PROBABILITY
AND
MATHEMATICAL STATISTICS

Vol. 15 (1995), pp. 301–310

# ON THE DEFINITION OF PROBABILITY DENSITIES AND SUFFICIENCY OF THE LIKELIHOOD MAP

BY

D. A. S. FRASER (North York), PHILIP McDUNNOUGH (Toronto), AMIR NADERI (Tehran) and ANDRÉ PLANTE (Montréal)

*Abstract.* The definition and sufficiency of the likelihood map are discussed in detail. It is shown that the likelihood map is sufficient without any restriction on the family of probability measures and that it is minimal sufficient subject to weak conditions. The paper provides a general foundation for the use of the likelihood function and the likelihood function map in statistical inference.

## 1. INTRODUCTION

This paper gives the formal definition of likelihood map (see [5], Chapter 8 in [14], and Chapter 4 in [15]) and shows how this definition avoids certain measure theoretic difficulties met with other methods relating the concept of likelihood with that of minimal sufficiency.

Fisher [5]–[8] informally defines the generalized likelihood function statistic as $L(\theta, x) = c(x) f_\theta(x)$, where $f_\theta(x)$ is either a density or a probability function, and, for each $x$, $c$ is an arbitrary constant independent of the parameter $\theta$. Fisher also relates $L(\theta, x)$ to his concepts of sufficiency and of exhaustive statistic. Several authors who formalized and extended Fisher's results (Neyman [23], Halmos and Savage [17], Lehmann and Scheffé [21], Bahadur [1]) do not use the term *likelihood function* presumably because the arbitrary constant $c$ makes it difficult to think of $L(\theta, x)$ as a function of $\theta$ for a fixed $x$; rather, they use the Radon–Nikodym derivative $q(x, \theta)$ with respect to a finite dominating measure as a key ingredient.

Barndorff-Nielsen et al. [2] also use $q(x, \theta)$ rather than $L(\theta, x)$, although they do use the term *likelihood function* in reference to $q(x, \theta)$; and they provide a striking example of the type of measure theoretic difficulty implicit in the standard approach. In their example, there exists a real-valued minimal sufficient statistic $T$ for a real-valued parameter, but one version of the $q(x, \theta)$ induces a much finer partition of the sample space than that induced by $T$.

Their partial remedy for this difficulty includes a continuity assumption for $q(x, \theta)$ with respect to $\theta$.

The likelihood map, unlike $q(x, \theta)$, is uniquely defined, and it provides a formalization of $L(\theta, x)$ and its properties. The equivalence class $\{cf_\theta(x)\colon c > 0\}$ is called the *likelihood function* corresponding to $x$, where $f_\theta(x)$ is an almost everywhere uniquely defined Radon–Nikodym derivative called *probability density*, to be defined in Section 2. The *likelihood map* is the function $x \mapsto \{cf_\theta(x)\colon c > 0\}$. The measurability of the likelihood map is established in Section 3, and its sufficiency and minimal sufficiency are proved in Section 4. Section 5 records some concluding remarks.

Let $\{P_\theta\colon \theta \in \Omega\}$ be a family of probability measures on $(X, B)$ dominated by a $\sigma$-finite measure $\mu$, and let, for each $\theta \in \Omega$, $f_\theta(x)$ be the Radon-Nikodym derivative of $P_\theta$ with respect to $\mu$. The equivalence class $\{cf_\theta(x)\colon c > 0\}$ is called the *likelihood function* corresponding to $x$; two functions of $\theta$ are thus *equivalent* if one is a positive multiple of the other. The likelihood map is then defined as the function from $X$ to the space $R^\Omega$ which is the set of real-valued functions on $\Omega$ and contains all the possible likelihood functions. The term *likelihood* was introduced by Fisher [5]. The definition of the likelihood map including its minimal sufficiency is given in [9], [10], [14], [15], but is implicit in [5].

