Ann. Inst. Statist. Math.
Vol. 52, No. 1, 173–183 (2000)

ON THE OPTIMALITY OF ESTIMATORS BASED ON P-SUFFICIENT STATISTICS

VASANT P. BHAPKAR*

Department of Statistics, College of Arts and Sciences, University of Kentucky, Lexington, KY 40506-0027, U.S.A.

(Received October 17, 1997; revised April 30, 1998)

Abstract. For estimation of functions involving only parameters of interest, in the presence of nuisance parameters, some optimality properties are established for partially sufficient (i.e. p-sufficient) statistics in two classes of regular probability models. The results are based on a characterization of regular unbiased estimating functions for parameters of interest in probability models for which a statistic exists such that its marginal distribution depends on unknown parameters only through the parameters of interest.

Key words and phrases: Fisher information, generalized Fisher information, partial sufficiency, ancillary statistic.

1. Introduction

If a statistic is sufficient with respect to the unknown parameter ω in a given probability model, then the choice of a suitable estimator of a given parametric function is usually confined to estimators based on the sufficient statistic alone. The Rao-Blackwell Theorem provides the theoretical justification, for this sufficiency reduction, when using the squared-error loss function or, more generally, using convex loss functions.

Suppose now that the unknown parameter ω is expressed as $(\theta;\phi)$ where θ is the parameter of interest and ϕ is the nuisance parameter. This paper establishes a generalization of Rao-Blackwell Theorem for estimation of functions of parameter θ, in the presence of nuisance parameter ϕ, for two broad classes of probability models under the Cramér-Rao type *regularity assumptions.* These two classes are specified as $\mathcal{M}_1$ and $\mathcal{M}_2$ in Section 2.

In particular, for the class $\mathcal{M}_1$, Theorem 2.1 in Section 2 extends the results due to Fraser (1956) to a larger class than the one considered by him, and it also provides a sharper version of the theorems than Fraser's; however, Theorem 2.1 requires the stronger assumptions of *regularity.* For the class $\mathcal{M}_2$, Theorem 2.2 in Section 2 provides a generalization of Rao-Blackwell Theorem which is even stronger compared to Theorem 2.1.

Theorems 2.1 and 2.2 may be interpreted as generalizations of Rao-Blackwell Theorem in the context of notions of sufficiency with respect to θ, in the presence of unknown ϕ, i.e., *partial sufficiency* (or simply p-sufficiency) with respect to θ in the terminology of Basu (1977, 1978).

---

* Vasant P. Bhapkar died on July 23, 1999 after a revised version of this paper had been accepted for publication. Editorial Committee thanks Professor Z. Govindarajulu for his careful proofreading of the manuscript.

There are several definitions of p-sufficiency in the statistical literature, viz. S-sufficiency in Basu (1977) and Barndorff-Nielsen (1978), G-sufficiency by Barnard (1963) as it has been termed by Barndorff-Nielsen (1978), M-sufficiency by Barndorff-Nielsen (1978), L-sufficiency by Rémon (1984) and sufficiency (with respect to θ), *ignoring* ϕ, as defined by Godambe (1980). The definition that is relevant to the classes $\mathcal{M}_1$ and $\mathcal{M}_2$ discussed in this paper is the one based on the generalized Fisher information measure, $\boldsymbol{I}_G(\theta;\omega)$, defined by (4.2) in Section 4. This generalization for the scalar case is due to Godambe (1984) and its matrix version has been used by Bhapkar (1989, 1990).

Section 2 gives the regularity assumptions and the statements of the two theorems. Proofs of the theorems are given in Section 3; they depend on Lemma 3.1 which is established in the Appendix. The p-sufficiency of S with respect to θ under conditions $\mathcal{M}_1$ and/or $\mathcal{M}_2$ is discussed in Section 4 in the context of the generalized Fisher information, $\boldsymbol{I}_G(\theta;\omega)$, with respect to θ. Some examples are also illustrated in this section. The concluding remarks in the last section point out the relationship between the conditions $\mathcal{M}_2$ and the definition of sufficiency with respect to θ, ignoring ϕ, given by Godambe (1980), when there exists a statistic which is ancillary with respect to θ.

2. Main results

Let P_ω^X denote the probability distribution of a random variable X with sample space χ with ω as the unknown parameter in space Ω. Assume that X has the probability density function (pdf) $p(x;\omega)$ with respect to a σ-finite measure μ.

We assume that the interest parameter θ and the nuisance parameter ϕ are variation independent, i.e. $\Omega = \Theta \times \Phi$, and the parametric spaces Θ and Φ are open subspaces of Euclidean spaces of dimensions d_1 and d_2, respectively.

Suppose now that the statistic (S,T) is sufficient with respect to the whole parameter ω; i.e. (S,T) is sufficient for the family of distributions $\{P_\omega^{(X)} : \omega \in \Omega\}$. Without loss of generality we may confine our attention to estimators based on (S,T) alone, for estimation of any given parametric function of ω, in view of the Rao-Blackwell Theorem (see, e.g. Theorem 4.2.1, Bickel and Doksum (1977)). For simplicity, assume then that the joint distribution $P_\omega^{(S,T)}$ of (S,T) has the pdf $p(s,t;\omega)$ with respect to a σ-finite measure μ.

