

SUFFICIENT STATISTICS AND DISCRETE EXPONENTIAL FAMILIES

BY J. L. DENNY

The University of Arizona

$\{P_\theta\}$ is a set of probabilities on a countable set χ such that $P_\theta(x) > 0$ for each x and θ . We prove that if $\{P_\theta\}$ is not an exponential family, then each sufficient statistic for n independent observations must be one-to-one, modulo permutations, on an infinite product set (which does not depend on the sufficient statistic).

The purpose of this note is to prove an elementary proposition which says, roughly, that if one has a class of discrete probability distributions and a sufficient statistic for n independent observations, then either one has an exponential family or else the statistic is equivalent to the order statistics, with positive probability.

χ denotes a countable set, Θ is the arbitrary parameter set, and $\{P_\theta : \theta \in \Theta\}$ is a set of probabilities satisfying $P_\theta(x) > 0$ for $x \in \chi$, $\theta \in \Theta$. If for fixed $\theta_0 \in \Theta$ there are $m < \infty$ real functions g_i on χ and $m + 1$ real functions c_i on Θ so that $dP_\theta/dP_{\theta_0}(x) = c_0(\theta) \exp\{\sum_1^m c_i(\theta)g_i(x)\}$ for all $\theta \in \Theta$ and $x \in \chi$, then $\{P_\theta\}$ is said to be an exponential family.

THEOREM. *Suppose $\{P_\theta\}$ is not an exponential family. Then there is an infinite subset $A \subset \chi$ with the following property. If T_n is a sufficient statistic for n independent observations from $\{P_\theta\}$, then*

$(x_{i_1}, \dots, x_{i_n}) \in A^n, (y_{i_1}, \dots, y_{i_n}) \in A^n$ and $T_n(x_{i_1}, \dots, x_{i_n}) = T_n(y_{i_1}, \dots, y_{i_n})$ implies $(x_{i_1}, \dots, x_{i_n}) = (y_{i_{\sigma(1)}}, \dots, y_{i_{\sigma(n)}})$ for at least one permutation σ , i.e., T_n is equivalent to the sample, modulo a permutation, on A^n .

Following the proof of the theorem we give an application which generalizes, in the discrete case, a previous result of the author [2]. E. B. Dynkin proved an analogous theorem for absolutely continuous distributions defined on an interval $I \subset R$ where the analogue of our set A is an open interval $U \subset I$ [4]. A corrected version of Dynkin's theorem appears on page 1461 of a paper of L. Brown [1]. However, Dynkin assumes, among other things, that the densities be continuously differentiable and it turns out that this assumption is crucial. For if the densities are only assumed to be Lipschitz, Dynkin's theorem breaks down [3]. In other words, families of distributions which are either discrete or else possess continuously differentiable densities are either exponential or else do not permit a true reduction of the data with no loss of information; this sort

Received July 12, 1971; revised December 27, 1971.

AMS 1970 subject classifications. Primary 62B05; Secondary 62E10.

Key words and phrases. Sufficient statistics, exponential families, Darms-Koopman families.

of informal remark cannot be made for families possessing densities assumed only to be continuous.

Finally, the converse to our theorem is false. For there is a one-parameter exponential family such that each sufficient statistic for n independent observations is equivalent to the sample, modulo a permutation. To verify this, choose $f: \chi \rightarrow H$, where H is a bounded countable set, linearly independent over the rationals, and f is one-to-one. Let P_{θ_0} be arbitrary and define $dP_{\theta}/dP_{\theta_0}(x) = c(\theta) \exp \{\theta \cdot f(x)\}$, $\theta \in R$. The assertion follows since $\sum_1^n f(x_{i_j})$ is a minimal sufficient statistic.

The theorem's proof rests on the following elementary

LEMMA. *Let Z^+ denote the positive integers and let $L(Z^+)$ be a real linear space of real functions defined on Z^+ ($L(Z^+)$ is a set of sequences). The following two properties are equivalent: (a) $L(Z^+)$ has infinite linear dimension; (b) there is a countably infinite set $A \subset Z^+$ such that if $L(A)$ denotes the real linear space of functions which are the restrictions to A of functions in $L(Z^+)$, then the weak closure of $L(A)$ is the space of all real functions on A .*

PROOF. Clearly (b) implies (a) and we prove the converse. Now $L(Z^+)$ has infinite dimension if and only if $k(n)$, the linear dimension of the space $L(n)$ of functions in $L(Z^+)$ restricted to $\{1, 2, \dots, n\}$, tends to infinity as n tends to infinity. Fix n and let $g_1, \dots, g_{k(n)} \in L(Z^+)$ be such that their restriction to $\{1, \dots, n\}$ is a basis for $L(n)$, $k(n) \leq n$. Then, there exists $i_1 < \dots < i_{k(n)} \leq n$ such that $g_1, \dots, g_{k(n)}$ is a basis for the linear space of functions in $L(n)$ restricted to $B_n = \{i_1, \dots, i_{k(n)}\}$. Choose the smallest $n_1 > n$ so that the linear dimension of $L(n_1)$ is at least $k(n) + 1$. We assert there is $g \in L(n_1)$ such that $g_1, \dots, g_{k(n)}, g$ are linearly independent on $B_{n+1} = B_n \cup \{n_1\}$. For suppose $g \in L(n_1)$ implies $g = \sum_1^{k(n)} b_j g_j$ on B_{n+1} . Since also $g = \sum_1^{k(n)} a_j g_j$ on $\{1, \dots, n_1 - 1\}$, we obtain by the uniqueness of the basis coefficients that $g = \sum_1^{k(n)} b_j g_j$ on $\{1, \dots, n_1\}$, the desired contradiction. Define thus an increasing sequence $\{B_n\}$, set $A = \cup B_n$, and let $\{x_{i_j}\}$ be an arbitrary real sequence defined on A . As $(x_{i_1}, \dots, x_{i_n})$ belongs to the linear space of real functions on $\{i_1, \dots, i_n\} \subset A$ for each n , we conclude $\{x_{i_j}\}$ belongs to the weak closure of $L(A)$.