In typical cases, the probability density function corresponding to a probability measure is not uniquely defined. As a result the likelihood function need not be unique and the likelihood map may not be uniquely defined nor have nice properties. In view of this ambiguity one questions which version of the likelihood map to use among the many possible versions available. Section 2.1 addresses this important issue. Consider a Euclidean space and let $P$ be a probability measure absolutely continuous with respect to Lebesgue measure $m$. Since $P$ may be regarded as a mass distribution, it is natural to consider

$$f(x) = \lim_{n \to \infty} \frac{P(S_n)}{m(S_n)}$$

for all sequences $\{S_n\}$ of closed spheres of radius $r_n$ centered at $x$ with $r_n \to 0$ as $n \to \infty$, as the mass density at $x$; for further details concerning existence and uniqueness, see, e.g., [3], Chapter 6. Naderi [22] generalized this concept to the case of metric spaces and developed a special choice of the Radon–Nikodym derivative to be used as *the* probability density; the results are summarized in Section 2.1; this covers Euclidean spaces and related manifolds, but instances in more general contexts need further investigation. In the remainder of the paper, we assume that the probability densities are uniquely determined. Although this assumption is required for developing the arguments used later in the paper, the way the uniqueness is achieved plays no role in the arguments.

In Section 2.2 we define more formally the likelihood map, and then in Section 3 we show that it is measurable. The term *sufficiency* was introduced by Fisher [6]. In order to check the sufficiency of statistics, various forms of what has come to be known as the factorization theorem were established by Fisher [5], [6], Neyman [23], and Halmos and Savage [17]. Sufficiency is defined in Section 4.1 and the likelihood map without conditions is shown (Theorem 2) to be sufficient in Section 4.2.

The concept of sufficiency leads naturally to that of minimal sufficiency, the theory of which was first developed rigorously by Lehmann and Scheffé [21] and Dynkin [4]. Pitcher [24] constructed a family of probability measures for which a minimal sufficient statistic did not exist. Landers and Rogge [19] showed that even for a dominated family of probability measures a minimal sufficient statistic need not exist. Minimal sufficiency is defined in Section 4.2 and the likelihood map is shown (Theorem 3) under weak conditions to be minimal sufficient.

Barndorff-Nielsen et al. [2] show that any statistic $T$ generating the same partition of the sample space as the likelihood function is minimal sufficient, in the Bahadur [1] sense. In the present generalities, Theorem 4 in Section 4.3 shows that this is also the case under similar regularity conditions. Moreover, we note that Theorem 4 is a direct consequence of Theorem 3. It is interesting to note that Theorem 3 covers cases such as the uniform $(0, \theta)$ which Theorem 4 does not. Finally, we note that it is often conceptually useful to approach the notion of sufficiency using what could be termed measurable partitions [26], [22]. In our context this will simply refer to a partition of the relevant measurable space, the elements of which are all measurable (i.e. elements of $B^*$ as defined in Section 2.1).

## 2. DEFINITIONS

**2.1. Definition of probability densities.** Let $X$ be a metric space and let $B$ be the smallest $\sigma$-field of subsets of $X$ that contain the open subsets of $X$. Let $\mu$ be a $\sigma$-finite measure on $B$ and let $\mu$ be outer regular in the sense that, for each $E \in B$, $\mu(E) = \inf\{\mu(G)\colon\ G \supset E,\ G \text{ open}\}$. Further, let $(X, B^*, \mu)$ be the completion of $(X, B, \mu)$ and let $P$ be a probability measure on $B^*$ absolutely continuous with respect to $\mu$.

A sequence $\{E_n\}$ of sets of $B^*$ is said to *converge regularly* to $x \in X$ if there is a number $\alpha > 0$ and a sequence $\{S_n\}$ of closed spheres of radius $r_n$ such that $x \in E_n \subset S_n$, $\mu(E_n) \geqslant \alpha\mu(S_n)$, for all $n \geqslant 1$, and $r_n \to 0$ as $n \to \infty$. A class $V \subset B^*$ is said to cover a set $E \subset X$ in the *sense of Vitali* if for every $a \in E$ there is a sequence of sets in $V$ that converges regularly to $a$. A class $V \subset B^*$ is said to have the *Vitali covering property* $[\mu]$ if, for any set $E \subset X$ and class $V_0 \subset V$ that covers $E$ in the sense of Vitali, there is a sequence $\{E_n\}$

of disjoint sets in $V_0$ satisfying

$$\mu(E - \bigcup_{n=1}^{\infty} E_n) = 0.$$

The class of closed sets in the Lebesgue completion of the Borel subsets of $R^k$ has the Vitali covering property $[m]$ with respect to Lebesgue measure $m$. For a proof, see [16] or [25].

