The Annals of Statistics
1988, Vol. 16, No. 2, 534–551

GENERALIZATIONS OF ANCILLARITY, COMPLETENESS AND SUFFICIENCY IN AN INFERENCE FUNCTION SPACE

BY CHRISTOPHER G. SMALL AND D. L. MCLEISH

University of Waterloo

In this paper we introduce E-ancillarity and complete E-sufficiency, natural extensions of the definitions of ancillarity and complete sufficiency to a space of estimating or inference functions. These are functions of both the data and the parameter. We begin either with a space of all such functions or with a subset defined to exploit special features of a model; for example, we allow restrictions to inference functions that are linear in the observations or linear in the parameter. Subsequently, a reduction analogous to complete sufficiency is carried out, and within the complete E-sufficient space of inference functions, one is chosen with properties that we deem desirable.

1. Introduction. This paper extends the concepts of ancillarity, sufficiency and completeness from statistics to *inference functions*, i.e., functions $\psi(\theta; X)$ of both a parameter θ and the data X. The extensions so obtained will be shown to be applicable in a wider context than the standard notions based upon distributional assumptions. The standard statistical concepts of ancillarity and sufficiency are defined in terms of the distributions or conditional distributions of statistics. However, they can also be motivated by properties of expectations. For example, suppose we observe X, a random vector in R^d, where the distribution P of X is restricted to a class $\mathscr{P}$ of probability measures. A statistic $T(X)$ is ancillary if its distribution does not depend on P. Alternatively, let $\mathscr{F}$ be a class of square-integrable functions of X. Let $\mathscr{A}$ be the class of all $g \in \mathscr{F}$, which are functions of X through $T(X)$. Then the ancillarity of T implies that $E_P g(X)$ is constant independent of P for all $g \in \mathscr{A}$. In fact, this condition is a generalization of ancillarity [cf. Lehmann (1981)] in the sense that an ancillary statistic generates such a class $\mathscr{A}$, whereas the converse is not true.

By Basu's theorem, if a statistic S is complete sufficient, it is independent of every ancillary statistic T. Again, this condition can be written in terms of the expectations over a class of functions; in particular, for every ancillary T and functions $g(T)$, $h(S)$, it follows that $\mathrm{Cov}_P[h(S), g(T)] = 0$.

Without loss of generality we can assume that $\mathscr{F}$ is a vector space of functions. It can be seen in this setting that the class of functions g in $\mathscr{F}$, which are functions of the ancillary T, is a linear subspace of the vector space $\mathscr{F}$. It also follows that the subspace of functions of the complete sufficient statistic S will be elements of the orthogonal complement of this linear vector space. This relationship is the prime motivation for a definition of ancillarity and sufficiency in this paper.

Received January 1987; revised June 1987.

AMS 1980 *subject classifications*. Primary 62A99; secondary 62B99.

Key words and phrases. Estimating function, score function, sufficiency, local sufficiency, ancillarity, completeness, nuisance parameter, Rao-Blackwell theorem.

Throughout this paper, we deal with a space of functions of both the data and a parameter θ, which may be used either as an estimating function to determine a point estimate or as the basis for constructing tests or confidence intervals for the parameter. Clearly, the score function $S(\theta; x) = \partial/\partial\theta \log f(x; \theta)$ has a special role to play in such an inference function space. The score function has several advantages not enjoyed by alternative generators of the minimal sufficient partition of the sample space. Under regularity, its expectation is zero and it is square integrable; in fact, its variance, the Fisher information, is related to the asymptotic variance of the maximum likelihood estimator more closely than is the variance of a finite-sample maximum likelihood estimator.

Similar attempts to provide a satisfactory extension of the notions of ancillarity and completeness in terms of expectations have been made before [cf. Lehmann (1981)]. These attempts have generally met with somewhat mixed success. Our theory differs in two notable ways from that of Lehmann; we apply the theory to linear vector spaces of estimating functions or, as we shall call them *inference functions*, functions $\psi(\theta; X)$ jointly of both the data and the parameter. The second apparently minor observation, but one that is critical to the theory, is to note that the generalization of ancillarity to be defined will require that the class of such inference functions be a closed linear space. The relevance of this will appear in the next section.

Other theories have been developed in the space of estimating functions [e.g., McLeish (1984)]. Godambe (1960) suggests an efficiency criterion $\mathrm{eff}(\psi; \theta)$ on a space of *unbiased* inference functions Ψ and selects one, ψ say, so as to maximize the efficiency. By unbiasedness of $\psi \in \Psi$ we mean that $E_\theta \psi(\theta; X) = 0$ for all θ. Whereas θ may not have an unbiased estimator, unbiased inference functions exist under fairly general circumstances. Note that provided there is sufficient regularity to ensure that derivatives and integrals can be exchanged, then the score function $S(\theta; X)$ will be unbiased in a one-parameter model. Similarly, the function $\psi(\theta; X) = f(X) - E_\theta f(X)$ will be unbiased. The efficiency $\mathrm{eff}(\psi; \theta)$ in the sense of Godambe (1960) provides a criterion for selection among such functions. If $\psi^*(\theta)$ maximizes $\mathrm{eff}(\psi; \theta)$ for every parameter value θ, then an estimator $\hat{\theta}^*$ could be chosen so that $\psi^*(\hat{\theta}^*) = 0$. Such an ordering of the inference function space has some obviously desirable features, especially if, as in Godambe's case, it circumvents the need to explicitly obtain the family of probability density functions. One might also hope that such a theory would circumvent the regularity requirements of maximum likelihood estimation and provide a satisfactory theory of inference under more general circumstances. Unfortunately, this theory has its own difficulties with regularity.

Godambe (1960) has suggested the efficiency criterion

$$\mathrm{eff}(\psi; \theta) = \frac{\{E_\theta\, \partial/\partial\theta\, \psi(\theta; X)\}^2}{E_\theta \psi^2(\theta; X)}$$

and has shown that within a wide class of unbiased inference functions, eff is maximized by inference functions of the form

$$\psi^*(\theta; X) = k(\theta) S(\theta; X),$$

where $k(\theta)$ is arbitrary. However, the following example illustrates that this efficiency criterion also has problems with nonregular examples.

Suppose X has distribution that is uniform on $[0, \theta]$. We wish to estimate the parameter θ. Note that the score function $S(\theta; x) = d/d\theta \log f(\theta; x)$ is defined for $x < \theta$ and equals $-1/\theta$ there. Thus there is no version of this derivative that is unbiased; the usual regularity conditions insuring that $E_\theta S(\theta; X) = 0$ fail, in this case. Consider the unbiased estimating functions defined for all $1 < n < \infty$ by

$$\psi_n(\theta; x) = \begin{cases} -C\left(1 - r\dfrac{x}{\theta}\right)^r, & x \leq \dfrac{\theta}{r}, \\ \left(\dfrac{x}{\theta} - \dfrac{1}{r}\right)^n, & x \geq \dfrac{\theta}{r}, \end{cases}$$

where $r = n^{1/2}$ and with C chosen so that the functions are unbiased. Then ψ_n are continuous functions of $x > 0$ such that the efficiencies $\text{eff}(\psi_n; \theta) \to \infty$ as $n \to \infty$. However, estimators corresponding to these estimating functions, obtained from the equations $\psi_n(\hat{\theta}; X) = 0$, satisfy $\hat{\theta} \to \infty$ as $n \to \infty$. A more reasonable estimator, such as the maximum likelihood estimator X, corresponds to the case $n = 1$.

