CHARACTERIZATIONS OF LINEAR SUFFICIENT STATISTICS

A Dissertation

Presented to

The Faculty of the Department of Mathematics

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

Richard Alan Redner

December 1977

ACKNOWLEDGEMENT

The author gratefully acknowledges the assistance of his advisor, H. Walker, in the preparation of this dissertation. In addition, the author wishes to thank H. P. Decell and B. C. Peters for their support and contribution to this work.

CHARACTERIZATIONS OF LINEAR SUFFICIENT STATISTICS

An Abstract

Presented to

The Faculty of the Department of Mathematics

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

Richard Alan Redner

December 1977

ABSTRACT

In this paper we characterize the continuous linear sufficient statistics for a dominated collection of measures on a Banach space. This is followed by a characterization of exponential families with emphasis on those measures on R^n whose densities with respect to Lebesgue measure are multivariate normal densities. Finally, the relation between Bayes sufficiency and sufficient statistics is studied.

TABLE OF CONTENTS

I. INTRODUCTION AND DEFINITIONS

A σ-algebra of subsets S of a space X is a collection of subsets with the property that if $A_i \in S$, $i = 1,2,\ldots,$ then $(\bigcup_{i=1}^{\infty} A_i) \in S$ and $(X \setminus A_1) \in S$. If C is a collection of subsets of a space X we will denote the smallest σ-algebra containing C by $\sigma(C)$. The Borel field of a topological space is the smallest σ-algebra which contains the open sets.

A measurable space is a set X together with a σ-algebra of subsets S of X. If (X,S) and (Y,R) are measurable spaces, we say that T is a measurable function from X to Y if for $A \in R$ then $T^{-1}(A) \in S$ (where $T^{-1}(A) = \{x \in X \mid T(x) \in A\}$) : we denote by $T^{-1}(R)$ the σ-algebra of subsets of preimages of elements of R under T. A measurable function will also be called a statistic. If f is a real valued Borel measurable function on (X,S) then the collection of symbols $f(\in)T^{-1}(R)$ will be used to mean that f is also measurable on $(X,T^{-1}(R))$.

A measure is a real non-negative countably additive function on the measurable sets of a measurable space. If μ is a measure on a measurable space (X,S) then (X,S,μ) is called a measure space. A measure μ is a finite measure if $\mu(X) < \infty$.

If μ and ν are measures on a measurable space then μ is absolutely continuous with respect to ν (denoted $\mu \ll \nu$) if for each measurable set A, $\mu(A) = 0$ whenever $\nu(A) = 0$. If $\mu \ll \nu$ and $\nu \ll \mu$ then we say that μ is equivalent to ν and this is denoted

$\mu \equiv \nu$. If the symbol $[\mu]$ follows an assertion about the points of a measurable space, then it is understood that the set for which the statement is not true is a measurable set with μ measure zero.

The Radon-Nikodym theorem will be a necessary tool in this investigation and can be stated as follows:

> A necessary and sufficient condition that $\mu << \nu$ is that there exist a non-negative Borel measurable function f on (X,S) such that
>
> $$\mu(E) = \int_E f d\nu \quad \text{for every} \quad E \in S.$$

The function f in the statement of the theorem is unique in the sense that if g is a Borel measurable function on (X,S) such that

$$\mu(E) = \int_E g d\nu \quad \text{for every} \quad E \in S$$

then $f = g[\nu]$. The function f is called the Radon-Nikodym derivative of μ with respect to ν and is often denoted $d\mu/d\nu$. We will write $[d\mu/d\nu]$ to denote the collection of Radon-Nikodym derivatives of μ with respect to ν.

The notion of absolute continuity of one measure with respect to another and equivalence of measures can be extended. If D is a collection of measures on a measurable space (X,S) then D is a dominated collection of measures if there is a measure ν on (X,S) (but not necessarily in D) with the property that if $\mu \in D$, then $\mu << \nu$. The collection of measures D is said to be equivalent to a collection of measures F on (X,S) (denoted $D \equiv F$) if, for each $A \in S$, $\mu(A) = 0$ for each $\mu \in D$ if and only if $\nu(A) = 0$ for each $\nu \in F$.

Throughout this paper $M(X)$ will denote the collection of finite measures on X defined on some fixed σ-algebra and D will denote a subcollection of $M(X)$ which is dominated by a σ-finite measure (i.e., a measure which makes X the countable union of sets of finite measures).

In [6] Halmos and Savage investigate sufficient statistics in a general setting. In the presense of a dominated collection of measures, their results offer an alternative to the conditional probability definition of statistical sufficiency. We will use this as our definition of sufficiency with T a statistic from (X,S) to (Y,R).

Theorem 1.1. (Halmos-Savage [6].) A necessary and sufficient condition that T be a sufficient statistic for D is that there exist a $\lambda \in M(X)$ such that $D \equiv \lambda$ and for $\mu \in D$ there exist $g_\mu \in [d\mu/d\lambda]$ such that $g_\mu(\epsilon)T^{-1}(R)$.

If D is a dominated collection of measures then there exists a countable subcollection $\{\nu_i\}_{i=1}^{\infty}$ of D such that $D \equiv \{\nu_i\}_{i=1}^{\infty}$ [6]. Define a measure λ on (X,S) by

$$\lambda(A) = \sum_{i=1}^{\infty} \alpha_i \nu_i(A)$$

for $A \in S$ and where $\alpha_i = 1/(2^i \ \nu_i(X))$. It follows that λ is equivalent to D and by the proof of Theorem 1.1 the following theorem holds.

Theorem 1.2. A necessary and sufficient condition that T be a sufficient statistic for D is that there exist a $g_\mu \in [d\mu/d\lambda]$ such that $g_\mu(\epsilon)T^{-1}(R)$ for each $\mu \in D$.

A homogeneous class of measures is a collection of measures whose members are pairwise equivalent.

Corollary 1.3. (Halmos-Savage [6].) Let D be a homogeneous set of measures with $\nu \in D$. A necessary and sufficient condition that T be a sufficient statistic for D is that T be sufficient for the pair of measures (ν,μ) for each $\mu \in D$.

Proof. Let $\lambda = \nu$ and use Theorem 1.2.

Theorem 1.4. (Halmos-Savage [6].) A necessary and sufficient condition that a statistic T be sufficient for a dominated collection of measures D is that there exist $g_{\mu,\nu} \in [d\mu/d(\mu + \nu)]$ such that $g_{\mu,\nu}(\epsilon)T^{-1}(R)$ for $\mu,\nu \in D$.

II. CHARACTERIZATIONS OF LINEAR SUFFICIENT STATISTICS

Throughout this chapter, if W is a Banach space, then $B(W)$ will denote the Borel field of W. Also, throughout this chapter, X and Y will denote Banach spaces and T will denote a continuous linear operator from X onto Y.

