STATISTICAL RESEARCH REPORT
Institute of Mathematics
University of Oslo

July 1978

ON COMPLETE SUFFICIENT STATISTICS AND UNIFORMLY MINIMUM VARIANCE UNBIASED ESTIMATORS

by

Erik N. Torgersen [1)]
University of California, Berkeley and University of Oslo

1) Research supported by Norges almenvitenskapelige forskningsråd
AMS 1970 subject classification. 62F10, 62G05

Key words and phrases. Random variables, Coherence,
Minimal Radon-Nikodym derivatives

S U M M A R Y

ON COMPLETE SUFFICIENT STATISTICS AND UNIFORMLY MINIMUM VARIANCE UNBIASED ESTIMATORS

Call a parameter estimable if it has an unbiased estimator with everywhere finite variance. Say that a model has property (RA,BL) if any estimable parameter has a UMVU estimator. Say that a model has property (BA) if it admits a quadratically complete and sufficient statistic.

By the Rado-Blackwell theorem (BA) $\Rightarrow$ (RA,BL) and Bahadur, 1957, showed that (RA,BL) $\Leftrightarrow$ (BA) for dominated models.

In 1964 Le Cam introduced the notion of the M-space of an experiment and thereby extended the usual notion of a bounded random variable. This M-space may be enlarged in order to permit the same extension of the concept of a real random variable. Within this extended framework the equivalence (RA,BL) $\Leftrightarrow$ (BA) holds without qualifications. Restricting ourselves to models where this extended notion of a random variable coinsides with the usual one, we shall see that the condition of dominatedness in Bahadur's theorem may be replaced by a weaker condition which is also applicable to many models in sampling theory.

1. INTRODUCTION

One of the most known, and deservedly so, theorems of mathematical statistics is the Rao-Blackwell theorem. If complete and sufficient statistics exist, then this theorem tells us not only what the UMVU estimators look like, but it also shows how UMVU estimators may be obtained from unbiased ones by conditioning.

The question then naturally arose whether there are experiments which do not allow (quadratically) complete statistics and still have the property that any parameter possessing an unbiased estimator with everywhere finite variance also has a UMVU estimator.

In this generality, and within the usual framework of mathematical statistics, the problem is still open. Bahadur [1], however, settled the problem for dominated models by showing that the answer is negative in this case.

We shall in the last part of this paper show that Bahadur's result extends to a wider class of experiments which includes several of the non dominated models encountered in sampling theory.

Before taking up this problem, however, we shall - following the footsteps of LeCam - consider an extension of the notion of a random variable. We believe that the discussion here shows that we more or less have placed ourselves in the role of a mathematician refusing to get involved with irrational numbers. Continuing this analogy one might tentatively consider Cauchy sequences of random variables for uniformities arising from statistical problems. It then turns out that such sequences may not converge. So where do we find the "irrational variables"? A clue, or rather a complete hint, is implicit in LeCam's paper [5]. Noting that limits of powerfunctions of tests need not be powerfunctions of any test, LeCam extended the notion of a bounded random variable. This was done by imbedding them in the space of bounded linear functionals on the band generated by the underlying probability distributions. Within this extended framework LeCam proved, among many other interesting results, that there is always a minimal algebra within the class of weakly closed and sufficient algebras.

Unbounded variables define linear functionals on the space of measures making the variables integrable and it is quite possible that we could have proceeded this way. A more direct line of attack, however, suggests itself by the fact that the linear functionals considered by LeCam may be represented as uniformly bounded families of random variables satisfying a coherence condition. Dropping the condition of uniform boundedness but keeping the coherence condition we arrive at our "irrational variables". Actually the space obtained that way might also be considered as the completion of the space of real random variables for a uniformity corresponding to everywhere convergence in probability.

Now comes a pleasant surprise. If we admit these new variables as estimators then the Rao-Blackwell theorem and its converse hold without exceptions. Furthermore, by restricting ourselves to experiments such that the new framework coinsides with the usual one, we obtain the generalization of Bahadur's result mentioned above.

These results were obtained by utilizing ideas in Bahadur [1] and in LeCam [5]. The starting point was the observation that a result (Proposition 6) in [1] might be reformulated in order to avoid the assumption of dominatedness. Bahadur showed there that if the model is dominated and the expectations of certain

Radon-Nikodym derivatives were UMVU estimable then the Radon-Nikodym derivatives themselves were the UMVU estimators. If the assumption of dominatedness is deleted, then the arguments of the proof show that the Radon-Nikodym derivatives coinside almost everywhere with UMVU estimators, provided we restrict the underlying distribution to a certain dominated set. If, furthermore, the Radon-Nikodym derivatives are minimal non negative within the extended space of random variables, then Bahadur's conclusion remain valid - provided the assumptions are adapted to the extended framework. The existence of these Radon-Nikodym derivatives is a consequence of the order completeness of the extended space of random variables.

Now we are almost through since the minimal sufficient algebra, whose existence was established in LeCam [5], is the smallest "weakly" closed algebra which contains the constants and all minimal Radon-Nikodym derivatives $dP_{\theta_2}/d(P_{\theta_1}+P_{\theta_2})$; $\theta_1, \theta_2 \in \Theta$. Here Θ is the parameter set and P_θ, for each θ in Θ, is the distribution of our observations when θ holds.

In the case where the extended framework coinsides with the traditional one we obtain the generalization of Bahadur's result described above.

2. AN EXTENSION OF THE CONCEPT OF A RANDOM VARIABLE

We shall in this paper consider experiments $\mathcal{E}$ of the form $\mathcal{E} = (\chi,\mathcal{A},P_\theta:\theta\in\Theta)$ where $(\chi,\mathcal{A})$ is the sample space (i.e. a measureable space) and $(P_\theta:\theta\in\Theta)$ is a family of probability measures on $\mathcal{A}$. The index set Θ is called the parameter set of $\mathcal{E}$. By some abuse terminology any function on Θ may be called a parameter. Often the sample space will be supressed in this notation and we may just write $\mathcal{E} = (P_\theta:\theta\in\Theta)$. Using the terminology established in LeCam [5] the band L of finite measures generated by the P_θ's is the L-space of $\mathcal{E}$ while the M-space, M, of $\mathcal{E}$ is the space of bounded linear functionals on L, i.e. $M = L^*$.

