case of testing hypotheses, there is shown a connection of this concept with that of the parametric sufficiency due to E. W. Barankin [1].

2. Introduction

For some time there have existed characterizations of the sufficiency of a statistic (or a σ-field in a sample space) from the standpoint of decision functions (see [2], [3], [4], and [5]). According to these characterizations, a statistic $t(x)$ is sufficient if and only if in a certain statistical problem the risk by a decision procedure through the observation of the sample x is not increased at all by the restriction to the observation of the statistic $t(x)$. We shall attempt here to give a parallel discussion in the case of parametric sufficiency, a concept introduced by Barankin [1]. A function $u(\theta)$ on a parameter space Θ is called a sufficient parameter if for any measurable set A the probability $P_\theta(A)$ of occurrence of the observed sample x in A when θ is the true parameter is a function of $u(\theta)$. Looking at "the function on the parameter space" more closely, we understand that this idea represents an amount of prior information. Let us consider this problem by example. Suppose a statistician is informed of nothing but the prior

Based on research supported by the National Science Foundation.

probabilities of two parts, ω and ω^c, of Θ before any statistical experiment takes place. The prior information given to the statistician could be considered as a function $u(\theta) = 0$ on ω; $u(\theta) = 1$ on ω^c and a probability distribution on $\{0, 1\}$. Thus a pair of a parametric σ-field $\mathcal{B}$ and a probability distribution on $\mathcal{B}$ is considered to be a kind of representation of prior information.

Suppose that a statistician is supplied with a partial prior information $\{\mathfrak{F}, \xi\}$, where $\mathfrak{F}$ is a σ-field generated by a finite disjoint partition $\Theta = \bigcup_{i=1}^{k} F_i$ of Θ and ξ is a probability distribution $\xi(F_1), \cdots, \xi(F_k)$. It seems to be reasonable that he will choose, as an optimal procedure in this situation, the procedure $\delta = \delta*$ (if it exists) which minimizes

$$\sum_{i=1}^{k} (\sup_{\theta \in F_i} r(\theta, \delta))\xi(F_i), \tag{2.1}$$

where $r(\theta, \delta)$ is a risk function of a procedure δ when θ is a true value. Such an optimality is a generalization of both the minimax property and the Bayes property.

In section 3 we give a definition of the *mean-max risk* which is a generalization of the formula (2.1), and we also give a useful inequality. In section 4 we define the optimality of procedures with respect to a partial prior information. In section 5 we give a definition of the sufficiency of a sub-σ-field. This section also contains an important theorem on the measurability of the risk function of the optimal procedure. In section 6 we restrict ourselves to the case of testing hypotheses, and give the main theorem that under some conditions a sub-σ-field $\mathcal{B}$ is sufficient if and only if the distribution of the sample x is $\mathcal{B}$-measurable for any fixed event A, that is, the sufficiency in our sense is equivalent to that in Barankin's sense. In the last section, we give some miscellaneous remarks.

3. Mean-max risk of a procedure

Consider a statistical game $(\Theta, \mathfrak{D}, r)$, where Θ is the space of the parameter θ, and $\mathfrak{D}$ is the space of procedures δ. The number $r(\theta, \delta)$ is a risk imposed on the statistician when he adopts a procedure δ, and θ is the true value. We shall associate with Θ a fixed σ-field $\mathcal{A}$ of subsets of Θ.

ASSUMPTION. *The risk $r(\theta, \delta)$ is a nonnegative function, and for each fixed δ it is $\mathcal{A}$-measurable and bounded in θ.*

Consider a sub-σ-field $\mathcal{B}$ of $\mathcal{A}$ and a prior distribution ξ defined on $\mathcal{A}$. The pair $(\mathcal{B}, \xi)$ is called a *partial prior information.* By this terminology we mean that the statistician will be informed of only the value of ξ on $\mathcal{B}$ before the experimental results are observed, so that he can use this information for the choice of procedures. For example, suppose that the statistician knows the complete symmetry of a die and by using this die he is going to allocate 6 different plants to 6 plots. In this case he knows that the chance of all allocations of the plants to the plots are the same. So he has a partial prior information $(1/6!, \cdots, 1/6!)$ for the 6! permutation of the allocation (or 6! parts of the parameter space).

DEFINITION. *For a sub-σ-field* $\mathcal{B}$ *of* $\mathcal{A}$ *and a prior probability measure* ξ *on* $\mathcal{A}$, *the* mean-max risk *is defined as*

$$r(\mathcal{B}, \xi, \delta) = \inf_{\mathcal{F}\subset\mathcal{B}} \sum_{i=1}^{k} (\sup_{\theta\in F_i} r(\theta, \delta))\cdot\xi(F_i) \tag{3.1}$$

where $\mathcal{F}$ *is a sub-σ-field generated by a finite* $\mathcal{B}$*-measurable disjoint partition* $\{F_1, F_2, \cdots, F_k\}$, $\bigcup_{j=1}^{k} F_j = \Theta$, $F_j \in \mathcal{B}$, $F_i \cap F_j = \emptyset$, $(i \neq j)$, *of* Θ.

According to Saks' definition [6] of the integral, we have

$$r(\mathcal{B}, \xi, \delta) = \int r(\theta, \delta)\xi(d\theta), \tag{3.2}$$

when $r(\theta, \delta)$ is $\mathcal{B}$-measurable on Θ. Hence, it always holds that

$$r(\mathcal{A}, \xi, \delta) = \int r(\theta, \delta)\xi(d\theta). \tag{3.3}$$

It follows directly from the definition of the mean-max risk that if $\mathcal{C}$ is a sub-σ-field of $\mathcal{B}$, then

$$r(\mathcal{C}, \xi, \delta) \geq r(\mathcal{B}, \xi, \delta) \tag{3.4}$$

for every ξ and δ.

We shall denote by $E_\xi[f(\cdot)|\mathcal{B}]$ the conditional expectation given $\mathcal{B}$ of a bounded $\mathcal{A}$-measurable function $f(\theta)$ of θ with respect to a prior distribution ξ.

LEMMA 1. *The following inequality holds:*

$$\tfrac{1}{2}\int |r(\theta, \delta) - E_\xi[r(\cdot, \delta)|\mathcal{B}]|\xi(d\theta) \leq r(\mathcal{B}, \xi, \delta) - \int r(\theta, \delta)\xi(d\theta). \tag{3.5}$$

PROOF. Since $\int_B r(\theta, \delta)\xi(d\theta) = \int_B E_\xi[r(\cdot, \delta)|\mathcal{B}]\xi(d\theta)$ for $B \in \mathcal{B}$, we have

$$\begin{aligned}\tfrac{1}{2}\int_B &|r(\theta, \delta) - E_\xi[r(\cdot, \delta)|\mathcal{B}]|\xi(d\theta)\\ &= \int_{B_+} (r(\theta, \delta) - E_\xi[r(\cdot, \delta)|\mathcal{B}])\xi(d\theta),\end{aligned} \tag{3.6}$$

where $B_+ = \{\theta: r(\theta, \delta) \geq E_\xi[r(\cdot, \delta)|\mathcal{B}]\} \cap B$. The fact that $\sup_{\theta\in B} r(\theta, \delta) \geq E_\xi[r(\cdot, \delta)|\mathcal{B}]$, ξ-almost everywhere on B, implies the following inequality for every $B \in \mathcal{B}$:

$$\begin{aligned}\int_{B_+} &(r(\theta, \delta) - E_\xi[r(\cdot, \delta)|\mathcal{B}])\xi(d\theta)\\ &\leq \int_{B_+} (\sup_{\theta\in B} r(\theta, \delta) - E_\xi[r(\cdot, \delta)|\mathcal{B}])\xi(d\theta)\\ &\leq \int_{B} (\sup_{\theta\in B} r(\theta, \delta) - E_\xi[r(\cdot, \delta)|\mathcal{B}])\xi(d\theta).\end{aligned} \tag{3.7}$$

