ADEQUATE SUBFIELDS AND SUFFICIENCY[1]

By Morris Skibinsky

Brookhaven National Laboratory

1. Introduction. The concept adequacy originated with Kolomgorov [9]. It has been treated on various levels of generality and with a diversity of structure, motivation, and nomenclature. See Bahadur [1], the definition of transitive and sufficient sequence, pp. 452, 453; Loéve [12], p. 351, the definition of conditional independence. Lehmann [10], p. 20, develops an "alternative criterion of sufficiency" for a parametric family of distributions when the "true" parameter is itself random. Raiffa and Schlaifer [13], pp. 32–34, who employ the term Baysian sufficiency, treat aspects of adequacy similar to those considered by Lehmann. At a different structural level, and with opposite motivation and somewhat greater detail, the same notion, as developed by Barankin [2] and Barankin and Kudō [3] is called (quasi-total) parametric sufficiency. Studies in yet another direction are undertaken by Hall, Wijsman, and Ghosh in [7], which explores the relationship between sufficiency, invariance and transitivity. The present paper is in the spirit of the Halmos-Savage-Bahadur approach to sufficiency ([8], [1]).

In Section 2, adequacy is defined in the abstract. It is then shown to be equivalent to sufficiency for an appropriate family of conditional probability measures. This equivalence was noted by Bahadur in [1]. Theorems 1 and 2 of Section 2 are, respectively, a new version of his Theorem 11.3 and a rigorous and somewhat more general version of the statement that follows it.

In Section 3, we show that for predicting an unobservable (real valued, square integrable) random variable Θ, from an observable random variable X, relative to squared difference loss, one need only consider statistics which are "adequate" for X with respect to Θ and the family of possible underlying probabilities.

There follows a summary of the more standard definitions and notation to be employed in subsequent paragraphs. Let $(\Omega, \mathcal{A})$ be a measurable space and $\mathcal{P}$ a family of probability measures P on $\mathcal{A}$.

We use $\mathcal{A}_1 \vee \mathcal{A}_2$ to denote the smallest σ-field which contains each member of two subclasses $\mathcal{A}_1$, $\mathcal{A}_2$ of $\mathcal{A}$, and $\sigma\{h : h \varepsilon \mathcal{H}\}$, to denote the smallest σ-field relative to which every member of a family $\mathcal{H}$ of $\mathcal{A}$-measurable functions is measurable. Let X be a function on Ω, f be a function on the range of X. We denote the composite function on Ω by fX. A set N is $[\mathcal{A}, \mathcal{P}]$ null if $N \varepsilon \mathcal{A}$ and $P(N) = 0$ for each $P \varepsilon \mathcal{P}$. We write $[\mathcal{A}, P]$ if $\mathcal{P}$ is a singleton $\{P\}$. Let p, q be $\mathcal{A}$-measurable functions on Ω. We write $p = q$ $[\mathcal{A}, \mathcal{P}]$ if there exists an $[\mathcal{A}, \mathcal{P}]$ null set N such that $p(\omega) = q(\omega)$ for all $\omega \varepsilon \Omega - N$. For sub-$\sigma$-fields $\mathcal{B}_0$, $\mathcal{B}$ of $\mathcal{A}$, we write $\mathcal{B}_0 \subset \mathcal{B}[\mathcal{A}, \mathcal{P}]$ if to each $B_0 \varepsilon \mathcal{B}_0$ there corresponds a $B \varepsilon \mathcal{B}$ such that

Received 23 September 1965; revised 13 September 1966.

[1] This work performed under the auspices of the U. S. Atomic Energy Commission.

$(B_0 \cap (\Omega - B)) \cup ((\Omega - B_0) \cap B)$ is $[\mathcal{A}, \mathcal{P}]$ null. $\mathcal{P}$ is dominated by a measure μ on $\mathcal{A}$ if every $[\mathcal{A}, \mu]$ null set is also $[\mathcal{A}, \mathcal{P}]$ null. We shall denote the conditional expectation under P of an $\mathcal{A}$-measurable and P-integrable function f, given a sub-σ-field $\mathcal{B}$, by $E_P(f \mid \mathcal{B})$. A sub-σ-field $\mathcal{B}$ of $\mathcal{A}$ is said to be sufficient for a family $\mathcal{P}$ of probability measures on $\mathcal{A}$ if to each $A \in \mathcal{A}$ there corresponds a $\mathcal{B}$-measurable function g_A, say, on Ω such that $P(AB) = \int_B g_A \, dP$, for all $B \in \mathcal{B}$ and all $P \in \mathcal{P}$, i.e., if for all $A \in \mathcal{A}$ and all $P \in \mathcal{P}$, $g_A = E_P(I_A \mid \mathcal{B}) \quad [\mathcal{A}, P]$. Let I_A denote the indicator function of a set A. A σ-field $\mathcal{B}$ is said to be separable if it is generated by a countable subcollection of its sets.