As any abstract L-space may be represented as some band of finite measures on some measurable space we have not excluded any type (in the sense of LeCam [5]) of experiments. Furthermore it is not difficult to see, using these representations, how the concepts below carry over to the general case. The uniformities considered in this paper might as well have been expressed in terms of families of non negative and normalized elements in abstract L-spaces. On the other hand the particular form permits representations in terms of measurable functions and we can keep our discussion within the usual framework of measure theory. Let us begin by an example indicating the need of an extended framework.

Example 1. Put $\chi = [0,1]$, $\mathcal{A}$ = the class of Borel subset of χ and $\Theta = \{-1\} \cap [0,1]$. Let P_θ, for each $\theta \in [0,1]$, be the Dirac measure in θ and let P_{-1} be the uniform distribution on $[0,1]$. This model is clearly complete and $P_{\theta_1} \wedge P_{\theta_2} = 0$ whenever $\theta_1 \neq \theta_2$. Nevertheless a real valued function g on Θ has an unbiased estimator with variance $\equiv 0$ if and only if $g|[0,1]$ is measurable and almost everywhere (Lebesgue) equal to some constant. More generally a real valued function g on Θ has an unbiased estimator v if and only if $g|[0,1]$ is measurable and $g(-1) = \int_0^1 g(\theta)d\theta$. If these conditions are satisfied then $v = g|[0,1]$ so that $\mathrm{Var}_{-1}v = \mathrm{Var}\, g(X)$ where X is uniformly distributed on $[0,1]$.

Although only families of real random variables are needed here we shall, as no extra effort is required, introduce the concepts for variables which are not necessarily real valued.

Let $(\mathcal{Y},\mathcal{B})$ be some measurable space and let $\mathcal{U}$ be a class of sub sets of Θ. We shall then say that a family $(f_\theta:\theta\in\Theta)$ of measurable functions from $\mathcal{E}$ (i.e. $(\chi,\mathcal{A})$) to $(\mathcal{Y},\mathcal{B})$ is $\underline{\mathcal{U}}$ <u>coherent</u> if there to each $U\in\mathcal{U}$ is a measurable function f_U from $\mathcal{E}$ to $(\mathcal{Y},\mathcal{B})$ so that $P_\theta^*(f_\theta \neq f_U) = 0$ when $\theta\in U$. $\mathcal{U}$ coherent families will be denoted as $f = (f_\theta:\theta\in\Theta)$, $g = (g_\theta:\theta\in\Theta),\ldots$.

Call a sub set U of Θ <u>dominated</u> if $(P_\theta:\theta\in U)$ is dominated and let $\mathcal{U}_f$, $\mathcal{U}_c$ and $\mathcal{U}_d$ be the classes of sub sets of Θ which are, respectively, finite, countable and dominated. We shall then say that f is <u>finitely coherent</u> or <u>countably coherent</u> or <u>dominatedly coherent</u> or <u>coherent</u> if f is, respectively, $\mathcal{U}_f$ coherent or $\mathcal{U}_c$ coherent or $\mathcal{U}_d$ coherent or $\{\Theta\}$ coherent.

Notions of coherence for variables taking their values in $[0,1]$ were introduced in Hasegawa and Perlman [4].

A measurable function s from $\mathcal{E}$ to $(\mathcal{Y},\mathcal{B})$ may be identified with the $\mathcal{U}$ coherent family $(s:\theta\in\Theta)$.

$\mathcal{U}$ coherent families f and g are called equivalent if $P_\theta^*(f_\theta \neq g_\theta) = 0$; $\theta\in\Theta$. [If $(\mathcal{Y},\mathcal{B})$ is Euclidean then $[f_\theta \neq g_\theta]$ is measurable and we may write P_θ instead of P_θ^*] . This notion of equivalence is clearly a proper equivalence relation.

It is not difficult to show that the notions of countable coherence and dominated coherence coinsides and that these notions coinsides with the notion of finite coherence when $(\mathcal{Y}, \mathcal{B})$ is Euclidean.

The experiment $\mathcal{E}$ will be called coherent if any finitely coherent family of real valued variables is coherent. Now any abstract M-space with unit may be represented as the class of continuous functions on some compact space. (See Kelley [6]). It follows that any experiment is, in the sense of LeCam [5], equivalent to a coherent one. If $\mathcal{E}$ is coherent then any finitely coherent family $f = (f_\theta : \theta \in \Theta)$ from $\mathcal{E}$ to a Euclidean space $(\mathcal{Y}, \mathcal{B})$ is coherent. $\mathcal{E} = (\chi, \mathcal{A} : P_\theta : \theta \in \Theta)$ is coherent if $(P_\theta : \theta \in \Theta)$ is dominated or if $\mathcal{A}$ consists of all sub sets of χ and each P_θ has countable support.

Let us in passing mention that LeCam's randomization criterion (Theorem 3 in [5]) may be expressed in terms of finitely coherent families as follows: Consider experiments $\mathcal{E} = (\chi, \mathcal{A}; P_\theta : \theta \in \Theta)$ and $\mathcal{F} = (\mathcal{Y}, \mathcal{B}; Q_\theta : \theta \in \Theta)$ and a non negative function ε on Θ. Then $\mathcal{E}$ is ε-deficient w.r.t. $\mathcal{F}$ in the sense of LeCam [5] if and only if there are finitely coherent families $M(B)$; $B \in \mathcal{B}$ from $\mathcal{E}$ to $[0,1]$ such that $2|Q_\theta(B) - \int M_\theta(B) dP_\theta| \leqq \varepsilon_\theta ; B \in \mathcal{B} : \theta \in \Theta$ and such that $M(\emptyset) = 0$, $M(\chi) = I$ and $M_\theta(\sum_{i=1}^\infty B_i) = \sum_{i=1}^\infty M_\theta(B_i)$ a.s. P_θ for all θ whenever the sets $B_1, B_2, \ldots$ are disjoint.