Hence, combining (3.6) with (3.7), we have

$$\begin{aligned}\tfrac{1}{2}\int_B &|r(\theta, \delta) - E_\xi[r(\cdot, \delta)|\mathcal{B}]|\xi(d\theta)\\ &\leq (\sup_{\theta\in B} r(\theta, \delta))\xi(B) - \int_B r(\theta, \delta)\xi(d\theta).\end{aligned} \tag{3.8}$$

Let $\mathcal{F}$ be a sub-σ-field of $\mathcal{B}$ generated by a finite $\mathcal{B}$-measurable disjoint partition

$\{B_1, B_2, \cdots, B_k\}$ of Θ. Since (3.8) holds for every B_i, we have, by substituting B_i for B and adding both sides of (3.8),

$$\tag{3.9} \frac{1}{2}\int_\Theta |r(\theta, \delta) - E_\xi[r(\cdot, \delta)|\mathcal{B}]|\xi(d\theta) \leq r(\mathcal{F}, \xi, \delta) - \int_\Theta r(\theta, \delta)\xi(d\theta).$$

This holds for every sub-σ-field $\mathcal{F}$ generated by a finite $\mathcal{B}$-measurable disjoint partition. Taking the infimum of the right side of (3.9), we have the required inequality.

LEMMA 2. *The function $r(\theta, \delta)$ is $\mathcal{B}$-measurable except for a set of ξ-measure zero if $r(\mathcal{B}, \xi, \delta) = \int r(\theta, \delta)\xi(d\theta)$.*

PROOF. The proof is clear from lemma 1.

4. The optimality with respect to partial prior information

DEFINITION. *Write $R(\mathcal{B}, \xi) = \inf_{\delta\in\mathfrak{D}} r(\mathcal{B}, \xi, \delta)$. A procedure $\delta^* \in \mathfrak{D}$ is called* optimal *with respect to a prior information* $(\mathcal{B}, \xi)$, *or simply* $(\mathcal{B}, \xi)$-optimal, *if δ^* satisfies $r(\mathcal{B}, \xi, \delta^*) = R(\mathcal{B}, \xi)$.*

This concept of optimality is similar to the modified minimax property defined by Wesler [7] from the slicing principle point of view. Let $\mathcal{O}$ be a sub-σ-field of $\mathcal{A}$ which consists only of the whole space Θ and the empty set. Clearly, optimal procedures with respect to $(\mathcal{O}, \xi)$ and $(\mathcal{A}, \xi)$ correspond to minimax and ξ-Bayes procedures, respectively.

It is quite reasonable that if two probability measures ξ and η on $\mathcal{A}$ coincide with each other on $\mathcal{B}$, then $r(\mathcal{B}, \xi, \delta) = r(\mathcal{B}, \eta, \delta)$. This property of the mean-max risk implies that the optimality with respect to $(\mathcal{B}, \xi)$ depends only on the marginal distribution of ξ on $\mathcal{B}$. In other words, the optimality with respect to $(\mathcal{B}, \xi)$ does not depend on the conditional probability measure of ξ, given $\mathcal{B}$. For instance, the minimax procedure does not depend on any prior distribution.

5. Definition of parametric sufficiency

DEFINITION. *A sub-σ-field $\mathcal{B}$ of $\mathcal{A}$ is said to be* parametric ξ-sufficient *with respect to $(\Theta, \mathcal{A}, \mathfrak{D}, r)$ (for the sake of brevity we shall simply call $\mathcal{B}$ a* ξ-sufficient σ-field *if no confusion occurs) if $R(\mathcal{B}, \xi) = R(\mathcal{A}, \xi)$. And if $\mathcal{B}$ is a ξ-sufficient σ-field for every prior probability measure ξ on $(\Theta, \mathcal{A})$, $\mathcal{B}$ is said to be* sufficient *with respect to $(\Theta, \mathcal{A}, \mathfrak{D}, r)$.*

It is a direct implication from the definition that if $\mathcal{C}$ is a sub-σ-field of a sub-σ-field $\mathcal{B}$ of $\mathcal{A}$ and $\mathcal{C}$ is a ξ-sufficient sub-σ-field of $\mathcal{A}$, then $\mathcal{B}$ is also a ξ-sufficient sub-σ-field of $\mathcal{A}$.

Concepts analogous to ξ-sufficiency have appeared implicitly in some previous papers. One such concept is that of the least favorable distribution: in a strictly determined statistical game, the ξ-sufficiency of the sub-σ-field $\mathcal{O}$ of $\mathcal{A}$ is equivalent to the fact that ξ is least favorable. Another example of this appeared in Blyth's paper [8] and Hodges-Lehmann's paper [9]. They considered statistical problems with two risk functions. According to them, if a procedure δ_0 minimizes an average risk $\alpha_1 \int r_1(\theta, \delta)\, d\xi_1 + \alpha_2 \int r_2(\theta, \delta)\, d\xi_2$ for some $\alpha_1 > 0$ and $\alpha_2 > 0$ and if

$$\int r_2(\theta, \delta_0)\, d\xi_2 = \sup_{\theta \in \Theta} r_2(\theta, \delta_0), \tag{5.1}$$

then δ_0 is a Bayes solution relative to ξ_1 (with respect to the risk $r_1(\theta, \delta)$) within the class of δ's for which $\sup_{\theta \in \Theta} r_2(\theta, \delta) \leq \sup_{\theta \in \Theta} r_2(\theta, \delta_0)$. To compare this result with our definition of ξ-sufficiency, we introduce a new parameter space $\Theta^* = \Theta \times \{1, 2\}$ and a risk function $r^*(\theta^*, \delta) = r^*((\theta, i), \delta) = r_i(\theta, \delta)$ on Θ^*, $i = 1$ and 2. Then, regarding $\xi = (\alpha_1, \alpha_2, \xi_1, \xi_2)$ as a prior distribution on Θ^*, the condition (5.1) will correspond to the ξ-sufficiency of the sub-σ-field {the empty set, $\Theta \times \{2\}$, (all measurable sets of Θ) $\times \{1\}$, Θ^*}. A similar consideration will be effective for the minimax procedure within a restricted class and for more general cases.

The following lemma is stated for the purpose of later use.