Let $\mathcal{B}$ and $\mathcal{C}$ denote sub-σ-fields of $\mathcal{A}$. A function $P_{\mathcal{B}}^{\mathcal{C}}$ on $\Omega \times \mathcal{B}$ is said to be a conditional probability on $\mathcal{B}$ given $\mathcal{C}$ and P if it is defined by taking $P_{\mathcal{B}}^{\mathcal{C}}(\cdot, B)$ to be a particular version of $E_P(I_B \mid \mathcal{C})$ for each $B \in \mathcal{B}$. $P_{\mathcal{B}}^{\mathcal{C}}(\cdot, B)$ is, of course, a $\mathcal{C}$-measurable function on Ω for each $B \in \mathcal{B}$. If in addition, its defining versions are such that $P_{\mathcal{B}}^{\mathcal{C}}(\omega, \cdot)$ is a probability measure on $\mathcal{B}$ for each $\omega \in \Omega$, then $P_{\mathcal{B}}^{\mathcal{C}}$ will be called a regular conditional probability on $\mathcal{B}$ given $\mathcal{C}$ and P.

We remark (see, e.g., A, p. 361 of [12]) that if $\mathcal{B}$ is generated by a finite or countable family of real valued random variables X, then for each sub-σ-field $\mathcal{C}$, there exists a regular condition probability $P_{\mathcal{B}}^{\mathcal{C}}$.

2. Equivalence of adequacy and sufficiency. Let $\mathcal{P}$ denote a family of probability measures on a σ-field $\mathcal{A}$ of subsets of a space Ω. Throughout, $\mathcal{B}$ and $\mathcal{C}$ will denote sub-σ-fields of $\mathcal{A}$, and $\mathcal{B}_0$ will denote a sub-σ-field of $\mathcal{B}$. In the following, we shall write

$$\mathcal{B}_0 \text{ suf } [\mathcal{P}; \mathcal{B}]$$

to mean that $\mathcal{B}_0$ is sufficient for the family of restrictions to $\mathcal{B}$ of the measures in $\mathcal{P}$.

DEFINITION 1. $\mathcal{B}_0$ is said to be *adequate* for $\mathcal{B}$ with respect to $\mathcal{C}$ and $\mathcal{P}$, symbolically,

$$\mathcal{B}_0 \text{ adq } [\mathcal{B}; \mathcal{C}, \mathcal{P}],$$

if $\mathcal{B}_0$ suf $[\mathcal{P}; \mathcal{B}]$ and if for each C in $\mathcal{C}$, P in $\mathcal{P}$ there exists a $\mathcal{B}_0$-measurable version of $E_P(I_C \mid \mathcal{B})$.

When $\mathcal{P}$ is a singleton $\{P\}$, we shall write $\mathcal{B}_0$ adq $[\mathcal{B}; \mathcal{C}, P]$. Note that in this case, the sufficiency requirement is trivially satisfied. A standard approximation argument (using monotone convergence theorem) shows that $\mathcal{B}_0$ adq $[\mathcal{B}; \mathcal{C}, \mathcal{P}]$ if and only if $\mathcal{B}_0$ suf $[\mathcal{P}; \mathcal{B}]$ and for each P in $\mathcal{P}$ there exists a $\mathcal{B}_0$-measurable version of $E_P(g \mid \mathcal{B})$ for each P-integrable $\mathcal{C}$-measurable function g.

