ON CONTINUOUS SUFFICIENT STATISTICS[1]

By J. L. Denny

Indiana University

1. Summary. In [1], [2] we construct for each integer $n \geqq 2$, a real-valued, bounded, uniformly continuous statistic defined on R^n, nondecreasing in each real argument, which is a minimal sufficient statistic for the family of all probability distributions defined on the Borel field β^n in R^n and dominated by Lebesgue measure λ_n. In this paper let $\{P_\theta\}$ be a family of probability distributions dominated by Lebesgue measure and defined on the restriction of β^n to a Borel set $A \subset R^n$. Let $f = (f_1, \cdots, f_k)$ and $g = (g_{11}, \cdots, g_{1n_1}, \cdots, g_{k1}, \cdots g_{kn_k})$ be continuous sufficient statistics for $\{P_\theta\}$ defined on A, with f_i and g_{ij} real-valued. If there are k functions $h_i: R^1 \to R^{n_i}$, $i = 1, \cdots, k$ so that $(g_{i1}, \cdots, g_{in_i}) = h_i \circ f_i$ a.e. (λ_n), then is g everywhere a continuous function of f, i.e., $g = h \circ f$ for continuous $h: f[A] \to g[A]$? If in addition $n_i = 1$, $i = 1, \cdots, k$ and each h_i is a 1-1 function, are f and g identical, i.e., $g = h \circ f$ for bicontinuous $h: f[A] \to g[A]$? Now if (1) A is connected, (2) A has a dense interior, and (3) almost every linear section of each f_i (and g_i in the second case) satisfy Lusin's condition (N), the answer to the above questions is affirmative (see Section 2 for definitions). But if at least one of (1), (2), or (3) is not satisfied, an affirmative answer is not in general possible (see Examples, Section 3). In Section 5 we show that this implies it is not possible to find a real-valued continuous minimal sufficient statistic f defined on R^n such that almost every linear section of f satisfies Lusin's condition (N), for some familiar probability distributions.

2. Definitions. If the set $A \subset R^1$ and the function $f: A \to R^1$ are such that $\lambda_1\{f[N]\} = 0$ for each Lebesgue set $N \subset A$ such that $\lambda_1\{N\} = 0$, then the function f is said to satisfy Lusin's condition (N) on A. The definition of sections of functions and sets will be found on p. 134 of [6]. Let the real-valued function f be defined on $A \subset R^n$, $n > 1$, and let $\mathrm{Pr}^j[A]$ be the image of A by each of the n projections of R^n onto R^{n-1}, $j = 1, \cdots, n$. A linear section of A at $x \in \mathrm{Pr}^j[A]$ will mean a subset of R^1. A linear section of f at $x \in \mathrm{Pr}^j[A]$ means a section of f defined on the linear section of A at $x \in \mathrm{Pr}^j[A]$. Let the Borel set $A \subset R^n$, $n > 1$, and let $B^j = \{x:$ the linear section of f at $x \in \mathrm{Pr}^j[A]$ satisfies Lusin's condition $(N)\}$. If B^j is a Lebesgue set with $\lambda_{n-1}\{\mathrm{Pr}^j[A] \sim B^j\} = 0$, $j = 1, \cdots, n$, we will say that almost every linear section of f satisfies Lusin's condition (N). Let $x^1 = (x_1^1, \cdots, x_n^1)$, $x^2 = (x_1^2, \cdots, x_n^2) \in R^n$. Let $P(x^1, x^2) \subset R^n$ denote the union of the possibly degenerate segments (see, e.g., p. 155 of [3]) joining $(x_1^1, \cdots, x_n^1)$ with $(x_1^2, x_2^1, \cdots, x_n^1)$, $(x_1^2, x_2^1, \cdots, x_n^1)$ with $(x_1^2, x_2^2, x_3^1, \cdots, x_n^1)$, $\cdots$, and

Received 17 September 1963.

[1] At different times, the author received support from the Office of Naval Research, Contract NONR-222(43) and the Air Force, Contract AFOSR-8163. Reproduction in whole or in part is permitted for any purpose of the United States Government.

$(x_1^2, \cdots, x_{n-1}^2, x_n^1)$ with $(x_1^2, \cdots, x_n^2)$. For $x^1, \cdots, x^m \varepsilon R^n$ we define $P(x^1, \cdots, x^m) = \bigcup_{j=1}^{m-1} P(x^j, x^{j+1})$.

3. Requisite lemmas. In the following two lemmas f and g are real-valued continuous functions defined on $[a, b] \subset R^1$, $B \subset [a, b]$ is a Lebesgue set such that $\lambda_1\{[a, b] \sim B\} = 0$, and h is a real-valued function defined on $f[B]$ such that $g = h \circ f$ on B.