It is quite natural that an attempt to provide an ordering on the space of inference functions is bound to fail to generate sensible estimators in some regular and nonregular examples. Godambe's efficiency criterion fails in the previous example because the regularity conditions that lead to unbiasedness of the score function for maximum likelihood estimation do not hold. There are various possible ways of measuring the sensitivity of an inference function $\psi(\theta; X)$ under changes in the underlying parameter. For example, in a one-parameter model we can measure the sensitivity through changes in the expectation, i.e., through the function $m_\psi(\eta, \theta) = E_\eta \psi(\theta; X)$. Specific functionals of m_ψ relevant to such a measure of sensitivity are finite differences such as $m_\psi(\eta, \theta)/(\eta - \theta)$ or the derivatives that these approximate $\partial/\partial\eta\, m_\psi(\eta, \theta)|_{\eta=\theta}$. More generally, we would expect that a reasonable measure of sensitivity around θ would be a linear functional of m_ψ such as

$$\int m_\psi(\eta, \theta)\, d\Lambda_\theta(\eta), \tag{1.1}$$

for some signed measure Λ_θ on the parameter space. Naturally, since inference functions define the same estimator when they are multiplied by a nonzero constant, (1.1) should be applied to a normalized version of ψ such as $\psi(\theta; X)/E_\theta^{1/2}[\psi^2(\theta; X)]$. With this substitution we obtain a class of information measures supplementing that of Godambe, viz.,

$$\mathcal{J}_\psi(\theta) = \frac{\left[\int E_\eta \psi(\theta; X)\, d\Lambda_\theta(\eta)\right]^2}{E_\theta \psi^2(\theta; X)}. \tag{1.2}$$

We will return to the problem of maximizing information measures of this form at the end of Section 2. A reasonable alternative to the use of information measures of the form (1.2) is a reduction of the data through sufficiency. It is our purpose to explore a theory on the estimating function space, which generalizes standard concepts of sufficiency.

2. Inference functions: Ancillarity and sufficiency. Let $\mathscr{X}$ be a sample space and $\mathscr{P}$ be a class of probability measures P on $\mathscr{X}$. For each $P \in \mathscr{P}$ we let V_P be the vector space of real-valued functions f defined on the sample space $\mathscr{X}$ such that $E_P[f(X)]^2 < \infty$. We introduce the usual inner product defined on V_P,

$$\langle f_1, f_2 \rangle_P = E_P\{f_1(X) f_2(X)\}.$$

Let θ be a real-valued function on the class of probability measures $\mathscr{P}$ and define the parameter space $\Theta = \{\theta(P); P \in \mathscr{P}\}$. Note that θ need not be a one-to-one functional. If it is, we call the model a *one-parameter model.*

The fundamental objects of our analysis will be *inference functions*, i.e., functions $\psi: \Theta \to \bigcup_P V_P$ such that $\psi(\theta(P)) \in V_P$ for all $P \in \mathscr{P}$. The function ψ is said to be *unbiased* if $\langle \psi(\theta(P)), 1 \rangle_P = 0$ for all $P \in \mathscr{P}$.

The particular value of P underlying the value of the parameter θ may be regarded as a nuisance parameter. The data are often such that we obtain little or no information on its value beyond the value of $\theta(P)$. Furthermore, the inference functions are allowed to depend on P only through $\theta(P)$ and it is therefore natural to require that the inner products on the space of inference functions have the same property. Two functions ψ and ϕ are said to have *constant covariant structure* if for all $P \in \mathscr{P}$ the inner product

$$\langle \psi(\theta(P)), \phi(\theta(P)) \rangle_P$$

is a function of P only through $\theta = \theta(P)$. In this circumstance we write $\langle \psi, \phi \rangle_\theta$ for this quantity as a function of θ. Let Ψ be a set of unbiased inference functions such that any pair of functions from Ψ have constant covariant structure. Clearly, Ψ can be made into a vector space by closing it under pointwise addition: $(\psi_1 + \psi_2)(\theta) = \psi_1(\theta) + \psi_2(\theta)$, and multiplication: $(k\psi)(\theta) = k(\theta)\psi(\theta)$, where we allow k to be a nonrandom function of $\theta \in \Theta$. Henceforth, by a "vector space" or "linear space" of inference functions, we shall mean a vector space in this general sense. So Ψ is endowed with a family of inner products $\langle \psi_1, \psi_2 \rangle_\theta$ for every $\theta \in \Theta$. We now introduce a topology on Ψ. The topology is most easily characterized by convergence; $\psi_n \to \psi$ if and only if $\langle \psi_n - \psi, \psi_n - \psi \rangle_\theta \to 0$ for all $\theta \in \Theta$. The topology determined by this notion of convergence, we call the *weak-square* topology. For the purposes of subsequent analysis, an additional closure condition is imposed upon Ψ. We require that if ψ_n is a sequence of inference functions in Ψ such that the double limit $\lim_{nm} \|\psi_n - \psi_m\|_\theta = 0$, then there exists a function $\psi \in \Psi$ such that $\lim_n \|\psi_n - \psi\|_\theta = 0$. We will loosely refer to Ψ as a Hilbert space of inference functions and mean that the coordinate projections $\{\psi(\theta); \psi \in \Psi\}$ form a Hilbert space in V_P. The conditions on Ψ can be summarized by stating that

> *henceforth the inference function space Ψ shall be assumed to be a Hilbert space of unbiased functions, any pair of which has constant covariant structure.*

It is worth commenting at this stage that the assumption of constant covariance ensures that orthogonality between inference functions is well defined for nuisance parameter models. In these models it also ensures the existence of a well-defined concept of convergence, the importance of which is indicated later in Example 2.5. In the absence of this assumption, Proposition 2.14 (to be stated later) will not hold and so there will be no guarantee that a canonical inferential reduction in Ψ will be possible.

An ancillary statistic is one whose distribution is insensitive to changes in the parameter. However, our windows on this distribution are the inference functions ψ viewed through their expectations and we expect that changes in the parameter are primarily evident through changes in these expectations. From this perspective, the following definition is a natural one.

DEFINITION 2.1. An unbiased inference function $\phi \in \Psi$ is said to be *E-ancillary* if ϕ can be written as the weak-square limit of functions ψ such that

$$E_Q \psi(\theta) = 0, \tag{2.1}$$

for all $Q \in \mathscr{P}$, $\theta \in \Theta$.