In some of the theorems which follow it will be necessary to assume that $\ker T$ is complemented by some closed subspace (i.e., $X = \ker T \oplus S$, $T \cap S = 0$). This condition will not always be satisfied; however, if X is a Hilbert space then S can always be taken to be the orthogonal complement of $\ker T$.

Theorem 2.1. Let $X = \ker T \oplus S$ for some closed subspace S of X. A necessary and sufficient condition that T be a sufficient statistic for D is that there exist $\lambda \in M(X)$ such that $D \equiv \lambda$ and

$$(1) \qquad \ker T \subset \{y \mid g_\mu(x + y) = g_\mu(x),\ x \in X\}$$

for each $\mu \in D$ and some $g_\mu \in [d\mu/d\lambda]$.

Proof. If T is a sufficient statistic, then there exists a $\lambda \equiv D$ such that for some $g_\mu \in [d\mu/d\lambda]$, $g_\mu(\epsilon)T^{-1}(R)$ for each $\mu \in D$. Then there exists a $(Y,B(Y))$ measurable function f_μ such that $g_\mu = f_\mu \circ T$ for $\mu \in D$ ([2] pg. 69). If $y \in \ker T$, then we have

$$g_\mu(x + y) = f_\mu \circ T(x + y) = f_\mu \circ T(x) = g_\mu(x) \quad \text{for} \quad x \in X.$$

Conversely, suppose $D \equiv \lambda$ and for $\mu \in D$,

$$\ker T \subset \{y | g_\mu(x + y) = g_\mu(x),\ x \in X\}$$

for $g_\mu \in [d\mu/d\lambda]$. We need only show (according to Theorem 1.1) that $g_\mu(\epsilon)T^{-1}(R)$. Since for any collection of sets C in X,

$$\sigma(g^{-1}(C)) = g^{-1}(\sigma(C))$$

([9] pg.10) it suffices to show that if r is a real number, then there exists $B_r \in B(Y)$ such that $g_\mu^{-1}(-\infty,r) = T^{-1}(B_r)$. We will first show that

$$g_\mu^{-1}(-\infty,r) = T^{-1}(T[g_\mu^{-1}(-\infty,r) \cap S])$$

and then that

$$B_r = T[g_\mu^{-1}(-\infty,r) \cap S] \in B(Y).$$

If $x \in T^{-1}(T[g_\mu^{-1}(-\infty,r) \cap S])$, then $T(x) \in T[g_\mu^{-1}(-\infty,r) \cap S]$ and, hence, $T(x) = T(z)$ for some $z \in g_\mu^{-1}(-\infty,r) \cap S$. Since T is linear, $(x - z) \in \ker T$, so that $g_\mu(x) = g_\mu(x - z + z) = g_\mu(z) < r$ and $x \in g_\mu^{-1}(-\infty,r)$.

If $x \in g_\mu^{-1}(-\infty,r)$, then, since $X = \ker T \oplus S$, $x = k + s$ for $k \in \ker T$ and $s \in S$. It follows that $T(x) = T(s)$, $s - x \in \ker T$,

$$g_\mu(s) = g_\mu(s - x + x) = g_\mu(x) < r$$

and so $s \in g_\mu^{-1}(-\infty,r)$. From this it follows that

$$T(x) = T(s) \in T[g_\mu^{-1}(-\infty,r) \cap S]$$

and finally that $x \in T^{-1}(T[g_\mu^{-1}(-\infty,r) \cap S])$.

We will now show that $T[g_\mu^{-1}(-\infty,r) \cap S] \in B(Y)$. Let $T_S : S \to Y$ be the restriction of T to S and observe that T is a one-to-one continuous mapping of the Banach space S onto the Banach space Y. Since T_S satisfies the hypothesis of the open mapping theorem, T_S is a homeomorphism of S onto Y and therefore maps elements of $B(S)$ to elements of $B(Y)$. Since g_μ is measurable,

$$[g_\mu^{-1}(-\infty,r) \cap S] \in B(X) \cap S = B(S)$$

([5] page 25). It follows that

$$T[g_\mu^{-1}(-\infty,r) \cap S] = T_S[g_\mu^{-1}(-\infty,r) \cap S] \in B(Y).$$

In the following two theorems, we show that if the class of Radon-Nikodym derivatives contains continuous representatives, then under certain other conditions T is a sufficient statistic if and only if condition (1) in Theorem 2.1 holds for the continuous representative g_μ for $\mu \in D$. Furthermore, if $D \equiv \lambda$ and V is an open or countable set in X such that $\lambda(X \backslash V) = 0$, then the sufficiency of a statistic T depends only on $g_\mu|_V$ for $\mu \in D$. First, we will need Lemma 2.2 below.

For $V \in B(X)$, and for a measurable function f from X into the real numbers such that

$$\ker T \subset \{y \in X \mid x \in (V - y) \cap V,\ f(x + y) = f(x)\},$$

we define a function $\hat{f}$ on X by

$$\hat{f}(x) = \begin{cases} f(x + y) & \text{if for some } y \in \ker T,\ x + y \in V, \\ 0 & \text{otherwise.} \end{cases}$$

Lemma 2.2.

$$\hat{f}^{-1}(-\infty,r) = \begin{cases} [X\backslash(V \oplus \ker T)] \cup [(f^{-1}(-\infty,r) \cap V) \oplus \ker T] & \text{for } r > 0 \\ [(f^{-1}(-\infty,r) \cap V) \oplus \ker T] & \text{for } r \le 0 \end{cases}$$

where for $A,B \in B(x)$, $B \oplus A = \{x \in X | x = b + a$ for $b \in B, a \in A\}$.