LEMMA. *Let ξ be a prior probability measure on $\mathcal{A}$, and $\mathcal{B}$ a ξ-sufficient sub-σ-field of $\mathcal{A}$ with respect to $(\Theta, \mathcal{A}, \mathcal{D}, r)$. Let ω be a $\mathcal{B}$-measurable subset of Θ, $1 > \xi(\omega) > 0$, and $s(\theta)$ an $\mathcal{A}$-measurable function on ω such that $0 \leq s(\theta) \leq 1$ and*

$$E_\xi[s(\theta) | \mathcal{B}] = \text{constant } c(\neq 0, 1), \qquad \xi\text{-a.e. on } \omega. \tag{5.2}$$

We shall write

$$\begin{aligned} r_1(\delta) &= \frac{1}{1 - \xi(\omega)} \int_{\Theta - \omega} r(\theta, \delta) \xi(d\theta), \\ r_2(\delta) &= \frac{1}{c\xi(\omega)} \int_\omega r(\theta, \delta) s(\theta) \xi(d\theta), \\ r_3(\delta) &= \frac{1}{(1 - c)\xi(\omega)} \int_\omega r(\theta, \delta)(1 - s(\theta)) \xi(d\theta). \end{aligned} \tag{5.3}$$

Let $\Theta^* = \{1, 2, 3\}$, $\mathcal{A}^* =$ the σ-field of all subsets of Θ^*, $\mathcal{D}^* = \mathcal{D}$, $r^*(i, \delta) = r_i(\delta)$, $\mathcal{B}^* =$ {empty set, Θ^*, $\{1\}$, $\{2, 3\}$} and $\xi^*(1) = 1 - \xi(\omega)$, $\xi^*(2) = c\xi(\omega)$, $\xi^*(3) = (1 - c\xi(\omega)$. Then $\mathcal{B}^*$ is ξ^*-sufficient with respect to $(\Theta^*, \mathcal{A}^*, \mathcal{D}^*, r^*)$.

PROOF. For every disjoint finite $\mathcal{B}$-measurable partition $\{F_1, \cdots, F_k\}$ of ω, we have

$$\begin{aligned} &\sum_{i=1}^{k} (\sup_{\theta \in F_i} r(\theta, \delta)) \xi(F_i) \\ &\quad \geq \sum_{i=1}^{k} \xi(F_i) \max \left\{ \int_{F_i} r(\theta, \delta) s(\theta) \xi(d\theta) \Big/ \int_{F_i} s(\theta) \xi(d\theta), \right. \\ &\qquad\qquad \left. \int_{F_i} r(\theta, \delta)(1 - s(\theta)) \xi(d\theta) \Big/ \int_{F_i} (1 - s(\theta)) \xi(d\theta) \right\} \end{aligned} \tag{5.4}$$

$$\geq \max \left\{ \sum_{i=1}^{k} \xi(F_i) \int_{F_i} r(\theta, \delta)s(\theta)\xi(d\theta)/c\xi(F_i), \right.$$
$$\left. \sum_{i=1}^{k} \xi(F_i) \int_{F_i} r(\theta, \delta)(1 - s(\theta))\xi(d\theta)/(1 - c)\xi(F_i) \right\}$$
$$= \max \left\{ \frac{1}{c} \int_{\omega} r(\theta, \delta)s(\theta)\xi(d\theta), \frac{1}{1 - c} \int_{\omega} r(\theta, \delta)(1 - s(\theta))\xi(d\theta) \right\}$$
$$= \xi(\omega) \times \max \{r_2(\delta), r_3(\delta)\}.$$

Since $r^*(\mathcal{B}^*, \xi^*, \delta) = \xi^*(1)r_1(\delta) + \xi(\omega) \max \{r_2(\delta), r_3(\delta)\}$ and

$$(5.5) \qquad r(\mathcal{B}, \xi, \delta) = \inf_{\mathfrak{F}'} \sum_{j=1}^{k'} (\sup_{\theta \in F'_j} r(\theta, \delta))\xi(F'_j) + \inf_{\mathfrak{F}''} \sum_{i=1}^{k''} (\sup_{\theta \in F''_i} r(\theta, \delta))\xi(F''_i)$$

for finite partitions $\mathfrak{F}'$ of ω and $\mathfrak{F}''$ of $\Theta - \omega$, we have $r(\mathcal{B}, \xi, \delta) \geq r^*(\mathcal{B}^*, \xi^*, \delta)$. Since $r^*(\mathcal{A}^*, \xi^*, \delta) = \xi^*(1)r_1(\delta) + \xi^*(2)r_2(\delta) + \xi^*(3)r_3(\delta) = \int r(\theta, \delta)\xi(d\theta) = r(\mathcal{A}, \xi, \delta)$, we have $r(\mathcal{B}, \xi, \delta) \geq r^*(\mathcal{B}^*, \xi^*, \delta) \geq r^*(\mathcal{A}^*, \xi^*, \delta) = r(\mathcal{A}, \xi, \delta)$. From this inequality it is clear that the ξ-sufficiency of $\mathcal{B}$ in $\mathcal{A}$ implies the ξ^*-sufficiency of $\mathcal{B}^*$ with respect to $(\Theta^*, \mathcal{A}^*, \mathfrak{D}^*, r^*)$.

The following diagram is instructive for relations among the concepts of sufficiency and optimality:

$$(5.6) \qquad \begin{array}{ccc} & (A) & \\ & r(\mathcal{A}, \xi, \delta^*) \geq R(\mathcal{A}, \xi) & \\ (C) & \wedge\!\!| \qquad\qquad \wedge\!\!| & (B). \\ & r(\mathcal{B}, \xi, \delta^*) \geq R(\mathcal{B}, \xi) & \\ & (D) & \end{array}$$

In this diagram the equality symbols show us that:
(i) on (A), δ^* is a ξ-Bayes solution,
(ii) on (B), $\mathcal{B}$ is ξ-sufficient,
(iii) on (C), $r(\theta, \delta^*)$ is $\mathcal{B}$-measurable, ξ-a.e.,
(iv) on (D), δ^* is $(\mathcal{B}, \xi)$-optimal.
From these facts we have theorem 1.

THEOREM 1. *Let $(\Theta, \mathcal{A}, \mathfrak{D}, r)$ be a statistical problem. Suppose δ^* is a procedure in $\mathfrak{D}$ and $\mathcal{B}$ a sub-σ-field of $\mathcal{A}$.*

(i) *If $\mathcal{B}$ is ξ-sufficient and δ^* is $(\mathcal{B}, \xi)$-optimal, then $r(\theta, \delta^*)$ is $\mathcal{B}$-measurable, ξ-a.e., and δ^* is a ξ-Bayes solution.*

(ii) *If $r(\theta, \delta^*)$ is $\mathcal{B}$-measurable and δ^* is a ξ-Bayes solution, then $\mathcal{B}$ is ξ-sufficient and δ^* is $(\mathcal{B}, \xi)$-optimal.*

(iii) *If $\mathcal{B}$ is a ξ-complete sub-σ-field (that is, all sets of ξ-measure zero in $\mathcal{A}$ belong to $\mathcal{B}$), then $\mathcal{B}$ is ξ-sufficient and δ^* is $(\mathcal{B}, \xi)$-optimal if and only if $r(\theta, \delta^*)$ is $\mathcal{B}$-measurable and δ^* is a ξ-Bayes solution.*

As a special case of theorem 1, we shall consider a strictly determined statistical game and put $\mathcal{B} = \mathcal{O}$. Then we obtain the following statement: (i) If ξ is least favorable and δ^* is minimax, then $r(\theta, \delta^*)$ is constant, ξ-a.e., and δ^* is a ξ-Bayes procedure. (ii) If $r(\theta, \delta)$ is constant and δ^* is a ξ-Bayes procedure, then ξ is least favorable and δ^* is minimax (cf. [10], theorems 3.9 and 3.10).

The next theorem is more interesting.

