ON UNBIASED ESTIMATES OF UNIFORMLY MINIMUM VARIANCE

By R. R. BAHADUR

Indian Statistical Institute, Calcutta

1. Summary

In a given statistical framework let T be the class of all estimates that are the uniformly minimum variance estimates of their respective expected values. Let T_b denote the class of bounded estimates in T. The main conclusions of the paper may then be outlined as follows. (i) There exists a statistic such that T_b is the class of all bounded functions of this statistic; moreover, every real valued function of this statistic is in T. It follows, in particular, that if t is in T_b and u is a real valued function of t, then u is in T. (ii) T contains an unbiased estimate of every estimable parameter if and only if the framework admits a complete sufficient statistic. In this case, as is well known, T is the class of all real valued functions of the complete sufficient statistic.

Conclusion (ii) can also be stated as follows. Suppose that in the given framework the maximum possible reduction of the sample space by means of sufficient statistics has already been carried out. Then either each estimable parameter has a unique unbiased estimate, or there exist estimable parameters that do not admit unbiased estimates of uniformly minimum variance.

A more precise statement and discussion of the above conclusions is deferred to later sections. The conclusions are established under the mild restriction that the sample space is, or may be taken to be, a subset of the m dimensional Cartesian space ($1 \leqslant m \leqslant \infty$), and that the alternative distributions of the sample point admit density functions with respect to a fixed σ-finite measure. It is shown by an example that the restriction to bounded estimates is essential to conclusion (i).

2. Introduction

Let X be a sample space of points x, and suppose that x is distributed in X according to some unknown one of a given set P of probability measures p. Let g be a real parameter, that is, a real valued function on P. g is said to be estimable if there exists at least one unbiased estimate of g, say $t(x)$, such that the variance of t is finite for each p. Suppose that g is estimable. A particular unbiased estimate of g, t_0 say, is said to be efficient at p_0 if the variance of t_0 does not exceed that of any other unbiased estimate when x is distributed according to p_0; t_0 is uniformly efficient if t_0 is efficient at each p in P.

A theory of uniformly efficient estimation was developed by Lehmann and Scheffé (1950, 1955). This theory includes the following application of a theorem of Rao and Blackwell. Suppose that there exists a sufficient statistic, $y = s(x)$ say, such that the set of alternative distributions of y is complete, that is to say, there exist no unbiased estimates of zero that depend on x only through s except for the trivial estimate $t(y) \equiv 0$. Such a statistic is called a complete sufficient statistic. In this case, according to Theorem 5.1 of Lehmann and Scheffé (1950), every estimable g admits a uniformly efficient estimate; moreover, the uniformly efficient estimate of a given g can be characterised as the unbiased estimate that depends on x only through s and is of finite variance for each p, or, alternatively, as the conditional expectation of t given y, where t is any unbiased estimate such that the variance of t is finite for each p.

As may be seen from examples (cf., Girshick, Mosteller and Savage (1946); Halmos (1946); Rao (1946, 1949); Lehmann and Stein (1950); Lehmann and Scheffé (1950, 1955)) the theorem just stated provides a powerful technique for the discovery (or verification, in case a likely estimate is on hand) of the uniformly efficient estimate of a given g. It can also be seen from examples, however, that this technique is not always available, that is to say, a complete sufficient statistic may not exist in a given case (cf., Lehmann and Scheffé (1950); Lehmann and Stein (1950)). The main object of this paper is to show, under certain technical qualifications, that this technique *is* available whenever *every* estimable g admits a uniformly efficient estimate. Some related conclusions, which are trivially true if a complete sufficient statistic exists, but which happen to be valid in general, are also established (cf., para 1 of Section 1).

One of the qualifications referred to above is necessitated by the possible existence of unbiased estimates of zero that are of infinite variance. This difficulty is excluded simply by letting completeness of a statistic y mean that there exist no nontrivial unbiased estimates of zero that depend on x only through y and are of finite variance for each p. It is easily seen that this definition of completeness is adequate for application to minimum variance estimation theory. Another source of difficulty is that it may be impossible to carry out the maximum reduction of the given sample space by means of sufficient statistics (cf., Pitcher, 1957). This difficulty is removed by assuming that (*) X is, or may be taken to be, a Borel set of the m-dimensional Cartesian space ($1 \leqslant m \leqslant \infty$), and that each p in P admits a probability density function with respect to a fixed σ-finite measure. This assumption is valid in most experiments of statistical interest, including sequential ones.

The following Section 3 discusses efficiency at a point. A geometric characterisation of estimates efficient at a given point is obtained (Theorem 1). It is pointed out that this characterisation yields necessary and sufficient conditions (a) in order that the Rao-Blackwell method of improving a given estimate always lead to the efficient estimate, and (b) in order that the sequence of Bhattacharyya lower bounds to the variance always tend to the exact bound. These conclusions concerning efficiency at a point are used in Sections 4 and 5 to establish the main theorems of the paper (Theorems 4 and 5).