Let $V \subset B^*$ be a class that has the Vitali covering property $[\mu]$. Let $P(E_n)/\mu(E_n) = 0$ if $\mu(E_n) = 0$, and define

$$\bar{D}(P, \mu, V)(x) = \sup\{\limsup_n P(E_n)/\mu(E_n)\},$$

$$\underline{D}(P, \mu, V)(x) = \inf\{\liminf_n P(E_n)/\mu(E_n)\},$$

where the expression in braces denotes the limit superior (limit inferior) for a sequence $\{E_n\}$ of sets in $V$ converging regularly to $x$, and the supremum (infimum) is taken among all sequences of sets in $V$ converging regularly to $x$. The probability measure $P$ is said to be *differentiable* at $x$ with respect to $V$ and $\mu$ if

$$\underline{D}(P, \mu, V)(x) = \bar{D}(P, \mu, V)(x) < \infty.$$

The common value is called the *derivative* of $P$ at $x$ with respect to $V$ and $\mu$ and is denoted by $D(P, \mu, V)(x)$. The following results hold:

(i) $D(P, \mu, V)(x)$ exists a.e. $[\mu]$.

(ii) Let $f(x) = D(P, \mu, V)(x)$ if $P$ is differentiable at $x$ with respect to $V$ and $\mu$, and let $f(x) = 0$ otherwise; then $f$ is $B^*$-measurable and $P(E) = \int_E f(x)d\mu(x)$ for every $E \in B^*$.

For a detailed proof see, e.g., [22].

Now let $m$ be Lebesgue measure. As the probability measure $P$ may be regarded as a mass distribution, it is natural to consider

$$f(x) = \lim_{n \to \infty} P(S_n)/m(S_n)$$

for all sequences $\{S_n\}$ of closed spheres of radius $r_n$ centered at $x$ with $r_n \to 0$ (as $n \to \infty$), and view it as the mass density at $x$. By the above results, $f(x)$ exists a.e. $[m]$. In fact, as the class of closed sets has the Vitali covering $[m]$, if

$$f(x) = \lim_{n \to \infty} P(E_n)/m(E_n)$$

for all sequences $\{E_n\}$ of closed sets converging regularly to $x$, then, by the above results, $f(x)$ exists a.e. $[m]$.

This motivates us to give a formal definition of *probability density* in a metric space $X$; in particular, this avoids non-uniqueness. Let $F$ be the class of all closed sets in $B^*$, and assume that $F$ has the Vitali covering property $[\mu]$;

then $D(P, \mu, F)$ exists a.e. $[\mu]$. We put $f(x) = D(P, \mu, F)(x)$ if defined, and $f(x) = 0$ otherwise, and we infer that $f$ is measurable and $P(E) = \int_E f(x) d\mu(x)$ for every $E \in B^*$. This $f$ is said to be the *probability density* of $P$ with respect to $\mu$. Note that a local limiting operation has been used to define a unique determination of the Radon–Nikodym derivative of $P$ with respect to $\mu$; this is called *the* probability density of $P$ with respect to $\mu$. Under the usual definition of probability density, any Radon–Nikodym derivative of $P$ with respect to $\mu$ is a probability density of $P$ with respect to $\mu$ and there is non-uniqueness quite generally with this usual definition.

