This is to certify that the

dissertation entitled

Sufficiency In The Presence
Of Nuisance Parameters

presented by

Nupun Andhivarothai

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Statistics

R.V Ramamoorthi
Major professor

Date Nov. 5, 1990

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.

DATE DUE	DATE DUE	DATE DUE

SUFFICIENCY IN THE PRESENCE OF NUISANCE PARAMETERS

By

Nupun Andhivarothai

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Statistics and Probability

1990

ABSTRACT

SUFFICIENCY IN THE PRESENCE OF NUISANCE PARAMETERS

by

Nupun Andhivarothai

This dissertation is devoted to the study of the concept of Sufficiency in the presence of nuisance parameters. We mainly investigate the notion of Partial Sufficiency proposed by Hajek in 1965. Decision theoretic aspects of Hajek's definition is investigated and we prove a converse to a Rao-Blackwell type theorem in the context of partial sufficiency. We next extend the concept to one experiment being partially sufficient to another experiment. Finally, we give some examples and applications to illustrate the concepts studied.

To my parents

and

my sisters, Jinnarat and Sawanee

Acknowledgements

I wish to express my sincere thanks to Professor R.V. Ramamoorthi for his guidance, patience and encouragement during the preparation of this dissertation.

I would like to thank Professor Habib Salehi and Professor Joseph Gardiner for serving in my committee. Special thanks are due to Professor Dennis Gilliland for not only serving in my committee but also for introducing me to statistical consulting,which has greatly enhanced my carreer opportunities.

I would like to thank the Department of Statistics and Probability and also the Department of Family Medicine for the financial support during my graduate studies at Michigan State University. Finally, I would like to express my deep appreciation to my parents and my sisters for their patience, encouragement and support.

Contents

Chapter 1

Introduction and Summary

Let X be a random variable distributed as $P_{\theta,\sigma}$ where θ and σ are unknown parameters. Classically, a reduction of X is achieved via a sufficient statistic for θ, σ. That this reduction does not entail any loss of information is established by the Rao–Blackwell theorem which shows that for any decision problem, the decision rules based on the sufficient statistic form an essential complete class. A variety of converses to the Rao–Blackwell theorem also show that the Fisher–Neyman definition of sufficiency is appropriate if we are looking for a reduction of X that would be as effective as X for all decision problems. However, it very often happens that we are interested in only a subset of the set of all decision problems. A typical case would be when we are interested in making inferences on the parameter θ and are indifferent to the value of σ. In such a situation θ would be referred to as the parameter of interest and σ as the nuisance parameter. There have been many attempts at defining a sufficient statistic of part of the parameters, in the case just mentioned, a sufficient statistic for the parameter θ in the presence of a nuisance parameter σ [Neyman and

Pearson (1936), Fraser (1965), Kolmogorov (1942), Hájek (1965)].

In this study, we extend the concept of "partial sufficiency" introduced by Hájek (1965) and give a result which is a converse to a Rao– Blackwell type theorem.

We next turn our attention to the problem of comparison of two experiments. Let $\mathcal{E}$ and $\mathcal{F}$ be two experiments parameterized by (θ, σ). Bohnenblust, Shapley and Sherman defined the notion of $\mathcal{E}$ being more informative that $\mathcal{F}$ in terms of risk functions obtainable in the experiments. Blackwell extended the concept of a sufficient statistic and defined $\mathcal{E}$ being sufficient for $\mathcal{F}$ in terms of the existence of Markov kernels. Blackwell then showed that "more informative" and "sufficient" are equivalent.

Blackwell's theory involves sufficiency for both parameters (θ, σ), more specifically, it needs the consideration of a loss function that would depend on both θ and σ. However, when σ is a nuisance parameter, it seems appropriate to consider a loss function that depends only on θ. These considerations motivate our study of "partial sufficiency" of experiment in Chapter 2.

We extend the concept of partial sufficiency to two experiments in Chapter 3. The notions of $\mathcal{E}$ being more informative than $\mathcal{F}$, say, for θ and $\mathcal{E}$ being partially sufficient for $\mathcal{F}$ are introduced. The equivalence of these two concepts is proved. A criterion in determining $\mathcal{E}$ being partially sufficient for $\mathcal{F}$ in terms of sufficiency of reduced experiments is also established.

To conclude this study, in Chapter 4 we give some examples to illustrate the concept of partially sufficient statistic and some application of results in the earlier chapters.

Chapter 2

Partial Sufficiency

This chapter is devoted to the study of a notion of partial sufficiency introduced by Hájek in 1965. We first establish the notation and preliminaries, and then prove the main result which is a converse to a Rao–Blackwell type theorem.

2.1 Notation and Preliminaries

A statistical experiment is the triplet $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ where $\mathcal{X}$ is a set, $\mathcal{A}$ is a σ–algebra of subsets of $\mathcal{X}$ and $\mathcal{P}$ is a family of probability measures on $(\mathcal{X}, \mathcal{A})$. $\mathcal{P}$ will be endowed with the σ–algebra $\mathcal{C}$, which is the smallest σ–algebra generated by $P \mapsto P(A)$, $A \in \mathcal{A}$. Subsets of $\mathcal{P}$ will be equipped with the relative σ–algebra. We will assume that the family $\mathcal{P}$ is parameterized by $\Theta \times \Sigma$, that is, there is a 1-1 function $(\theta, \sigma) \mapsto P_{\theta,\sigma}$ from $\Theta \times \Sigma$ onto $\mathcal{P}$.

Thus for us an experiment is given by $(\mathcal{X}, \mathcal{A}, P_{\theta,\sigma} : (\theta, \sigma) \in \Theta \times \Sigma)$ where $(\mathcal{X}, \mathcal{A})$ is the sample space and $\Theta \times \Sigma$ would be referred to as the parameter space. A decision problem consists of a measurable space – "Action Space" $(A, \underset{\sim}{A})$ and a "loss function"

$L(\theta, \sigma, a)$ from $\Theta \times \Sigma \times A \to \Re$, which is a measurable in (θ, σ, a). By a decision rule δ, we mean a function $\delta : \mathcal{X} \times \underset{\sim}{\mathcal{A}} \to [0, 1]$ such that

i) For all $A \in \underset{\sim}{\mathcal{A}}$, $\delta(x, A)$ is, as a function of x, $\underset{\sim}{\mathcal{A}}$ measurable.

ii) For each $x \in \mathcal{X}$, $\delta(x, \cdot)$ is a probability measure on $\underset{\sim}{\mathcal{A}}$.

If a decision rule δ in i) above is measurable with respect to a sub σ–algebra $\mathcal{B}$ of $\mathcal{A}$, then we shall refer to δ as a $\mathcal{B}$ measurable decision rule.