THEOREM 2. *Suppose $\mathcal{B}$ is a sub-σ-field of $\mathcal{A}$ and there exists a ξ-Bayes procedure δ^* in $\mathfrak{D}$. If $\mathcal{B}$ is sufficient with respect to $(\Theta, \mathcal{A}, \mathfrak{D}, r)$, then $r(\theta, \delta^*)$ is $\mathcal{B}$-measurable, ξ-a.e.,*

PROOF. Let

$$\begin{aligned} \omega_1 &= \{\theta : r(\theta, \delta^*) > E_\xi[r(\cdot, \delta^*)|\mathcal{B}]\}, \\ \omega_2 &= \{\theta : r(\theta, \delta^*) = E_\xi[r(\cdot, \delta^*)|\mathcal{B}]\}, \\ \omega_3 &= \{\theta : r(\theta, \delta^*) < E_\xi[r(\cdot, \delta^*)|\mathcal{B}]\}, \end{aligned} \tag{5.7}$$

and

$$\omega = \{\theta : \xi(\omega_2 \cup \omega_3|\mathcal{B}) = 0\} (\in \mathcal{B}). \tag{5.8}$$

Since

$$\xi(\omega \cap (\omega_2 \cup \omega_3)) = \int_\omega \xi(\omega_2 \cup \omega_3|\mathcal{B})\xi(d\theta) = 0, \tag{5.9}$$

we have

$$\begin{aligned} &\int_{\omega \cap \omega_1} \{r(\theta, \delta^*) - E_\xi[r(\cdot, \delta^*)|\mathcal{B}]\}\xi(d\theta) \\ &= \int_\omega \{r(\theta, \delta^*) - E_\xi[r(\cdot, \delta^*)|\mathcal{B}]\}\xi(d\theta) - \int_{\omega \cap (\omega_2 \cup \omega_3)} \{r - E_\xi[r|\mathcal{B}]\}\xi(d\theta) \\ &= \int_\omega \{r(\theta, \delta^*) - E_\xi[r(\cdot, \delta^*)|\mathcal{B}]\}\xi(d\theta) = 0. \end{aligned} \tag{5.10}$$

Therefore $\xi(\omega \cap \omega_1) = 0$, and so $\xi(\omega) = \xi(\omega \cap \omega_1) + \xi(\omega \cap (\omega_2 \cup \omega_3)) = \xi(\omega \cap \omega_1) = 0$, which means that $\xi(\omega_2 \cup \omega_3|\mathcal{B}) > 0$, ξ-a.e.

Take the indicator function $\chi(\theta)$ of the set $\omega_2 \cup \omega_3$ and consider a probability measure $\eta(\sigma)$ on $\mathcal{A}$:

$$\eta(\sigma) = \int_\sigma \xi(\omega_2 \cup \omega_3|\mathcal{B})^{-1}\chi(\theta)\xi(d\theta), \qquad \sigma \in \mathcal{A}. \tag{5.11}$$

For any $\mathcal{B}$-measurable set τ we have

$$\begin{aligned} \eta(\tau) &= \int_\tau \xi(\omega_2 \cup \omega_3|\mathcal{B})^{-1}\chi(\theta)\xi(d\theta) \\ &= \int_\tau \xi(\omega_2 \cup \omega_3|\mathcal{B})^{-1}E_\xi[\chi|\mathcal{B}]\xi(d\theta) \\ &= \int_\tau \xi(\omega_2 \cup \omega_3|\mathcal{B})^{-1}\xi(\omega_2 \cup \omega_3|\mathcal{B})\xi(d\theta) \\ &= \xi(\tau). \end{aligned} \tag{5.12}$$

Therefore, two measures ξ and η coincide with each other on $\mathcal{B}$, and so we have

$$r(\mathcal{B}, \xi, \delta) = r(\mathcal{B}, \eta, \delta) \qquad \text{for every} \quad \delta \in \mathfrak{D}. \tag{5.13}$$

On the other hand,

$$(5.14)\qquad \begin{aligned} r(\mathcal{A}, \eta, \delta) &= \int r(\theta, \delta)\, d\eta \\ &= \int r(\theta, \delta)\chi(\theta)\xi(\omega_2 \cup \omega_3|\mathcal{B})^{-1}\xi(d\theta) \\ &= \int_{\omega_2 \cup \omega_3} r(\theta, \delta)\xi(\omega_2 \cup \omega_3|\mathcal{B})^{-1}\xi(d\theta). \end{aligned}$$

By the definition of ω_2 and ω_3, we have

$$(5.15)\qquad \begin{aligned} &\int_{\omega_2 \cup \omega_3} r(\theta, \delta^*)\xi(\omega_2 \cup \omega_3|\mathcal{B})^{-1}\xi(d\theta) \\ &\leq \int_{\omega_2 \cup \omega_3} E_\xi[r(\theta, \delta^*)|\mathcal{B}]\xi(\omega_2 \cup \omega_3|\mathcal{B})^{-1}\xi(d\theta) \\ &= \int E_\xi[r(\theta, \delta^*)|\mathcal{B}]\xi(\omega_2 \cup \omega_3|\mathcal{B})^{-1}\chi(\theta)\xi(d\theta) \\ &= \int E_\xi[r(\theta, \delta^*)|\mathcal{B}]\, d\eta \\ &= \int E_\xi[r(\theta, \delta^*)|\mathcal{B}]\, d\xi = r(\mathcal{A}, \xi, \delta^*), \end{aligned}$$

where the equality sign in the second row holds if and only if $\xi(\omega_3) = 0$. Thus we have

$$(5.16)\qquad r(\mathcal{A}, \eta, \delta^*) \leq r(\mathcal{A}, \xi, \delta^*),$$

where the equality sign holds if and only if $\xi(\omega_3) = 0$. Here the reader should notice that $\xi(\omega_3) = 0$ is equivalent to $\xi(\omega_1) = 0$.

Since $\mathcal{B}$ is sufficient by assumption, we have

$$(5.17)\qquad \begin{aligned} R(\mathcal{B}, \xi) &= R(\mathcal{A}, \xi), \\ R(\mathcal{B}, \eta) &= R(\mathcal{A}, \eta), \end{aligned}$$

and from (5.13) we also have $R(\mathcal{B}, \xi) = R(\mathcal{B}, \eta)$. Hence $R(\mathcal{A}, \xi) = R(\mathcal{A}, \eta)$. Since δ^* is ξ-Bayes in $\mathfrak{D}$, $r(\mathcal{A}, \xi, \delta^*) = R(\mathcal{A}, \xi)$, and hence $r(\mathcal{A}, \xi, \delta^*) = R(\mathcal{A}, \eta) \leq r(\mathcal{A}, \eta, \delta^*)$. Therefore, it follows from (5.16) and the above inequality that $r(\mathcal{A}, \xi, \delta^*) = r(\mathcal{A}, \eta, \delta^*)$. This shows that $\xi(\omega_1) = \xi(\omega_3) = 0$, that is,

$$(5.18)\qquad r(\theta, \delta^*) = E_\xi[r(\theta, \delta^*)|\mathcal{B}], \qquad \xi\text{-a.e.}$$

COROLLARY. *If $\mathcal{B}$ is sufficient with respect to $(\Theta, \mathcal{A}, \mathfrak{D}, r)$, and is ξ-complete in $\mathcal{A}$, then the ξ-Bayes property of a procedure in $\mathfrak{D}$ is equivalent to $(\mathcal{B}, \xi)$-optimality.*

PROOF. The implication of $(\mathcal{B}, \xi)$-optimality from ξ-Bayes property is easily seen from theorem 2, whereas the inverse implication follows from theorem 1.