Proof. Suppose that $r > 0$. Let $x \in \hat{f}^{-1}(-\infty,r)$ and $r > t = \hat{f}(x)$. If there exists $y \in \ker T$ so that $x + y \in V$ then $t = \hat{f}(x) = f(x + y)$. Since $x = (x + y) + (-y)$ and $x + y \in [f^{-1}(-\infty,r) \cap V]$, it follows that $x \in [f^{-1}(-\infty,r) \cap V] \oplus \ker T$. If for every $y \in \ker T$, $x + y \notin V$, then $x \in X\backslash(V \oplus \ker T)$. Hence

$$\hat{f}^{-1}(-\infty,r) \subset [X\backslash(V \oplus \ker T)] \cup [(f^{-1}(-\infty,r) \cap V) \oplus \ker T].$$

Conversely, if $x \in [(f^{-1}(-\infty,r) \cap V) \oplus \ker T]$, then there exists $y \in \ker T$ so that $x + y \in f^{-1}(-\infty,r) \cap V$. Then $\hat{f}(x) = f(x + y) = t < r$, so $x \in \hat{f}^{-1}(-\infty,r)$. If $x \in [X\backslash(V \oplus \ker T)]$, then $\hat{f}(x) = 0$, so $x \in \hat{f}^{-1}(0)$. Thus

$$\hat{f}^{-1}(-\infty,r) \supset [X\backslash(V \oplus \ker T)] \cup [(f^{-1}(-\infty,r) \cap V) \oplus \ker T].$$

Similarly, if $r \le 0$, then

$$\hat{f}^{-1}(-\infty,r) = [(f^{-1}(-\infty,r) \cap V) \oplus \ker T].$$

Let D be a dominated collection of measures on $(X,B(X))$. Define (as in Chapter 1) a measure λ on $(X,B(X))$ by $\lambda(A) = \sum_{i=1}^{\infty} \alpha_i \nu_i(A)$ for $A \in B(x)$, where $\{\nu_i\}_{i=1}^{\infty}$ is a countable equivalent subcollection of D and $\alpha_i = (2^{-i}\ \nu_i(X)^{-1})$, $i = 1,2,\ldots$.

Let V be an open set in X such that $\lambda(X\backslash V) = 0$. Suppose that if $B \in B(X)$ and $\lambda(B \cap V) = 0$ then $\lambda(B \cap V - y) = 0$ for $y \in \ker T$. Suppose further that if C is a non-empty set in the relative topology of V then $\lambda(C) > 0$.

Theorem 2.3. Suppose that for each $\mu \in D$, $[d\mu/d\lambda]$ contains a representative f_μ which is continuous on V. A necessary and sufficient condition that T be a sufficient statistic for D is that

$$\ker T \subset \{y | x \in (V - y) \cap V \text{ and } f_\mu(x + y) = f_\mu(x)\}.$$

Proof. Suppose $\ker T \subset \{y | x \in (V - y) \cap V$ and $f_\mu(x + y) = f_\mu(x)\}$. Define $\hat{f}_\mu$ as in the paragraph preceding Lemma 2.2. Then $\hat{f}_\mu$ is a measurable function. It suffices to show $\hat{f}_\mu^{-1}(-\infty,r) \in B(X)$. Since V is an open set and $f_\mu|_V$ is continuous (relative topology on V), then $[f_\mu^{-1}(-\infty,r) \cap V) \oplus \ker T]$ and $V \oplus \ker T$ are open sets in X. So by Lemma 2.2, $\hat{f}^{-1}(-\infty,r) \in B(X)$. It follows that $\hat{f}_\mu \in [d\mu/d\lambda]$, since $f_\mu \in [d\mu/d\lambda]$ and $f_\mu = \hat{f}_\mu$, except perhaps on a set of λ measure zero.

Define $h_\mu: T(V) \to R^1$ by $h_\mu(z) = f_\mu(x)$ for $z = Tx$. Then h_μ is continuous on $T(V)$ since $h_\mu^{-1}(a,b) = T[f^{-1}(a,b) \cap V]$ is an open set in Y. Extending h_μ to all of Y by letting $h_\mu(z) = 0$ for $z \in Y\backslash T(V)$ we have that h_μ is Borel measurable and $\hat{f}_\mu = h_\mu \circ T$. Hence, T is a sufficient statistic.

Conversely, if T is sufficient for D, then for $\mu \in D$ there exists $g_\mu \in [d\mu/d\lambda]$ such that $\ker T \subset \{y | g_\mu(x + y) = g_\mu(x), x \in X\}$ and $g_\mu = f_\mu[\lambda]$.

Let $y \in \ker T$, $B = \{x | f_\mu(x) \neq g_\mu(x)\} \in B(X)$ and, $\lambda(B \cap V - y) = \lambda(B \cap V) = \lambda(B) = 0$. Define $\hat{f}_\mu$ by

$$\hat{f}_\mu|_V = f_\mu|_V$$
$$\hat{f}_\mu|_{X\setminus V} = g_\mu|_{X\setminus V} \quad .$$

Since

$$\begin{aligned} \hat{f}_\mu(x) &= g_\mu(x) && \text{for} \quad x \in X\setminus(B \cap V), \\ g_\mu(x) &= g_\mu(x + y) && \text{for} \quad x \in X, \\ \text{and} \quad g_\mu(x + y) &= \hat{f}_\mu(x + y) && \text{for} \quad x \in X\setminus(B \cap V - y), \end{aligned}$$

it follows that $\hat{f}_\mu(x) = \hat{f}_\mu(x + y)$, except on some set

$$C \subset (B \cap V) \cup (B \cap V - y),$$

such that $\lambda(C) = \lambda(C \cap V) = 0$. Since $\lambda(C) = 0$ and $(V - y) \cap V$ is open in V it follows that $C \cap [(V - y) \cap V]$ contains no non-empty open set of $V \cap (V - y)$. Then $\hat{f}(x) = \hat{f}(x + y)$ for $x \in (V - y) \cap V$. Hence $f(x) = f(x + y)$ for $x \in (V - y) \cap V$.

The following is an example, which has the property that a linear statistic T is a sufficient statistic yet the kernel condition does not hold for the continuous representative of the derivative.

Example 2.4. Let T be the linear function from the plane onto the x-axis defined by $T(x,y) = x$. Let V be the two closed unit squares centered at $(1/2, 1/2)$ and $(3/2, 5/2)$. Let g be the Radon-Nikodym derivative of two measures P_1 and P_2, which are equivalent to Lebesgue measure restricted to the two squares and zero elsewhere, defined by

$$g(x,y) = \begin{cases} 1 & \text{if } 0 \le x \le 1, \ 0 \le y \le 1 \\ 2 & \text{if } 1 \le x \le 2, \ 2 \le y \le 3 \\ 0 & \text{otherwise .} \end{cases}$$

We see that T is sufficient by Theorem 2.3, taking V to be the interior of the squares. However, the kernel condition of Theorem 2.3 is not satisfied for the continuous representative on the closed squares.

Theorem 2.5. Let V be a countable set and $\lambda \equiv D$ defined as in Theorem 2.3. Let $\lambda(X\backslash V) = 0$ and $\lambda(x) > 0$ for $x \in V$, and suppose that for each $\mu \in D$, $[d\mu/d\lambda]$ contains a representative f_μ which is continuous on V. A necessary and sufficient condition that T be a sufficient statistic is that $\ker T \subset \{y | x \in (V - y) \cap V, \ f_\mu(x + y) = f_\mu(x)\}$.