**2.2. Definition of the likelihood map.** Let $R_+^\Omega$ be the class of all functions from $\Omega$ to $R_+ = [0, \infty)$, and let $R_*^\Omega$ be defined by

$$R_*^\Omega = \{R(g) \colon g \in R_+^\Omega\},$$

where $R(g) = \{cg \colon c > 0\}$. Elements of $R_*^\Omega$ are rays from the origin in $R_+^\Omega$. Note that, for every $a > 0$, $R(ag) = R(g)$. For each $x \in X$, $f_\theta(x)$ is in $R_+^\Omega$; thus $R(f_\theta(x))$ is a ray in $R_*^\Omega$. It is called the *likelihood function corresponding to* $x$ or, simply, the *likelihood function.* The *likelihood map* is a function from $X$ to $R_*^\Omega$ which maps a point $x$ in $X$ to the likelihood function corresponding to $x$. As indicated in the Introduction we are assuming that there is a well-defined determination of the probability densities; correspondingly, the likelihood map is well defined.

## 3. MEASURABILITY OF THE LIKELIHOOD MAP

Let $L$ be the likelihood map from $X$ to $R_*^\Omega$, and let $R_{**}^\Omega$ be the class of all subsets $E$ of $R_*^\Omega$ with $L^{-1}(E) \in B^*$. Then $L$ is measurable $B^*/R_{**}^\Omega$. The following theorem shows that $R_{**}^\Omega$ contains the class of all sets of the form

$$C = \{R(g) \colon g \in A\},$$

called *measurable cone cylinder sets with base* $A$, where $A$ is a measurable cylinder whose base lies in the class of all finite-dimensional open, closed-open, and open-closed intervals.

THEOREM 1. (i) *Let $T$ be a transformation on the class of measurable cylinders defined by*

$$T(A) = \bigcup_{c>0} \{ca \colon a \in A\},$$

*and let $A_*$ be the class of all measurable cylinders whose bases lie in the class of all finite-dimensional open, closed-open, and open-closed rectangles. If $A \in A_*$, then $T(A) \in \sigma(A_*)$, where $\sigma(A_*)$ is the minimal $\sigma$-field over $A_*$.*

(ii) *Let $L$ be the likelihood map from $X$ to $R_*^\Omega = \{R(g) \colon g \in R_*^\Omega\}$. If $C$ is a measurable cone cylinder with a base in $A_*$, then $L^{-1}(C) \in B^*$.*

Proof. Let $A = \{g \in R_+^\Omega : (g_{\theta_1}, \ldots, g_{\theta_k}) \in I\}$ be a measurable cylinder with base $I$ of the form $I = I_1 \times \ldots \times I_k$, where $I_i = (\alpha_i, \beta_i]$, $0 \leq \alpha_i < \beta_i \leq \infty$, $i = 1, \ldots, k$. Further, let $I_r = rI = rI_1 \times \ldots \times rI_k$ for some rational number $r > 0$, and let $A_{I_r}$ be a measurable cylinder with base $I_r$. Then $T(A) = \bigcup_r A_{I_r}$. For let $x \in T(A)$. Then $x = c_0 g^0$ for some $c_0 > 0$ and $g^0 \in A$. Hence $g^0_{\theta_i} \in I_i = (\alpha_i, \beta_i]$, $i = 1, \ldots, k$, and thus $g^0_{\theta_i}/\beta_i \leq 1$ and $g^0_{\theta_i}/\alpha_i > 1$, $i = 1, \ldots, k$, where $g^0_{\theta_i}/\beta_i = 0$ for $\beta_i = \infty$, $g^0_{\theta_i}/\alpha_i = \infty$ for $\alpha_i = 0$. Let $g^0_{\theta_i}/\alpha_i = 1 + \varepsilon_i$, $i = 1, \ldots, k$, where $\varepsilon_i \in (0, \infty]$, and define $\varepsilon = \min\{\varepsilon_i, i = 1, \ldots, k\}$. Then

$$[1, 1+\varepsilon) \subset \bigcap_{i=1}^{k} [g^0_{\theta_i}/\beta_i, g^0_{\theta_i}/\alpha_i).$$