Operations on coherent families may be carried out as follows: Let for each t in a countable set T, f_t be a $\mathcal{U}$-coherent family from $\mathcal{E}$ to $(\mathcal{Y}_t, \mathcal{B}_t)$. Suppose also that ϕ is a measurable function from $\prod_t (\mathcal{Y}_t, \mathcal{B}_t)$ to $(\mathcal{Y}, \mathcal{B})$. Then $\phi(f_{t,\theta} : t \in T)$; $\theta \in \Theta$ is again $\mathcal{U}$-coherent and may be denoted by $\phi(f_t : t \in T)$ or by similar suggestive notations.

It is easily seen that such operations respect equivalence, i.e. that $\phi(f_t:t\in T)$ and $\phi(\tilde{f}_t:t\in T)$ are equivalent when f_t and $\tilde{f}_t$ are equivalent for each t. Here ϕ may be replaced by a U-coherent family of $(\mathcal{Y},\mathcal{B})$ measurable functions on the experiment with sample space $\prod_t(\mathcal{Y}_t,\mathcal{B}_t)$ which is induced by $(f_t:t\in T)$.

Leaving these generalities we shall in the sequel restrict attention to the set V of equivalence classes of finitely coherent families of real valued random variables.

Let $v,w\in V$ and let $\alpha,\beta\in R$. Then we may define elements $\alpha v+\beta w, vw$ and αv in V by:

$$\alpha v+\beta w = (\alpha v_\theta+\beta w_\theta:\theta\in\Theta)$$
$$vw = (v_\theta w_\theta:\theta\in\Theta)$$
$$\alpha v = (\alpha v_\theta:\theta\in\Theta).$$

Define also a relation $\geqq$ on V by defining $v\geqq w$ to mean $P_\theta(v_\theta\geqq w_\theta) = 1;\theta\in\Theta$.

It may then be checked that these operations are well defined and that V becomes an order complete vectorlattice and an algebra over the reals with unit 1 being the equivalence class of $(1;\theta\in\Theta)$. Also if $v_1,v_2,\ldots$ are in V and if ϕ is a measurable function from $R\times R\times\ldots$ to R then $\phi(v_1,v_2,\ldots)$ is a well defined element of V.

Example 2. Suppose $P_{\theta_1}\wedge P_{\theta_2}=0$ when $\theta_1\neq\theta_2$. Then any family $(v_\theta:\theta\in\Theta)$ of measurable functions is finitely coherent. Thus if g is a real valued function on Θ then $(g(\theta):\theta\in\Theta)$, considered as a finitely coherent family of constant functions, defines an unbiased estimator of g with variance zero.

Integrals of elements in V w.r.t. measures μ in L may be defined as follows:

Let $v \in V$ and suppose $\mu \in L$. Then there is a countable sub set Θ_o of Θ so that μ is in the band generated by $P_\theta : \theta \in \Theta_o$. Let $(v_\theta : \theta \in \Theta)$ be a representation of v and let v_{Θ_o} be such that $v_\theta = v_{\Theta_o}$ a.s. P_θ when $\theta \in \Theta_o$. Then we put $\mu(v) = \int v d\mu = \int v_{\Theta_o} d\mu$ provided the last integral exists. If μ is a probability measure then we may write $E_\mu v$ instead of $\int v d\mu$. It is easily seen that neither the existence nor the value of this integral depends on how Θ_o, $(v_\theta : \theta \in \Theta)$ and v_{Θ_o} was chosen. We shall say that v is μ integrable if $\mu(|v|) < \infty$.

If $v_1, v_2, \ldots \in V$ and $\mu \in L$ then the joint distribution $\mu(v_1, v_2, \ldots)^{-1}$ of $(v_1, v_2, \ldots)$ under μ may be defined similarily and without ambiguity.

A few sub spaces of V of particular interest are:

$$M_p = \{v : v \in V \text{ and } \int |v|^p dP_\theta < \infty \text{ for all } \theta\};\ 1 \leqq p \leqq \infty$$

$$M = \{v : v \in V \text{ and } \sup_\theta [P_\theta \text{ essential } \sup |v_\theta|] < \infty\}$$

Clearly $M \subseteqq M_\infty \subseteqq M_p \subseteqq M_q \subseteqq M_1$ when $\infty \geqq p \geqq q \geqq 1$.

If $v \in M_p$ then the L_p norm of v_θ w.r.t. P_θ will be denoted by $\|v\|_{\theta,p}$. The uniformity on M_p generated by the norms $\| \ \|_{\theta,p}$; $\theta \in \Theta$ will be denoted by d_p. It is easily checked that M_p with the topology of dp is a locally convex and complete linear space having the space of measurable functions as a dense sub set.

If $v \in M$ then $v \in M_\infty$ and we may put $\|v\| = \sup_\theta \|v\|_{\theta,\infty}$. The space M equipped with this norm is an abstract M-space and is actually the M-space of the experiment. The space M_p, $1 \leqq p \leqq \infty$ is the completion of M for the uniformity d_p.

Closure of a set $U \subseteqq M_p$ w.r.t. dp will be denoted by $\overline{U}^p$. If $p=2$ then we may write $\overline{U}$ instead of $\overline{U}^2$.

If D is a sub set of V then we put $D_+ = \{v: v \in D, v \geqq 0\}$. A family $(v^\theta : \theta \in \Theta)$ of elements in V will be called D estimable if $E_\theta w \underset{\theta}{\equiv} E_\theta v^\theta$ for some $w \in D$. In particular a real valued function g on Θ is D estimable if it is of the form $g(\theta) \underset{\theta}{\equiv} E_\theta w$ where $w \in D$. If g is V estimable (M_2 estimable) then we may say that g is estimable (quadratically estimable).