The writer is indebted to his colleague C. R. Rao for pointing out that certain propositions closely related to the ones established here are treated in (Rao, 1952). An argument used in the paper just cited has enabled the writer to simplify the proofs of the main theorems of this paper.

In the formal exposition of the following sections it is convenient and of some theoretical advantage to discuss subfields rather than statistics or measurable transformations. Such technical terms as are used without explanation are defined in the first part of (Bahadur, 1954.) The relation between transformations, statistics, and subfields is discussed, for example, in (Bahadur and Lehmann, 1955) and in (Bahadur, 1955a), and conclusions concerning subfields established here can be formulated in other terms (*e.g.* as in Section 1) by reference to these papers. In particular, if (*) holds, then to all intents and purposes a subfield corresponds to a statistic, and an estimate measurable with respect to a subfield is an estimate that depends on the sample point only through the corresponding statistic.

3. Efficiency at a point

Let X be a set of points x, S a field of subsets of X, and P a set of probability measures p on S. The framework X, S, P will remain fixed throughout this section and the following ones. A *parameter* is a real valued function on P. An *estimate* is defined (without reference to any particular parameter) to be a real valued S-measurable function of x. An estimate t is an *unbiased* estimate of a parameter g if t is P-integrable and

$$E_p(t) = g(p) \text{ for each } p \text{ in } P, \qquad \dots \quad (3.1)$$

where E_p denotes expected value when p obtains.

In order to simplify the writing, this section and the following ones are written as if the empty set were the only S-measurable set of p measure zero for each p in P. If this condition is not satisfied in the given case, many of the definitions, arguments and conclusions to follow should be accompanied, strictly speaking, by a null set statement or qualification. For example, the assertion that t_1 is the 'only' estimate with a certain property is likely to mean that if t_2 also has the property then $p(t_2 = t_1) = 1$ for each p in P. Specifically, the following convention is observed throughout the paper. If A and B are sets in S, $A = B$ means that $(A-B)\bigcup(B-A)$ is p-null for each p in P; if t_1 and t_2 are estimates, $t_1 = t_2$ means that $\{x : t_1(x) = t_2(x)\}$ = the whole space X. The relations of inclusion and equality between classes of sets of X, and between classes of estimates, are to be interpreted, of course, in terms of this convention. The interested reader may verify that in this paper the adoption of this convention does not lead to any real difficulty, partly because much of the discussion is in terms of subfields.

It is assumed in this section that P is a dominated set. Let there be given a measure μ on S such that the following three conditions are satisfied: μ is σ-finite; each p in P is absolutely continuous with respect to μ; and each of the density functions $dp/d\mu$ is square integrable with respect to μ, that is, $\int (dp/d\mu)^2 d\mu < \infty$ for each p in P. In most applications, μ is a probability measure in P, or perhaps a mixture of the measures in P, but the following development is valid provided only that μ and P satisfy the conditions stated.

Let V denote the real linear space of all estimates t such that $\int_X t^2 d\mu < \infty$. For $t_1, t_2 \epsilon V$ write $(t_1, t_2) = \int_X t_1 \cdot t_2 d\mu$ and $\| t_1 \| = (t_1, t_1)^{\frac{1}{2}}$. We have

$$E_p(t) = (t, dp/d\mu) \text{ for all } t \epsilon V, p \epsilon P. \quad \ldots \quad (3.2)$$

Let W denote the subspace ($\equiv$ closed linear manifold) spanned by the set $\{dp/d\mu : p \epsilon P\}$.

Let us say that a parameter g is *μ-estimable* if there exists a $t \epsilon V$ that satisfies (3.1). According to (3.2), g is μ-estimable if and only if there exists a $t \epsilon V$ such that $(t, dp/d\mu) \equiv g(p)$. For any μ-estimable g let U_g denote the class of all estimates $t \epsilon V$ that are unbiased estimates of g. It is easily seen that U_g is a closed and convex subset of V.

An estimate t_0 is said to be a *μ-efficient* estimate of g if $t_0 \epsilon U_g$ and $\|t_0\| = \inf \{\|t\| : t \epsilon U_g\}$. Note that in the special case when μ is a probability measure in P, a μ-efficient estimate is simply an unbiased estimate of minimum variance at μ. Also, in this case, the set W is the subspace generated by the likelihood ratios relative to μ.

Theorem 1 below asserts the existence and uniqueness of the μ-efficient estimate, and gives a geometrical description of it. The mathematical content of the theorem is virtually the same as that of the principal theorem of (Stein, 1950), (cf. also Section 8 of Barankin, 1949), but the present statement and proof seem simpler. A similar treatment was given earlier by Basu (1953).