6. The case of testing hypotheses

Let $(X, \mathbf{A})$ be a measurable space, with the sample space X having an associated σ-field $\mathbf{A}$. And let the parameter space Θ, having an associated σ-field $\mathcal{A}$, be a collection of θ's, to each of which corresponds a probability measure P_θ on

$(X, \mathbf{A})$ in such a manner that, for any subset $A \in \mathbf{A}$ of X, $P_\theta(A)$ is an $\mathcal{A}$-measurable function on Θ. Let ω be an $\mathcal{A}$-measurable, nonempty and true subset of Θ and then consider a problem of testing a hypothesis "$\theta \in \omega$" against the alternative "$\theta \notin \omega$." By Φ we shall denote the set of all test functions φ, namely the set of all $\mathbf{A}$-measurable functions φ on X satisfying $0 \leq \varphi(x) \leq 1$. The problem described above will be denoted by $(X, \mathbf{A}, \Theta, \mathcal{A}, P_\theta, \omega)$. Here the risk function $r(\theta, \varphi)$ of φ is automatically understood as

$$(6.1) \qquad r(\theta, \varphi) = \begin{cases} E_\theta[\varphi] & \text{for} \quad \theta \in \omega, \\ 1 - E_\theta[\varphi] & \text{for} \quad \theta \notin \omega, \end{cases}$$

where E_θ stands for the average operator with respect to the probability distribution P_θ on $(X, \mathbf{A})$. As is easily seen, for any prior probability measure ξ on $(\Theta, \mathcal{A})$ there exists at least one ξ-Bayes test $\varphi*$.

Let $\mathcal{B}$ be a sub-σ-field of $\mathcal{A}$. Obviously $P_\theta(A)$ is $\mathcal{B}$-measurable for every $A \in \mathbf{A}$ if and only if $E_\theta[\varphi]$ is $\mathcal{B}$-measurable for every $\varphi \in \Phi$. We shall discuss below the relation between the $\mathcal{B}$-measurability of $P_\theta(A)$ and the sufficiency of $\mathcal{B}$ with respect to $(X, \mathbf{A}, \Theta, \mathcal{A}, P_\theta, \omega)$, provided that ω is $\mathcal{B}$-measurable.

First we shall observe a corollary of theorem 2.

COROLLARY. *If $\mathcal{B}$ is a sufficient sub-σ-field of $\mathcal{A}$ with respect to $(X, A, \Theta, \mathcal{A}, P_\theta, \omega)$ and ω is $\mathcal{B}$-measurable, then, for any ξ-Bayes test $\varphi*$, $E_\theta[\varphi*]$ is $\mathcal{B}$-measurable, ξ-a.e., and φ^* is $(\mathcal{B}, \xi)$-optimal whenever $\mathcal{B}$ is ξ-complete in $\mathcal{A}$.*

As preparation for obtaining the main theorem, we shall give some lemmas without proof, concerning the problem of testing simple hypotheses. In these lemmas we shall use notations Q_0, Q_1, Q_2, and so on, for measures defined on $(X, \mathbf{A})$, and E_i for the average operation with respect to Q_i $(i = 0, 1, \cdots)$. And moreover, by $(Q_i:Q_j)$ we mean the problem of testing a simple hypothesis Q_i against a simple alternative Q_j.

LEMMA 1. *For the problem $(Q_1:Q_2)$ there is a system $\{\varphi_\alpha\}_{0 \leq \alpha \leq 1}$ of most powerful test functions for the hypothesis Q_1 against Q_2 such that $E_1[\varphi_\alpha] = \alpha$ and $\varphi_\alpha(x) \leq \varphi_{\alpha'}(x)$ on X if $\alpha < \alpha'$. Moreover, for any such system $\{\varphi_\alpha\}$ we can choose a nonnegative function $k(\alpha) \leq \infty$ on $[0, 1]$ such that the inequalities*

$$(6.2) \qquad \begin{aligned} k(\alpha)E_1[(1 - \varphi_\alpha)f] &\geq E_2[(1 - \varphi_\alpha)f], \\ k(\alpha)E_1[\varphi_\alpha g] &\leq E_2[\varphi_\alpha g], \end{aligned}$$

hold for all nonnegative $\mathbf{A}$-measurable functions f and g.

LEMMA 2. *Let $\{\varphi_\alpha\}$ and $\{\psi_\alpha\}$ be systems of the most powerful test functions for the problems $\{Q_1:Q_2\}$ and $\{Q_1:Q_3\}$, respectively, which are the systems defined in lemma 1. If, for any $\beta \in [0, 1]$, there are nonnegative numbers $k(\beta) \leq \infty$ and $\alpha \in (0, 1]$ such that $k(\beta)$ satisfies the same condition for $\{\psi_\beta\}$ as does $k(\alpha)$ for $\{\varphi_\alpha\}$ in lemma 1 and*

$$(6.3) \qquad \begin{aligned} k(\beta)E_1[(1 - \varphi_\alpha)\psi_\beta] &= E_2[(1 - \varphi_\alpha)\psi_\beta], \\ k(\beta)E_1[\varphi_\alpha(1 - \psi_\beta)] &= E_2[\varphi_\alpha(1 - \psi_\beta)], \end{aligned}$$

then $\{\varphi_\alpha\}$ is, in turn, a system of the most powerful test functions for the problem $\{Q_1:Q_3\}$.

LEMMA 3. *With the same notation as in lemma 2, we suppose that* $Q_2 + Q_3$ *is absolutely continuous with respect to* Q_1. *Then it is a necessary and sufficient condition for* $Q_2 = Q_3$ *that there be a set* $\{\varphi_\alpha\}_{0 \leq \alpha \leq 1}$ *of* **A**-*measurable functions on* X *such that* $\{\varphi_\alpha\}_{0\leq\alpha\leq 1}$ *is a system of the most powerful test functions for* $\{Q_1 : Q_2\}$ *as well as for* $\{Q_1 : Q_3\}$, *and* $E_2[\varphi_\alpha] = E_3[\varphi_\alpha]$ *holds for all* $\alpha \in [0, 1]$.

THEOREM 3. *Denote by* T *a problem* $(X, \mathbf{A}, \Theta, \mathcal{A}, P_\theta, \omega)$, *where* $\{P_\theta : \theta \in \Theta\}$ *is mutually absolutely continuous. Let* $\mathcal{B}$ *be a sub-*σ*-field of* $\mathcal{A}$ *and* ω *a* $\mathcal{B}$*-measurable nonempty and true subset of* Θ.

(i) *If* $P_\theta(A)$ *is a* $\mathcal{B}$*-measurable function of* θ *for any* **A**-*measurable subset* A *of* X, *then* $\mathcal{B}$ *is sufficient with respect to* T.

(ii) *If* $\mathcal{B}$ *is sufficient with respect to* T, *then* $P_\theta(A)$ *is* $\mathcal{B}$*-measurable,* ξ*-a.e., as a function of* θ *for any fixed* **A**-*measurable subset* $A \subset X$, *and for any prior distribution* ξ *on* $(\Theta, \mathcal{A})$ *for which* $1 > \xi(\omega) > 0$.