For let $\gamma \in [1, 1+\varepsilon)$. Then, for each $i$, $1 \leq i \leq k$, we have $g^0_{\theta_i}/\beta_i \leq 1 \leq \gamma < 1+\varepsilon \leq 1+\varepsilon_i = g^0_{\theta_i}/\alpha_i$. Hence

$$\gamma \in \bigcap_{i=1}^{k} [g^0_{\theta_i}/\beta_i, g^0_{\theta_i}/\alpha_i).$$

There exists a rational number $r_0$ such that $c_0 \leq r_0 < c_0(1+\varepsilon)$. Define a non-negative function $h$ on $\Omega$ by $h_{\theta_i} = c_0 g^0_{\theta_i}/r_0$, $i = 1, \ldots, k$. Since $r_0/c_0 \in [1, 1+\varepsilon)$, it follows that

$$r_0/c_0 \in \bigcap_{i=1}^{k} [g^0_{\theta_i}/\beta_i, g^0_{\theta_i}/\alpha_i).$$

Thus $h_{\theta_i} \in (\alpha_i, \beta_i]$, $i = 1, \ldots, k$. Therefore

$$(x_{\theta_1}, \ldots, x_{\theta_k}) = c_0(g^0_{\theta_1}, \ldots, g^0_{\theta_k}) = r_0(h_{\theta_1}, \ldots, h_{\theta_k}) \in r_0 I.$$

Hence $x \in A_{I_{r_0}}$, and thus $x \in \bigcup_r A_{I_r}$. Now let $x \in \bigcup_r A_{I_r}$. Then $x \in A_{I_v}$ for some rational number $q > 0$. Hence $(x_{\theta_1}, \ldots, x_{\theta_k}) \in I_q = qI$, and thus $x_{\theta_i} \in qI_i = q(\alpha_i, \beta_i]$, $i = 1, \ldots, k$. Therefore $\alpha_i < x_{\theta_i}/q \leq \beta_i$, $i = 1, \ldots, k$. Define a non-negative function $y$ on $\Omega$ by $y = x/q$. Since $\alpha_i < y_{\theta_i} \leq \beta_i$, $i = 1, \ldots, k$, it follows that $(y_{\theta_i}, \ldots, y_{\theta_k}) \in I$. Hence $y \in A$, and thus $x = qy \in T(A)$. Therefore $T(A) = \bigcup_r A_{I_r}$. Then, since $A_{I_r}$ is in $A_*$ and $A_*$ generates $\sigma(A_*)$, the preceding implies that $T(A) \in \sigma(A_*)$.

A completely analogous argument shows that, for every measurable cylinder $A \in A_*$ with a base in the class of all finite-dimensional closed-open rectangles or with a base in the class of all finite-dimensional open rectangles, $T(A) \in \sigma(A_*)$.

To prove (ii), let $L_1$ be a function from $X$ to $R_+^\Omega$ which maps every $x$ in $X$ into $f_\theta(x)$ in $R_+^\Omega$, and let $L_2$ be a function from $R_+^\Omega$ to $R_*^\Omega$ which maps every $g$ in $R_+^\Omega$ into $R(g)$ in $R_*^\Omega$. Further, let $L$ be the likelihood map. Then it is obvious that $L = L_2 \circ L_1$. If $A = \{g: (g_{\theta_1}, \ldots, g_{\theta_k}) \in I\}$ is a measurable cylinder in $A_*$, then the set

$$L_1^{-1}(A) = \{x \in X: L_1(x) \in A\} = \{x \in X: (f_{\theta_1}(x), \ldots, f_{\theta_k}(x)) \in I\}$$

lies in $B^*$, since for each $\theta_i$, $1 \leqslant i \leqslant k$, $f_{\theta_i}(x)$ is $B^*$-measurable. Since $A_*$ generates $\sigma(A_*)$, it follows that the function $L_1$ is measurable $B^*/\sigma(A_*)$. Further, if $C = \{R(g)\colon g \in A\}$ is a measurable cone cylinder with base $A \in A_*$, then the set

$$L_2^{-1}(C) = \{cg\colon c > 0, g \in A\} = \{\bigcup_{c>0} cg\colon g \in A\}$$

lies in $\sigma(A_*)$ by (i). Since $L^{-1}(C) = L_1^{-1}(L_2^{-1}(C))$, it follows that $L^{-1}(C) \in B^*$. The proof of (ii) is complete.