Theorem 1: *Let g be a μ-estimable parameter. (i) The set $U_g \cap W$ consists of one estimate, t_0 say. (ii) $t_0 = \Pi t$ for every $t \epsilon U_g$, where Π is the orthogonal projection to W. (iii) t_0 is μ-efficient. (iv) t_0 is the only μ-efficient estimate of g, that is, if $t \epsilon U_g$ and $t \neq t_0$ then $\|t_0\| < \|t\|$.*

Proof: Since Π is a projection, we have

$$(\Pi t, u) = (t, \Pi u) \qquad \dots \quad (3.3)$$

and

$$\| \Pi t \| \leqslant \| t \|, \qquad \dots \quad (3.4)$$

with equality in (3.4) only if $\Pi t = t$. Moreover, since Π is the projection to W,

$$W = \{t : t \in V, \Pi t = t\} = \{\Pi t : t \in V\} \qquad \dots \quad (3.5)$$

(cf., Halmos (1951)).

Choose and fix a $t \in U_g$ and define

$$t_0 = \Pi t. \qquad \dots \quad (3.6)$$

Then $t_0 \in W$, by (3.5) and (3.6). Also , for each p,

$$\begin{aligned}(t_0, dp/d\mu) &= (\Pi t, dp/d\mu) \text{ by } (3.6)\\ &= (t, \Pi\, dp/d\mu) \text{ by } (3.3)\\ &= (t, dp/d\mu) \text{ by } (3.5)\\ &= g(p) \text{ since } t \in U_g, \qquad \dots \quad (3.7)\end{aligned}$$

so that $t_0 \in U_g$. Thus $t \in U_g$ and (3.6) imply $t_0 \in U_g \cap W$, and $\|t_0\| \leqslant \|t\|$, with equality only if $t_0 = t$ (cf., (3.4), (3.6)).

Since U_g is non-empty by hypothesis, this last conclusion implies, in particular, that $U_g \cap W$ is non-empty. We now show that $U_g \cap W$ cannot contain more than one estimate. Let t_0 and t_1 be functions in $U_g \cap W$. Then $t_1 - t_0 = u$ say, is a function in W such that $(u, dp/d\mu) = (t_1, dp/d\mu) - (t_0, dp/d\mu) = g(p) - g(p) = 0$ for each p in P, that is, u is orthogonal to each $dp/d\mu$; hence u is orthogonal to each t in W; in particular, u is orthogonal to itself, so that $u = 0$. This establishes part (i) of Theorem 1. Parts (ii), (iii) and (iv) are immediate consequences of part (i) and the conclusion of the preceding paragraph. This completes the proof of Theorem 1.

Remark 1 : Suppose that Π is the projection, not to W as in the statement and proof of Theorem 1, but to a subspace containing W. It can then be seen from the above proof that the following is still true : if $t \in U_g$, then $t_0 = \Pi t \in U_g$ and $\|t_0\| \leqslant \|t\|$, with strict inequality unless $t = t_0$.

A method of improving a given unbiased estimate, due to Rao (1945) and Blackwell (1947), is to replace it by its conditional expectation given a suitable statistic. This method is a special case of the one mentioned in remark 1 above. To be precise, suppose $\mu(X) < \infty$, and let S_0 be a subfield of S such that each $dp/d\mu$ (and therefore each t in W) is an S_0-measurable function. For any S-μ-integrable function t let Ct denote the conditional expectation of t given S_0, that is, Ct is the unique S_0-μ-integrable function such that $\int_A t d\mu = \int_A Ct d\mu$ for every S_0-measurable A. It follows easily from well known properties of conditional expectation that, regarded as an operator on V, C is the orthogonal projection to the subspace of S_0-measurable estimates (cf. Moy (1954), Bahadur (1955b)). Since this last subspace contains W, remark 1 applies to C.

It is interesting to examine whether the Rao-Blackwell method not only improves a given estimate but actually yields the μ-efficient one. The preceding discussion shows that this last will be the case, for every μ-estimable g, if and only if W is the class of all S_0-measurable functions in V. Consequently, in order that the Rao-Blackwell method always yields the μ-efficient estimate, S_0 must be the smallest field such that each $dp/d\mu$ is S_0-measurable. This necessary condition is, however, insufficient. Some necessary and sufficient conditions are given in (Bahadur, 1955b). One of these conditions may be stated as follows. Let us say that W is *algebraic* if it contains every constant estimate, and if for any bounded estimate t in W, t^2 is also in W; and that W is *bounded* if the set of bounded estimates in W is everywhere dense (in the L_2 sense) in W. Then, with S_0 the smallest field such that each $dp/d\mu$ is S_0-measurable, W is the set of all S_0-measurable estimates in V if and only if W is algebraic and bounded.

Remark 2 : As a counterpart to remark 1 above, let Π be the projection to a subspace of W. It then follows from Theorem 1 that corresponding to each μ-estimable g there exists a function in the subspace, t^* say, such that $\Pi t = t^*$ for each $t \in U_g$. Consequently, $\|t\| \geqslant \|t^*\|$ for each $t \in U_g$, the inequality being strict unless $t = t^*$.