The joint distribution is now assumed to be *regular* in the sense that the Cramér-Rao regularity conditions are satisfied for both the marginal distribution $P_\omega^{(S)}$ of S, with pdf $f(s;\omega)$ with respect to a σ-finite measure ν, and the conditional distribution $P_\omega^{(T|s)}$ of T, given s, with pdf $h(t \mid s;\omega)$ with respect to σ-finite measure η_s for almost all s.

The other fundamental assumption in this paper is that the marginal distribution of S depends on ω only through θ. Hence we write hereafter $f(s;\theta)$ as the pdf of S.

The specific Cramér-Rao regularity conditions on (S,T) are then listed as $\boldsymbol{R}$ in the Appendix.

Now the class $\mathcal{M}_1$ of distributions of (S,T) is the class which satisfies the following conditions:

$\mathcal{M}_1$: (i) The marginal distribution $P_\theta^{(S)}$ of S depends on ω only through θ,

(ii) the joint distribution $P_\omega^{(S,T)}$ of (S,T) satisfies regularity conditions $\boldsymbol{R}$, and

(iii) the conditional distribution of T, given s, depends on ω only through $\delta = \delta(\omega)$, where δ is differentiable and $\psi = (\theta,\delta)$ is a one-to-one transformation of $\omega = (\theta,\phi)$.

We have then the following result:

THEOREM 2.1. *Suppose the joint distribution* $P_\omega^{(S,T)}$ *satisfies conditions* $\mathcal{M}_1$. *If there exists an unbiased estimator* $U = U(S,T)$ *of a real-valued parametric function*

$w(\theta)$, where $w(\theta)$ is differentiable, such that $\mathrm{Var}_\omega U < \infty$, *then there exists* $V = V(S)$ *such that*

$$\text{(i)}\ \ E_\theta V = w(\theta)$$

and

$$\text{(ii)}\ \ \mathrm{Var}_\theta V \le \mathrm{Var}_\omega U, \qquad \textit{all} \quad \omega \in \Omega.$$

COROLLARY 2.1. *If conditions of Theorem 2.1 hold, then with any convex loss function to estimate a real-valued function $\beta(\theta)$, the risk function $R(\omega, V)$ of V satisfies the inequality $R(\omega, V) \le R(\omega, U)$ for all $\omega \in \Omega$.*

The proof is deferred to the following section.

The above theorem is a sharper version of the results due to Fraser (1956) in a number of ways, When $\delta = \phi$ in $\mathcal{M}_1$ Fraser's Theorem 3 (p. 841 (1956)) requires further assumption of *completeness* of the family $\{P_\theta^{(S)} : \theta \in \Theta\}$. Also the result attributed to Fraser, as interpreted by Basu (p. 360 (1977)), states in our notation that $\mathrm{Var}_\theta V \le \mathrm{Var}_\omega U$ for estimators $U = U(S,T)$ *which have variance depending on θ alone*; Theorem 2.1 above requires no such restriction on U. Finally, Theorem 2.1 holds even when $\delta \ne \phi$, especially when θ and δ are not necessarily variation independent. However we note that the theorem does need stronger assumptions of regularity of distributions and differentiability of parametric function to be estimated.

The other class $\mathcal{M}_2$ of distributions is the class which satisfies the following conditions:

$\mathcal{M}_2$: (i) and (ii) as in $\mathcal{M}_1$,

(iii) the family of conditional probability distributions $\{P_\omega^{(T|s)} : \phi \in \Phi\}$ of T, given s, is complete for every fixed $\theta \in \Theta$ for almost all s,

(iv) S and T are stochastically independent.

We have then the following:

THEOREM 2.2. *Suppose the joint distribution $P_\omega^{(S,T)}$ satisfies conditions $\mathcal{M}_2$. If there exists an unbiased estimator $U = U(S,T)$ of a real-valued parametric function $w(\theta)$, where $w(\theta)$ is differentiable such that* $\mathrm{Var}_\omega U < \infty$, *then* $U = V(S)$ *with probability one for some function V.*

The proof is deferred to the following section.

It may be noted here that the condition (iv) in $\mathcal{M}_2$ is essential; some remarks are made in Section 5 regarding the role condition (iv) plays.

3. Proofs of Theorems

The proofs of the theorems in Section 2 are based on Lemma 3.1 below which provides a convenient representation of any *regular unbiased estimating function* (RUEF) for the parameter of interest θ.

Let C be the Hilbert space of real-valued functions $c = c(s, t; \omega)$ which satisfy

$$E_\omega c(S,T;\omega) = 0$$
$$E_\omega c^2(S,T;\omega) < \infty \tag{3.1}$$

for all $\omega \in \Omega$. Suppose $\mathcal{G} = \mathcal{G}(S,T)$ is the subspace of C spanned by real-valued functions $g = g(s,t;\theta)$ which depend of ω only through the parameter of interest θ. If g is differentiable with respect to elements of θ, and if the relation $\int gpd\mu = 0$ can be differentiated with respect to elements of ω, then we refer to g as a RUEF(even in the vector case $d_1 \geq 1$).