LEMMA 1. *If $f(a) = f(b)$, if f is greater than or equal to $f(a)$ on neighborhoods of a and b, and if f satisfies Lusin's condition (N), then $g(a) = g(b)$.*

PROOF. Since $f[B]$ is dense in $f[[a, b]]$, we need consider only the case where each neighborhood of a and each neighborhood of b contains at least one point where f is strictly greater than $f(a)$. Fix $\epsilon > 0$ and choose $\alpha \varepsilon (a, b)$ so that $f(\alpha) > f(a)$ and $|g(a) - g(y)| < \epsilon$ for $y \varepsilon (a, \alpha)$. Choose $\beta \varepsilon (\alpha, b)$ so that $f(\beta) > f(b)$, $|g(b) - g(y)| < \epsilon$ for $y \varepsilon (\beta, b)$, and $f[(\beta, b)] \subset f[[a, \alpha]]$. There exists $\alpha_1 \varepsilon B \cap (a, \alpha)$ and $\beta_1 \varepsilon B \cap (\cap, b)$ such that

$$f(\alpha_1) = f(\beta_1). \tag{1}$$

For if (1) is not true then $f[B \cap (a, \alpha)] \cap f[B \cap (\beta, b)]$ is empty and so $f[B \cap (\beta, b)] \subset f[([a, b] \sim B) \cap [a, \alpha]]$. Now $f[([a, b] \sim B) \cap [a, \alpha]]$ and $f[([a, b] \sim B) \cap (\beta, b)]$ have measure zero and since $f(\beta) > f(b)$ a contradiction follows. Thus by (1), $g(\alpha_1) = h(f(\alpha_1)) = h(f(\beta_1)) = g(\beta_1)$, and the continuity of g gives the assertion.

LEMMA 2. *If $f(a) = f(b)$ and if f satisfies Lusin's condition (N) then $g(a) = g(b)$.*

PROOF. By Lemma 1 we need consider only the case where at least one point, say a, can be approximated by points $y, z \varepsilon [a, b]$ which are arbitrarily close to a and so that $f(y) < f(a) < f(z)$ (if $f \leqq f(a)$ about a and b use Lemma 1 on $-f$). Fix $\epsilon > 0$ and choose intervals (a, α) and (β, b) so that $|g(a) - g(y)| < \epsilon$ for $y \varepsilon (a, \alpha)$ and $|g(b) - g(y)| < \epsilon$ for $y \varepsilon (\beta, b)$. The assumptions grant α_1, $\alpha_2 \varepsilon (a, b)$ so that $\alpha_1 \varepsilon (a, \alpha)$, $\alpha_2 \varepsilon (\alpha_1, \alpha)$ and $f(\alpha_1) = f(a)$, $f > f(a)$ on (α_1, α_2). Let $x_1 = \sup\{y : y \varepsilon (\alpha_2, b], f(y) = f(a)$, for each interval $[w, y]$ f is strictly greater than $f(a)$ at some point in $[w, y]\}$. An ab contrario argument shows the supremum is over a nonempty set. Evidently x_1 belongs to the above set and $f(x_1) = f(a)$. Thus there is $[x_2, x_3] \subset [\alpha_2, x_1]$ such that $f(x_3) = f(a), f > f(a)$ on $[x_2, x_3)$, and $|g(x_1) - g(y)| < \epsilon$ for $y \varepsilon (x_2, x_3]$. Lemma 1 then implies $g(\alpha_1) = g(x_3)$ and hence $g(\alpha_1) = g(x_3)$ and hence $|g(a) - g(x_1)| < 2\epsilon$. If $x_1 = b$ continuity of g gives the assertion. If $x_1 < b$ then f is easily checked to be less than or equal to $f(a)$ on $[x_1, b]$. Since $f(x_1) = f(b)$, Lemma 1 applied to $-f$ gives $g(x_1) = g(b)$. Again $|g(b) - g(a)| < 2\epsilon$ and the assertion is proved.

LEMMA 3. *Let $A \subset R^1$ be an interval and let $B \subset A$ be a Lebesgue set such that $\lambda_1\{A \sim B\} = 0$. Let f and g be continuous real-valued functions defined on A and let h be a real-valued function defined on $f[B]$ such that $g = h \circ f$ on B. If f satisfies Lusin's condition (N) then $g = \varphi \circ f$ on A for continuous φ defined on $f[A]$.*

PROOF. Lemma 2 implies $g = \varphi \circ f$ *on A for some function φ*. For each $[a, b] \subset A$

the restriction of φ to $f[[a, b]]$ is continuous (see, eg., p. 95 of [5]). The convexity of the set $f[A]$ then ensures the continuity of φ on $f[A]$.