The preceding remark describes a general method of obtaining lower bounds to the variance of unbiased estimates, of which various bounds in the literature are special cases. In particular, in case $P = \{p_\theta : \theta \text{ in a real interval}\}$, $dp_\theta = f(x, \theta)d\lambda$ where λ is a σ-finite measure, and $\mu = p_{\theta_0}$, then under suitable regularity conditions (cf., Stein (1950)),the n-th Bhattacharyya bound to the variance, $b_n(g)$ say, may be obtained from remark 2 by letting Π be the projection to the subspace spanned by $h_0, h_1, \ldots,$ and h_n, where $h_r(x) = r$-th partial derivative of $f(x, \theta)/f(x, \theta_0)$, evaluated at $\theta = \theta_0$, $r = 0, 1, \ldots, \ldots n$. It follows from the derivation outlined here that $\lim_{n \to \infty} b_n(g) =$ variance of the θ_0-efficient estimate of g(when θ_0 obtains), for every θ_0-estimable g, if and only if each likelihood ratio $f(x, \theta)/f(x, \theta_0)$ is in the subspace spanned by $\{h_r : r = 0, 1, 2, \ldots\}$. The reader may verify that this condition is always satisfied if we can write $f(x, \theta) \equiv \alpha(x) . \beta(\theta). \exp[\theta s(x)]$, $-\infty < \theta < \infty$.

4. Uniform efficiency

In this section we consider the framework $X = \{x\}$, S, and $P = \{p\}$ of the preceding section. It is not assumed for the present, however, that P is a dominated set.

Let U denote the class of all S-measurable estimates $t(x)$ such that

$$E_p(t^2) < \infty \text{ for each } p \in P. \qquad \ldots \quad (4.1)$$

A parameter g is said to be *estimable* if U contains an unbiased estimate of g; g is *boundedly estimable* if there exists a bounded unbiased estimate of g. Clearly, each boundedly estimable parameter is estimable. If g is estimable, a particular unbiased estimate of g, t_0 say, is *uniformly efficient* if for any unbiased $t \epsilon U$ we have

$$E_p(t_0^2) \leqslant E_p(t^2) \text{ for each } p \text{ in } P.$$

Now let S_0 be an arbitrary subfield of S. We shall say that S_0 is *complete* if $t = 0$ is the only S_0-measurable unbiased estimate of zero in U; S_0 is boundedly complete if $t = 0$ is the only bounded S_0-measurable unbiased estimate of zero.

Theorem 2: *If there exists a sufficient and complete subfield, then every estimable parameter admits a uniformly efficient estimate.*

The proof of this theorem is virtually the same as that of Theorem 5.1 of Lehmann and Scheffé (1950) and so is omitted. It follows from the omitted proof that if S_0 is sufficient and complete, the uniformly efficient estimate of a given estimable g is unique, and can be described as the S_0-measurable unbiased estimate in U, or alternatively, as Ct, where t is any estimate satisfying (3.1) and (4.1) and C denotes conditional expectation given S_0.

The above descriptions of the uniformly efficient estimate suggest that there can be at most one subfield S_0' that is sufficient and complete. It can be seen from the following theorem that this is indeed the case, since the necessary and sufficient subfield, if it exists, is unique (cf. Bahadur, 1954).

Theorem 3: *If S_0 is sufficient and boundedly complete, then S_0 is necessary.*

It may be noted here that this theorem does not presuppose the existence of the necessary and sufficient subfield.

Proof: Let S^0 be an arbitrary but fixed sufficient subfield. We have to show that $S_0 \subset S^0$. Consider a set $A \epsilon S_0$. Let χ denote the characteristic function of A, that is, $\chi = 1$ on A and $= 0$ on $X - A$. Let ϕ be the conditional expectation of χ given S^0, and ψ the conditional expectation of ϕ given S_0. Then

$$0 \leqslant \phi \leqslant 1, \quad 0 \leqslant \psi \leqslant 1 \qquad \dots \quad (4.2)$$

and

$$E_p(\psi) = E_p(\phi) = E_p(\chi) = p(A) \text{ for } p \epsilon P. \qquad \dots \quad (4.3)$$

Write $t = \chi - \psi$. Then t is a bounded unbiased estimate of zero, by (4.2) and (4.3), and t is S_0-measurable since χ and ψ are. Consequently, by bounded completeness of S_0, $t = 0$, that is, $\chi = \psi$. Since $\chi^2 = \chi$, this last relation implies

$$\chi \cdot \psi = \chi. \qquad \dots \quad (4.4)$$

It now follows from (4.4), the definition of ψ, and the S_0-measurability of χ, by a well-known property of conditional expectation, that

$$E_p(\chi) = E_p(\chi \cdot \psi) = E_p(\chi \cdot \phi) \text{ for } p \in P. \quad \dots \quad (4.5)$$

We see from (4.3) and (4.5) that

$$E_p[\chi(1-\phi)] = E_p[(1-\chi)\phi] = 0 \text{ for } p \in P. \quad \dots \quad (4.6)$$

It follows from (4.2) and (4.6) that $\phi = \chi$. Since ϕ is S^0-measurable by the definition of ϕ, χ is S^0-measurable, so that A is in S^0. Thus $A \in S_0$ implies $A \in S^0$. This completes the proof of Theorem 3.

