The Annals of Mathematical Statistics
1969, Vol. 40, No. 4, 1474–1476

NOTES

NOTE ON A THEOREM OF DYNKIN ON THE DIMENSION OF SUFFICIENT STATISTICS[1]

By J. L. Denny

The University of Arizona

1. Summary. We show that the existence of a continuous minimal sufficient statistic not equivalent to the order statistics, for $n \geqq 2$ independent observations, is not a sufficient condition for the family of densities, assumed to be Lipschitz, to be an exponential family. This result is intended to be compared with a theorem of Dynkin (p. 24 of [3]) which asserts that the existence of a sufficient statistic not equivalent to the order statistics implies that the family of densities is an exponential family, provided that the densities possess continuous derivatives.

2. Result. Let $I \subset R$ be an interval and let $\{p(\cdot, \theta)\}$ be a family of positive probability densities defined on I. Following L. Brown (p. 1461 of [1]) we will say that a sufficient statistic $\phi: I^n \rightarrow R^k$ for $n \geqq 2$ independent observations is trivial if there is a nonempty open set $U \subset I^n$ and a Borel set $B \subset U$ where the Lebesgue measure $\lambda_n(U \sim B) = 0$ so that the restriction of ϕ to B is one-one. A sufficient statistic ϕ is then said to be nontrivial if it is false that ϕ is trivial. The definition of nontrivial involves local properties of ϕ, and to say that a sufficient statistic is nontrivial is to say that for each nonempty open set $U \subset I^n$ the restriction of ϕ to U is not statistically equivalent to the restriction of the order statistics to U. The following theorem of Dynkin [3], corrected by L. Brown, appears on p. 1461 of [1]:

THEOREM. *Let $\{p(\cdot, \theta)\}$ be a family of probability densities on an interval I such that for each θ, $p(\cdot, \theta)$ is continuous on I, is bounded away from zero on I, and is continuously differentiable on I. Suppose there is a nontrivial sufficient statistic ϕ for θ on the basis of n independent observations. Then $\{p(\cdot, \theta)\}$ is a ρ-parameter exponential family where $n > \rho$.*

Our purpose is to show that Dynkin's theorem depends on the assumption that the densities possess continuous derivatives, in the sense that Dynkin's theorem does not generalize to densities which are Lipschitz. As noted by a referee, the definition of Lipschitz function is not standardized and the following definition is included: a function $f: I \rightarrow R$ is Lipschitz if there is a finite constant K so that for each $x, y \in I$, $|f(x) - f(y)| \leqq K|x - y|$. It may be helpful to recall that if f is Lipschitz then the derivative f' exists and is bounded almost everywhere (λ_1) and, moreover, if a function $g: I \rightarrow R$ possesses a finite derivative $g'(x)$ for each

Received 1 August 1968.

[1] This research was supported in part by the National Science Foundation under GP-8741.

$x \varepsilon I$ then g is Lipschitz if and only if g' is a bounded function. The construction in the following theorem depends on the fact that Lipschitz functions may have pleasant measure-theoretic properties and yet be nowhere locally one-one with a nonvanishing derivative almost everywhere, unlike continuously differentiable functions.

THEOREM. *Let $I \subset R$ be a bounded interval. There exists a family of probability densities $\{p(\cdot, \theta): \theta \varepsilon \Theta\}$ such that* (i) *for each $\theta \varepsilon \Theta\, p(\cdot, \theta)$ is bounded away from zero and is Lipschitz*; (ii) *for each Borel set $B \subset I$ such that $\lambda_1(B) > 0$, $\{p(\cdot, \theta)\}$ is not representable as an exponential family when restricted to B*; (iii) *for each integer $n \geqq 2$ there is a continuous Euclidean-valued nontrivial minimal sufficient statistic ϕ for the family. Moreover, for each n ϕ may be taken to be of the form $\phi(y_1, \cdots, y_n) = \sum_{i=1}^{n} \phi_{in}(y_i)$, where the ϕ_{in} are real-valued continuous functions on $(-1, 1)$ and the y_i are defined below.*

PROOF. We assume for notational convenience that the closure of I is $[0, 1]$. Let Θ be a family of bounded real-valued continuous functions on $(-1, 1)$ such that (i) $\theta \varepsilon \Theta$ implies the derivative of θ exists and is continuous on $(-1, 1)$; (ii) if $C \subset (-1, 1)$ is a Borel set of positive measure and if $\Theta(C)$ is the family of functions which are the restrictions to C of the members of Θ then the smallest real linear space containing $\Theta(C)$ is not finite-dimensional (as a referee observed, the family of functions $\{x^n: n = 0, 1, \cdots\}$ meets the requirements on Θ; a proof of this fact may be given by choosing a compact $C' \subset C$ of positive measure, noticing that the smallest linear space containing $\{x^n\}$ restricted to C' is an algebra which separates points of C', and using the Stone-Weierstrass theorem). Let continuous $f: I \rightarrow (-1, 1)$ satisfy the following: (i) f is Lipschitz and $\lim f(x) = 0$ as $x \downarrow 0$; (ii) for almost every $x \varepsilon I$ $|f'(x)| = 1$; (iii) f is nowhere locally monotone (the canonical construction of such functions is as follows: let $A \subset I$ be a Borel set such that for each nonempty open $U \subset [0, 1]\, \lambda_1(A \cap U) > 0$, $\lambda_1(A^c \cap U) > 0$, and define $f(x) = \int_0^x (I_A - I_{A^c})\, d\lambda_1$ where I_c is the indicator function of the set C).