LEMMA 4. *Let $A \subset R^n$ be a product of n open intervals and let $B \subset A$ be a Lebesgue set such that $\lambda_n\{A \sim B\} = 0$. Let f and g be continuous real-valued functions defined on A and let h be a real-valued function defined on $f[B]$ such that $g = h \circ f$ on B. If almost every linear section of f satisfies Lusin's condition (N) then $g = \varphi \circ f$ on A for continuous φ defined on $f[A]$.*

PROOF. For fixed y^1, $y^2 \varepsilon A$ and $\epsilon > 0$ there are $x^1, \cdots, x^m \varepsilon A$ such that: (1) $\|y^1 - x^1\| < \epsilon$, $\|y^2 - x^m\| < \epsilon$ and the linear sections of f on $P(x^1, \cdots, x^m)$ satisfies Lusin's condition (N); (2) each linear section of $P(x^1, \cdots, x^m)$ has all its linear Lebesgue measure on $B \cap P(x^1, \cdots, x^m)$; (Fubini theorem). Lemma 4 ensures the assertion ($f[A]$ is still convex).

REMARK. No analogue of this result seems available—for the entire set A—when continuous f and g take their values in R^n, $n \geqq 2$. For letting $A = (-\frac{1}{2}, \frac{1}{2}) \times (0, 1)$, $f(x, y) = (x^2y, xy^2)$, and $g(x, y) = (x, y)$, we have $g = h \circ f$ on $A \sim \{0\} \times (0, 1)$ but not on A.

From Lemma 4 and a well-known property of connected open sets in R^n (see, eg., problem 4, p. 90 of [3]) we obtain

COROLLARY 1. *Let the Borel set $A \subset R^n$ have a connected dense interior and let $B \subset A$ be a Lebesgue set with $\lambda_n\{A \sim B\} = 0$. Let f and g be continuous real-valued functions defined on A and let h be a real-valued function defined on $f[B]$ such that $g = h \circ f$ on B. If almost every linear section of f satisfies Lusin's condition (N) then $g = \varphi \circ f$ on A for continuous φ defined on $f[A]$.*

Corollary 1 yields

COROLLARY 2. *If in Corollary 1, A is only assumed to be a connected Borel set with a dense interior, the conclusions of Corollary 1 hold.*

REMARK. Corollaries 1 and 2 hold for non-Lebesgue sets. There are connected non-Lebesgue sets with connected dense interiors—start with a connected open set in R^2 with a frontier of positive measure.

EXAMPLES. Here is a real-valued continuous function f defined on [0, 1] such that $f(0) = f(1) = 0, f(\frac{1}{2}) = \frac{1}{2}$, and the restriction of f to a Borel set $B \subset [0, 1]$ with $\lambda_1\{B\} = 1$ is one-one. With $g(x) = x$ for $x \varepsilon [0, 1]$ we have $g = k \circ f$ on B. Let ψ be a real-valued, continuous, and strictly increasing function on [0, 1], $\psi(0) = 0, \psi(1) = 1$, and $\psi'(x) = 0$ a.e. (λ_1), and let $h = (1 - \psi)/2(1 - \psi(1/2))$. Let the Borel set $D \subset [0, 1]$ with $\lambda_1\{D\} = 1$, $\lambda_1\{h[D]\} = 0$, (see, eg., p. 271 of [7]) and let $B' = ([0, \frac{1}{2}) \sim h[D]) \cup (D \cap [\frac{1}{2}, 1])$. Let the Borel set $B \subset B'$ with $\lambda_1\{B\} = 1$ and define $f(x) = x$ for $x \varepsilon [0, \frac{1}{2})$, $f(x) = h(x)$ for $x \varepsilon [\frac{1}{2}, 1]$.

The reader can quickly construct two real-valued continuous functions on $(0, 1) \cup (1, 2)$ so that (2) and (3) of Section 1 are satisfied and Lemma 3 is false.

Let $C \subset [0, 1]$ be a Cantor-type set with $\lambda_1\{C\} > 0$ and let $y \varepsilon [0, 1] \sim C$. Let $D = ([0, 1] \times C) \cup ([0, 1] \times \{y\}) \cup (\{0\} \times [0, 1]) \cup (\{1\} \times [0, 1])$. The reader can easily find continuous real-valued f and g on D so that $g = h \circ f$ a.e.

(λ_2) but not everywhere, and every linear section of f and g satisfy Lusin's condition (N).