The following theorem is a converse of Theorem 3.

Theorem 4: *Suppose that P is a dominated set. If every boundedly estimable parameter admits a uniformly efficient estimate, then a sufficient and complete subfield exists.*

The proof of this theorem (as also of Theorem 5 below) is postponed to the following section.

In the general case, let T denote the class of all uniformly efficient estimates, i.e. $t \in T$ if and only if $t \in U$ and t is the uniformly efficient estimate of g defined by (3.1). Now, if a complete sufficient subfield exists, T can be characterised as the class of all estimates in U that are measurable with respect to this subfield. This suggests that in general (irrespective of whether a complete sufficient subfield exists) T can perhaps be characterised as the class of all estimates in U that are measurable with respect to some (not necessarily sufficient) subfield. A related conjecture is that if t is in T, and $u \in U$ is a function of t, then u is in T. It is shown in Section 6 by an example that neither conjecture is valid in general. The following theorems show, however, that both conjectures are 'nearly' valid. The near validity of the second conjecture was demonstrated earlier by Rao (1952).

Let T_b denote the class of bounded estimates in T. Note that T_b is non-empty, since it certainly contains every constant.

Theorem 5: *There exists a subfield S_0 such that (i) T_b is the class of all bounded S_0-measurable estimates, (ii) every S_0-measurable estimate in U is also in T, and (iii) S_0 is necessary and complete.*

Theorem 6: *If $\phi(z_1, z_2, \dots z_k)$ is a Borel measurable function of k real variables, and if $u = \phi(t_1, t_2, \dots, t_k)$ is in U, where $t_1, t_2, \dots$ and t_k are in T_b, then u is in T.*

Proof: Since the t_i are S_0-measurable by Theorem 5, and ϕ is Borel measurable, it follows that u is an S_0-measurable function in U, and so $u \in T$ by Theorem 5. This completes the proof. It is clear that Theorem 6 and its proof can be generalised, e.g. to pointwise limits of functions of estimates in T_b.

In connection with Theorems 5 and 6 it may be worth while to note that if X is a finite set then $T_b = T$. A less trivial (and perhaps less useful) sufficient condition for the same relation is that P be a finite set.

5. Proof of theorems 4 and 5

We shall first establish Theorem 5. It will be seen that a continuation of the same argument then yields Theorem 4. It is convenient to present the argument as a series of propositions, as follows.

The classes U, T, and T_b of estimates are defined in the preceding section. Let S_0 be the smallest field of sets of X such that each estimate in T_b is S_0-measurable. Further, let N_1 denote the class of all unbiased estimates of zero, and N_2 the subclass of all unbiased estimates of zero that are also in U.

Proposition 1 : An estimate $t \epsilon U$ is in T if and only if $t.z \epsilon N_1$ for every $z \epsilon N_2$.

Proof: For a proof of this well known and important proposition see, for example, Lehmann and Scheffé (1950), Theorem 5.3.

Proposition 2 : T_b is an algebraic linear manifold.

Proof: We have to verify that (i) T_b contains the estimate $t(x) \equiv 1$, (ii) $u, v \epsilon T_b$ implies $\alpha u + \beta v \epsilon T_b$ for all constants α and β, and (iii) $t \epsilon T_b$ implies $t^2 \epsilon T_b$. These verifications can be based on proposition 1. For example, to establish (iii) consider a $z \epsilon N_2$. Then $t \,.\, z$ is in N_1, by proposition 1. Since t is bounded, it follows that $t \,.\, z$ is in fact in N_2. Hence $t \,.\, (t \,.\, z) = t^2 \,.\, z$ is in N_1, by proposition 1. Since $z \epsilon N_2$ is arbitrary, it follows from proposition 1 that $t^2 \epsilon T$, and so $t^2 \epsilon T_b$, since t^2 is bounded. This proof of (iii) is due to Rao (1952).

Now choose and fix a finite measure μ on S, and for any S-measurable function $f(x)$ let $\|f\| = (\int\limits_X f^2 \, d\mu)^{\frac{1}{2}}$.

Proposition 3 : If f is S_0-measurable, and $\|f\| < \infty$, there exists a sequence $\{t_n\}$ in T_b such that $\lim\limits_{n\to\infty} \|f - t_n\| = 0$.

Proof: For the purposes of this proof only, let V denote the normed linear space of all S_0-measurable f with $\|f\| < \infty$, with the usual identification of functions that differ on a set of μ-measure zero. Since μ is an arbitrary measure, this last identification is not necessarily the same as the one used elsewhere in this paper. At any rate, since μ is a finite measure, and since each t in T_b is a (literally) bounded S_0-measurable function, we have $T_b \subset V$. Let Y be the closure of T_b. We have to show that $Y = V$.