PROOF. Assertion (i) is clear from the definitions of the mean-max risk and sufficiency of $\mathcal{B}$ and the $\mathcal{B}$-measurability of ω.

For (ii), suppose that $\mathcal{B}$ is sufficient in $\mathcal{A}$ and $P_\theta(A_0)$ is not $\mathcal{B}$-measurable, ξ-a.e., for some **A**-measurable subset A_0 of the sample space X, that is,

$$\xi\{\theta : P_\theta(A_0) \neq E_\xi[P.(A_0)|\mathcal{B}]\} > 0. \tag{6.4}$$

Without any loss of generality we may assume that

$$\xi\{\theta \in \omega : P_\theta(A_0) \neq E_\xi[P.(A_0)|\mathcal{B}]\} > 0. \tag{6.5}$$

We shall show here that it is possible to take an $\mathcal{A}$-measurable function $s(\theta)$ on ω such that $0 \leq s(\theta) \leq 1$ and

$$E_\xi[s(\theta)|\mathcal{B}] = \tfrac{1}{2}, \qquad \xi\text{-a.e. on } \omega, \tag{6.6}$$

and

$$\int_\omega s(\theta) P_\theta(A_0)\xi(d\theta) < \tfrac{1}{2} \int_\omega P_\theta(A_0)\xi(d\theta). \tag{6.7}$$

For any $\mathcal{B}$-measurable nonnegative function $k(\theta)$ on ω, let us write

$$S_k = \{\theta \in \omega : P_\theta(A_0) < k(\theta) E_\xi[P.(A_0)|\mathcal{B}]\} \tag{6.8}$$

and

$$T_k = \{\theta \in \omega : P_\theta(A_0) > k(\theta) E_\xi[P.(A_0)|\mathcal{B}]\}. \tag{6.9}$$

Denote by $\mathcal{K}$ the collection of all $k(\theta)$ such that $\xi(S_k|\mathcal{B}) \leq \frac{1}{2}$ holds ξ-a.e. We can easily see that $\mathcal{K}$ is not empty, because $k \equiv 0$ belongs to $\mathcal{K}$. Since for any k_1 and k_2 in $\mathcal{K}$

$$\xi(S_{k_1 \vee k_2}|\mathcal{B}) = \max\{\xi(S_{k_1}|\mathcal{B}), \xi(S_{k_2}|\mathcal{B})\}, \qquad \xi\text{-a.e.}, \tag{6.10}$$

we have $k_1 \vee k_2 = \max\{k_1, k_2\} \in \mathcal{K}$. Therefore we have a max k_α for any chain $k_1 \prec k_2 \prec \cdots \prec k_\alpha \prec \cdots$ of elements of $\mathcal{K}$, where the notation $k_\nu \prec k_\mu$ means that $k_\nu(\theta) \leq k_\mu(\theta)$ ξ-almost everywhere on ω and $\xi(S_{k_\mu} - S_{k_\nu}|\mathcal{B}) > 0$, ξ-a.e. By Zorn's lemma we can find a maximal element k_0 in $\mathcal{K}$ which belongs also to $\mathcal{K}$, that is,

$$\xi(S_{k_0}|\mathcal{B}) \leq \tfrac{1}{2}, \qquad \xi\text{-a.e.}, \tag{6.11}$$

and there exists no $k \in \mathcal{K}$ such that $k_0 < k$. Write

$$c(\theta) = \frac{\frac{1}{2} - \xi(S_{k_0}|\mathcal{B})}{\xi(\Theta - T_{k_0} - S_{k_0}|\mathcal{B})} \tag{6.12}$$

if the denominator does not equal 0, and let $c(\theta) = 0$ if the denominator is zero, and define

$$s(\theta) = \begin{cases} 1, & \text{on } S_{k_0}, \\ c(\theta), & \text{on } \Theta - S_{k_0} - T_{k_0}, \\ 0, & \text{on } T_{k_0}, \end{cases} \tag{6.13}$$

which is our desired function.

Write, for every A in $\mathbf{A}$,

$$Q_1(A) = \frac{1}{1 - \xi(\omega)} \int_{\omega^c} P_\theta(A)\xi(d\theta), \tag{6.14}$$

$$Q_2(A) = \frac{2}{\xi(\omega)} \int_\omega s(\theta)P_\theta(A)\xi(d\theta), \tag{6.15}$$

$$Q_3(A) = \frac{2}{\xi(\omega)} \int_\omega (1 - s(\theta))P_\theta(A)\xi(d\theta), \tag{6.16}$$

and

$$Q_0(A) = \tfrac{1}{2}(Q_2(A) + Q_3(A)) = \frac{1}{\xi(\omega)} \int_\omega P_\theta(A)\xi(d\theta). \tag{6.17}$$

These Q_0, Q_1, Q_2, and Q_3 are all probability measures on $(X, \mathbf{A})$.

Consider a problem T^* of testing a simple hypothesis Q_1 against a composite alternative $\{Q_2$ or $Q_3\}$. By the lemma in section 5, the sub-σ-field $\mathcal{B}^* = \{$the empty set, $\{1\}$, $\{2, 3\}$, $\Theta^* = \{1, 2, 3\}\}$ is ξ^*-sufficient with respect to T^*, where $\xi^* = (\xi^*(1), \xi^*(2), \xi^*(3))$, $\xi^*(1) = 1 - \xi(\omega)$, $\xi^*(2) = \xi^*(3) = \frac{1}{2}\xi(\omega)$. However, the assumption (6.5) and the definition of Q_1, Q_2, Q_3 are independent of the value $\xi(\omega)$ as long as we have $0 < \xi(\omega) < 1$. From this fact it follows that the sufficiency of $\mathcal{B}$ with respect to T implies the ξ^*-sufficiency of $\mathcal{B}^*$ with respect to T^* for every ξ^* with $\xi^*(1) > 0$, $\xi^*(2) > 0$ and $\xi^*(3) > 0$. In the case where $\xi(\omega) = 0$ or 1, it is obvious that $\mathcal{B}^*$ is ξ^*-sufficient with respect to T^*. Therefore, $\mathcal{B}^*$ is sufficient with respect to T^*.

From the above argument, our theorem is reduced to the following lemma.

LEMMA 4. *Suppose that Q_1, Q_2, and Q_3 are mutually absolutely continuous. If the σ-field $\mathcal{B}^*$ defined above is sufficient with respect to the problem T^*, then Q_2 coincides with Q_3.*

PROOF. Suppose that $\mathcal{B}^*$ is sufficient and that Q_2 does not coincide with Q_3. Let $\{\varphi_\alpha\}$, $0 \le \alpha \le 1$, be a system of the most powerful tests of level α for the problem T_1 of testing a simple hypothesis Q_1 against a simple alternative Q_2 and satisfying the condition that $\alpha < \alpha'$ implies $\varphi_\alpha(x) \le \varphi_{\alpha'}(x)$. We shall take another system $\{\psi_\alpha\}$, $0 \le \alpha \le 1$, of the most powerful tests of level α for the problem T_2 of testing a simple hypothesis Q_1 against a simple alternative Q_0 and satisfying a similar condition: $\alpha < \alpha'$ implies $\psi_\alpha(x) \le \psi_{\alpha'}(x)$.