Denote by $\mathcal{L}_A$ the set of all bounded loss functions defined on $\Theta \times \Sigma \times A$. If $L \in \mathcal{L}_A$ and δ is a decision rule, the "risk function" of δ (with respect to L) is the function on $\Theta \times \Sigma$ defined by

$$R_L(\theta, \sigma, \delta) = \int_{\mathcal{X}} \int_A L(\theta, \sigma, a)\delta(x, da)dP_{\theta,\sigma}(x).$$

We shall throughout treat θ as the "parameter of interest" and σ as the nuisance parameter. This treatment may be formalized in one way by considering only the following subset of $\mathcal{L}_A$, $\mathcal{L}_A^o = \{L \in \mathcal{L}_A : L$ depends on (θ, σ) through θ only$\}$.

Let $\mathcal{B}$ be a sub σ–algebra of $\mathcal{A}$. If P is a probability measure on $(\mathcal{X}, \mathcal{A})$, then for any bounded $\mathcal{A}$ measurable function f, $E_P(f|\mathcal{B})$ will denote any version of the conditional expectation of f given $\mathcal{B}$, under the measure P. If $\mathcal{P}_0$ is a family of probability measures on $(\mathcal{X}, \mathcal{A})$ then $\mathcal{B}$ is called sufficient for $\mathcal{P}_0$ if for any bounded $\mathcal{A}$ measurable function f there exists a $\mathcal{B}$ measurable function g such that $g = E_P(f|\mathcal{B})$ $[P]$ for all $P \in \mathcal{P}_0$.

We will assume throughout that

i) $\mathcal{X}$ is a Borel subset of a complete seperable metric space and $\mathcal{A}$ is the relativized Borel σ–algebra.

ii) $\{P_{\theta,\sigma} : \theta \in \Theta, \sigma \in \Sigma\}$ are all mutually absolutely continuous.

As mentioned earlier θ is the parameter of interest and σ is the nuisance parameter.

2.2 Partial Sufficiency

Definition 2.1 (Hájek (1965)) *$\mathcal{B}$ is said to be H–sufficient for θ in $\{P_{\theta,\sigma} : \theta \in \Theta, \sigma \in \Sigma\}$ if*

i) *$\mathcal{B}$ is θ–oriented, that is, for each θ, $\mathcal{B}$ is ancillary for the family $\overline{\mathcal{P}}_\theta = \{P_{\theta,\sigma} : \sigma \in \Sigma\}$, i.e. $P_{\theta,\sigma_1}(B) = P_{\theta,\sigma_2}(B)$ for σ_1, σ_2 in Σ, and $B \in \mathcal{B}$.*

ii) *For each θ, there exists a probability measure ξ_θ on $\overline{\mathcal{P}}_\theta$ such that $\mathcal{B}$ is sufficient for $\{P_{\xi_\theta} : \theta \in \Theta\}$, where P_{ξ_θ} is the marginal probability measure on $(\mathcal{X}, \mathcal{A})$ defined by*

$$P_{\xi_\theta}(A) = \int P_{\theta,\sigma}(A) d\xi_\theta(\sigma).$$

Definition 2.2 *$\mathcal{B}$ is said to be partially sufficient for θ, if $\mathcal{B}$ contains a H–sufficient σ–algebra.*

Theorem 2.1 *Let $\mathcal{B}$ be partially sufficient for θ in $\{\mathcal{X}, \mathcal{A}, P_{\theta,\sigma} : (\theta, \sigma) \in \Theta \times \Sigma\}$. Then given any decision problem $(A, \underline{\mathcal{A}})$ and an $\mathcal{A}$–measurable decision rule δ, there exists a $\mathcal{B}$–measurable decision rule δ^* such that for all $\theta \in \Theta$, and $\sigma \in \Sigma$,*

$$\int \delta^*(x, E) dP_{\theta,\sigma}(x) = \iint \delta(x, E) dP_{\theta,\sigma}(x) d\xi_\theta(\sigma).$$

Proof. Since $\mathcal{B}$ is partially sufficient for θ, there is a σ algebra $\mathcal{B}_0 \subset \mathcal{B}$ which is H–sufficient for θ. $\mathcal{B}_0$ is sufficient for $\{P_{\xi_\theta} : \theta \in \Theta\}$ and since $(\mathcal{X}, \mathcal{A})$ is standard Borel there exists an omnibus version of the conditional probability given $\mathcal{B}_0$. That is there is a function Q from $\mathcal{X} \times \mathcal{A} \mapsto [0,1]$ such that

(a) $Q(x, A)$ is $\mathcal{B}_0$–measurable for all $A \in \mathcal{A}$

(b) $Q(x, \cdot)$ is a probability measure on $(\mathcal{X}, \mathcal{A})$ for all x

and

(c) $\int\int Q(x, A) dP_{\theta,\sigma}(x) d\xi_\theta(\sigma) = P_{\xi_\theta}(A)$ for all $A \in \mathcal{A}$.

Given any decision problem $(A, \underset{\sim}{\mathcal{A}})$, and a decision rule δ, define δ^* by

$$\delta^*(x, E) = \int \delta(y, E) Q(x, dy), E \in \underset{\sim}{\mathcal{A}}.$$

By (a) $\delta^*(x, E)$ is a $\mathcal{B}_0$–measurable decision rule. Further, since $\mathcal{B}_0$ is θ–oriented $\int \delta^*(x, E) dP_{\theta,\sigma}(x)$ is constant in σ and hence for each $\theta \in \Theta$

$$\int \delta^*(x, E) dP_{\theta,\sigma}(x) = \int\int \delta^*(x, E) dP_{\theta,\sigma}(x) d\xi_\theta(\sigma) = \int\int \delta(x, E) dP_{\theta,\sigma}(x) d\xi_\theta(\sigma).$$

□

The next theorem is an analogue of the Rao–Blackwell theorem in the context of partial sufficiency and appears as Theorem 2.2 in Hájek (1965).

Theorem 2.2 (Hájek (1965)) *Let $\mathcal{E} = \{\mathcal{X}, \mathcal{A}, P_{\theta,\sigma} : (\theta, \sigma) \in \Theta \times \Sigma\}$ be an experiment and $\mathcal{B}$ be partially sufficient for θ in $\mathcal{E}$. Let $(A, \underset{\sim}{\mathcal{A}})$ be a decision space. Then given any decision rule δ, there exists a $\mathcal{B}$–measurable decision rule δ^* such that for all loss functions $L \in \mathcal{L}_A^o$, we have for each θ*

$$\sup_{\sigma \in \Sigma} R_L(\theta, \sigma, \delta^*) \leq \sup_{\sigma \in \Sigma} R_L(\theta, \sigma, \delta). \tag{2.1}$$

Proof. Let δ be any decision rule and δ^* be any $\mathcal{B}$–measurable decision rule satisfying the conclusion of Theorem 2.1. We then have

$$\int_{\mathcal{X}}\int_A f(a)\delta^*(x,da)dP_{\theta,\sigma}(x) \;=\; \int_\Sigma\int_{\mathcal{X}}\int_A f(a)\delta(x,da)dP_{\theta,\sigma}(x)d\xi_\theta(\sigma)$$

whenever f is of the form $I_E(a)$, $E \in \underset{\sim}{A}$. A standard induction argument via simple function yields

$$R_L(\theta,\sigma,\delta^*) = \int_\Sigma R_L(\theta,\sigma,\delta)d\xi_\theta(\sigma).$$

so that

$$\sup_{\sigma\in\Sigma} R_L(\theta,\sigma,\delta^*) = \int_\Sigma R_L(\theta,\sigma,\delta)d\xi_\theta(\sigma) \le \sup_{\sigma\in\Sigma} R_L(\theta,\sigma,\delta).$$

□

2.3 Main Theorem

We now move to a converse of Theorem 2.1 which is the main theorem in this chapter.