We let $\mathscr{A}$ be the set of E-ancillary functions in Ψ. By construction it is closed and it is easy to see that it is a linear subspace of Ψ. The necessity of closing the space we will discuss in Example 2.5.

EXAMPLE 2.2. Let $\mathscr{P}$ be a one-parameter model. Suppose T is an ancillary statistic in the usual sense that its distribution does not depend on the parameter $\theta \in \Theta$. Then for any values of the parameters θ, θ' and inference function $\psi \in \Psi$, if $\psi(\theta)$ is $\sigma(T)$-measurable for each θ,

$$E_{\theta'}[\psi(\theta)] = E_\theta[\psi(\theta)] = 0.$$

Thus the inference functions in Ψ that are functions of an ancillary statistic form an E-ancillary class in that they are in $\mathscr{A}$.

In a sense, the sufficient statistics contain the orthogonal complement to the set of ancillary statistics.

DEFINITION 2.3. Let $\mathscr{S}$ be a subset of Ψ. We say that $\mathscr{S}$ is an *E-sufficient* subset of inference functions if the condition that $\langle \psi, \phi \rangle_\theta = 0$ for all $\theta \in \Theta$ and for all $\psi \in \mathscr{S}$ implies that ϕ is an E-ancillary function.

THEOREM 2.4. *Let Ψ be the space of* all *unbiased inference functions in a one-parameter model. Suppose T is a sufficient statistic for θ. Let $\mathscr{S}$ be the space of all $\sigma(T)$-measurable inference functions in Ψ. Then $\mathscr{S}$ is E-sufficient.*

PROOF. Suppose ϕ is any element of ψ such that $\langle \phi, \psi \rangle_\theta = 0$ for all $\theta \in \Theta$ and for all $\psi \in \mathscr{S}$. It will be shown that $\phi \in \mathscr{A}$, and therefore that $\mathscr{S}$ is E-sufficient. For each $\theta \in \Theta$, let $\mathrm{supp}(\theta) = \mathrm{supp}(\theta(P))$ be the support of the distribution of P. Note that since T is sufficient, $\mathrm{supp}(\theta)$ is a $\sigma(T)$-measurable set. Denote by I_θ the indicator function of $\mathrm{supp}(\theta)$. Now, by assumption,

$\langle \psi, \phi \rangle_\theta = \langle \psi, \phi I_\theta \rangle_\theta = 0$ for all θ, $\psi \in \mathscr{S}$. Then setting $\psi(\theta) = E_\theta[\phi(\theta) I_\theta | T]$ we see that

$$E_\theta[\phi(\theta) | T] = 0 \quad \text{a.s. on supp}(\theta).$$

But by the definition of sufficiency,

$$0 = E_\theta[\phi(\theta) | T] = E_\eta[\phi(\theta) | T] \quad \text{a.s. on supp}(\theta) \cap \text{supp}(\eta).$$

So

$$\begin{aligned} E_\eta[\phi(\theta) I_\theta] &= E_\eta[\phi(\theta) I_\theta I_\eta] = E_\eta\{I_\eta E_\eta[\phi(\theta) I_\theta | T]\} \\ &= E_\eta\{I_\theta I_\eta E_\theta[\phi(\theta) | T]\} = 0, \end{aligned}$$

and, consequently, $\phi(\theta) I_\theta$ is an E-ancillary function. However, as $\|\phi - \phi I_\theta\|_\theta = 0$ and $\mathscr{A}$ is weak-square closed, it follows that $\phi \in \mathscr{A}$. □

The reader should notice that the proof given previously depends upon the fact that the conditional expectation of an unbiased inference function is unbiased, and therefore is an element of Ψ. When Ψ is a more restricted space, the subspace of $\sigma(T)$-measurable functions can even be empty and not E-sufficient. Fortunately, we shall see that E-sufficient subspaces still exist and remain appropriate classes for inference in the restricted setting.

Although it is natural to require that a linear subspace $\mathscr{A}$ of a Hilbert space be closed, we now also provide a simple example showing the necessity of closing this space in our definition of E-ancillarity.

EXAMPLE 2.5. Let X be a random variate with the location exponential probability density $f(x; \theta) = e^{-(x-\theta)}$, for $x > \theta$, $\theta \in \Theta = R$. Suppose we redefine $\mathscr{A}$ as the set of functions ψ satisfying (2.1). Then it is easy to see that a function $\psi \in \mathscr{A}$ if and only if $\psi(\theta; x) = 0$ almost everywhere $x \in R$. On the other hand, there is no set of inference functions $\mathscr{S}$, even $\mathscr{S} = \Psi$, such that $\langle \psi, \phi \rangle_\theta = 0$ for all $\psi \in \mathscr{S}$ implies that $\phi \in \mathscr{A}$, since an equation of this form can only control the values of $\phi(\theta; x)$ for $x > \theta$. Thus *the whole space of all unbiased inference functions is not E-sufficient.* This unsatisfactory result is a direct consequence of not closing the space $\mathscr{A}$ and disappears under Definition 2.1.

Basu's theorem [Basu (1958)] indicates that the space of functions of a complete sufficient statistic forms the orthogonal complement of the space of ancillary statistics. This motivates the following definition.

DEFINITION 2.6. A subset $\mathscr{S}$ of Ψ is said to be a *complete* E-sufficient subset if the statement $\phi \in \mathscr{A}$ is equivalent to the statement

$$\langle \psi, \phi \rangle_\theta = 0, \quad \text{for all } \theta \in \Theta,\ \psi \in \mathscr{S}. \tag{2.2}$$

If a complete E-sufficient subset exists, it must be unique and closed in the weak-square topology. Henceforth $\mathscr{S}$ shall denote the complete E-sufficient subset. It should also be noted that the subsets $\mathscr{A}$ and $\mathscr{S}$ are linear subspaces of Ψ.