## 4. SUFFICIENCY

**4.1. Background.** The term *sufficiency* was introduced by Fisher [6]. The factorization theorem was proved in various forms by Fisher [6], Neyman [23], and Halmos and Savage [17]. Following Halmos and Savage, we define measurable function $s(x)$ to be *sufficient* if for each $E$ in $B^*$ there is a determination of the conditional probability $P_\theta(E|t)$ which is $\theta$-free.

**4.2. Sufficiency of the likelihood map.** We now proceed to prove that the likelihood map from $(X, B^*)$ to $(R_*^\Omega, R_{**}^\Omega)$ is sufficient.

THEOREM 2. *The likelihood map is sufficient for* $P_* = \{P_\theta\colon \theta \in \Omega\}$.

Proof. Since the family $P_*$ of probability measures is dominated by $\mu$, there is a countable subfamily $P_*^0 = \{P_\theta\colon \theta \in \Omega_*\}$ of $P_*$ which is equivalent to $P_*$ (see [17]). Let $\Omega_* = \{\theta_n^*\} \subset \Omega$, and define a probability measure $P$ on $B^*$ by

$$P(E) = \sum_{n=1}^{\infty} P_{\theta_n^*}(E)/2^n.$$

Let $f(x)$ be the probability density of $P$ with respect to $\mu$. Then

$$f(x) = \sum_{n=1}^{\infty} P_{\theta_n^*}(x)/2^n \quad \text{for every } x \in N_*^c,$$

where $N_*$ is a set of $\mu$-measure zero. Let $N_0 = \{x\colon f(x) = 0\}$. Then

$$P(N_0) = \int_{N_0} f(x)\,d\mu(x) = 0.$$

Since $P(N_0) = 0$, and $P$ is equivalent to $P_*$, it follows that $P_\theta(N_0) = 0$ for every $\theta \in \Omega$. Let $N = N_* \cup N_0$. Then, since $\mu$ dominates $P_*$ and $N_0$ is a null set for $P_*$, $N$ is a null set for $P_*$. For each $\theta \in \Omega$, define the function $g_\theta(t)$ on $R_*^\Omega$ by

$$g_\theta(t) = \begin{cases} f_\theta(x)/f(x) & \text{for all } x \in N^c \cap L^{-1}(t), \\ 0 & \text{otherwise.} \end{cases}$$

Let $a \in L^{-1}(t) \cap N^c$ and note that $L^{-1}(t) \cap N^c$ is the set of all points $x$ in $N^c$ for which there is a function $h(x, a) \neq 0$, independent of $\theta$, such that $f_\theta(x) = h(x, a) f_\theta(a)$ for all $\theta \in \Omega$. Hence, for every $x \in N^c \cap L^{-1}(t)$,

$$f(x) = \sum_{n=1}^{\infty} \frac{f_{\theta_n^*}(x)}{2^n} = \sum_{n=1}^{\infty} h(x, a) \frac{f_{\theta_n^*}(a)}{2^n} = h(x, a) f(a),$$

and thus $f_\theta(x)/f(x) = f_\theta(a)/f(a)$. Therefore the function $g_\theta(t)$ is well defined. The function $g_\theta(t)$ is $R_{**}^\Omega$-measurable. For let $B \in \sigma(R_+)$, where $\sigma(R_+)$ is the minimal $\sigma$-field over $R_+$, and define the function $h_\theta$ on $N^c$ by $h_\theta(x) = f_\theta(x)/f(x)$. Then $g_\theta \circ L = h_\theta$, and $L^{-1}(g_\theta^{-1}(B)) = h_\theta^{-1}(B)$. Since $h_\theta$ is $B^*$-measurable, it follows that $h_\theta^{-1}(B) \in B^*$. Hence $L^{-1}(g_\theta^{-1}(B)) \in B^*$, and thus $g_\theta^{-1}(B) \in R_{**}^\Omega$. Therefore $g_\theta(t)$ is $R_{**}^\Omega$-measurable. Moreover, $f_\theta(x) = g_\theta(L(x)) f(x)$ for any $x \in N^c$. Hence $f_\theta(x) = g_\theta(L(x)) f(x)$ a.e. $[\mu]$, and thus $L$ is sufficient for $P_*$ by the factorization theorem (see, e.g., [20]).