Theorem 2.3 *Let $\mathcal{E} = \{\mathcal{X},\mathcal{A},P_{\theta,\sigma} : (\theta,\sigma)\in\Theta\times\Sigma\}$ be an experiment and let $\mathcal{B}$ be a sub σ–algebra of $\mathcal{A}$. If $\mathcal{B}$ satisifies* **Condition A,**

Conditon A: *For each θ, there exists a probability measure ξ_θ on Σ such that, for any decision space $(A,\underset{\sim}{A})$ given any decision rule δ, there exists a $\mathcal{B}$–measurable decision rule δ^* satisfying for all loss functions $L \in \mathcal{L}_A^o$*

$$R_L(\theta,\sigma,\delta^*) = \int R_L(\theta,\sigma,\delta)d\xi_\theta(\sigma).$$

Then $\mathcal{B}$ is partially sufficient for θ in $\mathcal{E}$.

Proof. Choose $(A, \underline{A})$ to be $(\mathcal{X}, \mathcal{A})$ and for each $A \in \mathcal{A}$, let $L_A(\theta, \sigma, a) = I_A(a)$, where $I_A(\cdot)$ denotes the indicator function, and set $\delta(x, E) = I_E(x)$. We then have from **Condition A** that there exists a $\mathcal{B}$–measurable decision rule δ^* such that

$$\int \delta^*(x, A) dP_{\theta,\sigma}(x) = \int P_{\theta,\sigma}(A) d\xi_\theta(\sigma)$$

for all $A \in \mathcal{A}$.

Since the left hand side in the above equation does not depend on σ, we in fact have

$$P_{\xi_\theta}(A) \;=\; \int \delta^*(x, A) dP_{\xi_\theta}(x).$$

Now define

$$\begin{aligned}
\delta_1^* &= \delta^* \\
\delta_2^*(x, A) &= \int \delta_1^*(y, A)\delta(x, dy) \\
\vdots &= \quad \vdots \\
\delta_n^*(x, A) &= \int \delta_{n-1}^*(y, A)\delta(x, dy).
\end{aligned}$$

An easy argument shows that

$$P_{\xi_\theta}(A) \;=\; \int \delta_n^*(x, A) dP_{\xi_\theta}(x) \tag{2.2}$$

for all n and $A \in \mathcal{A}$.

Let us define

$$\begin{aligned}
\delta_0^*(x, A) &= \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \delta_k^*(x, A) \qquad \text{when it exists} \\
&= P(A) \qquad\qquad\qquad\quad\;\; \text{otherwise,}
\end{aligned}$$

where P is an arbitrary probability measure on $(\mathcal{X}, \mathcal{A})$.

By Hopf's ergodic theorem in Neveu (1965), for each $\theta \in \Theta$

$$\delta_0^*(x, A) = E_{P_{\xi_\theta}}(I_A | \mathcal{B}_\theta) \text{ a.e. } [P_{\xi_\theta}]$$

where

$$\mathcal{B}_\theta = \{B : \delta_0^*(x, B) = I_B \quad [P_{\xi_\theta}]\}.$$

If we set $\mathcal{B}_0 = \{B : \delta_0^*(x, B) = I_B \; [P_{\xi_\theta}]$ for all $\theta \in \Theta\}$ then δ_0^* is $\mathcal{B}_0$ measurable, we have

$$\delta_0^*(x, A) \quad = \quad E_{P_{\xi_\theta}}(I_A | \mathcal{B}_0).$$

This shows that $\mathcal{B}_0$ is sufficient for $\{P_{\xi_\theta} : \theta \in \Theta\}$.

We next note that for $B \in \mathcal{B}_0$,

$$\delta_0^*(x, B) = I_B \quad [P_{\xi_\theta}]$$

and from the assumption on i) page 4, we have $\delta_0^*(x, B) = I_B \; [P_{\theta,\sigma}]$ for all σ so that

$$\int \delta_0^*(x, B) dP_{\theta,\sigma}(x) = P_{\theta,\sigma}(B).$$

On the other hand $\mathcal{B}_0$ measurability of δ_0^* yields

$$\int \delta_0^*(x, B) dP_{\theta,\sigma}(x) = P_{\xi_\theta}(B).$$

So that $P_{\theta,\sigma}(B)$ is constant in σ, thereby establishing that $\mathcal{B}_0$ is θ–oreinted.

This shows that $\mathcal{B}_0$ is H–sufficient and since $\mathcal{B}_0 \subset \mathcal{B}$ we have that $\mathcal{B}$ is partially sufficient for θ. □

Remark: We feel that Theorem 2.3 while interesting is still rather weak. This is because given a decision rule δ, we require a decision rule δ^* which would be as good as δ for all loss function in $\mathcal{L}^o_A$. A more reasonable condition would be to allow δ^* to depend on the loss function L and prove **Condition A.** However, we are unable to establish such a result.

2.4 Invariance

In the last section we studied the notion of partial sufficiency that was proposed by Hájek in 1965. In the same paper, he demonstrated that in situations where the nuisance parameter is generated by a group of transfomations on the sample space. The maximal invariant is partially sufficient. In this section we present Hájek result, since it provides a wide class of examples of paritally sufficient statistics. More specific examples will be given in a later chapter.

Let X be a random variable with a probability distribution P_θ, θ is the parameter of interest, $(\mathcal{X}, \mathcal{A})$ is a sample space of X. Suppose $P_\theta \in \mathcal{P} = \{P_\theta : \theta \in \Theta\}$, $\mathcal{P}$ is the family of probability distribution which is dominated by a σ–finite measure μ. Let $G = \{g\}$ be a group of 1-1 transformation from $\mathcal{X}$ onto $\mathcal{X}$. Let $A \in \mathcal{A}$, put

$$P_{\theta,g}(A) = P_\theta(g^{-1}A) \tag{2.3}$$

We will assume the following conditions: **Condition B:** Let $\mathcal{G}$ be a σ–algebra of subsets of G, and assume the followings:

i) Let μg be a measure such that $\mu g(A) = \mu(gA)$ and $\mu g \ll \mu$ for all $g \in G$

ii) Let $p_\theta(x, g)$ be a density of $P_{\theta,g}$ with respect to μ and that $p_\theta(x, g)$ is $\mathcal{A}\times\ \mathcal{G}$ –measurable

iii) Functions $\phi_h(g) = hg$ and $\psi_g(h) = gh$ are $\mathcal{G}$–measurable.

iv) There exists an invariant probability measure ν on $\mathcal{G}$, that is, $\nu(Bg) = \nu(gB)$ $= \nu(B)$ for all $g \in G$ and $B \in \mathcal{G}$.