Since T_b is a linear manifold of bounded functions, including the constants (cf. proposition 2), it is clear that Y is a bounded subspace, and that Y contains the constants. We shall now show that if f is a bounded function in Y then f^2 is in Y. Given a bounded f in Y, let $\{t_n\}$ be a sequence in T_b such that

$$\lim_{n\to\infty} \| t_n - f \| = 0. \qquad \dots \quad (5.1)$$

Let k be a positive integer, and let c_k be a constant such that $|t_k(x)| \leqslant c_k$. We then have $\|t_k t_n - t_k f\| \leqslant c_k \|t_n - f\|$ for each n. Hence, by (5.1),

$$\lim_{n\to\infty} \| t_k\, t_n - t_k f \| = 0. \qquad \dots \quad (5.2)$$

Since T_b is an algebraic linear manifold (cf. proposition 2), and since $t_k t_n = [(t_k+t_n)^2 - t_k^2 - t_n^2]/2$, we have $t_k\, t_n \,\epsilon\, T_b$. It now follows from (5.2) that

$$t_k\, f \,\epsilon\, Y \qquad \dots \quad (5.3)$$

for every k. Let c be a constant such that $|f(x)| \leqslant c$. We then have $\|t_k f - f^2\| \leqslant c \|t_k - f\|$ for every k. Hence

$$\lim_{k\to\infty} \|t_k\, f - f^2\| = 0 \qquad \dots \quad (5.4)$$

by (5.1). It follows from (5.3) and (5.4), as desired, that $f^2 \,\epsilon\, Y$.

Thus Y is an algebraic and bounded subspace. It follows (Bahadur, 1955b) that there exists a subfield of S_0, say S^0, such that Y is μ-equivalent to the set of all S^0-measurable functions f with $\|f\| < \infty$. Since $T_b \subset Y$, it follows from the definition of S_0 that this S^0 must be μ-equivalent to S_0, and hence $Y = V$. This completes the proof of proposition 3.

Proposition 4 : If $t \,\epsilon\, U$, and t is S_0-measurable, then $t \,\epsilon\, T$.

Proof: Let t be an S_0-measurable function in U. Choose and fix a p in P and a z in N_2. Since $E_p(t^2) < \infty$, it follows from proposition 3 with $\mu = p$ that there exists a sequence $\{t_n\}$ in T_b such that $E_p(t_n - t)^2 \to 0$. Since $E_p(z^2) < \infty$, it follows that $E_p(t_n z) \to E_p(tz)$. However, $E_p(t_n z) = 0$ for each n, by proposition 1. Hence $E_p(tz) = 0$. Since $p \,\epsilon\, P$ and $z \,\epsilon\, N_2$ are arbitrary, it follows from proposition 1 that $t \,\epsilon\, T$, and this completes the proof.

Proposition 5 : S_0 is necessary and complete.

Proof: Let A be a set in S_0 and let χ denote the characteristic function of A. Then $\chi \,\epsilon\, T$ by proposition 4. Consequently, if S^0 is any sufficient subfield, χ must be S^0-measurable, for otherwise the conditional expectation of χ given S^0 would be a different and better unbiased estimate of $g(p) \equiv p(A)$, by the theorem of Rao and Blackwell. Hence $A \,\epsilon\, S^0$ whenever S^0 is sufficient. Since $A \,\epsilon\, S_0$ is arbitrary, it follows that S_0 is necessary. To show that S_0 is complete, let $t \,\epsilon\, U$ an S_0-measurable unbiased estimate of zero. Then t is the uniformly efficient estimate of zero, by proposition 4. Hence $t = 0$. This completes the proof.

We may note here that Theorem 5 is an immediate consequence of the definition of S_0 and propositions 4 and 5. It remains therefore to establish Theorem 4. This will be done by showing that if P is dominated, and T contains an unbiased estimate of every boundedly estimable parameter, then S_0 is sufficient.

Let Q be a countable subset of P such that Q is equivalent to P, i.e. $p(A) = 0$ for all p in Q implies $p(A) = 0$ for all p in P. The existence of such a set Q is assured by Lemma 7 of Halmos and Savage (1949). Let $Q = \{p_1, p_2, \ldots\}$ be an enumeration of Q, let $c_1, c_2, \ldots$ be constants such that

$$c_i > 0, \ \sum_i c_i < \infty \qquad \ldots \quad (5.5)$$

and define

$$\mu(A) = \sum_i c_i p_i(A) \qquad \ldots \quad (5.6)$$

for A in S. We shall then say that μ is a pivotal measure based on Q.

Proposition 6 : For each q in Q, $dq/d\mu$ is S_0-measurable.