3. COMPLETENESS SUFFICIENCY AND UMVU ESTIMATORS

An element $v \in V$ will be called an unbiased estimator of a real valued function g on Θ if $E_\theta v \underset{\theta}{\equiv} g(\theta)$. A very important role is played by the unbiased estimators of zero. The set of all unbiased estimators of zero is denoted by N i.e.

$$N = \{v:\ v \in V \ \text{ and } \ E_\theta v \underset{\theta}{\equiv} 0\}.$$

An estimator of a real valued parameter g is here called uniformly minimum variance unbiased (UMVU) if it is unbiased and if the variance is everywhere finite and everywhere at most equal to the variance of any other unbiased estimator. Denoting the set of all UMVU estimators by T we have:

$$T = \{t:\ t \in M_2 \ \text{ and } \ \mathrm{Var}_\theta t \leqq \mathrm{Var}_\theta v \text{ for all } \theta$$
$$\text{whenever } \ v \in M_2 \ \text{ and } \ E_\theta v \underset{\theta}{\equiv} E_\theta t\}.$$

A parameter having a UMVU estimator will be called UMVU estimable.

Lehmann and Scheffe's fundamental result on UMVU estimators, [6], carries over to this framework without difficulties. Thus:

Theorem 3. $T = \{t : t \in M_2 \text{ and } t\circ(N \cap M_2) \subseteq N\}$.

Proceeding as in Bahadur [1] we get:

Corollary 4. T is linear and d_2 complete, and any $t \in T$ is uniquely determined by its expectation: $\theta \rightsquigarrow E_\theta t$.

Remark: The uniqueness property corresponds to (quadratically) completeness in usual statistical terminology.

Corollary 5. $(T \cap M_\infty)\cdot T \subseteq T$ so that $T \cap M$ and $T \cap M_\infty$ are both algebras containing the constants.

Remark: If V_1 and V_2 are sub sets of V then $V_1 \circ V_2$ denotes the set $\{v_1 \circ v_2 : v_1 \in V_1, v_2 \in V_2\}$.

Of particular interest are the sub algebras of V generated by sub σ-algebras of $\mathcal{A}$. If $\mathfrak{B}$ is a sub σ-algebra of $\mathcal{A}$ then the space of bounded $\mathfrak{B}$ measurable functions will be denoted by $\mathcal{M}(\mathfrak{B})$. Permitting ourselves some abuse of notations we shall also write $\mathcal{M}(\mathfrak{B})$ for the algebra of equivalenceclasses in V determined by functions in $\mathcal{M}(\mathfrak{B})$.

Before proceeding let us make a few remarks on sub algebras of M. If, in general, W is a sub space of M containing the constants and which is either a vector lattice or an algebra then its closure for the $w(M,L)$ topology (which is also the closure for the Mackey topology for the pairing (M,L)) is both a vector lattice and an algebra. Furthermore if $w_1, w_2, \ldots \in W$ and if ϕ is a measurable function from $R \times R \ldots$ to R which is bounded on $[-\|w_1\|, \|w_1\|] \times [-\|w_2\|, \|w_2\|] \times \ldots$ then $\phi(w_1, w_2, \ldots)$ is in the $w(M,L)$ closure of W. Consider so elements $w_1, w_2, \ldots$ in the closure $\bar{W}$ of W for d_2, and a measurable function ϕ from $R \times R \times \ldots$ to R such that $\phi(w_1, w_2, \ldots) \in M_2$. Then, by standard arguments, $\phi(w_1, w_2, \ldots) \in \bar{W}$. [This is clear if $\phi : R^k \to R$ is uniformly continuous and bounded. Thus, by approximation, indicators of closed sets in R^k, and hence all measurable indicator functions in R^∞, have the same property. By approximation by step functions this extend to all bounded measurable ϕ's and finally to all ϕ's such that $\phi(w_1, w_2, \ldots) \in M_2$.]

When applying these considerations to $T \cap M$ we may utilize:

Corollary 6. $T \cap M$ is $w(M,L)$ closed.

Proof: Suppose the net t_n in $T \cap M$ converges to t in M for the $w(M,L)$ topology. Let $z \in N \cap M_2$. Then $E_\theta z t_n = 0$ so that $E_\theta z t = \lim_n E_\theta z t_n = 0$. Hence, by Theorem 3, $t \in T$. □

Thus $T \cap M$ is a vector lattice. The fact that also $T \cap M_\infty$ is a vector lattice will follow from Proposition 7 below.

Two interesting sub spaces of T are:

$$T_h = \{t : t \in M_2 \text{ and } \phi(t) \in T \text{ whenever } \phi : R \to R \text{ is such that } \phi(t) \in M_2\}$$

and

$$T_a = \{t : t \in M_2 \text{ and } t \text{ is independent of every ancillary event}\}.$$

The analogous classes within the usual framework were treated by Bondesson in [3] and by Bahadur in [2] . Following Bahadur we shall call estimators in T_h, hereditary. This terminology is justified by the definition as well by the following result which is adapted from Bahadur [2].

Proposition 7. If $t_1, t_2, \ldots \in T_h$ and ϕ is a measurable function from $R\times R\times\ldots$ to R such that $\phi(t_1,t_2,\ldots)\in M_2$ then $\phi(t_1,t_2,\ldots)\in T_h$. Furthermore:

$$\overline{T\cap M} = \overline{T\cap M_\infty} = T_h \subseteq T_a \subseteq T = \bar{T}.$$

Remark: See Bahadur [2] for examples where $T\cap M \neq T_h$, $T_h \neq T_a$ and $T_a \neq T$.