We shall say that an event $A \in \mathcal{A}$ is G–invariant, if $gA = A$ for all $g \in G$. We can see that the set G–invariant events is a sub σ–algebra $\mathcal{B} \subset \mathcal{A}$, and we say that if f is $\mathcal{B}$ measurable iff $f(g(x)) = f(x)$ for $g \in G$.

Theorem 2.4 (Hájek (1965)) *Let $P_\theta \in \mathcal{P}$ and define $P_{\theta,g}$ by Equation 2.3 for each θ. Under* **Condition B**, *the sub σ–algebra $\mathcal{B}$ of G–invariant events is partially sufficient for θ.*

Proof. It is enough to show that $\mathcal{B}$ satisfies Definition 2.1

i) Since $\mathcal{B}$ is a sub σ–algebra of G–invariant events, for $B \in \mathcal{B}$

$$P_{\theta,g}(B) = P_\theta(B).$$

ii) for $A \in \mathcal{A}$, let

$$P_{\nu_\theta}(A) = \int P_{\theta,g}(A) d\nu(g)$$

we have

$$\begin{aligned} P_{\nu_\theta}(A) &= \int \left[\int_A p_\theta(x,g) d\mu \right] d\nu(g) \\ &= \int_A \int p_\theta(x,g) d\nu(g) d\mu \end{aligned}$$

with $\overline{p}_\theta(x) = \int p_\theta(x,g) d\nu(g)$

$$P_{\nu_\theta}(A) = \int_A \overline{p}_\theta(x) d\mu \qquad (2.4)$$

Let P_0 be some probability measure such that $P_\theta \ll P_0 \ll \mu$ and define $\overline{p}_0(x)$ by $\overline{p}_0(x) = \int p_0(x,g)d\nu(g)$.

$$P_{\nu_\theta}(A) \quad = \quad \int_A \frac{\overline{p}_\theta(x)}{\overline{p}_0(x)}\overline{p}_0(x)d\mu \tag{2.5}$$

It follows from Theorem 3.3 of Hájek (1965) that $\overline{p}_\theta(x)/\overline{p}_0(x)$ is $\mathcal{B}$–measurable and it is also a density of P_{ν_θ} with respect to P_{ν_0}. By Lemma 1 page 401 of Billingsley (1979), it follows that $\mathcal{B}$ is sufficient for $\{P_{\nu_\theta}\}$. □

Chapter 3

Comparison of Experiments in the Presence of Nuisance Parameters

3.1 Preliminaries

In this chapter we study the extension of the concept of partial sufficiency to two experiments. Let $\mathcal{E} = \{\mathcal{X}, \mathcal{A}, P_t : t \in T\}$ and $\mathcal{F} = \{\mathcal{Y}, \mathcal{B}, Q_t : t \in T\}$ be two experiments. Following Blackwell (1951), we define:

Definition 3.1 *$\mathcal{E}$ is more informative than $\mathcal{F}$ if for any decision problem $(A, \underline{A})$ and loss function $L(t, a)$, if given any decision rule δ in $\mathcal{F}$, there exists a decision rule δ^* in $\mathcal{E}$ such that the risk functions satisfy, for all $t \in T$*

$$R(t, \delta^*) \leq R(t, \delta).$$

Definition 3.2 *$\mathcal{E}$ is sufficient for $\mathcal{F}$, if there exists a Markov kernel M from $\mathcal{E}$ to $\mathcal{F}$ such that, for all $t \in T$ and $B \in \mathcal{B}$*

$$\int M(x, B) dP_t(x) = Q_t(B)$$

Blackwell(1951) showed that Definition 3.1 and Definition 3.2 are equivalent when T is finite. When the experiments are dominated the equivalence continues to hold, see Feldman and Ramamoorthi (1984) for a proof.

3.2 Main Results

A direct analogue of definition 3.1 in the context of "partial sufficiency", in view of the remark in Section 2.3 of Chapter 2, is not available. However, motivated by Theorem 2.1 of Chapter 2, we define the following:

Let $\mathcal{E} = \{\mathcal{X}, \mathcal{A}, P_{\theta,\sigma}; (\theta, \sigma) \in \Theta \times \Sigma\}$ and $\mathcal{F} = \{\mathcal{Y}, \mathcal{B}, Q_{\theta,\sigma}; (\theta, \sigma) \in \Theta \times \Sigma\}$ be two experiments. As before we treate θ as the parameter of interest and σ as the nuisance parameter. For an action space $(A, \underline{\mathcal{A}})$, let $\mathcal{L}_A^o$ be the class of all bounded loss functions which do not depend on σ.

Definition 3.3 *$\mathcal{E}$ is more informative than $\mathcal{F}$ for θ, if for any decision problem, there exists a probability measure μ_θ on Σ such that, given any decision rule δ in $\mathcal{F}$, there exists δ^* in $\mathcal{E}$ such that for all $L \in \mathcal{L}_A^o$*

$$R_L(\theta, \sigma, \delta^*) \leq \int R_L(\theta, \sigma, \delta) d\mu_\theta(\sigma).$$

Remark: For any $L \in \mathcal{L}$, defining

$$L^1(\theta, \sigma, a) = \sup_a L(\theta, \sigma, a) - L(\theta, \sigma, a).$$

It is easy to see that the "$\leq$" in Definition 3.3 can be replaced by "="

Definition 3.4 *$\mathcal{E}$ is partially sufficient for $\mathcal{F}$ if there exists a Markov kernel $M(\cdot,\cdot)$ from $\mathcal{E}$ to $\mathcal{F}$ and probability measures μ_θ on Σ such that*

$$\int M(x,B)dP_{\theta,\sigma}(x) = \int Q_{\theta,\sigma}(B)d\mu_\theta(\sigma),$$

for all $(\theta,\sigma) \in \Theta \times \Sigma$ and $B \in \mathcal{B}$.

Theorem 3.1 *$\mathcal{E}$ is more informative than $\mathcal{F}$ for θ iff $\mathcal{E}$ is partially sufficient for $\mathcal{F}$.*

Proof.