Proof: Choose and fix a q in Q. It follows from (5.5) and (5.6) that we can write $\mu = cq + \lambda$ where c is a positive constant and λ is a measure. Hence $1 = c\,(dq/d\mu) + (d\lambda/d\mu)$, and so $0 \leqslant dq/d\mu \leqslant (1/c)$. Thus $dq/d\mu$ is a bounded measurable function.

Write $f = dq/d\mu$, and define $g(p) = E_p(f)$ for p in P. Then g is a boundedly estimable parameter. Consequently, by hypothesis, there exists an unbiased estimate of g in T, say t. In particular, we have $E_p(t) = E_p(f)$, $E_p(t^2) \leqslant E_p(f^2)$ for all p in P. These relations imply, using (5.5) and (5.6), that

$$E_p(t) = E_p(f) \text{ for } p \in Q; \quad \int_X t^2 d\mu \leqslant \int_X f^2 d\mu. \qquad \ldots \quad (5.6)$$

Suppose for the moment that p is restricted to the set Q. Since $dp/d\mu$ is bounded for each p in Q by the first paragraph of this proof, and since μ is a finite measure, Theorem 1 of Section 3 can be applied to determine μ-efficient estimates of μ- estimable parameters on Q. This application shows that $f = dq/d\mu$ is μ-efficient. Consequently, by (5.6), t is also the μ-efficient estimate of the same parameter. Hence $t = f$, by uniqueness of the μ-efficient estimate. Thus f is a bounded estimate in T, i.e. $f \in T_b$. It now follows from the definition of S_0 that $f = dq/d\mu$ is S_0-measurable, and the proof is complete.

Proposition 7 : S_0 is sufficient.

Proof: Let p_1 and p_2 be two measures in P. Choose and fix a countable equivalent subset Q that includes p_1 and p_2 and let μ be a pivotal measure based on Q. It follows from proposition 6 that $dp_i = f_i(x)\, d\mu$, where f_i is S_0-measurable, $i = 1, 2$.

Hence, by the factorization theorem for sufficient subfields, S_0 is sufficient for the set $\{p_1, p_2\}$.

Since p_1 and p_2 are arbitrary measures in P, it follows that S_0 is pairwise sufficient for P. Proposition 7 now follows from the theorem of Halmos and Savage (1949) that in the dominated case pairwise sufficiency is equivalent to sufficiency.

This completes the proof of Theorem 4. The proof shows that the conclusion of the theorem holds provided only that every parameter of the form $g(p) = E_p(dq/d\mu)$ admits a uniformly efficient estimate, where μ is a pivotal measure based on Q, q is a measure in Q, and Q is a countable equivalent subset of P.

6. An example

It is the object of this section to show by an example that a real valued function of a uniformly efficient estimate is not necessarily uniformly efficient. It follows from the theory developed in Sections 4 and 5 that in any such example a complete sufficient statistic does not exist; the class T of all uniformly efficient estimates cannot be characterised as the class of all square integrable functions of some statistic; there exist estimates in T that cannot be approximated (in any reasonable sense) by bounded estimates in T; and both the sample space X and the set P of alternative distributions on X must be infinite.

To construct such an example, let X be the set of points $x = 0, 1, 2, \ldots$ ad inf., and S the field of all sets of X. We shall choose the set P of alternative distributions of x so that $T =$ the class of all linear functions of x. Such a choice of P evidently furnishes the required example.

Let m be a probability distribution such that

$$E_m(x^2) < \infty, E_m(x^4) = \infty \quad \ldots \quad (6.1)$$

and such that

$$E_m(z_1) = E_m(z_2) = 0 \quad \ldots \quad (6.2)$$

where

$$z_1(x) = (-1)^x . x^2 \text{ for } x = 0, 1, 2, \ldots \quad \ldots \quad (6.3)$$

and

$$z_2(x) = \begin{cases} 1 & \text{if } x = 0 \\ (-1)^x . x & \text{if } x = 1, 2, \ldots \end{cases} \quad \ldots \quad (6.4)$$

It is not difficult to see that such a distribution m exists. Let $p_0 = m$. Next, let

$$p_1(x) = \begin{cases} 2/7 & \text{if } x = 0 \\ 4/7 & \text{if } x = 1 \\ 1/7 & \text{if } x = 2 \\ 0 & \text{otherwise} \end{cases} \quad \ldots \quad (6.5)$$

and for $k = 2, 3, \ldots$ let

$$p_k(x) = \begin{cases} k(k+1)/4k^2-2 & \text{if } x = k-1 \\ 2(k^2-1)/4k^2-2 & \text{if } x = k \\ k(k-1)/4k^2-2 & \text{if } x = k+1 \\ 0 & \text{otherwise.} \end{cases} \quad \ldots \quad (6.6)$$

Having thus defined p_θ for integral values of θ in $[0, \infty)$, extend the definition by linear interpolation. Specifically, for any θ with $0 \leqslant \theta < \infty$, let $k = k(\theta)$ be the greatest integer $\leqslant \theta$, let $\alpha = \alpha(\theta)$ be determined by $\alpha k + (1-\alpha)(k+1) = \theta$, $0 < \alpha \leqslant 1$, and define $p_\theta(x) = \alpha p_k(x) + (1-\alpha) p_{k+1}(x)$. Let $P = \{p_\theta : 0 \leqslant \theta < \infty\}$.