Proof. If T is a sufficient statistic, then, for $\mu \in D$, there exists $g_\mu \in [d\mu/d\lambda]$ such that $g_\mu = h_\mu \circ T$ for some Borel measurable function h_μ. Since $\lambda(x) > 0$ for $x \in V$, $f_\mu = g_\mu$ on V.

Conversely, suppose that

$$\ker T \subset \{y | x \in (V - y) \cap V, \ f_\mu(x + y) = f_\mu(x)\} .$$

Define $\hat{f}_\mu$ as in the paragraph preceding Lemma 2.2, and observe that $\hat{f}_\mu$ is measurable. Since $\hat{f}_\mu = f_\mu[\lambda]$, then $\hat{f}_\mu \in [d\mu/d\lambda]$. Define $h_\mu : Y \to R^1$ by

$$h_\mu(z) = \begin{cases} \hat{f}(x) & \text{if } Tx = z \text{ and } x \in V \\ 0 & \text{otherwise .} \end{cases}$$

Clearly h_μ is Borel measurable and $\hat{f}_\mu = h_\mu \circ T$. Hence T is a sufficient statistic.

We observe that the theorems of this chapter have immediate corollaries corresponding to Corollary 1.3 and Theorem 1.4. In particular, the homogeneous case is of special importance.

III. APPLICATION TO EXPONENTIAL FAMILIES

Let X be a Banach space and $\{P_\theta\}_{\theta\in\Theta}$ be a collection of measures defined on $B(X)$. We say that $\{P_\theta\}_{\theta\in\Theta}$ is an exponential family with respect to a σ-finite measure λ on $B(X)$ if there exist non-negative functions c and h defined on Θ and X respectively and functions Q and t defined on Θ and X respectively with range in an inner product space with inner product $\langle\cdot,\cdot\rangle$ such that

$$(dP_\theta/d\lambda)(x) \equiv p_\theta(x) = c(\theta)h(x) \exp \langle Q(\theta), t(x)\rangle .$$

Exponential families are an important class of measures in theoretical and applied statistics. Under suitable hypotheses, if there exist sufficient statistics for a collection of measures, then that class is an exponential family. For further details see Brown 1964 [4] and Anderson [1].

Let $\{P_\theta\}_{\theta\in\Theta}$ be an exponential family of measures on $B(X)$.

<u>Theorem 3.1</u>. If $X = \ker T \oplus S$ for some closed subspace S, V is an open convex set, and $\theta_0 \in \Theta$ such that

(1) $P_{\theta_0}(X\backslash V) = 0$

(2) t is continuous and Gateaux differentiable on V

(i.e., $\delta t(x,y) = \lim\limits_{\alpha\to 0} \dfrac{t(x + \alpha y) - t(x)}{\alpha}$ exists for $x \in V$)

(3) $h > 0[\lambda]$ on V

(4) if $C \subset B(V)$ and $\lambda(C) = 0$ then $\lambda(C - y) = 0$ for $y \in \ker T$.

(5) if C is not empty and open in V then $\lambda(C) > 0$.

A necessary and sufficient condition that T be a sufficient statistic for $\{P_\theta\}_{\theta\in\Theta}$ is that for $\theta \in \Theta$,

$$\ker T \subset \{y | \langle Q(\theta) - Q(\theta_0), \delta t(x;y)\rangle = 0, x \in V\} .$$

<u>Proof</u>: Observe that the conditions of Theorem 2.3 are satisfied and hence T is a sufficient statistic if and only if for $\theta \in \Theta$

$$\ker T \subset \{y | x \in (V - y) \cap V \text{ and } (dP_\theta/dP_{\theta_0})(x + y) = (dP_\theta/dP_{\theta_0})(x)\},$$

where $(dP_\theta/dP_{\theta_0})(x) = (C(\theta)/C(\theta_0))\exp\langle Q(\theta) - Q(\theta_0), t(x)\rangle$.

Equivalently, since V is convex and t is Gateaux differentiable,

$$\ker T \subset \{y | \alpha \in R', x \in (V - y) \cap V \text{ and } \langle Q(\theta) - Q(\theta_0), t(x + \alpha y) - t(x)\rangle = 0\},$$

$$\ker T \subset \{y | x \in V \text{ and } \langle Q(\theta) - Q(\theta_0), \lim_{\alpha\to 0} \frac{t(x + \alpha y) - t(x)}{\alpha}\rangle = 0\} ,$$

and finally

$$\ker T \subset \{y | \langle Q(\theta) - Q(\theta_0), \delta t(x;y)\rangle = 0, x \in V\}.$$

Theorem 3.1 will now be used to characterize linear sufficient statistics for families of probability measures having Wishart densities.

Let S denote the symmetric $n \times n$ matrices, V the positive definite elements of S and $\{W(\Omega_\theta)\}_{\theta\in\Theta}$ $(\Omega_\theta \in V$ for $\theta \in \Theta)$ a family of Wishart probability measures with $r \geq n$ degrees of freedom having densities

$$w(x,\Omega_\theta) = \begin{cases} K(\Omega_\theta^{-1})|X|^{\frac{1}{2}(r-n-1)} \exp(\mathrm{trace}(-\frac{1}{2}\,\Omega_\theta^{-1}x)), & x \in V \\ 0 & \text{otherwise.} \end{cases}$$

Theorem 3.2. A necessary and sufficient condition that T be a sufficient statistic for $\{W(\Omega_\theta)\}_{\theta\in\Theta}$ is that for some $\theta_0 \in \Theta$

$$\ker T \subset [\Omega_\theta^{-1} - \Omega_\theta^{-1}]^\perp \quad \text{for} \quad \theta \in \Theta$$

where $\perp$ denotes the orthogonal complement with respect to the trace inner product $\langle\cdot,\cdot\rangle$ defined on symmetric matrices.

Proof: Observe that V is open in S and $w(x,\Omega_\theta) > 0$ for $x \in V$ and $\theta \in \Theta$. The preliminary conditions of Theorem 3.1 are satisfied with λ = Lebesque measure on S. Hence a statistic T is sufficient if and only if

$$\text{trace}[(\Omega_\theta^{-1} - \Omega_{\theta_0}^{-1})(\delta I(x;y))] = 0$$

for $x \in V$, $\theta \in \Theta$ and $y \in \ker T$, or equivalently

$$\text{trace}[(\Omega_\theta^{-1} - \Omega_{\theta_0}^{-1})y] = 0 \quad \text{for} \quad \theta \in \Theta \quad \text{and} \quad y \in \ker T.$$

We will now consider a collection of discrete probability measures.