THEOREM 2.7. *Let Ψ be a space of all unbiased inference functions in the one-parameter model. Let $\mathscr{S}$ be a space of all members of Ψ that are $\sigma(T)$-measurable, where T is a complete sufficient statistic. Then $\mathscr{S}$ is complete E-sufficient.*

PROOF. We use the notation of Theorem 2.4. By Theorem 2.4 $\mathscr{S}$ is E-sufficient. Suppose $E_\eta\phi(\theta; X) = 0$ for all η. Define a function $h(t)$ as follows:

$$h(t) = E_\eta[\phi(\theta; X)|T = t], \quad \text{for } (T = t) \subset \operatorname{supp}(\eta).$$

It must first be shown that h is well defined on $\bigcup_\eta\{t: (T = t) \subset \operatorname{supp}(\eta)\}$. Note that sufficiency implies that I_η is $\sigma(T)$-measurable for all η. So either $(T = t) \subset \operatorname{supp}(\eta)$ or $(T = t) \cap \operatorname{supp}(\eta)$ is empty. Suppose η and η' are two values such that $(T = t) \subset \operatorname{supp}(\eta) \cap \operatorname{supp}(\eta')$. Now sufficiency of T implies that $E_\eta[\phi(\theta)|T = t] = E_{\eta'}[\phi(\theta)|T = t]$. So h is well defined. Furthermore, $E_\eta[h(T)] = E_\eta[\phi(\theta)] = 0$ for all η. Completeness of T now implies that $h(T) = 0$, η-a.s. for all η. Let $\psi \in \mathscr{S}$ and $\theta \in \Theta$. Then $\langle\psi, \phi\rangle_\theta = E_\theta\{E_\theta[\psi(\theta)\phi(\theta)|T]\} = E_\theta[\psi(\theta)h(T)] = 0$. By extension to the closure of the set of such ϕ, we see that $\langle\psi, \phi\rangle_\theta = 0$ for all $\phi \in \mathscr{A}$ and for all θ. □

Theorem 2.7 establishes the existence of a complete E-sufficient subspace for one-parameter models possessing a complete sufficient statistic. However, it will be shown in the next section that complete E-sufficient spaces will exist in many cases even when the model does not have a complete sufficient statistic for the parameter.

One of the advantages in viewing the data exclusively through the expected value and covariance structure of a set of inference functions is the ease with which global definitions extend to local ones. Local definitions of first- and second-order ancillarity and sufficiency defined through the asymptotic distribution of the statistics are available in the literature [e.g., McCullagh (1984) and Cox (1980)].

DEFINITION 2.8. An unbiased inference function $\psi(\theta)$ is *locally* (*kth-order*) *E-ancillary* if it can be written as a weak-square limit of functions $\psi \in \Psi$ such that

$$E_P\psi(\theta) = o(\theta(P) - \theta)^k, \quad \text{as } \theta(P) \to \theta, \tag{2.3}$$

for all θ and all P such that $\theta(P) \to \theta$.

By analogy with the concept of E-sufficiency, we have

DEFINITION 2.9. A subset $\mathscr{L}$ is said to be *locally* (*kth-order*) *E-sufficient* if for any $\phi \in \Psi$ such that $\langle\psi, \phi\rangle_\theta = 0$ for all $\psi \in \mathscr{L}$ and $\theta \in \Theta$ it follows that ϕ is locally (kth-order) E-ancillary. The subset $\mathscr{L}$ is *complete locally* (*kth-order*) *E-sufficient* if this condition is necessary and sufficient.

Henceforth by a locally E-sufficient set, we will understand a first-order locally E-sufficient set. Consider the standard one-parameter model. Let L

denote the likelihood function and suppose

$$\lim_{\varepsilon \to 0} E_\theta \left[\frac{\Delta^j L(\theta)}{L(\theta)} - \frac{L^{(j)}(\theta)}{L(\theta)} \right]^2 = 0, \tag{2.4}$$

for all $1 \leq j \leq k$ and for all θ, where $\Delta h(\theta) = \varepsilon^{-1}[h(\theta + \varepsilon) - h(\theta)]$. Suppose Ψ is the space of all inference functions. Then if the following functions are square integrable (so that they lie inside Ψ), the kth-order locally E-sufficient subspace is spanned by the functions

$$\frac{L^{(j)}(\theta)}{L(\theta)}, \qquad j = 1, 2, \dots, k. \tag{2.5}$$

For example, the complete second-order E-sufficient space is spanned by the functions $S(\theta) = \partial/\partial\theta \log L(\theta)$ and $S^2(\theta) - I(\theta)$, where $I(\theta)$ is the observed information $-\partial/\partial\theta\, S(\theta)$.

We now discuss conditions under which complete E-sufficient subspaces exist, and under which the space Ψ is closed under projections conducted pointwise in θ. We begin by a discussion of the role played by projection in our space.

Suppose an inference function ψ has been found to estimate $\theta \in \Theta$. For a variety of reasons, ψ may not be an appropriate estimating function for the problem because it may depend on unobserved data, or not lie in the E-sufficient subspace, or it may lack computational simplicity or robustness as insurance against misspecification of the model. In general, we might suppose that there is some subset Υ in Ψ that contains the candidate inference functions for the problem. Then we would wish to replace the function ψ with some $\tilde{\psi} \in \Upsilon$.

The method we suggest here is to choose $\tilde{\psi}$ to be an element of Υ that is closest in a sense to be defined to the original inference function ψ. We can write

$$\|\psi - \tilde{\psi}\|_\theta^2 = \langle \psi - \tilde{\psi}, \psi - \tilde{\psi} \rangle_\theta.$$

It seems reasonable to define $\tilde{\psi}(\theta)$ pointwise in θ to be the value that minimizes $\|\psi - \tilde{\psi}\|_\theta^2$ for all $\tilde{\psi} \in \Upsilon$. Whereas we can define a function $\tilde{\psi}$ pointwise in this way, there is no guarantee that this function will lie in Υ: i.e., there will not always be a $\tilde{\psi} \in \Upsilon$ that uniformly minimizes the distance for all θ. However, we shall see that there are classes of sets Υ for which the pointwise minimizing $\tilde{\psi}$ remains in the set.

Definition 2.10. Let $\Upsilon \subset \Psi$ and let Ψ^* be the space of *all* unbiased inference functions (so that $\Psi \subset \Psi^*$). We set $\Upsilon_\theta = \{\phi(\theta); \phi \in \Upsilon\}$. Then the set Υ is said to be a *product set* if $\bigcap_{\theta \in \Theta}\{\psi \in \Psi^*; \psi(\theta) \in \Upsilon_\theta\} = \Upsilon$.

Then the standard result follows.

Proposition 2.11. *The intersection of product sets is a product set.*

When Υ is a product set we write $\Upsilon = \times_{\theta \in \Theta} \Upsilon_\theta$. The weak-squared topology is the weakest topology having all coordinate projections continuous. It is well known that

$$\mathrm{Cl}(\Upsilon) = \mathrm{Cl}\Big(\bigtimes_{\theta\in\Theta} \Upsilon_\theta\Big) = \bigtimes_{\theta\in\Theta} \mathrm{Cl}(\Upsilon_\theta),$$

where Cl denotes the appropriate closure in each setting. From this and the definition of the ancillary subspace, we have

PROPOSITION 2.12. *The weak-square closure of a product set is a product set.*

COROLLARY 2.13. *Suppose Ψ is a product set. Then $\mathscr{A}$ is a product set.*

PROPOSITION 2.14. *Suppose Ψ is a product set. Then there exists a complete E-sufficient space $\mathscr{S}$ that is also a product set.*

PROOF. Let $\mathscr{S}$ be the space of all functions $\psi \in \Psi$ such that $\langle \psi, \phi \rangle_\theta = 0$ for every $\theta \in \Theta$ and every $\phi \in \mathscr{A}$. It can be seen that the space $\mathscr{S}$ is nonempty because the zero-function is in $\mathscr{S}$. As defined, $\mathscr{S}$ is a closed linear product set such that $\mathscr{S}_\theta$ is the orthogonal complement of $\mathscr{A}_\theta$ in Ψ_θ. The equivalence in the definition of complete E-sufficiency now follows. □

The following proposition now tells us that we can find an element of a closed product set that is closest to a given inference function in Ψ.