To determine the class of uniformly efficient estimates in the framework X, S, P defined above, let the classes N_1 and N_2 of unbiased estimates of zero be defined as in the second paragraph of Section 5. It follows easily from (6.3), (6.4), (6.5) and (6.6) that $E_\theta(z_i) = 0$ for $\theta = 0, 1, 2, \ldots$ and $i = 1, 2$. Hence $E_\theta(z_i) = 0$ for all θ, $i = 1, 2$. Hence $az_1 + bz_2$ is in N_1, for all constants a and b. Now consider an arbitrary z_3 in N_1. Let a and b be constants (possibly both zero) such that with $z_4 = z_3 - az_1 - bz_2$ we have $z_4(x) = 0$ for $x = 0$ and $x = 1$; it is easily seen from (6.3) and (6.4) that such constants exist. Then z_4 is an unbiased estimate of zero, and z vanishes for $x = 0$ and $x = 1$. It now follows from (6.5) and (6.6) that we must have $z_4(x) = 0$ for all x. Hence $z_3 = az_1 + bz_2$.

We conclude that N_1 is the class of all estimates of the form $az_1 + bz_2$. It follows hence by (6.1), (6.3) and (6.4) that N_2 is the class of all estimates of the form bz_2. Since z_2 is never zero, it now follows from proposition 1 that T is the class of all estimates of the form $az_1 + bz_2/z_2$, i.e. of the form $\alpha + \beta x$, since $z_1/z_2 \equiv x$. This completes the verification of the example.

References

Bahadur, R. R. (1954): Sufficiency and statistical decision functions. *Ann. Math. Stat.*, **25**, 423–462.

——— (1955a): Statistics and subfields. *Ann. Math. Stat.*, **26**, 490–497.

——— (1955b): Measurable subspaces and subalgebras. *Proc. Am. Math. Soc.*, **6**, 565–570.

Bahadur, R. R. and Lehmann, E. L. (1955): Two comments on 'sufficiency and statistical decision functions.' *Ann. Math. Stat.*, **26**, 139–142.

Barankin, E. W. (1949): Locally best unbiased estimates. *Ann. Math. Stat.*, **20**, 477–501.

Basu, D. (1953): *Some Contributions to The Theory of Statistical Inference.* D. Phil. thesis, University of Calcutta (unpublished).

Blackwell, D. (1947): Conditional expectation and unbiased sequential estimation. *Ann. Math. Stat.*, **18**, 105–110.

Girshick, M. A., Mosteller, F. and Savage, L. J. (1946): Unbiased estimates for certain binomial sampling problems with applications. *Ann. Math. Stat.*, **17**, 13–23.

Halmos, P. R. (1946): The theory of unbiased estimation. *Ann. Math. Stat.*, **17**, 34–43.

——— (1951) : *Introduction to Hilbert Space and the Theory of Spectral Multiplicity*, New York, Chelsea, 1–48.

HALMOS, P. R. and SAVAGE, L. J. (1949) : Application of the Radon-Nikodym theorem to the theory of sufficient statistics. *Ann. Math. Stat.*, 20, 225–241.

LEHMANN, E. L. and STEIN, C. (1950) : Completeness in the sequential case. *Ann. Math. Stat.*, 21, 376–385.

LEHMANN, E. L. and SCHEFFÉ, H. (1950) : Completeness, similar regions, and unbiased estimation. *Sankhyā*, 10, 305–340.

——— (1955) : Completeness, similar regions, and unbiased estimation. *Sankhyā*, 15, 219–236.

SHU-TEH CHEN MOY, (1954) : Characterisations of conditional expectation as a transformation on function spaces. *Pacific Journal of Mathematics*, 4, 47.

PITCHER, T. S. (1957) : Statistical problems that do not admit necessary and sufficient statistics or subfields. *Ann. Math. Stat.*, 28, 267–268.

RAO, C. R. (1945) : Information and accuracy attainable in the estimation of statistical parameters. *Bull. Cal. Math. Soc.*, 37, 81–91.

——— (1946) : Minimum variance and the estimation of several parameters. *Proc. Camb. Phil. Soc.*, 43, 280–283.

——— (1949) : Sufficient statistics and minimum variance estimates. *Proc. Camb. Phil. Soc.*, 45, 213–218.

——— (1952) : Some theorems on minimum variance estimation. *Sankhyā*, 12, 27–42.

STEIN, C. (1950) : Unbiased estimates of minimum variance. *Ann. Math. Stat.*, 21, 406–415.

Paper received: September, 1957.