The notion of adequacy as defined above is equivalent, in appropriate context, to Bahadur's transitive sequence of sub-σ-fields when $\mathcal{P}$ is singleton, and otherwise, to his transitive and sufficient sequence [1]. Particular examples will be found in [1] and [7]. Bahadur's Theorem 11.3 in essence states the equivalence of (i) and (iv) in Theorem 1, below. We reintroduce in modified form the family of measures which he considered. To each pair C, P with C in $\mathcal{C}$, P in $\mathcal{P}$

such that $P(C) > 0$, let correspond the probability measure on $\mathcal{A}$ defined by

$$P^C(A) = P(AC)/P(C), \qquad \forall \quad A \varepsilon \mathcal{A}.$$

Take $\mathcal{P}(\mathcal{C})$ to be the family of probability measures so obtained.

THEOREM 1. *The following statements are equivalent*:

(i) $\mathcal{B}_0$ suf $[\mathcal{P}(\mathcal{C}); \mathcal{B}]$.

(ii) *To each* B *in* $\mathcal{B}$ *there corresponds a* $\mathcal{B}_0$*-measurable function* h_B, *say, such that*

$$E_P(I_{B_0}h_B \mid \mathcal{C}) = E_P(I_{B_0B} \mid \mathcal{C}) \quad [\mathcal{C}, P], \qquad \forall \quad B_0 \varepsilon \mathcal{B}_0, \qquad B \varepsilon \mathcal{B}, \qquad P \varepsilon \mathcal{P}.$$

(iii) *To each* B *in* $\mathcal{B}$ *there corresponds a* $\mathcal{B}_0$*-measurable function* h_B, *say, such that*

$$h_B = E_P(I_B \mid \mathcal{B}_0 \vee \mathcal{C}) \quad [\mathcal{A}, P], \qquad \forall \quad P \varepsilon \mathcal{P}.$$

(iv) $\mathcal{B}_0$ adq $[\mathcal{B}; \mathcal{C}, \mathcal{P}]$.

PROOF. Suppose (i). By definition of sufficiency and of the family $\mathcal{P}(\mathcal{C})$, there corresponds to each B in $\mathcal{B}$, a $\mathcal{B}_0$-measurable function h_B such that

$$(2.1) \quad P(B_0BC) = \int_{B_0C} h_B\, dP, \qquad \forall \quad B_0 \varepsilon \mathcal{B}_0, \qquad B \varepsilon \mathcal{B}, \qquad C \varepsilon \mathcal{C}, \qquad P \varepsilon \mathcal{P}.$$

This statement, which by definition of conditional expectation is the equivalent of (ii), in turn clearly implies (i). Moreover,

$$\mathcal{B}_0 \vee \mathcal{C} = \sigma\{I_{B_0C}: B_0 \varepsilon \mathcal{B}_0, C \varepsilon \mathcal{C}\}.$$

Hence by standard arguments using the extension theorem for finite measures, the statement implies (iii). It is of course trivially implied by (iii).

Independently of assertions (i) through (iv) we have that

$$(2.2) \quad \int_{B_0B} E_P(I_C \mid \mathcal{B}_0)\, dP = \int_{B_0C} E_P(I_B \mid \mathcal{B}_0)\, dP,$$

$$\forall \quad B_0 \varepsilon \mathcal{B}_0, \qquad B \varepsilon \mathcal{B}, \qquad C \varepsilon \mathcal{C}, \qquad P \varepsilon \mathcal{P}.$$

This follows from elementary properties of conditional expectation and repeated application of Theorem 1 in [5] or Lemma 4.6 in [1]. Now (iii) implies that

$$h_B = E_P(I_B \mid \mathcal{B}_0) = E_P(I_B \mid \mathcal{B}_0 \vee \mathcal{C}) \quad [\mathcal{A}, P], \qquad \forall \quad P \varepsilon \mathcal{P}.$$

It follows that $\mathcal{B}_0$ suf $[\mathcal{P}; \mathcal{B}]$ and that the right-hand side (and consequently the left-hand side) of (2.2) is equal to $P(B_0BC)$ for all B_0 in $\mathcal{B}_0$, B in $\mathcal{B}$, C in $\mathcal{C}$, and P in $\mathcal{P}$. Hence (taking $B_0 = \Omega$),

$$\int_B E_P(I_C \mid \mathcal{B}_0)\, dP = P(BC), \qquad \forall \quad B \varepsilon \mathcal{B}, \qquad C \varepsilon \mathcal{C}, \qquad P \varepsilon \mathcal{P}.$$

Thus $E_P(I_C \mid \mathcal{B}_0)$ is a $\mathcal{B}_0$-measurable version of $E_P(I_C \mid \mathcal{B})$ and by Definition 1, (iv) holds.