Proof: By Corollary 4, $T = \bar{\bar{T}}$ and it is clear that $T\cap M \subseteq T\cap M_\infty$ and that $T_a \subseteq T$. Also $T_h \subseteq T$ since $\phi(t)\in M_2$ when $\phi(x) \underset{x}{\equiv} x$ and $T \in T_h$. By the arguments preceding Corollary 6, $\phi(t_1,t_2,\ldots)\in \overline{T\cap M}\subseteq T$ when $t_1,t_2,\ldots \in \overline{T\cap M}$ and $\phi(t_1,t_2,\ldots)\in M_2$. It follows in particular that $\overline{T\cap M}\subseteq T_h$. Consider so a t in T_h. Then $t^+ \cap m \in T\cap M$, $m=1,2,\ldots$. Thus, by monotone convergence, $t^+ \in \overline{T\cap M}$. Similarily $t^- \in \overline{T\cap M}$ so that $t\in \overline{T\cap M}$. Hence $\overline{T\cap M} = T_h$.

Let $0 \leq t \in T\cap M_\infty$. Let $z\in N\cap M_2$, $m\in\{1,2,\ldots\}$ and $\theta\in\Theta$. Then there is a sequence p_n, $n=1,2,\ldots$ of polynomials such that $p_n(x) \to x\cap m$ uniformly on $[-\|t\|_{\theta,\infty}, \|t\|_{\theta,\infty}]$.
Hence $E_\theta(t\wedge m)z = \lim_n E_\theta p_n(t)z = 0$ since $p_n(t)\in T$. Thus $t\cap m\in T$. Letting $m\to\infty$ we find that $t\in\overline{T\cap M} = T_h$ so that $T\cap M_\infty \subseteq \overline{T\cap M}$. Hence $T\cap M_\infty$ is a vector lattice and $\overline{T\cap M_\infty} = \overline{T\cap M} = T_h$.

It remains to show that $T_h \subseteq T_a$. Let $t \in T_h$ and let $A \in \mathcal{A}$ be such that $c = P_\theta(A)$ does not depend on θ. Let ψ be any bounded measurable function on R. Then $\psi(t) \in T$ and $c - I_A \in N \cap M_2$. Hence $E_\theta \psi(t) I_A = cE_\theta \psi(t) = (E_\theta I_A) E_\theta \psi(t)$.

□

Here is the reformulation of proposition 6 in Bahadur [1] described in the introduction.

<u>Proposition 8</u>. Let c be a finite non negative measure on Θ with minimal countable support Θ_o. Put $\mu = \sum_\theta C(\theta) P_\theta$ and let $0 \leq v \in M_2$ be such that v_{Θ_o} is in the Hilbert sub space of $L_2(\mu)$ generated by $dP_\theta/d\mu$;$\theta \in \Theta_o$. Suppose $\theta \rightsquigarrow E_\theta v$ is UMVU estimable by $t \in T$. Then $t_\theta = v_\theta$ a.s. P_θ;$\theta \in \Theta_o$.

If v is minimal in the sense that $v' \geq v$ whenever $v' \in V_+$ is such that $v'_\theta = v_\theta$ a.s. P_θ when $\theta \in \Theta_o$ then

$$t = v \text{ so that } v \in T.$$

<u>Remark 1.</u> Using the notations introduced in section 2, v_{Θ_o} is up to a set of μ measure zero determined by the property that $v_\theta = v_{\Theta_o}$ a.s. P_θ when $\theta \in \Theta_o$.

<u>Remark 2.</u> If $w \in V_+$ and Θ_o is countable then there is always a smallest $v \in V_+$ such that $v_\theta = w_\theta$ a.s. P_θ when $\theta \in \Theta_o$. v may be obtained by considering the whole set V' of elements v' in V_+ such that $v'_\theta = w_\theta$ a.s. P_θ when $\theta \in \Theta_o$. Let, for each $\theta \in \Theta$, v_θ be the P_θ essential infimum of v'_θ as v' runs through V'. Then $(v_\theta : \theta \in \Theta)$ determines the element v in V.

Proof of the proposition: $t_{\Theta_o} \in L_2(\mu)$ since $\infty > \int v^2 d\mu \geqq \int t^2 d\mu$. Now $\int v^2 d\mu \geqq \int t^2 d\mu = \int (t-v)^2 d\mu + \int v^2 d\mu + 2\int v(t-v)d\mu$ so that $0 \geqq \int (t-v)^2 d\mu + 2\int v(t-v)d\mu$. The assumption of unbiasedness of t implies that $E_\theta(t_{\Theta_o} - v_{\Theta_o}) = 0\ ;\theta\in\Theta_o$. Thus $t_{\Theta_o} - v_{\Theta_o} \perp dP_\theta/d\mu$ when $\theta\in\Theta_o$. By assumption $v_{\Theta_o} \in [dP_\theta/d\mu;\theta\in\Theta_o]$. Hence $\int v(t-v)\, d\mu = \int v_{\Theta_o}(t_{\Theta_o} - v_{\Theta_o})d\mu = 0$ so that $\int (t-v)^2 d\mu \leqq 0$. Thus $t_{\Theta_o} = v_{\Theta_o}$ a.s. μ so that $t_\theta = t_{\Theta_o} = v_\theta$ a.s. P_θ when $\theta\in\Theta_o$.

Suppose next that v is minimal. By what we have already shown: $|t_\theta| = v_\theta$ a.s. $P_\theta\ ;\theta\in\Theta_o$. By minimality: $v \leqq |t|$ so that $E_\theta t^2 \leqq E_\theta v^2 \leqq E_\theta |t|^2 = E_\theta t^2$. Hence $E_\theta t^2 \underset{\theta}{\equiv} E_\theta v^2$ so that v itself is a UMVU estimator.

□

The final step is achieved by noting that by arguments in LeCam [4] the minimal sufficient algebra is generated by the minimal non negative Radon-Nikodym derivatives $dP_{\theta_1}/d(P_{\theta_1}+P_{\theta_2})\ ;\theta_1,\theta_2\in\Theta$. Before we, for the sake of completeness, explain this let us shortly review the concept of a sufficient algebra and describe the associated conditional expectations.