Let $\mathcal{G}(S)$ be the subspace of $\mathcal{G}$ of functions g depending on (s,t) only through s, and suppose $\mathcal{G}_0(S)$ is the subspace (in $\mathcal{G}(S)$) of functions $g_0 = g_0(s;\theta)$ uncorrelated with the score function $\boldsymbol{l}_\theta(s;\theta) = \partial \log f(s;\theta)/\partial\theta$, assuming regularity conditions $\boldsymbol{R}$.

LEMMA 3.1 *Assume that the marginal distribution of S depends on ω only through θ and the joint distribution of (S,T) satisfies the regularity conditions $\boldsymbol{R}$. Then for any $g \in \mathcal{G}(S,T)$ we have the representation*

$$g(s,t;\theta) = \boldsymbol{b}'(s,t;\theta)\boldsymbol{l}_\theta(s;\theta) + g_0(s;\theta), \tag{3.2}$$

where $g_0 \in \mathcal{G}_0(S)$ and $E_\omega[\boldsymbol{b}(s,T;\theta) \mid s] = \boldsymbol{a}(\omega)$ for almost all s.

This lemma has been established by Bhapkar (1997); for the sake of completeness this proof is included in the Appendix. Lemmas A.2 and A.3 essentially prove Lemma 3.1 above.

PROOF OF THEOREM 2.1. Suppose now that $U = U(s,t)$ is to be used as an estimator of some parametric function of θ alone, and assume that $E_\omega U(S,T) = w(\theta)$, and $\mathrm{Var}_\omega U < \infty$. Then $g(s,t;\theta) = U(s,t) - w(\theta)$ is an UEF for θ.

In view of the characterization (3.2) we have

$$U(s,t) - w(\theta) = g = \boldsymbol{b}'(s,t;\theta)\boldsymbol{l}_\theta(s;\theta) + g_0(s;\theta)$$

for some $g_0 \in \mathcal{G}_0(S)$, where $E_\omega[\boldsymbol{b}(s,T;\theta) \mid s] = \boldsymbol{a}(\omega)$. Taking expectation with respect to the distribution of T, given s, we have

$$V(s;\omega) - w(\theta) = \boldsymbol{a}'(\omega)\boldsymbol{l}_\theta(s;\theta) + g_0(s;\theta), \tag{3.3}$$

where $V(s;\omega) = E_\omega[U(s,T) \mid s]$. Multiplying by $\boldsymbol{l}'_\theta(S;\theta)$ and taking expectation with respect to S, we get

$$E_\omega[V(S;\omega)\boldsymbol{l}'_\theta(S;\theta)] = \boldsymbol{a}'(\omega)\boldsymbol{I}^{(s)}(\theta), \tag{3.4}$$

since $\boldsymbol{l}_\theta(S;\theta)$ elements are uncorrelated with $\mathcal{G}_0(S)$.

Let $h^*(t \mid s;\delta) = h(t \mid s;\omega)$ be the conditional pdf of T in view of condition (iii) $\mathcal{M}_1$. In terms of the new parameter $\psi = (\theta,\delta)$, we have $p(s,t;\omega) = p^*(s,t;\psi) = f(s;\theta)h^*(t \mid s;\delta)$. Assuming that the UEF $g(s,t;\theta) = U(s,t) - w(\theta)$ is *regular*, the relation

$$\int g(s,t;\theta)p^*(s,t;\psi)d\mu(s,t) = 0$$

can be differentiated under the integral sign with respect to θ. Equivalently,

$$\int U(s,t)p^*(s,t;\psi)d\mu(s,t) = w(\theta),$$

$$\text{i.e.} \quad \int V^*(s;\delta)f(s;\theta)d\nu(s) = w(\theta), \tag{3.5}$$

can be differentiated with respect to θ, where $V^*(s;\delta) = V(s;\omega) = E_\omega[U(s,T) \mid s]$, in view of $\mathcal{M}_1$. Thus, we have from (3.5)

$$E_\psi[V^*(S;\delta)\boldsymbol{l}_\theta(S;\theta)] = \left[\frac{\partial_w(\theta)}{\partial\theta}\right],$$

and, hence,

$$\boldsymbol{I}^{(S)}(\theta)\boldsymbol{a}^*(\psi) = \left[\frac{\partial w(\theta)}{\partial\theta}\right]',$$

in view of (3.4), where $\boldsymbol{a}^*(\psi) = \boldsymbol{a}(\omega)$. Thus, $\boldsymbol{a}^*(\psi) = [\boldsymbol{I}^{(S)}(\theta)]^{-1}[\frac{\partial w(\theta)}{\partial\theta}]'$, and from (3.3) we have $V^*(s;\delta) = w(\theta) + [\boldsymbol{a}^*(\psi)]'\boldsymbol{l}_\theta(s;\theta) + g_0(s;\theta)$. But the right hand side term is now depending on ψ only through θ, while the left hand side depends of ψ only through δ. Thus, it follows that $E_\omega[U(s,T) \mid s] = E_\psi[U(s,T) \mid s] = V(s)$ independent of ψ, or equivalently of ω. Thus if $U(S,T)$ is unbiased for $w(\theta)$, so is $V(S)$; furthermore we immediately have

$$\mathrm{Var}_\omega V(S) \le \mathrm{Var}_\omega U(S,T), \qquad \text{all} \quad \omega \in \Omega. \qquad \square$$

PROOF OF THEOREM 2.2. Since $E_\omega U(S,T) = w(\theta)$, $g = g(s,t;\theta) = U(s,t) - w(\theta)$ is a RUEF for θ under the assumptions of the theorem.