PROPOSITION 2.15. *Let Υ be a closed product set in Ψ. Let $\psi \in \Psi$. Then there exists a $\tilde{\psi} \in \Upsilon$ such that for every $\theta \in \Theta$,*

$$\|\psi - \tilde{\psi}\|_\theta^2 = \inf_{\phi\in\Upsilon} \|\psi - \phi\|_\theta^2.$$

PROOF. For each $\theta \in \Theta$, let $\tilde{\psi}^\theta \in \Psi$ be chosen so that $\tilde{\psi}^\theta(\theta)$ lies in Ψ_θ and minimizes $E_\theta\{[\psi(\theta) - v]^2\}$ among all $v \in \Upsilon_\theta$. As Υ is closed, $\tilde{\psi}^\theta(\theta)$ can be chosen to lie in Υ_θ. The inference function $\tilde{\psi}$ defined by

$$\tilde{\psi}(\theta) = \tilde{\psi}^\theta(\theta)$$

will then be in Υ because Υ is a product set. □

PROPOSITION 2.16. *Suppose Υ is a linear subspace of Ψ that is a closed product set. Let $P_1 \ll P_2$ for all P_1 and P_2 in $\mathscr{P}$. Then $\tilde{\psi}$ is P-a.s. unique for all P.*

PROOF. As Υ_θ is closed and convex in V_θ, it follows that $\tilde{\psi}^\theta(\theta)$ is θ-a.s. unique for each $\theta \in \Theta$. Absolute continuity of P_1 and P_2 now implies that $\tilde{\psi}^\theta(\theta)$ is P-a.s. unique for all P. □

We now provide two common examples of linear subspaces that are also closed product sets.

EXAMPLE 2.17. Let Υ be the set of functions of Ψ that are $\sigma(T)$-measurable, where T is a measurable function on the sample space $\mathscr{X}$. Then, provided Ψ is a product set, it follows that Υ is a closed linear product space.

EXAMPLE 2.18. Suppose $X = (X_1, X_2, \dots, X_n)$ is a vector of random variables. Let Υ be the set of all functions in Ψ that are linear in the data for each value of $\theta \in \Theta$. Then if Ψ is a product set it follows that Υ is a closed linear product space. Note that we allow the coefficients of the linear functions to depend themselves nonlinearly upon the parameter θ.

In the remainder of this section we shall assume that we are working with a one-parameter model in which all distributions are absolutely continuous with respect to each other so that likelihood ratios exist.

Let λ be a reference measure such that $P \ll \lambda$ and $\lambda \ll P$ for all $P \in \mathscr{P}$. We define the likelihood

$$L(\theta) = dP/d\lambda,$$

where $\theta = \theta(P)$. Finally, when Ψ is a product space we shall let $\mathscr{A}$ and $\mathscr{S}$ denote the space of E-ancillary functions and the complete E-sufficient space, respectively. Both spaces will be closed linear product subspaces of Ψ. Projections into $\mathscr{A}$ and $\mathscr{S}$ will exist and be λ-almost surely unique. If ψ is any element of Ψ, then there is a λ-almost sure decomposition of ψ into $\psi = \psi_a + \psi_s$, where $\psi_a \in \mathscr{A}$ and $\psi_s \in \mathscr{S}$.

Proposition 2.14 shows that when Ψ is a product space, there exists a complete E-sufficient subspace. We now show how this space can be constructed. Define $\Psi_{\theta,\eta}$ to be the set of $Z \in \Psi_\theta$ such that $E_\eta|Z| < \infty$. Suppose $\Psi_{\theta,\eta} = \Psi_\theta$ for all θ and η. For each η and θ we define a linear functional $g(Z) = E_\eta Z$. Let us assume that these functionals are bounded. In terms of the original inference functions, this requires that

$$\sup\{|E_\eta\psi(\theta)|;\ \psi \in \Psi, \|\psi\|_\theta \le 1\} < \infty. \tag{2.6}$$

In this case the self-duality of Hilbert space implies that there exists an element of Z^* of Ψ_θ such that $E_\theta(Z^*Z) = g(Z)$ for all $Z \in \Psi_\theta$. Of course, the particular Z^* will depend on the values of η and θ. Let us make this dependence explicit with the subscript $Z^*_{\theta,\eta}$. When Ψ is a closed product space, there is a function $\psi_\eta \in \Psi$ such that $\psi_\eta(\theta) = Z^*_{\theta,\eta}$ almost surely. These functions, for all $\eta \in \Theta$ are the generators of the complete E-sufficient subspace. Note that when the function $L(\eta)/L(\theta) - 1$ is in Ψ it equals $\psi_\eta(\theta)$. In particular, we have the following result.

PROPOSITION 2.19. *Assuming Ψ is a product space and assuming* (2.6), *the complete E-sufficient subspace is the weak-square closure of linear combinations of the form*

$$\int_\Theta \psi_\eta(\theta)\, d\Lambda_\theta(\eta), \tag{2.7}$$

where, for each θ, Λ_θ is a signed measure on Θ with finite support.

We shall now consider the functions given in Proposition 2.19 in the context of the information measures of (1.2). Assume the conditions of Proposition 2.19 hold. Suppose that Λ_θ has finite support and that ψ maximizes J_ψ of (1.2). Let ϕ be such that $E_\eta\phi(\theta) = 0$ for all η. We define a function of a real variable t by $\kappa(t) = J_{\psi+t\phi}(\theta)$. As ψ maximizes (1.2) the function κ is seen to be maximised at $t = 0$. Then $\kappa'(0) = 0$, which implies that $\langle \psi, \phi \rangle_\theta = 0$. As this holds for all such ϕ, we conclude that ψ lies within the complete E-sufficient subspace. To show that the functions (2.7) are those maximizing (1.2), first note that (1.2) is unbounded unless (2.6) holds for all η in the support of Λ_θ. Then for ψ an arbitrary such function and for ψ^* defined by (2.7),

$$\int E_\eta\psi(\theta)\Lambda_\theta(d\eta) = \int E_\theta\left[\psi_\eta(\theta)\psi(\theta)\right]\Lambda_\theta(d\eta) = E_\theta\left[\psi^*(\theta)\psi(\theta)\right].$$

By the Cauchy–Schwarz inequality, this expectation is less than or equal to $\{E_\theta[\psi^*(\theta)]^2 E_\theta\psi^2(\theta)\}^{1/2}$, and therefore $J_\psi \leq E_\theta[\psi^*(\theta)]^2 = J_{\psi^*}$.