A weakly closed (i.e. w(M,L) closed) sub algebra of M containing the constants defines an experiment $\mathcal{E}|W = (P_\theta|W;\ \theta\in\Theta)$ in the sense of LeCam [5] with $L|W$ as L-space and W as M-space. This experiment is always at most as informative as $\mathcal{E}$ itself. If $\mathcal{E}|W$ is equally informative as $\mathcal{E}$ (i.e. the deficiency of $\mathcal{E}|W$ w.r.t. $\mathcal{E}$ is zero) then W is called <u>sufficient</u>. It was shown in LeCam's paper that W is sufficient if and only if there is a , necessarily unique, non negative linear projection Π of M onto W such that $\Pi(1) = 1$ and $P_\theta(\Pi(v)) \underset{\theta}{\equiv} P_\theta(v)$ for any $v \in M$. This projection has the additional property that $\Pi(wv) = w\Pi(v)$ whenever $w\in W$ and $v\in M$.

Also, by propositions 11 and 13 in LeCam [5], ordered experiments are equivalent if and only if they are pairwise equivalent.

Consider now a sufficient algebra W and the associated conditional expectation (projection) Π. Let $v \in M_1$. Then there is a sequence $v_1, v_2, \ldots$ in M converging to v in the d_1 uniformity. It is easily seen that $\Pi(v_n)$, $n=1,2,\ldots$ converges for the same uniformity to an element which does not depend on how the sequence in M converging to v was chosen. Denote this element by $\Pi(v)$. Then Π extended this way defines a non negative projection (conditional expectation) of M_1 onto the closure $\bar{W}^1$ of W for the d_1 uniformity such that:

$$P_\theta(\Pi(v)) \underset{\theta}{\equiv} P_\theta(v) \quad ; \quad v \in M_1$$

and

$$\Pi(wv) = w\Pi(v) \quad ; \quad w \in \bar{W}^1, \; v \in M_1$$

provided $wv \in M_1$.

Furthermore the restriction of Π to M_2 is a projection onto $\bar{W}$ such that $\Pi(wv) = w\Pi(v) \in \bar{W}$ whenever $v \in M_2$ and $w \in \bar{W}$.

The final spadework is contained in the following result which is derived from arguments in LeCam [5].

<u>Proposition 9</u>. Let C be a non negative measure on Θ with countable support and put $\mu = \sum_\theta C(\theta)P_\theta$. Suppose $C(\theta_0) > 0$ and that v is a minimal non negative version of $dP_{\theta_0}/d\mu$. Then v is contained in any sufficient algebra.

<u>Proof:</u> Let W be any sufficient algebra and let Π be the associated conditional expectation. The inequality $\mu \geqq C(\theta_0)P_{\theta_0}$ imply that $0 \leqq v \leqq C(\theta_0)^{-1}$. Hence $v \in M$ and $0 \leqq \Pi(v) \leqq C(\theta_0)^{-1}$.

Let $s \in M_+$. Then $P_{\theta_0}(s) = P_{\theta_0}(\Pi(s)) = \mu(\Pi(s)v) = \sum_\theta C(\theta)P_\theta(\Pi(s)v) =$
$= \sum_\theta C(\theta)P_\theta(\Pi(s)\Pi(v)) = \sum_\theta C(\theta)P_\theta(s\Pi(v)) = \mu(s\,\Pi(v))$. Hence $\Pi(v)$ is a version of $dP_{\theta_0}/d\mu$. By minimality $v \leqq \Pi(v)$. Hence, for any θ : $0 \leqq P_\theta(\Pi(v)-v) = P_\theta(\Pi(v))-P_\theta(v) = 0$. It follows that $v = \Pi(v) \in W$.

□

As a corollary we get the following characterization of the minimal sufficient algebra, which might also have been derived directly from the proofs in LeCam [5].

Corollary 10. Let for each $(\theta_1,\theta_2) \in \Theta \times \Theta$, u_{θ_1,θ_2} be a minimal non negative version of $dP_{\theta_2}/d(P_{\theta_1}+P_{\theta_2})$. Let H be the smallest $w(M,L)$ closed subalgebra of M which contains all functions u_{θ_1,θ_2} ; $\theta_1,\theta_2 \in \Theta$ and the constants. Then H is sufficient and is contained in any other sufficient algebra.

Remark. Clearly $0 \leqq u_{\theta_1,\theta_2} \leqq 1$ so that $u_{\theta_1,\theta_2} \in M$ and thus H is well defined.

Proof: Let α_1,α_2 be positive numbers and put $\lambda = \alpha_1 P_{\theta_1} - \alpha_2 P_{\theta_2}$. Then $||\lambda|| = (P_{\theta_1}+P_{\theta_2})(|\alpha_1(I-u_{\theta_1,\theta_2})-\alpha_2 u_{\theta_1,\theta_2}|)$ so that H is pairwise sufficient and hence, by proposition 11 in LeCam [5] ; H is sufficient. If W is another sufficient algebra then, by Proposition 9, $W \supseteq H$.

□

The minimal sufficient algebra will in the following be denoted by H and the associated conditional expectation (projection) by Π.

Our first, and not so surprising, result linking UMVU theory and sufficiency is:

Proposition 11. (The necessity of $T \cap M$.)

$$T \subseteq \bar{H} \quad \text{and} \quad T \cap M \subseteq H .$$

Proof: The last inclusion follows, since $\bar{H} \cap M = H$, from the first. Let $t \in T$. Then

$$E_\theta t^2 = E_\theta(t-\Pi(t))^2 + 2E_\theta\Pi(t)(t-\Pi(t)) + E_\theta\Pi(t)^2$$
$$= E_\theta(t-\Pi(t))^2 + E_\theta\Pi(t)^2 \geq E_\theta\Pi(t)^2.$$ Hence, since $\Pi(t) \in M_2$; $E_\theta t^2 \underset{\theta}{\equiv} E_\theta\Pi(t)^2$ so that $t = \Pi(t) \in \bar{H}$. □

Now all the pieces are here and putting them together we get the main result of this paper:

Theorem 12. Consider only non negative minimal Radon-Nikodym derivatives. Say that $v \in M_2$ is UMVU estimable if its expectation is. Then the following conditions are all equivalent:

(i) $dP_{\theta_1}/d(P_{\theta_1}+P_{\theta_2})$ is UMVU estimable ; $\theta_1,\theta_2 \in \Theta$.