Now for any RUEF g, we have the representation

$$g(s,t;\theta) = \boldsymbol{b}'(s,t;\theta)\boldsymbol{l}_\theta(s;\theta) + g_0(s;\theta)$$

in view of (3.2). We first show that under assumptions $\mathcal{M}_2$, $\boldsymbol{b}(s,t;\theta) = \boldsymbol{b}^*(t;\theta)$ a.e. (μ), for some $\boldsymbol{b}^*$.

Define $\boldsymbol{b}^*(t;\theta) = E_\omega[\boldsymbol{b}(S,t;\theta) \mid t] = E_\theta \boldsymbol{b}(S,t;\theta)$ in view of independence of S and T, and the condition (i) in $\mathcal{M}_2$. Let $\boldsymbol{b}_0(s,t;\theta) = \boldsymbol{b}(s,t;\theta) - \boldsymbol{b}^*(t;\theta)$. Then $E_\omega[\boldsymbol{b}_0(s,T;\theta) \mid s] = E_\omega \boldsymbol{b}_0(s,T;\theta) = E_\omega \boldsymbol{b}(s,T;\theta) - E_\omega \boldsymbol{b}^*(T;\theta) = \boldsymbol{a}(\omega) - \boldsymbol{a}(\omega) = \boldsymbol{0}$, since $E_\omega \boldsymbol{b}(s,T;\theta) = E_\omega[\boldsymbol{b}(s,T;\theta) \mid s] = \boldsymbol{a}(\omega)$ and $E_\omega \boldsymbol{b}^*(T;\theta) = E_\omega\{E_\omega[\boldsymbol{b}(S,T;\theta) \mid T]\} = E_\omega \boldsymbol{b}(S,T;\theta) = E_\omega\{E_\omega[\boldsymbol{b}(S,T;\theta) \mid S]\} = \boldsymbol{a}(\omega)$. But $E_\omega[\boldsymbol{b}_0(s,T;\theta) \mid s] = \boldsymbol{0}$ implies $\boldsymbol{b}_0(s,t;\theta) = \boldsymbol{0}$ a.e. (μ). Hence $\boldsymbol{b}(s,t;\theta) = \boldsymbol{b}^*(t;\theta)$.

Thus we have

$$U(s,t) - w(\theta) = g(s,t;\theta) = [\boldsymbol{b}^*(t;\theta)]'\boldsymbol{l}_\theta(s;\theta) + g_0(s;\theta), \tag{3.6}$$

for some $g_0 \in \mathcal{G}_0(S)$, where $E_\omega \boldsymbol{b}^*(T;\theta) = \boldsymbol{a}(\omega)$.

Multiplying both sides of (3.6) with $\boldsymbol{l}_\theta'(S;\theta)$, and taking expectation with respect to S, for fixed t, we get

$$E_\theta[U(S,t)\boldsymbol{l}_\theta'(S;\theta)] = [\boldsymbol{b}^*(t;\theta)]'\boldsymbol{I}^{(S)}(\theta). \tag{3.7}$$

But for every fixed t,

$$E_\theta[U(S,t)\boldsymbol{l}_\theta'(S;\theta)] = \left[\frac{\partial U^*}{\partial\theta}(t;\theta)\right], \tag{3.8}$$

in view of the regularity assumption for the UEF, where $U^*(t;\theta) = E_\theta U(S,t) = E_\theta[U(S,t) \mid t]$.

However $E_\omega U^*(T;\theta) = E_\omega U(S,T) = w(\theta)$, so that $E_w[U^*(T;\theta) - w(\theta)] = 0$, and then completeness implies $U^*(t;\theta) = w(\theta)$, a.e. (η).

Hence, in (3.8) we have

$$E_\theta[U(S,t)\boldsymbol{l}'_\theta(S;\theta)] = \left[\frac{\partial w(\theta)}{\partial \theta}\right].$$

Thus in (3.7), we have

$$\boldsymbol{b}^*(t;\theta) = [\boldsymbol{I}^{(S)}(\theta)]^{-1}\left[\frac{\partial w(\theta)}{\partial \theta}\right]',$$

and, hence $\boldsymbol{b}^*(t;\theta) = \boldsymbol{a}^*(\theta)$ for some $\boldsymbol{a}^*$. From (3.6), then, it follows that $U(s,t) = V(s)$ a.e. (μ) for some V. Thus, the theorem is proved.

4. Partial sufficiency of S

For the classes $\mathcal{M}_1$ and $\mathcal{M}_2$ of probability distributions considered in the earlier sections, S may be considered to be sufficient with respect to θ, when ϕ is unknown (i.e. p-sufficient for θ). It has been shown (Bhapkar (1997)) that S contains the whole *generalized* Fisher information with respect to θ that is available in (S,T), when the joint distribution satisfies the regularity conditions $\boldsymbol{R}$.