PROOF OF THE THEOREM. Without loss of generality let $\chi = Z^+$. Let $\{g_s: s \in S\}$ be a basis for the smallest real linear space $L(Z^+)$ containing the functions $\log dP_{\theta}/dP_{\theta_0}$, $\theta \in \Theta$. Then for n independent observations, the function $(i_1, \dots, i_n) \rightarrow \{\sum_{j=1}^n g_s(i_j): s \in S\}$, $i_j \in Z^+$, is a function of each sufficient statistic (at the suggestion of a referee we include a proof of this fact). Let then $T_n: (Z^+)^n \rightarrow \mathcal{Y}$, \mathcal{Y} a set, be a sufficient statistic: $dP_{\theta,n}/dP_{\theta_0,n}(i_1, \dots, i_n) = h(T_n(i_1, \dots, i_n), \theta)$, for some function h (the functions in the factorization theorem not depending on θ cancel out). If $g_t \in \{g_s: s \in S\}$ then $g_t = \sum_{k=1}^m a_k \log dP_{\theta_k}/dP_{\theta_0}$ and hence $\sum_{j=1}^n g_t(i_j) = \sum_{k=1}^m a_k \log h(T_n(i_1, \dots, i_n), \theta_k) = h_t(T_n(i_1, \dots, i_n))$, as claimed. Next, note that $\{P_{\theta}\}$ is an exponential family if and only if S is a finite set. Assume that S is not a finite set and let $A \subset Z^+$ be obtained by the lemma.

Each $f \in L(A)$ is of the form $f = \sum_1^m a_k g'_{s_k}$ where each g'_{s_k} is the restriction to A of an element of $\{g_s : s \in S\}$. For each $p \in A$ let $I_p : A \rightarrow \{0, 1\}$ be the indicator function of $\{p\}$. By the lemma I_p belongs to the weak closure of $L(A)$: $I_p(i) = \lim_m \sum_{k=1}^{r(m)} a_{m,k} g'_{m,s_k}(i)$ for each $i \in A$. Thus for $(i_1, \dots, i_n) \in A^n$, $\sum_{j=1}^n I_p(i_j) = \lim_m (\sum_{k=1}^{r(m)} a_{m,k} (\sum_{j=1}^n g'_{m,s_k}(i_j))) = \lim_m (\sum_{k=1}^{r(m)} a_{m,k} h_{m,s_k}(T_{n,A}(i_1, \dots, i_n)))$, where $T_{n,A}$ is the restriction of T_n to A^n . Thus for each $p \in A$, $\sum_{j=1}^n I_p(i_j) = \phi_{p,n}(T_{n,A}(i_1, \dots, i_n))$ for $(i_1, \dots, i_n) \in A^n$ (the usual measure-theoretic qualifications have been omitted in this discrete set-up). But $(i_1, \dots, i_n) \in A^n$ can be obtained from a permutation of $(k_1, \dots, k_n) \in A^n$ if and only if $\sum_{j=1}^n I_p(i_j) = \sum_{j=1}^n I_p(k_j)$ for each $p \in A$. Thus T_n is one-to-one, modulo a permutation, on A^n and the theorem is proved.

The proof of the following application is omitted.

COROLLARY. *Let $f_i, i = 1, \dots, k$ be real functions on χ such that (a) for each integer $n, (x_{i_1}, \dots, x_{i_n}) \rightarrow (\sum_1^n f_1(x_{i_j}), \dots, \sum_1^n f_k(x_{i_j}))$ is a sufficient statistic for n independent observations from $\{P_\theta\}$; (b) there is an integer p such that for each $i = 1, \dots, k$, at most p of the numbers $\{f_i(x_j) : j = 1, 2, \dots\}$ are linearly independent over the rationals. Then $\{P_\theta\}$ is an exponential family and $dP_\theta/dP_{\theta_0}(x) = c_0(\theta) \exp \{\sum_1^k c_i(\theta) f_i(x)\}$ for each $x \in \chi, \theta \in \Theta$.*

REFERENCES

- [1] BROWN, L. (1964). Sufficient statistics in the case of independent random variables. *Ann. Math. Statist.* **35** 1456-1474.
- [2] DENNY, J. L. (1967). Sufficient conditions for a family of probabilities to be exponential. *Proc. Nat. Acad. Sci. USA* **57** 1184-1187.
- [3] DENNY, J. L. (1969). Note on a theorem of Dynkin on the dimension of sufficient statistics. *Ann. Math. Statist.* **40** 1479-1476.
- [4] DYNKIN, E. B. (1951). Necessary and sufficient statistics for a family of probability distributions. *Select. Transl. Math. Statist. Prob.* **1** 23-41.