(i) Suppose $\mathcal{E}$ is more informative than $\mathcal{F}$ for θ. Choose $(A, \underset{\sim}{A})$ to be $(\mathcal{Y}, \mathcal{B})$ and let $\delta(y, E) = I_E(y)$, where $I_E(\cdot)$ is an indicator function. Then, from the assumption that $\mathcal{E}$ is more informative than $\mathcal{F}$ for θ, we have a decision rule δ^* in $\mathcal{E}$ such that, for all $L \in \mathcal{L}_A^o$

$$R_L(\theta,\sigma,\delta^*) = \int R_L(\theta,\sigma,\delta)d\mu_\theta(\sigma) \tag{3.1}$$

For each $B \in \mathcal{B}$, the loss function $L(\theta,\sigma,a) = I_B(a)$ is in $\mathcal{L}_A^o$, where $I_B(\cdot)$ is an indicator function. Using Equation (3.1), it is evident that δ^* satisfies

$$\int \delta^*(x,B)dP_{\theta,\sigma}(x) = \int Q_{\theta,\sigma}(B)d\mu_\theta(\sigma).$$

(ii) Suppose $\mathcal{E}$ is partially sufficient for $\mathcal{F}$.

Let M be the Markov kernel provided by the partial sufficiency of $\mathcal{E}$ to $\mathcal{F}$. Given any decision rule δ in $\mathcal{F}$, define δ^* by

$$\delta^*(x,E) = \int \delta(y,E)M(x,dy).$$

It is easily verified that δ^* satisfies

$$R_L(\theta,\sigma,\delta^*) \;=\; \int R_L(\theta,\sigma,\delta)d\mu_\theta(\sigma)$$

for all $L \in \mathcal{L}_A^o$. □

Let $\mathcal{E} = \{\mathcal{X}, \mathcal{A}, P_{\theta,\sigma}; (\theta,\sigma) \in \Theta \times \Sigma\}$ be an experiment. If $\mathcal{A}_0$ is a sub σ–algebra of $\mathcal{A}$, we will denote by $\mathcal{E}_0$ the experiment $\{\mathcal{X}, \mathcal{A}_0, P_{\theta,\sigma}; (\theta,\sigma) \in \Theta \times \Sigma\}$. Our next theorem relates partial sufficiency and sufficiency.

Theorem 3.2 *Let $\mathcal{E} = \{\mathcal{X}, \mathcal{A}, P_{\theta,\sigma}; (\theta,\sigma) \in \Theta \times \Sigma\}$ be an experiment and let the sub σ–algebra $\mathcal{A}_0$ be H–sufficient for θ.*

Similarly let $\mathcal{F} = \{\mathcal{Y}, \mathcal{B}, Q_{\theta,\sigma}; (\theta,\sigma) \in \Theta \times \Sigma\}$ be an experiment and let $\mathcal{B}_0$ be H–sufficient for θ.

Then $\mathcal{E}$ is partially sufficient for $\mathcal{F}$ iff $\mathcal{E}_0$ is sufficient for $\mathcal{F}_0$.

Proof. Suppose $\mathcal{E}$ is partially sufficient for $\mathcal{F}$.

Let δ be any decision rule in $\mathcal{F}_0$, then since δ is also a decision rule in $\mathcal{F}$, there exists a decision rule δ^* in $\mathcal{E}$ such that

$$R_L(\theta,\sigma,\delta^*) \;=\; \int R_L(\theta,\sigma,\delta)d\mu_\theta(\sigma)$$

for $L \in \mathcal{L}_A^o$.

However, since $\mathcal{B}_0$ is H–sufficient for $\{\mathcal{Y}, \mathcal{B}, Q_{\theta,\sigma} : (\theta,\sigma) \in \Theta \times \Sigma\}$ and $L \in \mathcal{L}_A^o$, $R_L(\theta,\sigma,\delta)$ is constant in σ, so that for all $(\theta,\sigma) \in \Theta \times \Sigma$

$$R_L(\theta,\sigma,\delta) \;=\; \int R_L(\theta,\sigma,\delta)d\mu_\theta(\sigma)$$

and hence we have

$$R_L(\theta,\sigma,\delta^*) \;=\; R_L(\theta,\sigma,\delta) \tag{3.2}$$

By H–sufficiency of $\mathcal{A}_0$, we have a decision rule δ_0^* in $\mathcal{E}_0$ such that

$$R_L(\theta,\sigma,\delta_0^*) = \int R_L(\theta,\sigma,\delta^*)d\xi_\theta(\sigma).$$

However by equation (3.2), we have that $R_L(\theta,\sigma,\delta^*)$ is a constant in σ so that

$$R_L(\theta,\sigma,\delta_0^*) = R_L(\theta,\sigma,\delta^*) = R_L(\theta,\sigma,\delta).$$

Conversely, suppose $\mathcal{E}_0$ is sufficient for $\mathcal{F}_0$, then

a) the Markov kernel $M_1(x,A) = I_A$ from $\mathcal{E}$ to $\mathcal{E}_0$ satisfies, for all $A \in \mathcal{A}_0$

$$\int M_1(x,A)dP_{\theta,\sigma}(x) = P_{\theta,\sigma}(A)$$

b) there exists a Markov kernel M_2 from $\mathcal{E}_0$ to $\mathcal{F}_0$ such that, for all $B \in \mathcal{B}_0$

$$\int M_2(x,B)dP_{\theta,\sigma}(x) = Q_{\theta,\sigma}(B)$$

c) there exists a Markov kernel M_3 from $\mathcal{F}_0$ to $\mathcal{F}$ such that, for all $B \in \mathcal{B}_0$

$$\int M_3(y,B)dQ_{\theta,\sigma}(y) = \int Q_{\theta,\sigma}(B)d\mu_\theta(\sigma)$$

It is easily verified that the Markov kernel

$$M(x,B) = \int M_3(y,B)M_2(x_2,dy)M_1(x_1,dx_2)$$

satisfies

$$\int M(x,B)dP_{\theta,\sigma}(x) = \int Q_{\theta,\sigma}(B)d\mu_\theta(\sigma). \tag{3.3}$$

□

In the presence of partially sufficient σ–algebras, the following theorem is of interest.

Theorem 3.3 *Let $\mathcal{E} = \{\mathcal{X}, \mathcal{A}, P_{\theta,\sigma}, (\theta,\sigma) \in \Theta \times \Sigma\}$ be an experiment and let $\mathcal{A}_0$ be H–sufficient for $\mathcal{A}$. Similarly let $\mathcal{F} = \{\mathcal{Y}, \mathcal{B}, Q_{\theta,\sigma}, (\theta,\sigma) \in \Theta \times \Sigma\}$ be an experiment and let $\mathcal{B}_0$ be H–sufficient for $\mathcal{B}$.*

Then the followings are equivalent:

i) Given any decision rule δ in $\mathcal{F}$, there exists δ^* in $\mathcal{E}$ such that for all $L \in \mathcal{L}_A^o$

$$\sup_{\sigma\in\Sigma} R_L(\theta,\sigma,\delta^*) \;\leq\; \sup_{\sigma\in\Sigma} R_L(\theta,\sigma,\delta).$$

ii) $\mathcal{E}_0$ is sufficient for $\mathcal{F}_0$.

Proof. *i)* implies *ii)*.