For the case of scalar θ (i.e. $d_1 = 1$) the generalized Fisher information with respect to θ in (S,T) is defined by (Godambe (1984))

$$I_G^{(S,T)}(\theta,\omega) = \inf_{u\in\mathcal{U}} E_\omega[\ell_\theta(S,T;\omega) - u(S,T;\omega)]^2; \tag{4.1}$$

here $\ell_\theta(s,t;\omega) = \partial \log p(s,t;\omega)/\partial\theta$ and $\mathcal{U}$ is the orthogonal complement in $\mathcal{C}$ of $\mathcal{G}$, the class of functions $g = g(s,t;\theta)$ in $\mathcal{C}$ which depend on ω only through θ. More generally for the vector θ (i.e. $d_1 \geq 1$) the generalized information matrix with respect to θ is defined as

$$\boldsymbol{I}_G^{(S,T)}(\theta;\omega) = \inf_{\boldsymbol{u}} E_\omega[\boldsymbol{l}_\theta(S,T;\omega) - \boldsymbol{u}[\boldsymbol{l}_\theta(S,T;\omega) - \boldsymbol{u}]'; \tag{4.2}$$

here now $\boldsymbol{u} = \boldsymbol{u}(s,t;\omega)$ is a vector with elements from $\mathcal{U}$. It has been shown by Bhapkar and Srinivasan (1994) that

$$\boldsymbol{I}_G^{(S,T)}(\theta;\omega) = E_\omega \boldsymbol{g}^*(S,T;\theta)\boldsymbol{g}^{*'}(S,T;\theta), \tag{4.3}$$

where $\boldsymbol{g}^*$ is the vector of projections of elements of $\boldsymbol{l}_\theta(s,t;\omega)$ on $\mathcal{G}$.

It can be shown (Bhapkar (1989)) that confining $\boldsymbol{u}$ in (4.2) to a sub-class of $\mathcal{U}$ consisting of functions $\boldsymbol{u}(s,t;\omega) = \boldsymbol{N}(\omega)\boldsymbol{l}_\theta(s,t;\omega)$ leads to the usual Fisher information with respect to θ (see, e.g., Efron (1977), Liang (1983), Zhu and Reid (1994)), defined by

$$\boldsymbol{I}^{(S,T)}(\theta;\omega) = \boldsymbol{I}_{\theta\theta} - \boldsymbol{I}_{\theta\phi}\boldsymbol{I}_{\phi\phi}^{-1}\boldsymbol{I}_{\phi\theta}, \tag{4.4}$$

when the Fisher information matrix is partitioned as

$$\boldsymbol{I}^{(S,T)}(\omega) = \begin{bmatrix} \boldsymbol{I}_{\theta\theta} & \boldsymbol{I}_{\theta\phi} \\ \boldsymbol{I}_{\phi\theta} & \boldsymbol{I}_{\phi\phi} \end{bmatrix}.$$

Of course the expression (4.4) is valid for the case of positive definite $\boldsymbol{I}_{\phi\phi}$; for a more general expression see Bhapkar (1990). In any case we have

$$\boldsymbol{I}_G^{(S,T)}(\theta;\omega) \leq \boldsymbol{I}^{(S,T)}(\theta;\omega) \tag{4.5}$$

as proved in Bhapkar (1989).

It turns out that for distributions in $\mathcal{M}_1$ we have

$$I_G^{(S,T)}(\theta;\omega) = I^{(S,T)}(\theta;\omega) = I^{(S)}(\theta), \tag{4.6}$$

while for distributions in $\mathcal{M}_2$ we have

$$I_G^{(S,T)}(\theta;\omega) = I^{(S)}(\theta) \leq I^{(S,T)}(\theta;\omega). \tag{4.7}$$

The properties (4.6) and (4.7) have been used to claim (Bhapkar (1990, 1997)) the partial sufficiency property of S, in classes $\mathcal{M}_1$ and $\mathcal{M}_2$, on the basis of generalized Fisher information with respect to θ.

Another justification for such a claim is provided by the fact (see Bhapkar (1997)) that in classes $\mathcal{M}_1$ and $\mathcal{M}_2$ the marginal score function of S, viz. $l_\theta(s;\theta) = \partial \log f(s;\theta)/\partial\theta$ happens to be the optimal unbiased estimating function for θ. Note that this optimality holds in the class of all regular unbiased estimating functions based on S and T, and hence more generally in the class of all regular unbiased estimating fuctions.

Now Theorems 2.1 and 2.2 in this paper have provided a further justification for this claim, especially in the estimation context.

Below now are given some examples of models where such p-sufficient statistics exist. First, refer to Examples 2–5 of Basu (1977) for illustrations of S-sufficient statistics, i.e. statistics wich satisfy $\mathcal{M}_1$ and $\delta = \phi$ in condition (iii). As an example of situation where $\delta \neq \phi$ we have the following example.