On the other hand, if (iv) holds, the left-hand side (and consequently the right-hand side) of (2.2) is equal to $P(B_0BC)$). Since by (iv), $\mathcal{B}_0$ suf $[\mathcal{P}; \mathcal{B}]$, there exists, corresponding to each B in $\mathcal{B}$, a $\mathcal{B}_0$-measurable function h_B such that (2.1) holds. Q.E.D.

We shall suppose now that to each P in $\mathcal{P}$ there corresponds a regular conditional probability $P_{\mathcal{B}}^{\mathcal{C}}$ and write

$$\mathcal{P}_{\mathcal{B}}^{\mathcal{C}} = \{P_{\mathcal{B}}^{\mathcal{C}}(\omega, \cdot): P \varepsilon \mathcal{P}, \omega \varepsilon \Omega\}.$$

Any class of probability measures so generated by any collection of regular conditional probabilities on $\mathcal{B}$ given $\mathcal{C}$ (one to each P in $\mathcal{P}$) will be called a version of $\mathcal{P}_{\mathcal{B}}^{\mathcal{C}}$.

The following theorem is a rigorous version of the heuristic statement which follows Theorem 11.3 of [1].

THEOREM 2[2]. *If corresponding to each P in $\mathcal{P}$ there exists a regular conditional probability $P_{\mathcal{B}}^{\mathcal{C}}$, then for any version of $P_{\mathcal{B}}^{\mathcal{C}}$*

$$\mathcal{B}_0 \textit{ sufficient for } \mathcal{P}_{\mathcal{B}}^{\mathcal{C}} \Rightarrow \mathcal{B}_0 \operatorname{adq} [\mathcal{B}; \mathcal{C}, \mathcal{P}].$$

If in addition, $\mathcal{B}_0$ and $\mathcal{B}$ are separable, the reverse implication holds for a suitably chosen version of $\mathcal{P}_{\mathcal{B}}^{\mathcal{C}}$.

PROOF. Let f be any $\mathcal{B}$-measurable P-integrable function on Ω. Then by an argument strictly analogous to A, p. 354 [12], if there exists a regular conditional probability $P_{\mathcal{B}}^{\mathcal{C}}$, we may write

$$E_P(f \mid \mathcal{C}) = \int_\Omega f\, dP_{\mathcal{B}}^{\mathcal{C}}, \qquad [\mathcal{C}, P],$$

where, of course, the exceptional set may depend on f.

Suppose that $\mathcal{B}_0$ is sufficient for $\mathcal{P}_{\mathcal{B}}^{\mathcal{C}}$. By definition of sufficiency there corresponds to each B in $\mathcal{B}$, a $\mathcal{B}_0$-measurable function h_B such that

$$P_{\mathcal{B}}^{\mathcal{C}}(\cdot, B_0B) = \int_{B_0} h_B\, dP_{\mathcal{B}}^{\mathcal{C}}, \qquad \forall \quad B_0 \varepsilon \mathcal{B}_0, \qquad P \varepsilon \mathcal{P}.$$

By the definition of regular conditional probability, the left-hand side above is for each P in $\mathcal{P}$ equal $[\mathcal{C}, P]$ to $E_P(I_{B_0B} \mid \mathcal{C})$. By the remark made at the outset of this proof, the right-hand side is for each P in $\mathcal{P}$ equal $[\mathcal{C}, P]$ to $E_P(I_{B_0}h_B \mid \mathcal{C})$. Thus (ii) of Theorem 1 holds and hence $\mathcal{B}_0 \operatorname{adq} [\mathcal{B}; \mathcal{C}, \mathcal{P}]$.

Now suppose that $\mathcal{B}_0$ and $\mathcal{B}$ are separable and that $\mathcal{B}_0 \operatorname{adq} [\mathcal{B}; \mathcal{C}, \mathcal{P}]$, i.e., that (ii) of Theorem 1 is true. By the initial remark of this proof, (ii) of Theorem 1 implies that to each triple (B_0, B, P) with components respectively in $\mathcal{B}_0$, $\mathcal{B}$, and $\mathcal{P}$, there corresponds a $[\mathcal{C}, P]$ null set $N_{B_0,B}(P)$ such that for all ω in $\Omega - N_{B_0,B}(P)$,

$$P_{\mathcal{B}}^{\mathcal{C}}(\omega, B_0B) = \int_{B_0} h_B\, dP_{\mathcal{B}}^{\mathcal{C}}(\omega, \cdot). \tag{2.3}$$