We shall show first that there are a $\beta \in (0, 1)$ and a $k \in (0, \infty)$ such that, for any $\alpha \in (0, 1)$, the two following inequalities hold with at least one of them being a strict inequality:

$$
\begin{aligned}
kE_1[(1 - \varphi_\alpha)\psi_\beta] &\leq E_0[(1 - \varphi_\alpha)\psi_\beta], \\
kE_1[\varphi_\alpha(1 - \psi_\beta)] &\geq E_0[\varphi_\alpha(1 - \psi_\beta)].
\end{aligned} \tag{6.18}
$$

The existence of a k for which the above formulas hold is guaranteed by lemma 1 (no trouble for $k = 0$ or ∞ occurs, because of the absolute continuity assumption). Suppose that for every $\beta \in (0, 1)$ there is an $\alpha \in (0, 1)$ such that both of the above formulas hold with the equality signs. Then by lemma 2, we can choose $\{\psi_\alpha\}$ as $\varphi_\alpha(x) = \psi_\alpha(x)$ for all α. On the other hand, the most powerful tests $\psi_\alpha(1 > \alpha > 0)$ for $T_2 = (Q_1: Q_0)$ are η^*-Bayes tests for T^*, where $\eta^* = (\eta_1^*, \eta_2^*, \eta_3^*)$, $\eta_i^* > 0$ $(i = 1, 2, 3)$. Since $\mathfrak{B}^*$ is sufficient with respect to T^*, it follows from theorem 2 that the risks at Q_2 and Q_3 are equal, and hence, $E_2[\psi_\alpha] = E_3[\psi_\alpha] = E_0[\psi_\alpha]$ for $0 < \alpha < 1$. Therefore, from lemma 3 we have $Q_2 = Q_0 = Q_3$, which contradicts our assumption.

Thus there is a $\beta \in (0, 1)$ such that for every $\alpha \in (0, 1)$

$$
\begin{aligned}
E_0[\varphi_\alpha(1 - \psi_\beta)] - E_0[(1 - \varphi_\alpha)\psi_\beta] \\
< k\{E_1[\varphi_\alpha(1 - \psi_\beta)] - E_1[(1 - \varphi_\alpha)\psi_\beta]\}.
\end{aligned} \tag{6.19}
$$

Therefore, we have

$$
\begin{aligned}
E_0[\psi_\beta] &> E_0[\varphi_\beta] - k\{E_1[\varphi_\beta] - E_1[\psi_\beta]\} \\
&= E_0[\varphi_\beta],
\end{aligned} \tag{6.20}
$$

and obviously,

$$
E_1[\psi_\beta] = E_1[\varphi_\beta] = \beta. \tag{6.21}
$$

Now we shall consider the closed convex subset

(6.22)

$$
C = \{(E_1[\varphi], 1 - E_2[\varphi], 1 - E_3[\varphi]): 0 \leq \varphi(x) \leq 1, \varphi(x): \mathbf{A}\text{-measurable}\}
$$

of the 3-dimensional Euclidean space, and two points $p = (E_1[\varphi_\beta], 1 - E_2[\varphi_\beta], 1 - E_3[\varphi_\beta])$ and $q = (E_1[\psi_\beta], 1 - E_2[\psi_\beta], 1 - E_3[\psi_\beta])$ in C. By (6.21), p and q have the equal first coordinates. Denote by π the plane which is orthogonal to the first coordinate axis and passes through p and q. Inequality (6.20) makes it possible to determine a pair of positive numbers η_2 and η_3 such that $\eta_2 + \eta_3 < 1$ and

$$
\eta_2 E_2[\psi_\beta] + \eta_3 E_3[\psi_\beta] > \eta_2 E_2[\varphi_\beta] + \eta_3 E_3[\varphi_\beta]. \tag{6.23}
$$

Let $\varphi*$ be a test function such that the point

$$
p^* = (E_1[\varphi^*], 1 - E_2[\varphi^*], 1 - E_3[\varphi^*]) \tag{6.24}
$$

in C is located on the plane π and p^* is a supporting point on π in the direction (η_2, η_3), that is,

$$
E_1[\varphi^*] = E_1[\varphi_\beta] = \beta \tag{6.25}
$$

and

$$\eta_2 E_2[\varphi^*] + \eta_3 E_3[\varphi^*] = \max\,(\eta_2 E_2[\varphi] + \eta_3 E_3[\varphi]\colon 0 \le \varphi(x) \le 1), \tag{6.26}$$

where φ is A-measurable and $E_1[\varphi] = \beta$. Since there is a nonnegative number η_1 such that p^* is also a supporting point of C in the direction (η_1, η_2, η_3) in the 3-dimensional Euclidean space, we have

$$\begin{aligned} &\eta_1 E_1[\varphi^*] + \eta_2(1 - E_2[\varphi^*]) + \eta_3(1 - E_3[\varphi^*]) \\ &\quad = \min\,\{\eta_1 E_1[\varphi] + \eta_2(1 - E_2[\varphi]) + \eta_3(1 - E_3[\varphi])\colon \\ &\qquad\qquad 0 \le \varphi(x) \le 1,\ \varphi\colon \text{A-measurable}\}. \end{aligned} \tag{6.27}$$

Therefore we have, by (6.21), (6.23), and (6.27),

$$\begin{aligned} &\eta_1 E_1[\varphi*] + \eta_2(1 - E_2[\varphi*]) + \eta_3(1 - E_3[\varphi*]) \\ &\quad \le \eta_1 E_1[\psi_\beta] + \eta_2(1 - E_2[\psi_\beta]) + \eta_3(1 - E_3[\psi_\beta]) \\ &\quad < \eta_1 E_1[\varphi_\beta] + \eta_2(1 - E_2[\varphi_\beta]) + \eta_3(1 - E_3[\varphi_\beta]). \end{aligned} \tag{6.28}$$

Since φ_β and φ^* are Bayes tests with respect to T^* and $\mathcal{B}^*$ is sufficient, it follows from theorem 2 that

$$E_2[\varphi^*] = E_3[\varphi^*] \quad \text{and} \quad E_2[\varphi_\beta] = E_3[\varphi_\beta]. \tag{6.29}$$

From (6.28) and (6.29), it follows that

$$(\eta_2 + \eta_3)E_2[\varphi^*] - \eta_1 E_1[\varphi^*] > (\eta_2 + \eta_3)E_2[\varphi_\beta] - \eta_1 E_1[\varphi_\beta]. \tag{6.30}$$

Combining this inequality with (6.25) gives

$$E_2[\varphi^*] > E_2[\varphi_\beta]. \tag{6.31}$$

This inequality shows, with (6.25), that φ_β is not the most powerful test function of level β for the problem T_1 of testing simple hypothesis Q_1 against the alternative Q_2. This is a contradiction.

7. Remarks

(1) A functional $F_\xi[f] = \inf_{\mathfrak{F}\subset\mathcal{B}} \sum_{i=1}^{k} (\sup_{\theta\in F_i} f(\theta))$, $\mathfrak{F} = \{F_1, \cdots, F_k\}$, of an $\mathcal{A}$-measurable function $f(\theta)$ on Θ is also defined as

$$F_\xi[f] = \inf_{\substack{u\,(\epsilon)\,\mathcal{B} \\ u \ge f}} \int u(\theta)\xi(d\theta), \tag{7.1}$$

so that $r(\mathcal{B}, \xi, \delta)$ might be regarded as an upper integral of the risk function $r(\theta, \delta)$ with respect to a sub-σ-field $\mathcal{B}$ of $\mathcal{A}$.