(ii) Each $v \in M$ is UMVU estimable

(iii) Each $v \in M_2$ is UMVU estimable

(iv) $T \cap M = H$

(v) $T = \bar{H}$

(vi) $\bar{H}$ is complete (i.e. $P_\theta(h) \underset{\theta}{\equiv} 0$ and $h \in \bar{H} \Rightarrow h = 0$)

(vii) M has a sufficient subalgebra W such that $\bar{W}$ is complete.

If one, and hence all, of these conditions are satisfied then:

$$T_h = T_a = T = \bar{H} .$$

Proof: As (iii) $\Rightarrow$ (ii) $\Rightarrow$ (i) is trivial it suffices to show that (i) $\Rightarrow$ (iv) $\Rightarrow$ (v) $\Rightarrow$ (vi) $\Leftrightarrow$ (vii) $\Rightarrow$ (iii).

(i) $\Rightarrow$ (iv) : Follows from Proposition 11, Corollary 10 and Proposition 9.

(iv) $\Rightarrow$ (v) : Follows from Proposition 11 and Corollary 4.

(v) $\Rightarrow$ (vi) : Follows from Corollary 4.

(vi) $\Leftrightarrow$ (vii): Follows from the fact that if W is a complete and sufficient algebra then $W=H$. [If $w \in W$ then $\Pi(w) \in H$ and $E_\theta w \underset{\theta}{\equiv} E_\theta \Pi(w)$ so that $w = \Pi(w) \in H$].

(vi) $\Rightarrow$ (iii): Suppose (vi) holds and let $v \in M_2$. Put $t = \Pi(v)$. Then $t \in \bar{H}$ and $E_\theta t \underset{\theta}{\equiv} E_\theta v$. Let $z \in N \cap M_2$. Then $\Pi(z) \in \bar{H} \cap N$. Hence, by completeness, $\Pi(z) = 0$ so that $E_\theta tz = E_\theta \Pi(tz) = E_\theta t\Pi(z) = 0$. Hence $t \in T$.

If these conditions are satisfied then, by (iv), $T = \overline{T \cap M} = T_h$ and the final statement follows from Proposition 7.

□

Let us return to the "traditional" framework. If G is any sub set of V then we shall denote by $\check{G}$ the set of equivalence classes in G which are determined by measurable functions, i.e. $\check{G} = \{g : g \in G$ and g is coherent$\}$. We thus have the sets $\check{M}_2$, $\check{N}$ and $\check{M}_2 \wedge \check{N}$. The set $\tilde{T}$ of "usual" UMVU estimators is, in general, larger than $\check{T}$. By the Lehmann-Scheffé theorem again: $\tilde{T} = \{t : t \in \check{M}_2$ and $t \circ (\check{M}_2 \cap \check{N}) \subseteq \check{N}\}$. Let also $\mathcal{B}$ be the σ-algebra induced by $\tilde{T} \cap M$ and put $\tilde{T}_h = \{t : t \in \check{M}_2$ and $\phi(t) \in \tilde{T}$ whenever

$\phi : R \to R$ is such that $\phi(t) \in \check{M}_2\}$.

Then, as shown by Bahadur [1], $\mathcal{B}$ is complete w.r.t. $\mathcal{A}$ and is contained in the completion (w.r.t. $\mathcal{A}$) of any sufficient σ-algebra. Furthermore, by Bahadur [2], $\check{T}_h$ is precisely the class of everywhere quadratically integrable and $\mathcal{B}$ measurable functions.

Denote by $\tilde{T}_a$ the set of "traditional" UMVU estimators which are independent of ancillary events.

<u>Example 13</u>. Take $[0,1]^2$ with the Borel class $\mathcal{A}$ as sample space. Put $\Theta = -1 \cup [0,1]$ and let P_θ ; $0 \leq \theta \leq 1$, be the uniform distribution on $\{(\theta,y) : 0 \leq y \leq 1\}$. Let P_{-1} have density $(x,y) \rightsquigarrow 2y$ w.r.t. the uniform distribution on $[0,1]^2$. This example satisfies the conditions of Theorem 12.

Consider now this example within the traditional framework. Let ρ be any differentiable function on R^2 such that $\rho(x,0) = \rho(x,1) = 0$, $0 \leq x \leq 1$ and $\int_0^1 \int_0^1 \rho(x,y)dxdy = 0$. Put $\delta(x,y) = \frac{\partial}{\partial y}\rho(x,y)$. Then δ is an unbiased estimator of zero. It follows that, for any UMVU estimator ϕ:

$$\int_0^1 \int_0^1 \phi(x,y)\left(\frac{\partial}{\partial y}\rho(x,y)\right)2y\,dxdy = 0 .$$

Furthermore it is easily seen that a UMVU estimator ϕ is essentially a function of x alone. Hence the last equation may be written $\iint \phi(x,y)\rho(x,y)dxdy = 0$. By varying ρ we see that ϕ = constant a.e. P_θ for all θ. Thus only the constants have UMVU estimators in the traditional set up.

The final theorem of this paper is - except for the last statement - essentially reformulations of results in Bahadur [1] and [2].

Theorem 14. The following five conditions are equivalent:

(i) Each $\check{M}$ estimable parameter is $\tilde{T} \cap \check{M}$ estimable.

(ii) Each $(\check{M}_2)_+$ estimable parameter is $\tilde{T}_+$ estimable.