Let $\hat{\mathcal{B}}_0 = \{B_{0i}: i = 1, 2, \cdots\}$ and $\hat{\mathcal{B}} = \{B_j: j = 1, 2, \cdots\}$ be countable generating subclasses of $\mathcal{B}_0$ and $\mathcal{B}$, respectively. (We take them to be fields without loss of generality.) Let

$$N(P) = \bigcup_{i,j} N_{B_{0i},B_j}(P), \qquad \forall \quad P \varepsilon \mathcal{P}.$$

[2] This theorem was discovered independently by Barankin and Kudō [3], subsequent to its discovery by Barndorff-Nielsen and the author [4].

$N(P)$ is $[\mathcal{C}, P]$ null for each P in $\mathcal{P}$ and clearly, for each fixed pair P, ω with P in $\mathcal{P}$ and ω in $\Omega - N(P)$, (2.3) holds for all B_0, B in $\hat{\mathcal{B}}_0$ and $\hat{\mathcal{B}}$, respectively. But h_B depends only on B. Hence for each of these fixed pairs, P, ω, (2.3) holds for all B_0 in $\mathcal{B}_0$ and all B in $\mathcal{B}$. This follows by a familiar argument involving straightforward applications of the monotone convergence theorem, the equivalence of the minimal monotone class and minimal σ-field over the same field, and the extension theorem for finite measures (e.g. [12] pages 60, 87, 124). It is now apparent that $\mathcal{B}_0$ is sufficient for the family

$$(2.4) \qquad \{P_{\mathcal{B}}^{\mathcal{C}}(\omega, \cdot):(P, \omega) \text{ such that } P \varepsilon \mathcal{P} \text{ and } \omega \varepsilon \Omega - N(P)\}.$$

For each P in $\mathcal{P}$, let ζ_P denote an arbitrary but fixed point of $\Omega - N(P)$. Now define the family

$$\mathcal{Q} = \{Q_{P,\omega}:P \varepsilon \mathcal{P}, \omega \varepsilon \Omega\}$$

of probability measures on $\mathcal{B}$ by taking $Q_{P,\omega}$ equal to $P_{\mathcal{B}}^{\mathcal{C}}(\omega, \cdot)$ or $P_{\mathcal{B}}^{\mathcal{C}}(\zeta_P, \cdot)$ for each P in $\mathcal{P}$ according as ω is in $\Omega - N(P)$ or ω is in $N(P)$. It is clear that $\mathcal{Q}$ is a version of $\mathcal{P}_{\mathcal{B}}^{\mathcal{C}}$. Moreover $\mathcal{Q}$ is precisely the family (2.4). Q.E.D.

Theorems 1 and 2 make available standard theorems on sufficiency for application to the theory of adequacy, e.g. [8], [11], [1], [6]. In particular let us consider the question of minimal adequacy. Let $\mathcal{B}^*$ be a sub-σ-field of $\mathcal{B}$.

DEFINITION 2. $\mathcal{B}^*$ is said to be *minimal adequate* for $\mathcal{B}$ with respect to $\mathcal{C}$ and $\mathcal{P}$, symbolically,

$$\mathcal{B}^* \text{ min adq } [\mathcal{B}; \mathcal{C}, \mathcal{P}],$$

if $\mathcal{B}^*$ adq $[\mathcal{B}; \mathcal{C}, \mathcal{P}]$ and $\mathcal{B}_0$ adq $[\mathcal{B}; \mathcal{C}, \mathcal{P}]$ (for a sub-σ-field $\mathcal{B}_0$ of $\mathcal{B}$) implies that $\mathcal{B}^* \subset \mathcal{B}_0 \quad [\mathcal{B}, \mathcal{P}]$.

COROLLARY 1. *If for h a bounded $\mathcal{B}_0$-measurable function, $E_P(I_C h) = 0$ for all C in $\mathcal{C}$ and all P in $\mathcal{P}$ implies that $h = 0 \quad [\mathcal{A}, \mathcal{P}]$*, then

$$\mathcal{B}_0 \text{ adq } [\mathcal{B}; \mathcal{C}, \mathcal{P}] \Rightarrow \mathcal{B}_0 \text{ min adq } [\mathcal{B}; \mathcal{C}, \mathcal{P}].$$

PROOF. The condition is equivalent to bounded completeness for $\mathcal{P}(\mathcal{C})$ on $\mathcal{B}$. By Theorem 1 and the well known theorem of Lehmann-Sheffé [11], the result follows.