3. Selecting inference functions in one-parameter models. Throughout this section we shall assume a one-parameter model in which all distributions are absolutely continuous with respect to each other. Furthermore, we shall assume that Ψ is the space of *all* unbiased square-integrable inference functions $\psi(\theta; X)$. It can be seen that Ψ is a product Hilbert space with the constant covariance condition trivially satisfied.

For the special case where the model is a linear one-parameter exponential family with complete sufficient statistic T, one natural parametrization is the expected value of T. In this case, the score function can be written in the form $S(\theta) = c(\theta)[T - \theta]$. Useful features of this parametrization are that $E_\eta S(\theta)$ is then seen to be linear in η and $c(\theta)$ is the Fisher information.

Here, we consider the problem of choosing ψ in the complete E-sufficient subspace so as to have $E_\eta\psi(\theta)$ linear in η. It is easy to see that if such a function exists, it must be unique (up to an arbitrary multiple that can depend on θ but not on the data). Suppose that ψ_1 and ψ_2 are two such functions both lying in $\mathscr{S}$. Then we can write

$$E_\eta\psi_1(\theta) = k_1(\theta)[\eta - \theta] \tag{3.1}$$

and

$$E_\eta\psi_2(\theta) = k_2(\theta)[\eta - \theta]. \tag{3.2}$$

It follows that $k_2\psi_1 - k_1\psi_2$ is an element of $\mathscr{A}$. But as a linear combination of elements of $\mathscr{S}$ it must itself be in $\mathscr{S}$, which requires that $k_2\psi_1 - k_1\psi_2 = 0$. If the complete E-sufficient subspace is generated by a complete sufficient statistic, we have seen that such a linearization is possible in the appropriate parametrization. However, the existence of such a linearized function for a given parametrization in a general setting needs to be considered.

Projection turns out to be a useful tool for this question because we need only find any linearized inference function and then project it into the complete E-sufficient subspace. For example, suppose $X_1, \ldots, X_n$ are i.i.d. random variables from a location model with density $f(x - \theta)$. If each X_i has mean θ and finite variance, then $\psi(\theta) = \bar{x} - \theta$ is an element of Ψ with the required linearity, although it will not always lie within the complete E-sufficient subspace. However, we can write $\psi = \psi_s + \psi_a$, where $\psi_s \in \mathscr{S}$ and $\psi_a \in \mathscr{A}$. Then ψ_s will also have expectation linear in η because $E_\eta \psi(\theta) = E_\eta \psi_s(\theta)$ for all η, $\theta \in \Theta$. Of course, if a complete sufficient statistic exists, then this construction is equivalent to calculating the conditional expectation of $\bar{X}$ on the complete sufficient statistic, and so is Rao-Blackwellization of $\bar{X}$. The root of the resulting inference function will therefore be the unique UMVUE for θ.

The procedure can also be applied in cases where no complete sufficient statistic or UMVUE exists, because Proposition 2.14 guarantees the existence of a complete E-sufficient subspace under quite general conditions. To study projection in this more general setting, note first that we can project $\psi = \bar{X} - \theta$ into the space of inference functions that are measurable with respect to a minimal sufficient statistic. Minimal sufficient statistics exist under fairly general conditions and projection in this step amounts to calculating the conditional expectation of $\bar{X}$ given a minimal sufficient statistic. Now the E-ancillary component ofthis function can be seen to be an unbiased estimator of zero. Suppose we let T be the minimal sufficient statistic for the model. Then the $\sigma(T)$-measurable image of ψ under projection will be

$$\psi_T(\theta) = E(\bar{X}|T) - \theta, \tag{3.3}$$

which can be written as the sum of an E-sufficient component and an E-ancillary component. We select the former.

EXAMPLE 3.1. Suppose that $f(x) = (1/2)e^{-|x|}$. Let T be the vector of order statistics $X_{(1)}, \dots, X_{(n)}$. Then T is minimal sufficient. So $\psi(\theta) = \bar{X} - \theta$ already lies in the E-sufficient space generated by the minimal sufficient statistic. Does it lie in the complete E-sufficient subspace? For $n = 1$ we can write

$$X - \theta = \int_{-\infty}^{+\infty} \left[\frac{L(\eta)}{L(\theta)} - 1 \right] d\Lambda_\theta(\eta), \tag{3.4}$$

where $d\Lambda_\theta(\eta) = \exp(\theta - \eta)$ for $\eta \geq \theta$ and $d\Lambda_\theta = \exp(\theta + \eta)$ for $\eta < \theta$. So ψ_1 lies within the complete E-sufficient subspace. In fact, for $n = 1$, $\bar{X} = X$ is the unique uniformly minimum variance unbiased estimator. However, for $n > 2$, ψ is not in the E-sufficient subspace. To see this, let $n = 3$ and let $\phi(\theta) = \mathrm{X}_{(1)} - 2X_{(2)} + X_{(3)}$. Then ϕ is E-ancillary and positively correlated with ψ. Larger values of n are similar.

Secondarily to our stated purpose, this example gives some idea of the difficulty of searching for a UMVUE in models that are not exponential families. This is especially interesting in location models because, with the exception of normal models and the logarithm of gamma variates, they are never exponential families. Bondesson (1975) has investigated the existence of UMVUE's for location models and has used harmonic analysis to obtain some limited results.

He was able to show that UMVUE's usually do not exist in such models. Fortunately, complete E-sufficient functions that are linear in η can be found in more generality. All that is necessary to prove their existence is the existence of a function, linear in η, which does not lie within the space $\mathscr{A}$ of E-ancillary inference functions. The argument in the example is capable of generalization, which we summarize with earlier ideas in the following proposition. First we introduce a formal definition.

DEFINITION 3.2. An inference function $\psi(\theta)$ is said to be *E-linear* if we can write $E_\eta \psi(\theta) = c(\theta)(\eta - \theta)$ for some function $c(\theta)$.

PROPOSITION 3.3. If there exists an E-linear inference function ψ lying within the complete E-sufficient subspace, then it is the unique such function within that subspace up to a multiple $k(\theta)$. Furthermore, if T is a UMVUE for θ, then $\psi(\theta) = k(\theta)[T - \theta]$ for some choice of $k(\theta)$.

PROOF. Uniqueness has already been demonstrated. As noted by Lehmann and Scheffé (1950) and Rao (1952), if T is a UMVUE, then it is uncorrelated with every E-ancillary function. Therefore it must lie in $\mathscr{S}$. As T is unbiased, $T - \theta$ is E-linear, and from uniqueness, the result is proved. Note that the existence of an E-linear function in $\mathscr{S}$ does not imply the existence of a UMVUE. □

The analytical difficulty of projection into the complete E-sufficient subspace when the model does not admit a complete sufficient statistic or is not a location model can be seen. We now consider a linearization similar to the previous one but only in a local sense of the second-order properties of the expectation function. This will have the advantage of being easier to calculate for many models.