Example 4.1. Let $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ be a random sample from the p-variate normal distribution with mean $\boldsymbol{\mu}$ and known covariance matrix Σ. If $\boldsymbol{\mu}' = (\boldsymbol{\mu}_1', \boldsymbol{\mu}_2')$ and $\boldsymbol{\theta} = \boldsymbol{\mu}_1$, then $\boldsymbol{S} = \bar{\boldsymbol{X}}_1$ is p-sufficient, since the conditional distribution of $\boldsymbol{T} = \bar{\boldsymbol{X}}_2$ depends on $\boldsymbol{\omega} = \boldsymbol{\mu}$ only through $\boldsymbol{\delta} = \boldsymbol{\mu}_2 - \Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{\mu}_1$.

We note here that the condition $\mathcal{M}_1$ is satisfied. Furthermore $\boldsymbol{S} = \bar{\boldsymbol{X}}_1$ is not necessarily Fraser-sufficient (i.e. S-sufficient in the terminology of Basu (1977)) especially when $\boldsymbol{\mu}_2$ is restricted.

More generally, when the positive difinite covariance matrix Σ is unknown, and θ is the set of elements of both $\boldsymbol{\mu}_1$ and Σ_{11}, we have p-sufficiency of S, which is the set of elements of both $\bar{\boldsymbol{X}}_1$ and $\boldsymbol{A}_{11}$, when $n \geq p-1$, $\boldsymbol{A} = \sum_{i=1}^n (\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})'$ and $\boldsymbol{A}$ is partitioned as

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix}.$$

These examples, and especially example 4.1, bring out the significance of Theorem 2.1. It is enough for estimation of $w(\boldsymbol{\theta})$ to confine attention to estimators based on $\boldsymbol{X}_1$, which has distribution depending on $\boldsymbol{\theta}$ alone, provided the remaining part of $\boldsymbol{X}$, say $\boldsymbol{X}_2$, has no additional information with respect to $\boldsymbol{\theta}$ (in the sense that (4.6) holds with $S = \boldsymbol{X}_1$ and $T = \boldsymbol{X}_2$) in view of the fact that the conditional distribution of $\boldsymbol{X}_2$, given $\boldsymbol{X}_1$, depends only on δ which satisfies condition (iii) in $\mathcal{M}_1$.

As examples of models where conditions $\mathcal{M}_2$ are satisfied, we consider $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$, a random sample of size n from normal, inverse Gaussian or gamma distributions as listed in Table I of Yanagimoto and Yamamoto (1993).

Note however that Yanagimoto and Yamamoto assert only that the marginal score function of S is optimal in the class of regular estimating functions based on S alone, while the assertion in Bhapkar (1997) is stronger in the sense that it is optimal in the class of <u>all</u> regular unbiased estimating functions. Furthermore, we have property (4.7) and, also, Theorem 2.2.

It is interesting to note with respect to relation (4.7) the following details below for the normal distribution.

Example 4.2 Let $X = (X_1, \dots, X_n)$ be a random sample from normal distribution with variance $\theta = \sigma^2$ and mean ϕ. For $S = \sum(X_i - \bar{X})^2$, $T = \bar{X}$ we have p-sufficiency of S with respect to σ^2 (for $n \geq 2$) since conditions $\mathcal{M}_2$ are met. We have then

$$I_G^{(S,T)}(\theta; \omega) = I^{(S)}(\theta) = \frac{n-1}{2\theta^2}.$$

In this case it can be shown that

$$I^{(S,T)}(\theta; \omega) = I^{(X)}(\theta; \omega) = \frac{n}{2\theta^2};$$

another interesting feature to note is that for the case $n = 1$ we have $I(\theta; \omega) = 1/2\sigma^2$, while $I_G(\theta; \omega) = 0$.

5. Remarks

For the class of regular distributions meeting conditions $\mathcal{M}_1$ and/or $\mathcal{M}_2$, S may be considered p-sufficient with respect to θ in view of the justifications pointed out in the previous section. More generally, when data X provide sufficient statistic (S, T) with respect to ω, where (S, T) meets conditions $\mathcal{M}_1$ and/or $\mathcal{M}_2$, then S can be claimed to be p-sufficient in view of (4.6), (4.7) and the equality

$$I_G^{(X)}(\theta; \omega) = I_G^{(S,T)}(\theta; \omega), \tag{5.1}$$

which is obviously anticipated (see Bhapkar (1991) for a formal proof).

It may be noted here that the condition (iv) in $\mathcal{M}_2$ is essential. Although Lloyd (1987) had claimed that conditions $\mathcal{M}_2^*$ (i.e. $\mathcal{M}_2$ without (iv)) would lead to optimality of marginal score function of S as an optimal unbiased estimating function for θ, a counter-example was provided in Bhapkar (1995). See Bhapkar (1997) for discussion on this matter.

Conditions (iii) and (iv) in $\mathcal{M}_2$ may be represented in an equivalent form, in view of condition (iv), with (iii) replaced by (iii)$'$ the family of marginal probability distributions $\{P_\omega^{(T)} : \phi \in \Phi\}$ of T is complete.