(iii) Each $\check{M}_2$ estimable parameter is $\tilde{T}$ estimable and $\tilde{T}_h = \tilde{T}_a = \tilde{T}$.

(iv) $\mathcal{B}$ is sufficient (and hence minimal sufficient).

(v) There is a sufficient and quadratically complete sub σ-algebra of $\mathcal{A}$.

These conditions all imply the sixth condition:

(vi) Each $\check{M}_2$ estimable parameter is $\tilde{T}$ estimable.

If $\mathcal{E}$ is coherent then all six conditions are equivalent.

Remark. (iii) $\Longleftrightarrow$ (iv) $\Longleftrightarrow$ (v) is proved in Bahadur [2] while (i) and (ii) are merely rephrasings of these conditions. (v) $\Rightarrow$ (vi) is a consequence of the Rao-Blackwell theorem and the implication (vi) $\Rightarrow$ (v) for dominated experiments was established in Bahadur [1]. In proving the last statement we utilized the fact, proved by Siebert in [8], that a w(M,L) closed (in $\mathcal{A}$) and pairwise sufficient σ-algebra is sufficient when $\mathcal{E}$ is coherent.

As a complete proof is not long we include one here.

Proof: (iii) $\Rightarrow$ (ii): Let $v \in \check{M}_+$. By (iii) there is a $t \in \tilde{T}$ so that $E_\theta v \underset{\theta}{\equiv} E_\theta t$. Then, since $\tilde{T}_h = \tilde{T}$, t is $\mathcal{B}$ measurable and

$E_\theta I_B t = E_\theta I_B v \geqq 0$ when $B \in \mathcal{B}$. Hence $t \geqq 0$.

(ii) $\Rightarrow$ (i): Let $v \in \check{M}$. Then $Iv + \|v\| \in (\check{M}_2)_+$. Hence, by (ii), there are $t_1, t_2 \in \tilde{T}_+$ such that $E_\theta t_1 \underset{\theta}{\equiv} E_\theta v + \|v\|$ and $E_\theta t_2 \equiv -t_\theta v + \|v\|$. Thus $E_\theta (t_1 - t_2)/2 \equiv E_\theta v$. By uniqueness $t_1 - \|v\| = \|v\| - t_2$ so that $0 \leqq t_1, t_2 \leqq \|v\|$. Hence $(t_1 - t_2)/2 \in \tilde{T} \cap \check{M}$.

(i) $\Rightarrow$ (iv): Let, for each $A \in \mathcal{A}$, $\hat{I}_A \in \tilde{T} \cap \check{M}$ be such that $E_\theta I_A \equiv E_\theta \hat{I}_A$. Then $E_\theta I_B \hat{I}_A = E_\theta I_B I_A$ when $B \in \mathcal{B}$ so that, since $\hat{I}_A$ is $\mathcal{B}$measurable, $P_\theta(A|\mathcal{B}) = \hat{I}_A$ a.s. P_θ for each θ. It follows that $\mathcal{B}$ is sufficient.

(iv) $\Rightarrow$ (v): Follows from the quadratic completeness of $\mathcal{B}$.

(v) $\Rightarrow$ (iii): Suppose v holds, that $\mathcal{S}$ is sufficient and quadratically complete and that $v \in \check{M}_2$. Then, by the Rao-Blackwell theorem, $t = E(v|\mathcal{S}) \in \tilde{T}$ and $E_\theta t \equiv E_\theta v$. By the same theorem $\tilde{T}$ is precisely the class of functions in $\check{M}_2$ which are $\mathcal{S}$ measurable. Hence $\tilde{T} = \tilde{T}_h$.

The implication (v) $\Rightarrow$ (vi) is trivial so suppose that $\mathcal{E}$ is coherent and that (vi) holds. By coherence, $\tilde{T} = T$ and $\check{M}_2 = M_2$ so that (iii) of Theorem 10 holds. Hence, by Theorem 12, $T \cap M = H$ so that H is generated by $\{I_B : B \in \mathcal{B}\}$. Thus $\mathcal{B}$ is pairwise sufficient and, since it is $w(M,L)$ closed in $\mathcal{A}$ and $\mathcal{E}$ is coherent, it is actually sufficient. [If $A \in \mathcal{A}$ then, by pairwise sufficiency, $(P_\theta(A|\mathcal{B}) : \theta \in \Theta)$ is finitely coherent. Hence, since $\mathcal{E}$ is coherent, there is a test function δ so that $P_\theta(A|\mathcal{B}) = \delta$ a.s. P_θ for each θ. Finally δ is, since $\mathcal{B}$ is $w(M,L)$ closed in $\mathcal{A}$, $\mathcal{B}$ measurable]. Thus (iv) holds.

□

REFERENCES

[1] Bahadur, R.R. (1957). On unbiased estimates of uniformly minimum variance. Sankhyá, **18**, 211-244.

[2] Bahadur, R.R. (1976). A note on UMV estimates and ancillary statistics. In memorial volume dedicated to J. Hajék. Charles University, Prague.

[3] Bondesson, L. (1975). Uniformly minimum variance estimation in location parameter families. Annals of Statistics, **3**, 637-660.

[4] Hasegawa, M. and Perlman, M.D. (1974). On the existence of a minimal sufficient subfield. Ann. Statist., **2**, 1049-1055. Correction: Ann. Statist., **3**, 1371-1372.

[5] LeCam. L. (1964). Sufficiency and approximate sufficiency. Ann. Math. Statist., **35**, 1419-1455.

[6] Lehmann, E.L. and Scheffé, H. (1950). Completeness, similar regions, and unbiased estimation. Sankhyá **10**, 305-340.

[7] Kelley, J.L. (1955). General topology. D. Van Nostrand.

[8] Siebert, E. (1976). Klasseneigenschaften Statistischer Experimente und ihre Charakterisierung durch Kegelmasse. Thesis. Eberhard-Karls-Universität zu Tübingen.