Let $V = \{x \in R^n | x$ has non-negative integer components x_i, $i = 1,\ldots,n\}$. If $\lambda \in (0,+\infty)$, let P_λ be the measure on $(R^n, B(R^n))$ such that

(1) $P_\lambda(x) = e^{-n\lambda} \prod_{i=1}^{n} \dfrac{\lambda^{x_i}}{(x_i)!}$ if $x \in V$

(2) $P_\lambda(X\backslash V) = 0.$

This measure is the probability measure associated with an independent sample of size n of a random variable with the Poisson distribution. Using Theorem 2.3 directly we have the following theorem.

Theorem 3.3. A necessary and sufficient condition that T be a sufficient statistic for $\{P_\lambda\}_{\lambda\in(0,+\infty)}$ is that

$$\ker T \subset \{y \in R^n | (V - y) \cap V \neq \emptyset \text{ implies } \Sigma y_i = 0\}.$$

Proof: By Theorem 2.3, T is a sufficient statistic if and only if

$$\ker T \subset \{y \in R^n | x \in (V - y) \cap V,\ \lambda \in (0,+\infty) \text{ and } \frac{dP_\lambda}{dP_1}(x + y) = \frac{dP_\lambda}{dP_1}(x)\}$$

$$= \{y \in R^n | (V - y) \cap V \neq \emptyset \text{ implies } \prod_{i=1}^{n} \lambda^{y_i} = 1\} .$$

We observe that any scalar multiple of the mean and any linear map which is $1 - 1$ on V is a sufficient statistic.

IV. APPLICATIONS TO THE MULTIVARIATE NORMAL CASE

Let $\{P_i\}_{i=0}^m$ be a family of $m + 1$ probability measures defined on $(R^n, B(R^n))$ having normal densities

$$p_i(x) = (2\pi)^{-n/2}|\Omega_i|^{-1/2}\exp[-\tfrac{1}{2}(x - \eta_i)^T\Omega_i^{-1}(x - \eta_i)]$$

for $i = 0,1,\ldots,m$, where η_i and Ω_i are known and Ω_i is a positive definite operator on R^n. In the following paragraphs we will characterize, in a variety of ways, the linear sufficient statistics for such a collection of measures. We will then investigate the construction of sufficient statistics and give unbiased consistent estimators which are functions of the sufficient statistic.

Theorem 4.1. A necessary and sufficient condition that a $k \times n$ rank k matrix B be a sufficient statistic for $\{P_i\}_{i=0}^m$ is that

$$\ker B \subset \bigcap_{i=1}^{m}\{\ker(\Omega_i^{-1} - \Omega_0^{-1}) \cap [\Omega_i^{-1}\eta_i - \Omega_0^{-1}\eta_0]^\perp\} .$$

Proof: Write $(x - \eta_i)^T\Omega_i^{-1}(x - \eta_i)$ as

$$<\Omega_i^{-1}, xx^T>_{tr} - 2<\Omega_i^{-1}\eta_i, x> + <\Omega_i^{-1}\eta_i\eta_i^T, I_{n\times n}>_{tr}$$

where $I_{n\times n}$ is the identity matrix on R^n. By Theorem 3.1 if $1 \le i \le m$ then B is a sufficient statistic for $\{P_i, P_0\}$ if and only if

$$\ker B \subset \{y \in R^n | <\Omega_i^{-1} - \Omega_0^{-1}, \delta(II^T)(x;y)>_{tr} - 2<\Omega_i^{-1}\eta_i - \Omega_0^{-1}\eta_0, \delta I(x;y)> +$$

$$+ <\Omega_i^{-1}\eta_i\eta_i^T - \Omega_0^{-1}\eta_0\eta_0^T, 0>_{tr} = 0, x \in R^n\}$$

or equivalently,

$$\ker B \subset \{y \mid 2\langle \Omega_i^{-1} - \Omega_0^{-1}, yx^T\rangle_{tr} - 2\langle \Omega_i^{-1}\eta_i - \Omega_0^{-1}\eta_0, y\rangle = 0,\ x \in R^n\}$$

$$= \{\ker(\Omega_i^{-1} - \Omega_0^{-1}) \cap [\Omega_i^{-1}\eta_i - \Omega_0^{-1}\eta_0]^{\perp}\} .$$

<u>Theorem 4.2.</u> (Peters [4].) A necessary and sufficient condition that a $k \times n$ rank k matrix B be a sufficient statistic for $\{P_j\}_{j=0}^m$ is that for $j = 1,\ldots,m$

(a) $[\Omega_0 B^T(B\Omega_0 B^T)^{-1}B](\Omega_j - \Omega_0) = \Omega_j - \Omega_0$

(b) $[\Omega_0 B^T(B\Omega_0 B^T)^{-1}B](\eta_j - \eta_0) = \eta_j - \eta_0$.

<u>Proof</u>: Let $P^2 = P = \Omega_0 B^T(B\Omega_0 B^T)^{-1}B$ and observe that if B is a sufficient statistic then

$$\ker P = \ker B \subset \ker(\Omega_j^{-1} - \Omega_0^{-1}) = \ker[(\Omega_j - \Omega_0)\Omega_0^{-1}].$$

It follows that

$$\mathrm{Range}[\Omega_0^{-1}(\Omega_j - \Omega_0)] \subset \mathrm{Range}\ P^T$$

and hence that

$$B^T(B\Omega_0 B^T)^{-1}B\Omega_0\Omega_0^{-1}(\Omega_j - \Omega_0) = \Omega_0^{-1}(\Omega_j - \Omega_0)$$

which is equivalent to (a) .

Since

$$\ker P = \ker B \subset \ker(\Omega_j^{-1} - \Omega_0^{-1}) \cap [\Omega_j^{-1}\eta_j - \Omega_0^{-1}\eta_0]^{\perp} \subset [\Omega_0^{-1}(\eta_j - \eta_0)]^{\perp}$$

it follows that

$$\Omega_0^{-1}(\eta_j - \eta_0) \subset [\ker P]^{\perp} = \text{range } P^T .$$

Hence,

$$B^T(B\Omega_0 B^T)^{-1}B\Omega_0\Omega_0^{-1}(\eta_j - \eta_0) = \Omega_0^{-1}(\eta_j - \eta_0)$$

which is equivalent to (b).

Since all of the preceding arguments are reversible, (a) and (b) imply B is a sufficient statistic for $\{P_i\}_{i=0}^m$, completing the proof.

Theorem 4.3. A necessary and sufficient condition that a $k \times n$ rank k matrix B be a sufficient statistic for $\{P_i\}_{i=0}^m$ is that for $A \equiv (\sum_{i=1}^m (\Omega_i^{-1} - \Omega_0^{-1})^2 + (\Omega_i^{-1}\eta_i - \Omega_0^{-1}\eta_0)(\Omega_i^{-1}\eta_i - \Omega_0^{-1}\eta_0)^T)$, $\ker B \subset \ker A$.