Now a statistic T, which satisfies (iii)$'$ has been termed *ancillary* with respect to θ by Godambe (1980) (or p-ancillary in the complete sense in Bhapkar (1989)) *provided* the conditional distribution, given t, depends on ω only through θ. We note that this qualifying condition is met in $\mathcal{M}_2$ in view of conditions (i) and (iv). Thus, under conditions $\mathcal{M}_2$ we have not only S p-sufficient with respect to θ (according to our definition in Section 4), but we also have T p-ancillary with respect to θ. However, we note that the condition (iv) in $\mathcal{M}_2$ plays a crucial role here too. Without (iv), neither (iii) nor (iii)$'$, in conjunction with (i) and (ii) in $\mathcal{M}_2$, would necessarily lead to either p-sufficiency of S or p-ancillarity of T.

In the context of Example 4.2, Theorem 2.2 adds considerable force to the claim that $S = \sum(X_i - \bar{X})^2$ be considered p-sufficient for $\theta = \sigma^2$, when it is considered in conjunction with the property

$$I_G^{(X)}(\theta; \omega) = I^{(S)}(\theta) \tag{5.2}$$

and, also, the property of optimality of the marginal score function of S as a RUEF for θ.

It may also be noted here that S has also been shown to be G-sufficient, and also M-sufficient, by Barndorff-Nielsen (1978). In this light the facts mentioned in Section 4 provide a strong justification for considering S to be p-sufficient with respect to $\theta = \sigma^2$. While discussing Fisher's fiducial argument Seidenfeld (1992) had posed the question, "... in what sense S can be considered to be sufficient with respecct to σ^2 in the absence of any knowledge concerning μ..." The justification in this paper does provide one affirmative answer to this question.

Appendix

For the lemmas that are proved here, assume the regularity conditions $\boldsymbol{R}$, given below, for the joint distribution $P_\omega^{(S,T)}$ of (S,T), with the pdf $p(s,t;\omega)$ with respect to measure μ.

$\boldsymbol{R}$: (i) The marginal distribution $P_\theta^{(S)}$ of S has pdf $f(s;\theta)$ with respect to measure ν.

(ii) The relation $\int f(s;\theta)d\nu(s) = 1$ can be differentiated twice under the integral sign with respect to elements of θ,

(iii) The Fisher information matrix $\boldsymbol{I}^{(S)}(\theta)$ of S is positive definite.

(iv) The conditional distribution $P_\omega^{(T|S)}$ of T, given s, has pdf $h(t \mid s;\omega)$ with respect to measure $d\eta_s$ a.e. (ν).

(v) The relation $\int h(t \mid s;\omega)d\eta_s(t) = 1$ can be differentiated with respect to elements of ω twice under the integral sign.

(vi) The relation $\int g(s,t;\theta)p(s,t;\omega)d\mu(s,t) = 0$ can be differentiated under the integral sign with respect to elements of ω for any real-valued function g such that $E_\omega g = 0$ and $E_\omega g^2 < \infty$, where g is differentiable with respect to elements of θ.

Condition (vi) is essentially the regularity condition on the unbiased estimating function g for θ; see Godambe and Thompson (1974).

Let $\Pi = \Pi_1 \times \Pi_2$ be a probability measure over Ω, which is the product of measures Π_1 and Π_2 over Θ and Φ, respectively.

Consider the Hilbert space $\mathcal{C}$ of real-valued functions $c = c(s,t;\omega)$ which satisfy

$$E(c) \equiv \int c(s,t;\omega)dP_\omega^{(S,T)}(s,t)d\Pi(\omega) = 0$$

and

$$E(c^2) \equiv \int c^2(s,t;\omega)dP_\omega^{(S,T)}(s,t)d\Pi(\omega) < \infty;$$

then the norm of c is $[E(c^2)]^{1/2}$.

Define now

$$\mathcal{G} = \{g = g(s,t;\theta) : g(s,t;\theta) = c(s,t;\theta,\phi) \text{ a.e. } (\mu \times \Pi) \text{ for some } c \in \mathcal{C}\}$$
$$\mathcal{G}(S) = \{g = g(s;\theta) : g(s;\theta) = g^*(s,t;\theta) \text{ a.e. } (\mu \times \Pi_1) \text{ for some } g^* \in \mathcal{G}\}$$
$$\mathcal{G}_1(S) = \{g_1 = g_1(s;\theta) : g_1(s;\theta) = \boldsymbol{c}'(\theta)\boldsymbol{l}_\theta(s;\theta) \text{ and } g_1 \in \mathcal{G}(S)\}.$$

The following two lemmas are easy to prove.

LEMMA A.1. *Let $\mathcal{G}_0(S)$ be the orthogonal complement of $\mathcal{G}_1(S)$ in $\mathcal{G}(S)$. Define*

$$\mathcal{G}^*(S,T) = \Big\{g^* = g^*(s,t;\theta) : g^* \in \mathcal{G} \text{ and}$$
$$k^*(s;\theta) \equiv \int g^*(s,t;\theta)dP_\omega^{(T|s)}(t)d\Pi_2(\phi) \in \mathcal{G}_1(S) \text{ a.e. } (\nu \times \Pi_1)\Big\}.$$

Then $\mathcal{G} = \mathcal{G}_0(S) \oplus \mathcal{G}^*(S,T)$.