**4.3. Minimal sufficiency.** The theory of minimal sufficiency was initiated as exhaustiveness by Fisher [6] and followed later by Lehmann and Scheffé [21] and Dynkin [4]. A sufficient statistic $s(x)$ is said to be *minimal sufficient* if it is a function of any other sufficient statistics a.e. (with respect to $\{P_\theta\colon \theta \in \Omega\}$). We now examine the minimal sufficiency of the likelihood map:

THEOREM 3. *Assume that $\Omega_0$ is a countable subset of $\Omega$, and that for each $\theta \in \Omega$ there is a sequence $\{\theta_n\}$ of points in $\{\Omega_0\}$ satisfying one of the following two equivalent conditions:*

(i) $f_{\theta_n} \to f_\theta$ *in $\mu$-measure,*

(ii) $f_{\theta_n} \to f_\theta$ *in $L^1$.*

*Then the likelihood map $L$ from $X$ to $R_*^{\Omega_0}$ is minimal sufficient for $P_* = \{P_\theta\colon \theta \in \Omega\}$, and the partition of $X$ induced by $L$ is measurable.*

Let $P_0$ denote the class of probability measures indexed by $\Omega_0$. The proof can be divided into several steps: $L$ is minimal sufficient for $P_0$; $L$ is sufficient for $P$; every null set for $P_0$ is a null set for $P$; $L$ is minimal sufficient for $P$; the partition of $X$ induced by $L$ is measurable. The proof follows the pattern used by Lehmann and Scheffé [21] in their construction of the minimal sufficient statistic for a dominated family of probability measures. For details in the present notation see the more general framework by Naderi [22].

If $X$ is a Euclidean space, then the class of all real-valued integrable functions on $X$ is a separable metric space with distance defined by $d(g, h) = \int_X |g-h| \, d\mu(x)$. Since every subspace of a separable metric space is separable, the class of probability densities is a separable metric space. Thus, in this case, the assumptions of Theorem 3 are satisfied. The next result is a direct consequence of Theorem 3. The assumptions, however, are stronger than those of Theorem 3 and do not cover, for example, the uniform $(0, \theta)$, $\theta \in \Omega$.

THEOREM 4. *Let $\Omega$ be a separable metric space, and let, for each $x \in X, f_\theta(x)$ be continuous in $\theta$. Then the likelihood map L from X to $R_+^\Phi$, where $\Phi$ is any countable dense subset of $\Omega$, is minimal sufficient for P, and the partition of X induced by L is measurable.*

Proof. For each $\theta \in \Omega$, there is a sequence $\{\theta_n\}$ of points in $\Phi$ which converges to $\theta$. Since $f_\theta$ is continuous in $\theta$, it follows that $f_{\theta_n}(x) \to f_\theta(x)$ for all $x \in X$. Hence $f_{\theta_n} \to f_\theta$ a.e., and thus $f_{\theta_n} \to f_\theta$ in $\mu$-measure, and the results follow from Theorem 3.

## 5. CONCLUDING REMARKS

The concept of sufficiency has played a central role in the area of statistical inference. There has, unfortunately, been some confusion surrounding the existence of a best or minimal sufficient statistic. In this article, we have provided results which show the existence of such a statistic under very broad conditions. In particular, we give a rigorous foundation connecting sufficiency with the notion of the likelihood map. The key element here is an appropriate definition of likelihood based on a precise definition of a probability density. Finally, we note that since our results are derived in the context of general metric spaces, they may be applied to many problems arising in the study of inference for stochastic processes.

We express our appreciation to many readers for helpful comments on two earlier drafts. Partial support by the Natural Sciences and Engineering Research Council of Canada is gratefully acknowledged.