Let δ be any decision rule in $\mathcal{F}_0$. For $L \in \mathcal{L}$ we have for all $(\theta,\sigma) \in \Theta \times \Sigma$

$$R_L(\theta,\sigma,\delta) \;=\; \sup_{\sigma\in\Sigma} R_L(\theta,\sigma,\delta). \tag{3.4}$$

Since δ is also a decision rule in $\mathcal{F}$, we have by *i)* a decision rule δ^* in *i)* such that

$$R_L(\theta,\sigma,\delta^*) \leq \sup_{\sigma\in\Sigma} R_L(\theta,\sigma,\delta^*) \leq R_L(\theta,\sigma,\delta).$$

Since $\mathcal{A}_0$ is paritally sufficient for $\mathcal{A}$, there exists δ_0^* in $\mathcal{E}_0$ such that

$$\begin{aligned} R_L(\theta,\sigma,\delta_0^*) &= \int R_L(\theta,\sigma,\delta^*)d\xi_\theta(\sigma) \\ &\leq \sup_{\sigma\in\Sigma} R_L(\theta,\sigma,\delta^*) \\ &\leq R_L(\theta,\sigma,\delta). \end{aligned}$$

So that $\mathcal{E}_0$ is sufficient for $\mathcal{F}_0$.

ii) implies *i)*.

Let δ be any decision rule in $\mathcal{F}$. Then there exists δ_1 in $\mathcal{F}_0$ and δ^* in $\mathcal{E}_0$ such that

$$R_L(\theta,\sigma,\delta_1) = \int R_L(\theta,\sigma,\delta_1)d\mu_\theta(\sigma) \leq \sup_{\sigma\in\Sigma} R_L(\theta,\sigma,\delta)$$

and

$$R_L(\theta,\sigma,\delta^*) \;\; \leq \;\; R_L(\theta,\sigma,\delta_1)$$

so that

$$\sup_{\sigma\in\Sigma} R_L(\theta,\sigma,\delta^*) \leq \sup_{\sigma\in\Sigma} R_L(\theta,\sigma,\delta).$$

Since δ^* is in $\mathcal{E}_0$ and hence in $\mathcal{E}$ this establishes *i)*. □

Chapter 4

Examples and Applications

We give a few examples to illustrate the notion of partial sufficiency and then we show the application of the theorems in Chapter 3. Some of these examples already appear in Hájek (1965). Others are new.

4.1 Examples

In this section, examples of partial sufficiency will be given in terms of a statistic T instead of the sub σ–algebra $\mathcal{B}$, induced by T . However, before giving examples one may recognize that for T to be H–sufficient or partially sufficient statistic for θ, it is necessary that T is θ–oriented or equivalently, we have a factorization of the form

$$p(x|\theta,\sigma) = g(T|\,\theta)f(x|T,\theta,\sigma) \tag{4.1}$$

where $p(x|\theta,\sigma)$ is a density function of $P_{\theta,\sigma}$.

It is also necessary that there exists ξ_θ,a probability measure on Σ such that the "mixed" density function,

$$p_{\xi_\theta}(x) = \int_\Sigma p(x|\theta,\sigma)d\xi_\theta(\sigma)$$

can be factored as

$$p_{\xi_\theta}(x) = G(T,\theta)F(x) \tag{4.2}$$

We now look at examples for partial sufficiency.

Example 4.1 (Hájek(1965)) Consider a sample $X = (X_1, X_2, \ldots, X_n)$ of size n, each i.i.d from $\mathcal{N}(\mu,\sigma^2)$. The statistic $T(X) = s^2 = \sum_{i=1}^n (X_i - \overline{X})^2/(n-1)$ is partially sufficient for σ^2 if $\sigma^2 \in (0,K)$ for K finite. To see this, we first factor the density function of X as follow:

$$p(x|\mu,\sigma^2) = A(\sigma)\exp\left[-\frac{ns^2}{2\sigma^2}\right]\exp\left[-\frac{n(\overline{X}-\mu)^2}{2\sigma^2}\right]$$

where $A(\sigma) = \left(\sqrt{2\pi}\sigma\right)^{-n}$.

Choose ξ_σ to be a normal distribution with mean 0 and variance $(K-\sigma^2)/n$. The "mixed" density function is:

$$\begin{aligned}
\overline{p}(x|\sigma) &= \int_\Re p(x|\mu,\sigma)d\xi_\sigma(\mu) \\
&= A(\sigma)\exp\left[-\frac{ns^2}{2\sigma^2}\right]\int_\Re \exp\left[-\frac{n(\overline{X}-\mu)^2}{2\sigma^2}\right]d\xi_\sigma(\mu) \\
&= A(\sigma)\exp\left[-\frac{ns^2}{2\sigma^2}\right]B(\overline{X})C(\sigma) \qquad (4.3)
\end{aligned}$$

where

$$\begin{aligned}
B(\overline{X}) &= \exp\left[-\frac{n\overline{X}^2}{2K}\right] \\
C(\sigma) &= \sigma
\end{aligned}$$

Remark: If we choose ξ_σ to be uniform distribution (the Lebesgue measure) over the whole real line, the proof Theorem 2.2 will break down since the left hand side of the Equation 2.1 is equal to infinity but not the right hand side. It seems unlikely that s^2 will be P–sufficient for σ^2, if $\sigma \in (0, \infty)$, however we do not have proof.

Example 4.2 (Neyman & Scott) Consider the data consisting of $2n$ observations $X_1, X_1', X_2, X_2', \ldots, X_n, X_n'$. Let X_i and X_i' be independent normal random variables with mean μ_i $(i = 1, 2, \ldots, n)$ and variance σ^2. The parameter of interest is σ^2, the nuisance parameter is the vector $\mu = (\mu_1, \mu_2, \ldots, \mu_n)$. Take $s^2 = \sum_{i=1}^n (X_i - X_i')^2$, $\overline{X}_i = (X_i + X_i')/2$ and with $A(\sigma) = \left(\sqrt{2\pi}\sigma\right)^{-n}$, we have

$$p(x|\mu, \sigma^2) = A(\sigma)\exp\left[-\frac{ns^2}{2\sigma^2}\right]\exp\left[-\frac{n(\overline{X}_i - \mu_i)^2}{2\sigma^2}\right]$$

The statistic s^2 is clearly σ-oriented and is partially sufficient if we take $\xi_\sigma(\mu_1, \mu_2, \ldots, \mu_n) = \prod_{i=1}^n \phi_\sigma(\mu_i)$, where $\phi_\sigma(\mu_i)$ is a normal density with mean 0, variance $(K - \sigma^2)/2$ and also assume that $\sigma \in (0, K)$, K is finite.

Example 4.3 Let $X = (X_1, X_2, \ldots, X_s)$ has a multimomial distribution with parameters n, $p_1, p_2, \ldots p_s$. The distribution of X is given by

$$P(X_1 = n_1, X_2 = n_2, \ldots X_s = n_s) = \frac{n!}{n_1! n_2! \ldots n_s!} p_1^{n_1} p_2^{n_2} \ldots p_s^{n_s}$$

where $\sum_{i=1}^s p_i = 1$ and $\sum_{i=1}^s n_i = n$.