Consider for a moment a general unbiased inference function $\psi(\theta)$. Differentiating twice the unbiasedness condition,

$$\frac{d^2}{d\theta^2}\int \psi(\theta)L(\theta)\,d\lambda = A + 2B + C = 0, \tag{3.5}$$

where

$$A = E_\theta \frac{d^2}{d\theta^2}\psi(\theta),$$

$$B = \frac{d}{d\eta}E_\eta \frac{d}{d\theta}\psi(\theta)\Big|_{\eta=\theta}$$

and

$$C = \frac{d^2}{d\eta^2}E_\eta \psi(\theta)\Big|_{\eta=\theta}.$$

A local linearity condition that we now consider is to require that (I) $A = 0$, (II) $B = 0$ and (III) $C = 0$. Note that "global" E-linearity defined earlier does not imply condition (I). In view of the identity given previously, equations (I)–(III) form two constraints. Thus the three conditions (I)–(III) are naturally linked to inference functions ψ, which are approximately linear both pointwise and in expectation in the parameter θ. When this approximate linearity is desirable, for example, when we wish a one-step estimator based on a linear approximation that is close to the root, we may choose from the second-order locally E-sufficient subspace of inference functions, of the form $k_1 S + k_2(S^2 - I)$ one which satisfies the conditions (I)–(III).

In general, in order to find $k_1(\theta)$, $k_2(\theta)$ it is convenient to find $k_2(\theta)$ in terms of $k_1(\theta)$ using (III) and then to use condition (II) to construct a first-order linear differential equation in $k_1(\theta)$.

Example 3.4. Let $X_1, X_2, \dots, X_n$ be independent, identically distributed variates with score function $S(\theta) = c(\theta)[\sum_{i=1}^n X_i - n\theta]$. Then there is a unique (up to multiplication by a factor constant in X, θ) function ψ in the second-order E-sufficient subspace satisfying (I)–(III), namely,

$$\psi(\theta) = \sum_{i=1}^{n} X_i - n\theta. \tag{3.6}$$

Somewhat more generally than the previous example, it is easy to see that $k_2(\theta)$ can be chosen equal to 0 in any model for which the unbiased inference functions spanning the second-order E-sufficient subspace, $S(\theta)$ and $S^2(\theta) - I(\theta)$, are orthogonal, since in this case,

$$\frac{d^2}{d\eta^2} E_\eta k_1(\theta) S(\theta)|_{\eta=\theta} = \langle k_1 S, S^2 - I \rangle = 0 \tag{3.7}$$

and we may choose

$$k_1(\theta) = \exp\left\{ -\frac{1}{2} \int \frac{E_\theta[I'(\theta)]}{E_\theta I(\theta)} d\theta \right\}. \tag{3.8}$$

so that condition (I) holds with $\psi = k_1 S$.

Thus, under the orthogonality of the functions S and $S^2 - I$, there is a simple nonrandom multiple of the score function that satisfies the conditions (I)–(III). Whereas members of the one-parameter exponential family parametrized by the mean as in Example 3.4 are clearly included among such functions, they are by no means the only distributions with these properties. The following is a standard *mixture* model, which we show also admits a multiple of the score function satisfying conditions (I)–(III).

Example 3.5. Let f and g be two probability density functions with common support and suppose we observe $X_1, X_2, \dots, X_n$, independent, identically distributed random variates from the mixture probability density $\theta f(x) + (1 - \theta) g(x)$, where $0 \le \theta \le 1$ is the mixture parameter. In this case the score function and information function can be written $S(\theta) = \sum_{i=1}^n S_i(\theta)$ and $I(\theta) =$

$\Sigma_{i=1}^{n} S_i^2(\theta)$, where

$$S_i(\theta) = \frac{f-g}{\theta f + (1-\theta) g}(X_i). \tag{3.9}$$

In this case, $E(S^3) = E(SI)$ and solving (3.8),

$$k_1(\theta) = \left[E_\theta I(\theta)\right]^{-1}. \tag{3.10}$$

However, it should be remarked that even within the one-parameter exponential family, the linearized element of the second-order E-sufficient subspace is not always just a multiple of the score function.

4. Nuisance parameter models. Suppose $X_1, \dots, X_n$ come from a location-scale model with p.d.f. given by $\sigma^{-n} f[(x_1 - \theta)/\sigma, \dots, (x_n - \theta)/\sigma]$. We treat the two parameters separately.

(a) *Location parameter.* Let $a(\theta) = [\Sigma_{i=1}^{n}(x_i - \theta)^2]^{1/2}$ and define Ψ_1 to be the space of all unbiased square-integrable functions of the vector

$$v = \left[(x_1 - \theta)/a(\theta), \dots, (x_n - \theta)/a(\theta)\right]. \tag{4.1}$$

This vector will lie on an n-dimensional unit sphere with distribution dependent only on f.

(b) *Scale parameter.* Let Ψ_2 be the space of all unbiased square-integrable inference functions of the vector $w = (x_1 - \bar{x}, \dots, x_n - \bar{x})$.

EXAMPLE 4.1. Consider, for example, the case where f is the standard normal product density. Then the t-statistic $t = n^{1/2}(\bar{x} - \theta)/s$ lies within the complete E-sufficient subspace because it is complete for fixed θ over the normal family with mean η and fixed variance σ^2 and the conditional distribution of v given t is parameter free. We will show that *the student t-statistic is the unique E-linear function lying in the complete E-sufficient subspace of* Ψ_1.

We have that v will be uniformly distributed on the n-dimensional sphere

$$S^n = \left\{ v = (v_1, \dots, v_n) \colon \sum_{i=1}^{n} v_i^2 = 1 \right\}. \tag{4.2}$$

Define latitudes on S^n by setting

$$A_s = \left\{ v \in S \colon \sum_{i=1}^{n} v_i = s \right\}, \tag{4.3}$$

for $-n^{1/2} \le s \le +n^{1/2}$. Then for any value η of the location parameter, the density of v on the sphere defined by centering at θ will be constant on the latitudes A_s for all s. Therefore any function $\phi(\theta) = \phi(\theta; t)$ with the property that its average value on every latitude A_s is zero will be E-ancillary. In view of this, every E-sufficient function for θ will be a function of the data through

$\Sigma_{i=1}^{n} v_i$ and therefore a function of the usual student t-statistic. It is interesting to note that although the t-statistic is the E-linear element up to a multiple in the complete E-sufficient subspace, it is not the local E-sufficient function that can be shown to be a multiple of v.