## REFERENCES

[1] R. R. Bahadur, *Sufficiency and statistical decision functions*, Ann. Math. Statist. 25 (1954), pp. 423–462.

[2] O. Barndorff-Nielsen, J. Hoffmann-Jørgensen and L. Pedersen, *On the minimal sufficiency of the likelihood function*, Scand. J. Statist. 3 (1976), pp. 37–38.

[3] D. L. Cohn, *Measure Theory*, Birkhäuser, Boston 1980.

[4] E. B. Dynkin, *Necessary and sufficient statistics for a family of probability distributions*, Uspekhi Mat. Nauk 6 (1951), pp. 68–90.

[5] R. A. Fisher, *On the 'probable error' of a coefficient of correlation deduced from a small sample*, Metron I (1921), pp. 3–32. Reprinted in: *Contributions to Mathematical Statistics*, Wiley, New York 1950.

[6] — *On the mathematical foundations of theoretical statistics*, Philos. Trans. Roy. Soc. London Sec. A 222 (1922). Reprinted in: *Contributions to Mathematical Statistics*, Wiley, New York 1950.

[7] — *Two new properties of mathematical likelihood*, Proc. Roy. Soc. Ser. A 144 (1934), pp. 285–307.

[8] — *Statistical Methods and Scientific Inference*, Oliver and Boyd, London 1956.

[9] D. A. S. Fraser, *On sufficiency and the exponential family*, J. Roy. Statist. Soc. Ser. B 25 (1963), pp. 115–123.
[10] – *On the sufficiency and likelihood principle*, J. Amer. Statist. Assoc. 58 (1963), pp. 641–647.
[11] – *Sufficiency for regular models*, Sankhyā Ser. A 28 (1966), pp. 137–144.
[12] – *On sufficiency and conditional sufficiency*, ibidem 28 (1966), pp. 145–150.
[13] – *Sufficiency for selection models*, ibidem 28 (1966), pp. 329–334.
[14] – *Probability and Statistics, Theory and Applications*, Duxbury, New York – ITS, Toronto 1976.
[15] – *Inference and Linear Models*, McGraw-Hill, New York 1979.
[16] H. Hahn and A. Rosenthal, *Set Functions*, The University of New Mexico Press, Albuquerque 1948.
[17] P. R. Halmos and L. J. Savage, *Application of the Radon–Nikodym theorem to the theory of sufficient statistics*, Ann. Math. Statist. 20 (1949), pp. 225–241.
[18] C. A. Hayes and C. Y. Pauc, *Derivation and Martingales*, Springer, Berlin 1970.
[19] D. Landers and L. Rogge, *Sufficient σ-fields and minimal sufficient statistics. Two counterexamples*, Ann. Math. Statist. 43 (1972), pp. 2045–2049.
[20] E. L. Lehmann, *Theory of Point Estimation*, Wiley, New York 1983.
[21] – and H. Scheffé, *Completeness, similar regions, and unbiased estimation*, Sankhyā 10 (1950), pp. 305–340.
[22] A. Naderi, *Measurability and other properties of likelihood map*, Ph. D. Thesis, University of Toronto, 1988.
[23] J. Neyman, *Sur un teorema concernente le cosidette statistiche sufficienti*, Giorn. Ist. Ital. Att. 6 (1935), pp. 320–334.
[24] T. S. Pitcher, *Sets of measures not admitting necessary and sufficient statistics or subfields*, Ann. Math. Statist. 28 (1957), pp. 267–268.
[25] S. Saks, *Theory of the Integrals*, 2nd rev. ed., Hafner, New York 1937.
[26] S. D. Silvey, *Statistical Inference*, Chapman and Hall, London 1975.

D. A. S. Fraser
University of Toronto
Toronto, M5S 1A1 Canada

Philip McDunnough
University of Toronto
Toronto, M5S 1A1 Canada

Amir Naderi
Sharif University of Technology
Tehran, Iran

André Plante
Université de Québec à Montréal
Montréal, H3C 3P8 Canada

*Received on 8.3.1994*