The statistic $T(X) = (T_1(X_1), T_2(X_2), \ldots, T_s(X_s))$ is sufficient for $(p_1, p_2, \ldots p_s)$. Also the statistic $T_1(X_1) = n_1$ is partially sufficient for p_1 since the marginal distribution of $T_1(X_1)$ is binomial(n, p_1) which is independent of p_2, $p_3, \ldots, p_s$. Hence, it is p_1–*oriented.*

The factorization equation (4.2) holds if take ξ_{p_1} a point mass at

$$(p_2, p_3, \ldots, p_s) = \left(\frac{1-p_1}{s-1}, \frac{1-p_1}{s-1}, \ldots, \frac{1-p_1}{s-1}\right)$$

So we have

$$\begin{aligned} P(T(X_1) = n_1, X_2 = n_2, \ldots X_s = n_s)\xi_{p_1}(p_2, p_3, \ldots, p_s) &= \frac{n!}{n_1! \ldots n_s!} p_1^{n_1} \left(\frac{1-p_1}{s-1}\right)^{n-n_1} \\ &= \frac{n!(s-1)}{n_1! \ldots n_s!} \left(\frac{p_1}{1-p_1}\right)^{n_1} \left(\frac{1-p_1}{s-1}\right)^{n} \end{aligned}$$

The argument above works for any p_i for $i = 1, 2, \ldots, s$; therefore, we have $T_i(X_i))$ is partially sufficient for p_i; $i = 1, 2, \ldots, s$.

Example 4.4 A linear model is represented by $Y = X_1\tau + X_2\beta + \epsilon$, $E(\epsilon) = 0$, $Var(\epsilon) = \sigma^2\mathbf{I}_n$, where Y is an $n \times 1$ random vector, X_1 is $n \times k$ matrix with rank k, X_2 is $n \times p$ matrix with rank p and $\mathbf{I}_n$ is an $n \times n$ identity matrix. τ is a $k \times 1$ vector of parameters of interest, β is a $p \times 1$ vector of nuisance parameters, and ϵ is an $n \times 1$ random vector of errors from a normal distribution with mean $\mathbf{0}$ and covariance matrix $\sigma^2\mathbf{I}_n$. The density function of Y, with respect to Lebesgue measure, is

$$p(y \mid \tau, \beta) = (\sqrt{2\pi}\sigma)^{-n}\exp\left[-\frac{1}{2\sigma^2}\|y - X_1\tau - X_2\beta\|^2\right]$$

Let $\mathcal{M}(X_i)$ denote the space spanned by the columns of X_i and $\mathcal{M}(X_i)^\perp = \left\{x : x'y = 0 \text{ for } y \in \mathcal{M}(X_i)\right\}$, for $i = 1, 2$. Let P denote the orthogonal projection onto $\mathcal{M}(X_2)^\perp$, so $Q = I - P$ is the orthogonal projection onto $\mathcal{M}(X_2)$, with $y = Py + Qy$;

$$\begin{aligned} \|y - X_1\tau - X_2\beta\|^2 &= \|(Py - X_1\tau) + (Qy - X_2\beta)\|^2 \\ &= \|Py - X_1\tau\|^2 + \|Qy - X_2\beta\|^2 \end{aligned}$$

Hence,

$$p(y\,|\,\tau,\beta) \;=\; (\sqrt{2\pi}\sigma)^{-n}\exp\left[-\frac{1}{2\sigma^2}\left\{\|Py - X_1\tau\|^2 + \|Qy - X_2\beta\|^2\right\}\right]$$

Py is *τ–oriented* since $E(Py) = PX_1\tau$ and $Var(Py) = \sigma^2 P$. Choose $\xi_\tau(\beta) = 1_{\{0\}}(\beta)$, therefore, the statistic Py is partially sufficient for τ.

Example 4.5 (Maximal Invariance) Consider a sample $X = (X_1, X_2, \dots, X_n)$ of size n, each i.i.d from $\mathcal{N}(\mu, \sigma^2)$. The transformation $g_a(X_1, X_2, \dots, X_n) = (X_1 + a, X_2 + a, \dots, X_n + a)$ of the group $G = \{g_a : a \in \Re\}$ transforms the sample space $\Re^n$ onto itself. It is associated with the group $\overline{G} = \{\overline{g_a} : a \in \Re\}$ of $\overline{g_a}(\mu, \sigma) = (\mu + a, \sigma)$ of parameter space onto itself. The group $\overline{G}$ leaves σ invariant.

The maximal invariant for the problem of estimating σ with respect to the group G is the different statistic

$$D = (X_2 - X_1, X_3 - X_2, \dots, X_n - X_1).$$

The statistic s, as a function of D, is invariant and partially sufficient for the problem of estimating σ in the sense that

i) s is σ–oriented, and

ii) the statistic s is sufficient for σ.

Example 4.6 (Invariance) Let X be a random vector taking value in $\Re^2$. Let $X \sim N(\mu, I)$, $\mu \in \Re^2$.

For the parameter $\mu \in \Re^2$, to know $\mu \in \Re^2$, we need to know the norm of μ and the angle from a fixed array, say, $\hat{\mu} = (\|\mu\|/\sqrt{2}, \|\mu\|/\sqrt{2})$, represented by Γ, an orthogonal matrix, such that $\Gamma\hat{\mu} = \mu$. If we are only interested in estimating $\|\mu\|$, we may treat Γ as a nuisance parameter.

In terms of invariance, let the group $G = \mathcal{O}_2 =$ all orthogonal transformation on $\Re^2$ and $P_{\hat{\mu},\Gamma} = N(\Gamma\hat{\mu}, I)$. The statistic $\|X\|$ is sufficient for $\|\mu\|$, it is also G-invariant since $\|\Gamma X\| = \|X\|$. Clearly, there exists an invariant probability measure on $\mathcal{O}_2$ since $\mathcal{O}_2$ is compact. By Theorem 2.4 $\|X\|$ is partially sufficient for $\|\mu\|$.

4.2 Application of Comparison of Experiments in the Presence of Nuisance Parameter

4.2.1 Comparison of Normal Experiments with Unknown Mean and Unknown Variance

Let $\mathcal{E}_i$ be a normal experiment with unknown mean μ and unknown variance σ^{2k_i} where $k_i(>0)$ is a known constant; i = 1,2. Suppose that we are only interested in making inferences on the parameter σ with regardless to the value of μ–that is, μ is the nuisance parameter and σ is the parameter of our interest. We want to determine for what value of k_1 and k_2 that $\mathcal{E}_1$ is more informative than $\mathcal{E}_2$ for σ.

The next theorem will answer this question.