Proof: According to Theorem 4.1 it suffices to show that

$$\ker A = \bigcap_{i=1}^m \{\ker(\Omega_i^{-1} - \Omega_0^{-1}) \cap [\Omega_i^{-1}\eta_i - \Omega_0^{-1}\eta_0]^{\perp}\} .$$

Let $x \in \ker A$. Then $x^TAx = 0$, which implies that $x^T(\Omega_i^{-1} - \Omega_0^{-1})^2 x = 0$ and $x^T(\Omega_i^{-1}\eta_i - \Omega_0^{-1}\eta_0)(\Omega_i^{-1}\eta_i - \Omega_0^{-1}\eta_0)^T x = 0$ for $i = 1,2,\ldots,m$. But then $(\Omega_i^{-1} - \Omega_0^{-1})x = 0$ and $x^T(\Omega_i^{-1}\eta_i - \Omega_0^{-1}\eta_0) = 0$ for $i = 1,2,\ldots,m$ and so $x \in \bigcap_{i=1}^m \{\ker(\Omega_i^{-1} - \Omega_0^{-1}) \cap [\Omega_i^{-1}\eta_i - \Omega_0^{-1}\eta_0]^{\perp}\}$.

Conversely, let $x \in \bigcap_{i=1}^m \{\ker(\Omega_i^{-1} - \Omega_0^{-1}) \cap [\Omega_i^{-1}\eta_i - \Omega_0^{-1}\eta_0]^{\perp}\}$ then $(\Omega_i^{-1} - \Omega_0^{-1})^2 x = 0$ and $(\Omega_i^{-1}\eta_i - \Omega_0^{-1}\eta_0)^T x = 0$ and so $Ax = 0$ and $x \in \ker A$.

In the next theorem we will use the fact that there exists a non-singular matrix M such that $M\Omega_0 M^T = I$ and, hence, the affine transformation $x \to M(x - \eta_0)$ provides a change of variables that allows one to assume that $\eta_0 = 0$ and $\Omega_0 = I$. We observe through the following lemma that this entails no loss of generality since we can easily recover the sufficient statistic for the original collection of measures.

Lemma 4.4. Suppose that T is a sufficient statistic for $\{P_i\}_{i=0}^m$. Let $\hat{P}_i$ denote the measure corresponding to P_i and the change of variables $y = Mx - z$, where M is a non-singular matrix and z is an arbitrary element of R^n. It follows that TM^{-1} is a sufficient statistic for $\{\hat{P}_i\}_{i=0}^m$.

Proof: According to Theorem 4.1, it suffices to show (since $\hat{P}_i$ is a normal measure with mean $M\eta_i - z$ and covariance matrix $M\Omega_i M^T$ for $i = 0,1,\ldots,m$) that $\ker TM^{-1} \subset \ker(M^{T^{-1}}(\Omega_i^{-1} - \Omega_0^{-1})M^{-1})$ and $\ker TM^{-1} \subset [M^{T^{-1}}(\Omega_i^{-1}\eta_i - \Omega_0^{-1}\eta_0) - M^{T^{-1}}(\Omega_i^{-1} - \Omega_0^{-1})M^{-1}z]^{\perp}$.

Let $y \in \ker TM^{-1}$. Then, since $\ker TM^{-1} = M(\ker T)$, let $y = Mx$ where $x \in \ker T$. Then $M^{T^{-1}}(\Omega_i^{-1} - \Omega_0^{-1})M^{-1}(Mx) = 0$ since $x \in \ker(\Omega_i^{-1} - \Omega_0^{-1})$ and $y^T(M^{T^{-1}}(\Omega_i^{-1}\eta_i - \Omega_0^{-1}\eta_0) - M^{T^{-1}}(\Omega_i^{-1} - \Omega_0^{-1})M^{-1}z)$ $= x^T(M^TM^{T^{-1}}(\Omega_i^{-1}\eta_i - \Omega_0^{-1}\eta_0) - x^T(M^TM^{T^{-1}}(\Omega_i^{-1} - \Omega_0^{-1})M^{-1}z = 0$ and TM^{-1} is a sufficient statistic for $\{\hat{P}_i\}_{i=0}^m$.

Theorem 4.5. (Decell [4].) If $\eta_0 = 0$ and $\Omega_0 = I$, then a necessary and sufficient condition that there exist a $k \times n$ rank k matrix B sufficient for $\{P_i\}_{i=0}^m$ is that there exist a rank k orthogonal projection Q such that

$$(I - Q)[\eta_1|\eta_2| \ldots |\eta_m|\Omega_1 - I| \ldots |\Omega_m - I] = z ,$$

where z is the $n \times (n + 1)\cdot m$ zero matrix.

Proof: If a $k \times n$ matrix B of rank k is sufficient for $\{P_i\}_{i=0}^m$, we may assume without loss of generality that $BB^T = I$ since B is a sufficient statistic for $\{P_i\}_{i=0}^m$ if and only if KB is a sufficient statistic for each non-singular $k \times k$ matrix K. One can choose K so that $KBB^TK^T = (KB)(KB)^T = I$.

For $i = 1,2,\ldots,m$ Theorem 4.2 implies that

$$\Omega_i B^T(B\Omega_i B^T)^{-1} = IB^T(BIB^T)^{-1} = B^T$$

so that

$$(B\Omega_i B^T)^{-1} = B\Omega^{-1}B^T \quad \text{and} \quad \Omega_i B^T(B\Omega_i B^T)^{-1}B = B^TB \ .$$

Right multiplication of the latter equation by $\Omega_i B^TB$ will establish that

$$\Omega_i B^TB = B^TB\Omega_i B^TB \ ,$$

and by symmetry it follows that

$$\Omega_i B^TB = B^TB\Omega_i \ .$$

Since $\eta_0 = 0$ and $\Omega_0 = I$, Theorem 4.2 implies

$$\eta_i - B^TB\eta_i = 0$$

and

$$\Omega_i - B^TB\Omega_i = I - B^TB \ .$$

Since $BB^T = I$ it follows that $B^T = B^+$ (where $(\cdot)^+$ denotes the generalized inverse of $(\cdot)$) and hence that $Q = B^TB = B^+B$ is the orthogonal projection onto the range of B^T. Clearly Q has rank k and we conclude that $(I - Q)\eta_i = 0$ and $(I - Q)(\Omega_i - I) = 0$ $i = 1,2,\ldots,m$ and the condition follows.