When $\mathcal{P}$ is dominated by a σ-finite measure we have the following two corollaries, using Theorem 1 and Theorems 6.2, 6.4 of [1], respectively.

COROLLARY 2. *For each $\mathcal{B} \subset \mathcal{A}$ there is a $\mathcal{B}_0 \subset \mathcal{B}$ such that* $\mathcal{B}_0$ min adq $[\mathcal{B}; \mathcal{C}, \mathcal{P}]$.

REMARK. We note that when $\mathcal{P}$ is a singleton $\{P\}$, then $\sigma\{E_P(I_C \mid \mathcal{B}):C \varepsilon \mathcal{C}\}$ min adq $[\mathcal{B}; \mathcal{C}, P]$.

COROLLARY 3. *Let $\mathcal{B}^* \subset \mathcal{B}_0 \subset \mathcal{B}$. Then* $\mathcal{B}^*$ adq $[\mathcal{B}; \mathcal{C}, \mathcal{P}]$ *if and only if* $\mathcal{B}^*$ adq $[\mathcal{B}_0; \mathcal{C}, \mathcal{P}]$ *and* $\mathcal{B}_0$ adq $[\mathcal{B}; \mathcal{C}, \mathcal{P}]$.

3. Application to least squares prediction. Let X and Θ be random variables defined on $(\Omega, \mathcal{A})$ with underlying probability measure P on $\mathcal{A}$ which is known *a priori*, only to be a member of $\mathcal{P}$. Let t be a measurable function on the range of X, and $\mathcal{B}_0$, $\mathcal{B}$, and $\mathcal{C}$ the σ-fields generated by $T = tX$, X, and Θ, respectively. We shall say that T is an adequate statistic for X with respect to Θ and $\mathcal{P}$ (T adq $[X; \Theta, \mathcal{P}]$), if $\mathcal{B}_0$ adq $[\mathcal{B}; \mathcal{C}, \mathcal{P}]$. Suppose only X to be observable and that while $\mathcal{P}$ is known, the "true" member P which applies is not. For simplicity, suppose Θ to be real valued and square integrable relative to each P in $\mathcal{P}$. We are concerned below with predicting Θ from X, relative to a squared difference loss. In this problem, a measurable function f, say, on the range of X is sought such that the risk

$$E_P(fX - \Theta)^2 = R(f, P), \quad \text{say},$$

is small regardless of which P in $\mathcal{P}$ obtains. It is well known that

$$\inf_f R(f, P) = R(\xi_P, P) = R(P), \quad \text{say},$$

where $\xi_P X = E_P(\Theta \mid X)$ $[\mathcal{B}, P]$. When $\mathcal{P}$ is a singleton (i.e. when P is known), we can use $\xi_P X$ as a predictor of Θ to incur minimal risk. Note that if T adq $[X; \Theta, P]$, then this optimal solution is equal $[\mathcal{B}, P]$ to a function of T. When $\mathcal{P}$ is not a singleton some other stratagem must be employed. For example one might estimate P from X in some reasonable way and use $\xi_P X$, with the estimate substituted for P, to predict Θ. For examples see [14]. The risk associated with such a predictor will of course exceed or at best equal $R(P)$ for each P in $\mathcal{P}$, but the predictor itself is bonifide since it does not depend upon P.

The theorem proved below may be paraphrased roughly as follows. Let T be adequate for X with respect to Θ and $\mathcal{P}$, then to each predictor of Θ from X (of the above described type or not), there corresponds a predictor which is a function only of T with risk uniformly (in $\mathcal{P}$) bounded above by the risk of the first predictor. It follows that in seeking predictors of Θ relative to mean squared loss, we may restrict consideration to functions of adequate statistics.