Theorem 4.1 *Let $\mathcal{E}_i$ be a normal experiments with unknown mean μ and unknown variance σ^{k_i} where $k_i > 0$ and $i = 1, 2$. If $k_1 > k_2$ and $\sigma \in (0, K]$, $K < \infty$, then $\mathcal{E}_1$ is more informative than $\mathcal{E}_2$ for σ.*

Proof. Let s_i^2 be a sample variance obtained from n_i observation from $\mathcal{E}_i$ for $i = 1, 2$. By Example 4.1 s_i^2 is a partially sufficient statistic for σ^{2k_i}; $i = 1, 2$. Let $\mathcal{E}_0^1$ and $\mathcal{E}_0^2$ be two experiments derived from the partially sufficient statistics s_1^2 and s_2^2, respectively.

Then by Theorem 3 of Goel and DeGroot (1979), $\mathcal{E}_0^1$ is sufficient for $\mathcal{E}_0^2$ if $k_1 > k_2 > 0$. So by Theorem 3.1 and Theorem 3.2, we have if $k_1 > k_2 > 0$ then $\mathcal{E}_1$ is more informative than $\mathcal{E}_2$ for σ iff $\mathcal{E}_0^1$ is sufficient for $\mathcal{E}_0^2$. And this concludes the proof. □

Remark: Goel and DeGroot (1979) proved the above theorem for the case of μ assumed to be known but no restriction on σ. The condition that σ is bounded which we required in proving the above theorem, is a legitimate assumption that can be impose here.

4.2.2 Comparison of Linear Normal Experiments With A Known Nonsingular Covariance Matrix

Let $\mathcal{E}_i = \mathcal{L}(X_i\tau + Z_i\beta, \sigma^2 I_{n_i})$ be a linear normal experiment which is represented by $Y = X_i\tau + Z_i\beta + \epsilon$, where Y is an $n_i \times 1$ random vector, X_i is $n_i \times k$ matrix with rank k, Z_i is $n_i \times p$ matrix with rank p and I_{n_i} is an $n_i \times n_i$ identity matrix. τ is a $k \times 1$ vector of parameters of interest, β is a $p \times 1$ vector of nuisance parameters, and ϵ_i is an $n_i \times 1$ random vector of errors from a normal distribution with mean 0 and covariance matrix $\sigma^2 I_{n_i}$; $i = 1, 2$.

By Example 4.4 $P_i y_i$ is a paritally sufficient statistic for τ; $i = 1, 2$. Let $\mathcal{E}_0^i = \mathcal{L}(P_i X_i \tau, \sigma^2 P_i)$ be a linear experiment base on the partially sufficient statistic $P_i y_i$ where P_i is an orthogonal projection matrix onto $\mathcal{M}(Z_i)^\perp$. Note that P_i may be represented in the form $P_i = I_{n_i} - Z_i(Z_i'Z_i)^{-1}Z_i'$; $i = 1, 2$.

Theorem 4.2 *Assume the above setup, $\mathcal{E}_1$ is more informative than $\mathcal{E}_2$ for τ iff $X_1'P_1X_1 - X_2'P_2X_2$ is non negative definite.*

Proof. i) Suppose $\mathcal{E}_1$ is more informative than $\mathcal{E}_2$ for τ. By Theorem 3.1 and Theorem 3.2, we have $\mathcal{E}_0^1$ is sufficient for $\mathcal{E}_0^2$. It follows that from Rao–Blackwell Theorem that $Var(c'\hat{\tau}_1) \leq Var(c'\hat{\tau}_2)$; $c \in \mathcal{M}(X_2'P_2') \subset \mathcal{M}(X_1'P_1')$ and $c'\hat{\tau}_i$ is the UMVUE of $c'\tau$. Since the UMVUE and BLUE coincide, $Var(c'\hat{\tau}_i) = c'(X_i'P_i'P_iP_iX_i)^-c = c'(X_i'P_iX_i)^-c$. By Lemma 2 of Stęniak, Wang and Wu (1984), we have $X_1'P_1X_1 - X_2'P_2X_2$ is non negative definite.

ii) Suppose $X_1'P_1X_1 - X_2'P_2X_2$ is non–negative definite. Let $C = X_1'P_1X_1 - X_2'P_2X_2$. Denote y_0^i as a random vector representing $\mathcal{E}_0^i$; $i = 1, 2$. Let $\mathcal{E}_0^*$ be a "fictitious experiment" such that $X_*'X_* = C$, X_* is a design martix of $\mathcal{E}_0^*$. Let y_0^* be a random vector representing $\mathcal{E}_0^*$ and suppose y_0^* and y_0^2 are independent. Then it follows that $(P_1X_1)'y_0^1$ is sufficient for τ and $(P_2X_2)'y_0^2 + X_*'y_0^*$ is sufficient for τ under the combination of experiment $\mathcal{E}_0^2$ and $\mathcal{E}_0^*$. But $(P_1X_1)'y_0^1$ has the same distribution as $(P_2X_2)'y_0^2 + X_*'y_0^*$. Hence $\mathcal{E}_0^1$ is sufficient for $\mathcal{E}_0^2$. By Theorem 3.1 and Theorem 3.2, $\mathcal{E}_1$ is more informative than $\mathcal{E}_2$ for τ. $\square$

Bibliography

1. Billingsley, P., *Probability and Measure* John Wiley, New York, 2nd, 1979.

2. Blackwell, D. (1951), Comparison of Experiments. *Proceedings of the Second Symposium on Mathematical Statistics and Probability,* **1**:92-102.

3. Goel, P. K., DeGroot (1979), Comparison of Experiments and Information Measures. *Annals of Statistics,* **7**:1066-1077.

4. Feldman, D., Ramamoorthi, R. V., A Decision Theoretic Proof of Blackwell Theorem. *Technical Report,* Department of Statistics & Probability, Michigan State University, 1984.

5. Hájek, J. (1965), On Basic Concepts of Statistics. *Proceedings of the Fifth Symposium on Mathematical Statistics and Probability,* **1**:139-162.

6. Fraser, D. A. S. (1956), Sufficient Statistics With Nuisance Parameters, *Annals of Mathematical Statistics,* **27**:838-842.

7. Kolmogorov, A. N. (1942), Sur l'estimation statistique des parameters de la loi de Gauss. *Izv. Akad. Nauk SSSR Ser. Mat.,* **6**:3–32.

8. Neveu, J., *Mathematical Foundations of Calculus of Probability* Holden–Day, San Franciso, 1965.

9. Neyman, J., Scott, E. L. (1948), Consistent Estimates Based on Partially Consistent Observations, *Econometrica,* **16**:1-32.

10. Neyman, J., Pearson, E. S. (1936), Sufficient Statistics and Uniformly Most Powerful Tests of Statistical Hypotheses, *Statistical Research Memoirs of the University of London,* **1**:133-137.

11. Stępniak, C., Wang, S. and Wu, C. F. (1984), Comparison of Linear Experiments With Known Covariances. *Annals of Statistics,* **12**:358-365.