LEMMA A.2. *Let* $\mathcal{G}(S,T)$ *be defined by*

$$\mathcal{G}(S,T) = \{g = g(s,t;\theta) : g \in \mathcal{G} \text{ and } g = \boldsymbol{b}'(s,t;\theta)\boldsymbol{l}_\theta(s;\theta) \text{ such that } E_\omega[\boldsymbol{b}(s,T;\theta) \mid s] = \boldsymbol{a}(\omega) \text{ a.e. } (\nu \times \Pi)\}.$$

Then (i) $\mathcal{G}(S,T) \subset \mathcal{G}^*(S,T)$, *and* (ii) *if* $g^* \in \mathcal{G}^*(S,T)$ *and* $g^* \perp \mathcal{G}(S,T)$, *then* $k^*(s;\theta) = 0$ *a.e.* $(\nu \times \Pi_1)$.

Then we have the following:

LEMMA A.3. *If* Π *is the one-point distribution at* ω, *then* $\mathcal{G}^*(S,T) = \mathcal{G}(S,T)$.

PROOF. If g^* is defined as in Lemma A.2, then $k^*(s;\omega) = 0$, where $k^*(s;\omega) = E_\omega[g(s,T;\theta) \mid s]$, in view of Lemma A.2. Thus, $g^* \in \mathcal{G}(T \mid S)$, where

$$\mathcal{G}(T \mid S) = \{g = g(s,t;\theta) : g \in \mathcal{G}, E_\omega[g(s,T;\theta) \mid s] = 0, \text{ a.e. } (\nu)\}.$$

However $\mathcal{G}(T \mid S) \subset \mathcal{G}(S,T)$, so that $g^* \perp \mathcal{G}(T \mid S)$, since $g^* \perp \mathcal{G}(S,T)$. Hence $g^* = 0$ and the lemma follows.

REFERENCES

Barnard, G. A. (1963). Some logical aspects of the fiducial argument, *J. Roy. Statist. Soc. Ser. B*, **25**, 111–115.

Barndorff-Nielson, O. (1978). *Information and Exponential Families in Statistical Theory*, Wiley New York.

Basu, D. (1977). On the elimination of nuisance parameters, *J. Amer. Statist. Assoc.*, **72**, 355–366.

Basu, D. (1978). On partial sufficiency: A review, *J. Statist. Plann. Inference*, **2**, 1–13.

Bhapkar, V. P. (1989). Conditioning on ancillary statistics and loss of information in the presence of nuisance parameters, *J. Statist. Plann. Inference*, **21**, 139–160.

Bhapkar, V. P. (1990). Conditioning, marginalization and Fisher information functions, *Probability Statistics and Design of Experiments* (ed. R. R. Bahadur), 123–136, Wiley Eastern Limited, New Delhi.

Bhapkar, V. P. (1991). Loss of information in the presense of nuisance parameters and partial sufficiency, *J. Statist. Plann. Inference*, **28**, 185–203.

Bhapkar, V. P. (1995). Completeness and optimality of marginal likelihood estimating equations, *Comm. Statist. Theory Methods*, **24**, 945–952.

Bhapkar, V. P. (1997). Estimating functions, partial sufficiency and Q-sufficiency in the presence of nuisance parameters, *Selected Proceedings of the International Symposium on Estimating Functions*, IMS Lecture Notes-Monograph Series, Vol. **32**, 83–104, Hayward, California.

Bhapkar, V. P. and Srinivasan, C. (1994). On Fisher information inequalities in the presence of nuisance parameters, *Ann. Inst. Statist. Math.*, **46**, 593–604.

Bickel, P. J. and Doksum, K. A. (1977). *Mathematical Statistics*, Holden-Day, Inc., San Francisco.

Efron, B. (1977). The efficiency of Cox's likelihood function for censored data, *J. Amer. Statist. Assoc.*, **72**, 557–565.

Fraser, D. A. S. (1956). Sufficient statistics with nuisance parameters, *Ann. Math. Statist.*, **27**, 838–842.

Godambe, V. P. (1980). On sufficiency and ancillarity in the presence of nuisance parameter, *Biometrika*, **67**, 155–162.

Godambe, V. P. (1984). On ancillarity and Fisher information in the presence of a nuisance parameter, *Biometrika*, **71**, 626–629.

Godambe, V. P. and Thompson, M. E. (1974). Estimating equations in the presence of a nuisance parameter, *Ann. Statist.*, **2**, 568–571.

Liang, K. (1983). On information and ancillarity in the presence of a nuisance parameter, *Biometrika*, **70**, 607–612.

Lloyd, C. J. (1987). Optimality of marginal likelihood estimating equations, *Comm. Statist. Theory Methods*, **16**, 1733–1741.

Rémon, M. (1984). On a concept of partial sufficiency: *L*-sufficiency, *International Statistical Review*, **52**, 127–135.

Seidenfeld, T. (1992). R. A. Fisher's fiducial argument and Bayes' theorem, *Statist. Sci.*, **7**, 358–368.

Yanagimoto, T. and Yamamoto, E. (1993). A criterion of sensitivity of an estimating function, *Comm. Statist. Theory methods.*, **22**, 451–460.

Zhu, Y. and Reid, N. (1994). Information, ancillarity, and sufficiency in the presence of nuisance parameters, *Canad J. Statist.*, **22**, 111–123.