(2) Under certain conditions, the $\mathcal{B}$-measurability of an $\mathcal{A}$-measurable function is equivalent to the $\mathcal{B}$-measurability, ξ-a.e., for any prior distribution ξ on Θ. Therefore, in such cases, the assertion of theorem 3 is simply that $\mathcal{B}$ is sufficient if and only if $P_\theta(A)$ is $\mathcal{B}$-measurable for any A-measurable subset A of the sample space. For example, if $\mathcal{B}$ is induced by a statistic in the Bahadur sense (see [11]), and the induced σ-field in the range of the statistic contains every singleton, then every ξ-almost $\mathcal{B}$-measurable set for any ξ is $\mathcal{B}$-measurable.

(3) From theorem 3 we can get the following statement: under the assumption that the space $\{P_\theta\}$ of distributions is mutually absolutely continuous, the sufficiency of $\mathcal{B}$ with respect to every decision problem with a bounded $\mathcal{B} \times \mathbf{S}$ measurable loss function $L(\theta, s) \geq 0$ implies the $\mathcal{B}$-measurability of $P_\theta(A)$, ξ-a.e., for any prior measure ξ and for any set $A \in \mathbf{A}$, where $\mathbf{S}$ is a σ-field of subsets of the action space.

Inversely, if $P_\theta(A)$ is $\mathcal{B}$-measurable, then $\mathcal{B}$ is sufficient with respect to every decision problem with a bounded $\mathcal{B} \times \mathbf{S}$ measurable loss function $L(\theta, s) \geq 0$.

This kind of assertion is parallel to the characterization of the sufficiency of a statistic due to Blackwell [2] and also to Le Cam [12].

(4) It is well known that for a set S of a 2-dimensional Euclidean space there are two probability measures Q_1 and Q_2 on a measurable space $(X, \mathbf{A})$ such that $S = \{(\int \varphi(x)Q_1(dx), \int \varphi(x)Q_2(dx)) : \varphi \in \Phi\}$, if and only if (i) S is closed and convex, (ii) $(0, 0)$ and $(1, 1) \in S$, (iii) $S \subset [0, 1; 0, 1]$, and (iv) S is symmetric with respect to the point $(\frac{1}{2}, \frac{1}{2})$. For the n-dimensional space we do not know a nice necessary and sufficient condition for a convex set S to be the range of some n-dimensional vector measure ($n \geq 3$). However, our lemma 4 gives a partial solution to this problem. Suppose that $n = 3$ and S, the range set of 3-dimensional vector measure, has only one common point with each coordinate axis, and let π be a plane parallel to the second and third coordinate axes. If every section of S by each of such a plane π is contained in the relative first quadrant, then these sections lie entirely on the plane "the second coordinate = the third coordinate," so that S collapses from three dimensions to two dimensions.

(5) As an example of a sufficient parameter, we can consider the estimable parameters in the linear statistical model

$$\mathbf{X}(n \times 1) = A(n \times k)\boldsymbol{\beta}(k \times 1) + \boldsymbol{\epsilon}(n \times 1), \tag{7.2}$$

where $\mathbf{X}$ and $\boldsymbol{\epsilon}$ are random vectors, A a known matrix, and $\boldsymbol{\beta}$ an unknown vector. Here we assume that the distribution of $\boldsymbol{\epsilon}$ is normal with mean zero-vector and covariance matrix $\sigma^2 I$, I = unit matrix, σ^2 unknown constant. In this problem, $(\boldsymbol{\beta}, \sigma^2)$ is a parameter, and σ^2 together with a system of linearly independent estimable parameters are sufficient. (This example is due to Goro Ishii).

(6) Let $\mathcal{B}$ be a sub-σ-field of $\mathcal{A}$, and $\mathbf{A}(\mathcal{B})$ the family of all $\mathbf{A}$-measurable subsets A of X for which $P_\theta(A)$ is $\mathcal{B}$-measurable. For this family $\mathbf{A}(\mathcal{B})$, analogous assertions to the family of ancillary events in Basu's paper [13] hold. If a sub-σ-field $\mathbf{B}$ of $\mathbf{A}$ is contained in $\mathbf{A}(\mathcal{B})$, then $\mathcal{B}$ is sufficient with respect to every problem of statistical decisions with sample space $(X, \mathbf{B})$. In the case where $\mathcal{B}$ is induced by a function $u(\theta)$ of the parameter θ and $\mathbf{B}$ is induced by a statistic $t(x)$, we could say that $u(\theta)$ is sufficient for the statistic $t(x)$. For example, in the model (7.2) σ^2 is a sufficient parameter for the statistic $t = X'(I - P_A)X$, where P_A is a projection operator of R^n onto the hyperplane spanned by the column vectors of the matrix A. Although t is partially sufficient for σ^2 in Fraser's sense [14] in this case, such an inverse statement is not always true. Our concept of

the parametric sufficiency of $u(\theta)$ for $t(x)$ corresponds to Basu's concept [15] of "φ-free" of $t(x)$ if $\theta = (\varphi, u(\theta))$.

I wish to thank Professor E. W. Barankin for valuable conversations on my problem.

REFERENCES

[1] E. W. Barankin, "Sufficient parameters: solution of the minimal dimensionality problem," *Ann. Inst. Statist. Math.*, Vol. 12 (1960), pp. 91–118.
[2] D. Blackwell, "Equivalent comparison of experiments," *Ann. Math. Statist.*, Vol. 24 (1953), pp. 265–272.
[3] R. R. Bahadur, "A characterization of sufficiency," *Ann. Math. Statist.*, Vol. 26 (1955), pp. 286–292.
[4] H. Kudō, "Dependent experiments and sufficient statistics," *Natur. Sci. Rep. Ochanomizu Univ.*, Vol. 4 (1953), pp. 151–163.
[5] ———, "On sufficiency and completeness of statistics," *Sûgaku*, Vol. 8 (1957), pp. 129–138. (In Japanese.)
[6] S. Saks, *Theory of the Integral*, Monografie Matematyczne, Warszawa, 1937.
[7] O. Wesler, "Invariance theory and a modified minimax principle," *Ann. Math. Statist.*, Vol. 30 (1959), pp. 1–20.
[8] Colin R. Blyth, "On minimax statistical decision procedures and their admissibility," *Ann. Math. Statist.*, Vol. 22 (1951), pp. 22–42.
[9] J. L. Hodges, Jr. and E. L. Lehmann, "The use of previous experience in reaching statistical decisions," *Ann. Math. Statist.*, Vol. 23 (1952), pp. 396–407.
[10] Abraham Wald, *Statistical Decision Functions*, New York, Wiley, 1950.
[11] R. R. Bahadur, "Sufficiency and statistical decision functions," *Ann. Math. Statist.*, Vol. 25 (1954), pp. 423–462.
[12] L. Le Cam, "Sufficiency and approximate sufficiency," *Ann. Math. Statist.*, Vol. 35 (1964), pp. 1419–1455.
[13] D. Basu, "The family of ancillary statistics," *Sankhyā*, Vol. 21 (1959), pp. 247–256.
[14] E. L. Lehmann, *Testing Statistical Hypotheses*, New York, Wiley, 1959.
[15] D. Basu, "Problems relating to the existence of maximal and minimal elements in some families of statistics (subfields)," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1966, Vol. I, pp. 41–50.