4. A Theorem. From Corollary 2 we have the

THEOREM. *Let $\{P_\theta\}$ be a family of probability distributions defined on a connected Borel set $A \subset R^n$ with a dense interior. Let $f = (f_1, \cdots, f_k)$ and $g = (g_{11}, \cdots, g_{1n_1}, \cdots, g_{k1}, \cdots, g_{kn_k})$ be continuous sufficient statistics for $\{P_\theta\}$ such that $f_i: A \to R^1$ and $g_{ij}: A \to R^1$. Let $B \subset A$ be a Lebesgue set with $\lambda_n\{A \sim B\} = 0$ and let $h_i: f_i[B] \to R^{n_i}$ so that $(g_{i1}, \cdots, g_{in_i}) = h_i \circ f_i$ on B. If almost every linear section of each f_i satisfies Lusin's condition (N) then $g = h \circ f$ on A for continuous $h: f[A] \to g[A]$.*

COROLLARY. *If in the above theorem $n_i = 1$, $i = 1, \cdots, k$, $f_i = \psi_i \circ g_i$ on B for $\psi_i: g[B] \to R^1$ and if almost every linear section of each g_i satisfies Lusin's condition (N) then $g = h \circ f$ on A for bicontinuous $h: f[A] \to g[A]$.*

5. Applications. Let f be a real-valued continuous minimal sufficient statistic for the uniform distribution. That is, if $g(x_1, \cdots, x_n) = (\min_i x_i, \max_i x_i)$ then $g = h \circ f$ a.e. (λ_n) for some h. We show that not almost every linear section of f satisfies Lusin's condition (N). Arguing ab contrario, by the Theorem, $g = h \circ f$ for continuous h.

Assume $x_1 \equiv \min_i x_i < \max_i x_i \equiv x_n$ and let $x_1 \varepsilon [a_1, b_1]$, $x_n \varepsilon [a_n, b_n]$ with $b_1 < a_n$.

Let $C = [a_1, b_1] \times \{x_2\} \times \cdots \times \{x_{n-1}\} \times [a_n, b_n]$. Then $f(x) \neq f(y)$ for $x, y \varepsilon C$ implies $g(x) \neq g(y)$, i.e., $f = \psi \circ g$ on C for some ψ. It follows (see, eg., p. 95 of [5]) that on C $g = h \circ f$ for bicontinuous $h: f[C] \to g[C]$. This is a contradiction—$f[C]$ is 1-dimensional and $g[C]$ is 2-dimensional (see, eg., p. 24 of [4]).

Let $D \subset R^n$ contain a connected Borel set B with a dense interior and let $\{P_\theta\}$ be a family of probability distributions defined on the restriction of β^n to D and dominated by Lebesgue measure. Assume that $g(x_1, \cdots, x_n) = (\sum_{i=1}^n x_i^{j_1}, \cdots, \sum_{i=1}^n x_i^{j_k})$, $n \geqq k > 1$, $j_i < j_{i+1}$ positive integers, is a sufficient statistic for $\{P_\theta\}$ and if f is any minimal sufficient statistic for $\{P_\theta\}$ then $g = h \circ f$ a.e. (λ_n) on B, for a function h. If f is a real-valued continuous minimal sufficient statistic for $\{P_\theta\}$ then not almost every linear section of f satisfies Lusin's condition (N). For example, there is not a real-valued continuous minimal sufficient statistic defined on R^n for the normal distribution, meeting the Lusin condition (N). The proof, which mimics the preceding except for a routine use of Jacobians, is left to the reader.

REMARK. If the Borel set $A \subset R^n$ is such that $\lambda_n\{R^n \sim A\} = 0$ then the probability space $(A, \beta^n(A), \lambda_n)$, where $\beta^n(A)$ is the restriction of β^n to A, is equivalent for all statistical purposes to $(R^n, \beta^n, \lambda_n)$. Using [2] one can construct such an equivalent probability space for which there is a statistic possessing all the properties described in Section 1 and in addition the property that every linear section satisfies Lusin's condition (N). Suppose for arbitrary $\epsilon > 0$ we are given not the value of the statistic but only that the value lies in $(a, a + \epsilon)$

for suitable $a \varepsilon R^1$. Then in both cases, $(R^n, \beta^n, \lambda_n)$ and $(A, \beta^n(A), \lambda_n)$, we have no information.

6. Acknowledgment. This paper contains part of the author's dissertation written at the University of California, Berkeley, under the direction of Professor E. W. Barankin. To him and to Professors L. M. LeCam and A. P. Morse go my thanks for encouragement.

REFERENCES

[1] DENNY, J. L. (1962). Minimal dimension of sufficient statistics. Unpublished doctoral dissertation, Univ. of California, Berkeley.

[2] DENNY, J. L. (1962). A real-valued continuous function on R^n almost everywhere 1-1. To appear in *Fund. Math.*

[3] DIEUDONNE, J. (1960). *Foundations of Modern Analysis*. Academic Press, New York.

[4] HUREWICZ, WITOLD and WALLMAN, HENRY (1941). *Dimension Theory*. Princeton Univ. Press.

[5] KELLEY, JOHN L. (1955). *General Topology*. Van Nostrand, Princeton.

[6] LOÈVE, MICHEL (1963). *Probability Theory* (3rd ed.). Van Nostrand, Princeton.

[7] SAKS, STANISLAW (1937). *Theory of the Integral* (2nd ed.). Hafner, New York.