To estimate location parameters for nonnormal densities $f(x)$, we can show that functions of the form

$$(4.4)\qquad \int \left[\frac{\int_0^\infty v^{n-1} f\,[v(x_1-\theta)+\varepsilon,\ldots,v(x_n-\theta)+\varepsilon]\,dv}{\int_0^\infty v^{n-1} f\,[v(x_1-\theta),\ldots,v(x_n-\theta)]\,dv} - 1 \right] d\Lambda_\theta(\varepsilon)$$

lie inside the complete E-sufficient subspace provided the second moment is finite. So a locally E-sufficient function found by dividing by ε and letting $\varepsilon \to 0$ will have a root $\hat\theta$ satisfying

$$(4.5)\qquad \int_0^\infty v^{n-1} \frac{\partial}{\partial \varepsilon} f\left[v\left(x_1-\hat\theta\right)+\varepsilon,\ldots,v\left(x_n-\hat\theta\right)+\varepsilon\right]\Big|_{\varepsilon=0}\,dv = 0.$$

Consider now the case of the scale parameter of the normal model. In this case, it is easy to establish that functions in the complete E-sufficient subspace in Ψ_2 will be functions of the data through the sample variance. It can also be seen that the inference function

$$(4.6)\qquad \psi(\sigma) = \frac{\Sigma_{i=1}^{n}(x_i-\bar{x})^2}{n-1} - \sigma^2$$

is both E-linear in σ^2 and locally E-sufficient. The resulting estimator for σ^2 will therefore be the bias corrected sample variance. In the more general setting, where the distribution is not assumed to be normal, we note that a locally E-sufficient function for σ can be found as the score function for σ based upon the marginal distribution of the maximal location invariant statistic.

We end this section by examining the elimination of nuisance parameters by conditioning on statistics. Suppose there exists a family of statistics T_θ for every $\theta \in \Theta$ such that:

(I) For every θ_0 and for every θ_1 the statistic T_{θ_0} is a complete statistic for the nuisance parameters $\xi_1,\ldots,\xi_p$ in the model $\{P;\,\theta(P)=\theta_1\}$. By this we mean that if a function h satisfies $E_P h(T_{\theta_0}) = 0$ for all P such that $\theta(P)=\theta_1$, it follows that $P[h(T_{\theta_0})=0]=1$ for all such P.

Consider an inference function $\psi \in \Psi$. It follows from the unbiasedness of ψ and condition (I) that $E_\theta[\psi(\theta)|T_\theta]=0$ almost surely. Thus ψ is conditionally unbiased given T. It also follows immediately from condition (I) and the assumption that Ψ has constant covariance structure that conditionally on T any two functions ψ_1 and ψ_2 have constant covariance structure. By this we mean that $E_P[\psi_1(\theta(P))\psi_2(\theta(P))|T_\theta]$ depends upon P only through $\theta(P)$. Thus the space of functions Ψ is then seen to be a space of inference functions conditionally on the family of statistics T_θ. Furthermore, if ϕ is E-ancillary in Ψ with finite expectation, it follows from the completeness condition (I) that $\phi(\theta;X)$ will be conditionally E-ancillary in the sense that $E_P[\psi(\theta;X)|T_\theta]=0$, T-a.s.

Suppose that a conditional complete E-sufficient subspace exists in the conditional model given T. Consider now an inference function ψ that lies within the conditional complete E-sufficient subspace for every realization of T, and consider a function ϕ, which in the unconditional sense, is E-ancillary. From the previous remarks, the function ϕ will be conditionally E-ancillary given T and therefore conditionally orthogonal to ψ,

$$E_P\left[\psi(\theta(P))\phi(\theta(P))|T_{\theta(P)}\right] = 0, \tag{4.7}$$

for all P. Taking the expectation of (4.7), we observe that ψ and ϕ are unconditionally orthogonal. As ϕ is an arbitrary E-ancillary function it follows that ψ lies within the complete E-sufficient subspace of Ψ. Thus we have proved the following.

PROPOSITION 4.2. *If ψ lies within the conditional complete E-sufficient subspace of Ψ given T, then it follows that ψ also lies unconditionally within the complete E-sufficient subspace.*

The relevance of this result to the consideration of nuisance parameters is that in a number of cases by conditioning upon an appropriate statistic, a nuisance parameter can be eliminated. In such conditional models, it becomes fairly easy to construct elements of the complete E-sufficient subspace using the methods of Section 2. Proposition 4.2 then shows that such a function is also an element of the complete E-sufficient subspace in the unconditional model. A condition that guarantees the elimination of nuisance parameters is the following.

(II) For every $\theta_0 \in \Theta$, the statistic T_{θ_0} is sufficient for the nuisance parameters $\xi_1, \dots, \xi_p$ in the model $\{P\colon \theta(P) = \theta_0\}$.

Henceforth assume that both (I) and (II) hold. Consider all functions of the form

$$\psi(\theta) = \int\left[\frac{dQ_{X|T_\theta}}{dP_{X|T_\theta}} - 1\right] d\Lambda_{\theta, T_\theta}(Q), \tag{4.8}$$

where $\theta(P) = \theta$ and Λ is a signed measure of finite support defined on subsets of $\mathscr{P}$. Note that although this appears superficially to depend on the value of P, the sufficiency condition (II) indicates that the function is dependent only on $\theta(P)$. We wish to show that if functions of the form (4.8) are square integrable, then they lie in the complete E-sufficient space of inference functions. It is enough to note that by construction (4.8) lies within the conditional complete E-sufficient subspace. By Proposition 4.2 the result that ψ is within the unconditional complete E-sufficient subspace follows.

We conclude with a few remarks about cases where Ψ is not a product space and so a complete E-sufficient subspace is not guaranteed to exist. In such cases, the program suggested in this paper of E-sufficiency reduction cannot be carried out in the straightforward approach of Section 2. Nevertheless, there may exist functions ψ that are orthogonal to all E-ancillary functions, and one of these may be an appropriate function for inference in the context. The situation differs mainly in the more restrictive choices available and in the failure to decompose any inference function cleanly into an E-sufficient and an E-ancillary component orthogonal to each other.

REFERENCES

BASU, D. (1958). On statistics independent of sufficient statistics. *Sankhyā* **20** 223–226.

BONDESSON, L. (1975). Uniformly minimum variance estimation in location parameter families. *Ann. Statist.* **3** 637–660.

COX, D. R. (1980). Local ancillarity. *Biometrika* **67** 279–286.

GODAMBE, V. P. (1960). An optimum property of regular maximum likelihood estimation. *Ann. Math. Statist.* **31** 1208–1211.

LEHMANN, E. L. (1981). An interpretation of completeness and Basu's theorem. *J. Amer. Statist. Assoc.* **76** 335–340.

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

MCCULLAGH, P. (1984). Local sufficiency. *Biometrika* **71** 233–244.

MCLEISH, D. L. (1984). Estimation for aggregate models: The aggregate Markov chain. *Canad. J. Statist.* **12** 265–282.

RAO, C. R. (1952). Some theorems on minimum variance estimation. *Sankhyā* **12** 27–42.

DEPARTMENT OF STATISTICS AND
ACTUARIAL SCIENCE
UNIVERSITY OF WATERLOO
WATERLOO, ONTARIO N2L 3G1
CANADA