Conversely, if the condition holds, let B be any $k \times n$ rank k matrix such that $B^T = B^+$ and $\text{range}(B^T) = \text{range}(Q)$. Then clearly $B^+B = Q$ and $BB^+ = I$. By the symmetry of $I - Q$ and $\Omega_i - I$, we conclude that $\Omega_i B^T B = B^T B \Omega_i$ and, hence, that

$$\begin{aligned} Q = B^+B &= B^+B\Omega_i B^T(B\Omega_i B^T)^{-1}B \\ &= \Omega_i B^+BB^T(B\Omega_i B^T)^{-1}B \\ &= \Omega_i B^T(B\Omega_i B^T)^{-1}B \ . \end{aligned}$$

In addition, $\Omega_i B^T(B\Omega_i B^T)^{-1} = B^T$. Clearly B satisfies the conditions of Theorem 4.2.

Definition. We will say that a rank k orthogonal projection Q generates a sufficient statistic for $\{P_i\}_{i=0}^m$ provided that, for any $k \times n$ rank k matrix B such that $B^+B = Q$, B is a sufficient statistic for $\{P_i\}_{i=0}^m$.

Corollary 4.6. (Decell [4].) If $M = [\eta_1 | \dots | \eta_m | \Omega_1 - I | \dots | \Omega_m - I]$ then

(a) $Q = MM^+$ generates a sufficient statistic for $\{P_i\}_{i=0}^m$

(b) $k = \text{rank}(MM^+) = \text{tr}(MM^+)$ is the smallest integer for which there exists a rank k orthogonal projection generating a sufficient statistic for $\{P_i\}_{i=0}^m$.

Proof: Since $(I - MM^+)M = 0$ MM^+ generates a sufficient statistic. Let k be the smallest integer for which there exists a rank k orthogonal projection P generating a sufficient statistic for $\{P_i\}_{i=0}^m$. According to the definition of M, $(I - P)M = 0$ so that $PM = M$ and

$PMM^+ = MM^+$. However, $PMM^+ = MM^+$ implies that the range$(MM^+) \subset$ range(P) so that the minimality of k and the fact that MM^+ is an orthogonal projection imply that range(MM^+) = range(P) and hence that $MM^+ = P$.

Dropping the requirement that $\Omega_0 = I$ and $\eta_0 = 0$ we have the following.

Corollary 4.7. Let

$$A = \sum_{i=1}^{m}\{(\Omega_i^{-1} - \Omega_0^{-1})^2 + (\Omega_i^{-1}\eta_i - \Omega_0^{-1}\eta_0)(\Omega_i^{-1}\eta_i - \Omega_0^{-1}\eta_0)^T\},$$

then A^+A generates a sufficient statistic for $\{P_i\}_{i=0}^{m}$.

Proof: Since A is a symmetric matrix, it follows that A^+A is the orthogonal projection onto $[\ker A]^\perp$. If B is any $k \times n$ rank k matrix such that $B^+B = A^+A$ then $\ker B = \ker(A^+A) = \ker A$. Hence, B is a sufficient statistic for $\{P_i\}_{i=0}^{m}$ by Theorem 4.3.

Given a $k \times n$ rank k matrix B, let $V = \Omega_0 B^T(B\Omega_0 B^T)^{-1}$. Let $E_i(f(z))$ denote the expectation of the random variable f with respect to the measure P_i, $i = 0,1,\ldots,m$. Let $x_1,\ldots,x_N$ be identically distributed independent n-dimensional random variables such that x_1 is normally distributed. Define functions $H:R^{k\times N} \to R^n$ and $G:R^{k\times N} \to R^{n\times n}$ by

$$H(y_1,\ldots,y_N) = (I - VB)\eta_0 + V(\frac{1}{N}\sum_{j=1}^{N} y_j)$$

and

$$G(y_1,\ldots,y_N) = (I - VB)\Omega_0 + VB\Omega_0(I - VB)^T + V[\frac{1}{N-1}\sum_{j=1}^{N} (y_j - \bar{y})(y_j - \bar{y})^T]V^T$$

where $\bar{y} = \frac{1}{N}\sum_{j=1}^{N} y_j$.

Theorem 4.8. If B is a sufficient statistic for $\{P_i\}$ then $H(Bx_1,\ldots,Bx_N)$ and $G(Bx_1,\ldots,Bx_N)$ are unbiased consistent estimators of the mean and covariance of the measure P_i, $i = 0,1,\ldots,m$.

Proof:

$$E_i\{(I - VB)\eta_0 + V(\frac{1}{N}\sum_{j=1}^{N} Bx_j)\} = (I - VB)\eta_0 + VB\eta_i ,$$

which is $(I - VB)\eta_j + VB\eta_j = \eta_j$ by Theorem 4.2.

$$E\{(I - VB)\Omega_0 + VB\Omega_0(I - VB)^T + V[\frac{1}{N-1}\sum_{j=1}^{N} (B(x_j - \overline{x}))(B(x_j - \overline{x}))^T]V^T\}$$

$$= (I - VB)\Omega_0 + VB\Omega_0(I - VB)^T + VB\Omega_i B^T V^T$$

which, by Theorem 4.2, is equal to

$$(I - VB)\Omega_i + VB\Omega_i(I - VB)^T + VB\Omega_0(VB)^T$$

$$= (I - VB + VB)\Omega_i(I - VB + VB)^T$$

$$= \Omega_i$$

The consistency of these estimators follows from the continuity of the matrix operations.

The significance of this last theorem is that these estimators of the original parameters are functions of B and are $B^{-1}(R^k)$ measurable. Halmos and Savage in [6] indicate that "under suitable hypothesis, if there exists a maximum likelihood estimate R of some parameter, then R depends only on $Tx = y$ where T is a sufficient statistic (i.e., the estimator of R is T^{-1} measurable)." Theorem 5 then is in this sense a typical result concerning estimators of parameters and sufficient statistics.

It should be noted that although Theorems 4.1, 4.2 and 4.8 are stated for finite collections of measures, they clearly hold for arbitrary collections of measures.

V. BAYES SUFFICIENCY AND ITS RELATION TO STATISTICAL SUFFICIENCY

In this chapter we describe the relationship between sufficiency as described in this paper and Bayes sufficiency. We will no longer assume that a statistic T is linear. We will assume that T is a measurable function from a measurable space (X,S) <u>onto</u> a measurable space (Y,R).

Let $\{P_i\}_{i=1}^{m}$, $2 \leq m \leq \infty$, be a homogeneous collection of measures on (X,S), and define a collection of measures $\{h_i\}_{i=1}^{m}$ by $h_i(A) = P_i(T^{-1}(A))$ for $A \in R$. Observe that $\{h_i\}_{i=1}^{m}$ is a homogeneous set of measures.