THEOREM 3. *Let Θ be a square integrable random variable relative to each P in $\mathcal{P}$ and suppose that $T = tX$ is an adequate statistic for X with respect to Θ and $\mathcal{P}$. Then to each measurable function f on the range of X (with fX square integrable for each P in $\mathcal{P}$), there corresponds a measurable function g on the range of t such that*

$$R(gt, P) \leqq R(f, P), \qquad \forall \quad P \in \mathcal{P},$$

with equality holding for any P if and only if $fX = E_P(fX \mid T)$ $[\mathcal{B}, P]$.

PROOF. By (iii) of Theorem 1, there exists a measurable function g, say, on the range of T such that

$$gT = E_P(fX \mid T) = E_P(fX \mid T, \Theta) \quad [\mathcal{A}, P], \qquad \forall \quad P \in \mathcal{P}.$$

Using a well known inequality (e.g. see lemma at the end of Section 2 in [5]) the square integrability of fX and Θ implies the integrability of the products, $(fX)\cdot\Theta$ and $(gT)\cdot\Theta$. It follows (Theorem 1 of [5] or Lemma 4.6 of [1]) that

$$E_P((fX)\cdot\Theta) = E_P(\Theta\cdot E_P((fX) \mid T, \Theta))) = E_P((gT)\cdot\Theta), \qquad \forall \quad P \in \mathcal{P}.$$

Hence

$$R(f, P) - R(gt, P) = E_P(fX)^2 - E_P(gT)^2, \qquad \forall \quad P \in \mathcal{P}.$$

By the well known inequality cited above, this difference is non-negative for each P in $\mathcal{P}$ and by the device already once employed, may be written as $E_P(fX - gT)^2$. Q.E.D.

Observe that if a statistic $T^* = t^*X$ exists which is minimal adequate for X with respect to Θ and $\mathcal{P}$ then taking note of Corollary 3 there exists by the above theorem, a measurable function g^* on the range of t^* such that

$$g^*T^* = E_P(gT \mid T^*) = E_P(fX \mid T^*) \quad [\mathcal{B}, P], \qquad \forall \quad P \varepsilon \mathcal{P},$$

and such that

$$R(g^*t^*, P) \leqq R(gt, P) \qquad \forall \quad P \varepsilon \mathcal{P},$$

equality holding if and only if $gT = g^*T$ $[\mathcal{B}, P]$.

We remark also that Theorem 3 is easily modified to hold for $R(f, P) = E_PL((fX - \Theta))$ where L is convex.

REFERENCES

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

[2] Barankin, E. W. (1960). Sufficient parameters: solution of the minimal dimensionality problem. *Ann. Inst. Statist. Math.* **12** 91–118.

[3] Barankin, E. W. and Kudō, H. (1965). A general theorem on sufficiency. To be published in the *Boletin de la Sociedad Matemática Mexicana.*

[4] Barndorff-Nielsen, O. and Skibinsky, M. (1963). Adequate subfields and almost sufficiency. Applied Mathematics Publication 329, Brookhaven National Laboratory.

[5] Blackwell, D. (1947). Conditional expectation and unbiased sequential estimation. *Ann. Math. Statist.* **18** 105–110.

[6] Burkholder, D. L. (1961). Sufficiency in the undominated case. *Ann. Math. Statist.* **32** 1191–1200.

[7] Hall, W. J., Wijsman, R. A., and Ghosh, J. K. (1965). The relationship between sufficiency and invariance with applications in sequential analysis. *Ann. Math. Statist.* **36** 575–614.

[8] Halmos, P. R. and Savage, L. J. (1949). Application of the Radon-Nikodym theorem to the theory of sufficient statistics. *Ann. Math. Statist.* **20** 225–241.

[9] Kolmogorov, A. (1942). Sur L'estimation Statistique des Parametres de la Loi de Gauss. *Bull. Acad. Sci. U.R.S.S. Ser. Math.* **6** 3–32.

[10] Lehmann, E. L. (1959). *Testing Statistical Hypotheses.* Wiley, New York.

[11] Lehmann, E. L. and Scheffé, H. (1950). Completeness, similar regions, and unbiased estimation—Part I. *Sankhyā* **10** 305–340.

[12] Loève, M. (1960). *Probability theory* (2nd edition). Van Nostrand, New York.

[13] Raiffa, H. and Schlaifer, R. (1961). Applied Statistical Decision Theory. Division of Research, Graduate School of Business Administration. Harvard University.

[14] Robbins, H. (1964). The empirical Bayes approach to statistical decision problems. *Ann. Math. Statist.* **35** 1–20.