For each $\theta \varepsilon \Theta$, $x \varepsilon I$ define $p(x, \theta) = c(\theta)\exp\theta(f(x))$ where $c(\theta)$ is the normalizing constant. Then $p(\cdot, \theta)$ is Lipschitz and bounded away from zero. To say that $\{p(\cdot, \theta)\}$ is not representable as an exponential family when restricted to B is to say that there do not exist $m < \infty$ Borel functions $g_1, \cdots, g_m$ defined on B so that for each $\theta \varepsilon \Theta$ there is Borel $B(\theta) \subset B$ with $\lambda_1(B \sim B(\theta)) = 0$ and $\theta(f(x)) = \sum_{i=1}^{m} c_i(\theta) g_i(x)$ for each $x \varepsilon B(\theta)$ and real constants $c_i(\theta)$. Suppose for some such Borel B with $\lambda_1(B) > 0$ the g_i exist. Then there is compact $C \subset B$ with $\lambda_1(C) > 0$ so that each g_i restricted to C is continuous. Choose a Borel set $D \subset C$ such that each point of D is a point of density of D. Then $\theta(f(x)) = \sum_{i=1}^{m} c_i(\theta) g_i(x)$ for each $x \varepsilon D$ and the family of functions $\theta(f(\cdot))$, $\theta \varepsilon \Theta$, on D generates a finite-dimensional vector space. This leads to a contradiction since $\lambda_1(D) > 0$ implies $\lambda_1(f(D)) > 0$. Now $T(x_1, \cdots, x_n) = (f(x_1), \cdots, f(x_n))$ is a sufficient statistic and it is easy to verify that $F(x_1, \cdots, x_n) = (y_1, \cdots, y_n)$, where y_i is the ith smallest $f(x_j)$, is a minimal sufficient statistic. To prove F is nontrivial it suffices to prove that T is nontrivial and to prove that T is nontrivial it suffices to prove: if $C \subset I$ is a compact nondegenerate interval and Borel $E \subset C$ satisfies $\lambda_1(C \sim E) = 0$ then it is false that the restriction of f to E is one-one. Suppose on the contrary for some such C and E, f is one-one on E. Let $A_1 = \{x:f'(x) = 1\} \cap E$ and $A_2 = \{x:f'(x) = -1\} \cap E$. Then a computation yields $\lambda_1(f(A_1 \cup A_2)) = \lambda_1(C)$ and thus there is x_0 and y_0 in C so that $f(x_0) - f(y_0) = \lambda_1(C)$. We assume $y_0 < x_0$, for the other case is analagous. Also $f(x_0) - f(y_0) \leqq \lambda_1(f([y_0, x_0] \cap A_1)) = \lambda_1([y_0, x_0] \cap A_1) < < \lambda_1([y_0, x_0]) \leqq \lambda_1(C)$ where the first inequality follows from Theorem 6.6 on p. 280 of [5]. This is the contradiction.

Finally, if $C_1 \subset f(I)$ and $\lambda_1(C_1) = \lambda_1(f(I))$ then $\lambda_1(f^{-1}(C_1)) = \lambda_1(I)$ since $|f'(x)| = 1$ almost everywhere. Then by Fubini's theorem, if $B_1 \subset F(I^n)$ satisfies $\lambda_n(F(I^n) \sim B_1) = 0$ then $\lambda_n(F^{-1}(B_1)) = \lambda_n(I^n)$. Let continuous $\phi:F(I^n) \rightarrow R$ have the property that there is Borel $B_1 \subset F(I^n)$ such that $\lambda_n(B_1) = \lambda_n(F(I^n))$ and the restriction of ϕ to B_1 is one-one. Then $\phi \circ F$ is the claimed real-valued continuous nontrivial minimal sufficient statistic. In particular let $\phi(y_1, \cdots, y_n) = \sum_{i=1}^{n} \phi_{in}(y_i)$ be one of the functions used by A. N. Kolmogorov in [4]. Such functions are almost everywhere one-one (see p. 103 of [6], and for another example of almost everywhere one-one functions see [2]). This completes the proof.

REFERENCES

[1] Brown, L. (1964). Sufficient statistics in the case of independent random variables. *Ann. Math. Statist.* **35** 1456–1474.

[2] Denny, J. L. (1964). A continuous real-valued function on E^n almost everywhere 1–1. *Fund. Math.* **55** 95–99.

[3] Dynkin, E. B. (1951). Necessary and sufficient statistics for a family of probability distributions. *Select. Transl. Math. Statist. Prob.* **1** 23–41.

[4] Kolmogorov, A. N. (1957). On the representation of continuous functions of several variables by finite superposition of continuous functions of one variable and addition. *Dokl. Akad. Nauk SSSR.* **114** 369–373.

[5] Saks, S. (1937). *Theory of the Integral* (2nd ed.) Hafner, New York.

[6] Sprecher, D. A. (1966). On admissibility in representations of functions of several variables as finite sums of functions of one variable. *Numerical Solutions of Partial Differential Equations.* Academic Press, New York.