If $\{\alpha_i\}_{i=1}^{m}$ is a set of prior probabilities for $\{P_i\}_{i=1}^{m}$ and π_i is the i^{th} population, then the Bayes classification rule on the random variable x can be stated as follows:

Classify $x \in \pi_i$ if and only if

$$\alpha_i \frac{dP_i}{dP_j}(x) \geq \alpha_j \quad \text{for} \quad j > i$$

and

$$\alpha_i \frac{dP_i}{dP_j}(x) > \alpha_j \quad \text{for} \quad j < i.$$

The Bayes classification on x using the random variable y can be stated as follows:

Classify $x \in \pi_i$ if and only if

$$\alpha_i \frac{dh_i}{dh_j}(Tx) \geq \alpha_j \qquad j > i$$

$$\alpha_i \frac{dh_i}{dh_j}(Tx) > \alpha_j \qquad j < i.$$

A statistic T is Bayes sufficient for $\{P_i\}_{i=1}^m$ if for each set of positive prior probabilities $\{\alpha_i\}_{i=1}^m$ the probability of misclassification using the Bayes classification rule for the random variable x is equal to the probability of misclassification using the Bayes classification rule for y.

We consider first the case $m = 2$.

Lemma 5.1. The probability of misclassification using x equals that of using y if and only if for $u \equiv P_1$,

$$(1) \quad \mu(\{x|\alpha_1 \frac{dh_1}{dh_2}(Tx) \geq \alpha_2 \text{ and } \alpha_1 \frac{dP_1}{dP_2}(x) < \alpha_2\}) = 0$$

$$(2) \quad \mu(\{x|\alpha_1 \frac{dh_1}{dh_2}(Tx) < \alpha_2 \text{ and } \alpha_1 \frac{dP_1}{dP_2}(x) > \alpha_2\}) = 0 .$$

Proof. The equivalence follows from the fact that conditions 1 and 2 imply that the Bayes rules, using x and y, are equal except on a set of measure zero.

Lemma 5.2. If T is a sufficient statistic for $\{P_1, P_2\}$ in the Bayes sense then T is a sufficient statistic for $\{P_1, P_2\}$.

Proof. By Lemma 5.1,

$$\{x|\alpha_1 \frac{dP_1}{dP_2}(x) \geq \alpha_2\} = \{x|\alpha_1 \frac{dh_1}{dh_2}(Tx) \geq \alpha_2\}[\mu] .$$

So for each rational number α let

$$B_\alpha = \{x| \frac{dP_1}{dP_2}(x) < \alpha \text{ and } \frac{dh_1}{dh_2}T(x) \geq \alpha\} .$$

Then $\mu(B_\alpha) = 0$ and hence $\mu(\underset{\alpha \text{ rat}}{\cup} B_\alpha) = 0$. Hence, for $A \in B(x)$

$$\int_A \frac{dP_1}{dP_2}\, d\mu = \int_A \frac{dh_1}{dh_2}\, T\, d\mu$$

and

$$\frac{dP_1}{dP_2} = \frac{dh_1}{dh_2} \circ T\ [\mu].$$

This means that

$$\frac{dP_1}{dP_2}\ (\epsilon) T^{-1}(B(Y))\ .$$

Theorem 5.3. A statistic T is sufficient for $\{P_i\}_{i=1}^m$, $2 \le m \le \infty$, if and only if T is a sufficient statistic in the Bayes sense.

Proof. If T is a sufficient statistic, then $\frac{dP_i}{dP_j} = \frac{dh_i}{dh_j} \circ T$ and so the Bayes classification using x is the same as Bayes classification using Y. Conversely, if T is sufficient for $\{P_i\}_{i=1}^m$ in the Bayes sense, then T is sufficient for $\{P_i, P_j\}$ in the Bayes sense. Hence $\frac{dP_i}{dP_j}\ (\epsilon) T^{-1}(B(Y))$ and so T is a sufficient statistic for $\{P_i\}_{i=1}^m$.

The following example shows that if a statistic T does not increase the probability of misclassification for a single set of prior probabilities, then T is not necessarily a sufficient statistic.

Example 5.4. Let

$$p_1(x,y) = \begin{cases} x + y & \text{for } (x,y) \in [0,1] \times [0,1] \\ 0 & \text{otherwise} \end{cases}$$

$$p_2(x,y) = \begin{cases} 1 - x + y & \text{for } (x,y) \in [0,1] \times [0,1] \\ 0 & \text{otherwise} \end{cases}$$

be the densities with respect to Lebesque measures on R^2 of P_1 and P_2 respectively. Then for $(x,y) \in [0,1] \times [0,1]$, $\frac{dP_1}{dP_2}(x,y) > 1$ if and only if $x > 1/2$. Now if the prior probabilities are equal, then the projection $P(x,y) = (x,0)$ minimizes the probability of error, since $\frac{dh_1}{dh_2} \circ T > 1$ if and only if $x > 1/2$. However, P is not a sufficient statistic by Theorem 2.2 and the fact that

$$\frac{dP_1}{dP_2}(1/4,\ 1/4) = 1/2 \neq 1 = \frac{dP_1}{dP_2}(1/4,\ 1/2).$$

Although this example shows the conditions in Lemma 5.2 cannot in general be relaxed, in the case of normal densities we will state the following theorem due to Peters [8] without proof.

Theorem 5.4. If a $k \times n$ linear statistic B of rank k does not increase the probability of error for a finite collection of normal densities and a fixed set of priors for which the Bayes decision regions are not empty, then B is a sufficient statistic.

References

1. Anderson, E. B., "Sufficiency and Exponential Families for Discrete Sample Spaces," J. Amer. Stat. Ass., 65 (1970), 1248.

2. Breiman, L., Probability. Menlo Park, California: Addison-Wesley Publishing Co., 1964.

3. Brown, L. D., "Sufficient Statistics in the Case of Independent Random Variables," Annals of Math. Stat., 35 (1968), 1456.

4. Decell, H. P., Peters, B. C., and Redner, R., Characterizations of Linear Sufficient Statistics, Report #59, N.A.S.A. Contract #NAS-9-15000, Houston, Texas, August 1976, Houston, Texas: University of Houston Printing, 1976.

5. Halmos, P. R., Measure Theory. New York: Van Nostrand Reinhold Co., 1950.

6. Halmos, P. R. and Savage, L. J., "Application of the Radon-Nikodym Theorem to the Theory of Sufficient Statistics," Annals of Math. Stat., 20 (1949), 225-241.

7. Johnson, N. L. and Kotz, S., Distributions in Statistics: Continuous Multivariate Distributions. New York: John Wiley and Sons, Inc., 1972.

8. Peters, B. C., private communications, 1976.

9. Tucker, Howard G., Probability and Mathematical Statistics, New York: Academic Press, 1967.