SUFFICIENT CONDITIONS FOR A FAMILY OF PROBABILITIES TO BE EXPONENTIAL

BY J. L. DENNY

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA, (RIVERSIDE)

Communicated by J. Neyman, March 6, 1967

We make the following statement precise under fairly weak conditions: in an experiment, if we summarize $n$ statistically independent observations $(x_1, \ldots, x_n)$ in $m < n$ real numbers $(y_1, \ldots, y_m)$, where $y_j = \sum_{i=1}^{n} f_i(x_i)$ and the $f_i$ are given functions, and if we assume we have lost no information by the summary, then the family of probabilities associated with the experiment must be an exponential family.

Let $(\mathcal{X}, \mathcal{A}, \{P_t: t \in T\})$ be fixed, where $T$ is a set, $\mathcal{A}$ is a sigma-algebra of subsets of $\mathcal{X}$, and $\{P_t\}$ is a family of probabilities, which satisfy $P_t(A) = 0$ if and only if $P_t(A) = 0$ for $(t, t', A) \in T \times T \times \mathcal{A}$. We say that $\{P_t\}$ is an exponential family if for a fixed $t_0 \in T$ there are $p + 1$ real-valued functions $c_j$ on $T$ and $p$ real-valued Borel functions $\varphi_j$ on $\mathcal{X}$, $\varphi_j^{-1}(B) \in \mathcal{A}$ when $B \subset R$ is a Borel set, so that

$$P_t(A) = \int_A c_0(t) \exp \left( \sum_{j=1}^{p} c_j(t) \varphi_j(x) \right) P_{t_0}(dx)$$

for $(t, A) \in T \times \mathcal{A}$.

Familiar examples of exponential families are the multivariate Gaussian and the Poisson distributions, and examples of nonexponential families are the Cauchy and the Weibull distributions.

Let $B \subset R^n$ be a Lebesgue set (union of a Borel set and a subset of a null Borel set), and for each $x \in B$ let $B(x)$ be a Lebesgue set for which the Lebesgue measure of the symmetric difference of $B$ and $B(x)$ is zero.

Lemma 1. If each $x \in B$ is a point of density of $B$ and if $C = \{x + y: x \in B, y \in B(x)\}$ is a Lebesgue set, then there is nonvoid open $U$ and a null set $N$ so that $C \cup N \supset U$.

Proof: By Mueller's proof the algebraic sum $B + B$ is open. It is sufficient to prove that each point of $B + B$ is a point of density of $C$ and this is a consequence of the equivalence of $B$ and $B(x)$.

Lemma 2. Let $U \subset R^n$ be an open neighborhood of the origin. Let the Lebesgue set $B \subset U$ satisfy $B \cup N = U$, where $N$ is a null set, and also let $B$ and $B(x), x \in B$, satisfy the hypotheses of Lemma 1. If $\varphi: B \cup (\{B(x): x \in B\}) \cup (\{x + B(x): x \in B\}) \rightarrow R$ satisfies the condition that $\varphi$ is a bounded Borel function when restricted to each bounded Borel subset of its domain and that $\varphi(x + y) = \varphi(x) + \varphi(y)$ when $x \in B$ and $y \in B(x)$, then there is a Borel subset $C$ of the domain of $\varphi$, which carries all the measure of the domain and a real linear map $\varphi^-$ on $R^n$ so that $\varphi = \varphi^-$ on $C$.

Proof: Let $D$ be the closure of the domain of $\varphi$. Define the functions $\tilde{\varphi}$ and $\tilde{\psi}$ on $D \cap (U + U)$ by $\tilde{\varphi}(x) = \text{ess} \lim_{y \rightarrow x} \varphi(y)$ and $\tilde{\psi}(x) = \text{ess} \lim_{y \rightarrow x} \varphi(y)$. It is elementary to show that $\tilde{\varphi}(x + y) \leq \tilde{\varphi}(x) + \tilde{\varphi}(y), \tilde{\psi}(x + y) \geq \tilde{\psi}(x) + \tilde{\psi}(y)$, and then that $\tilde{\varphi}(x) = \tilde{\psi}(x), x \in D \cap (U + U)$. Because of the boundedness of $\varphi$, there is a
real linear map $\varphi^-$ on $R^n$ so that $\varphi^-(x) = \mathcal{J}(x)$ on $D \cap (U + U)$, and it then follows that $\varphi^- = \varphi$ almost everywhere.

If $M \subset R^n, \varphi: M \rightarrow R,$ and $\varphi$ restricted to each Borel subset of $M$ is a Borel function, then we say that $\varphi$ is locally Borel. If $h$ maps $x$ into $R^p$, then $\psi(h): x^* \rightarrow R^p$

is the function whose value at $(x_1, \ldots, x_n) \in x^*$ is $\sum_{i=1}^{n} h(x_i)$. The product measure $P^*_i$ is defined on the product sigma-algebra $\mathfrak{B}^*.

**LEMMA 1.** Let $g$ and $f_1, \ldots, f_m$ be real-valued Borel functions on $x, f = (f_1, \ldots, f_m)$, satisfying (i) $g(x)$ is a bounded set; (ii) for some integer $k$ if $A \in \mathfrak{B}$ and $P^i_k(A)$ is positive, then the inner Lebesgue measure of $\psi_k(f)(A)$ is positive; (iii) there is a locally Borel function $\varphi: M \rightarrow R$ satisfying $\psi_n(g) = \varphi \circ \psi_n(f)$ on $A_n \in \mathfrak{B}$ with $P^i_k(A_n) = 1$ for $n = 4k$. Then there are real constants $a_0, \ldots, a_m$ so that $P_i(g = a_0 + \sum_{j=1}^{m} a_j f_j) = 1$.

**Proof:** Condition (iii) ensures that $P^i_k(A) > 0$ implies that there is a Borel subset $B$ of the density points of $\psi_k(f)(A)$ so that $P^i_k((\psi_k(f))^{-1}(B)) \geq P^i_k(A)$. This implies that the infimum of the real subset of inner Lebesgue measure of $\psi_k(f)(A)$: $A \in \mathfrak{B}, P^i_k(A) = 1$ is positive and is obtained at some $A \in \mathfrak{B}$ for which we may assume that $\psi_k(f)(A)$ is a Borel set. Hence, by the theorem of Fubini and Lemma 1, it follows that $\psi_{2k}(f)(A_{2k}) \cup N = U$ where $A_{2k} \in \mathfrak{B}$ is a set of probability one, $N$ is a null set, and $U$ is an open set. For the same reasons concerning $\psi_k(f)$, we may assume $\psi_{2k}(f)(A_{2k})$ is a Borel set, each point of which is a density point. Condition (iii) evidently holds on $A_n \cap (A_{2k} \times A_{2k})$, and we can translate $f$ so that simultaneously $U$ is an open neighborhood of the origin and for some $(x, y) \in A_n \cap (A_{2k} \times A_{2k})$ for which the section $x'$ of $A_n \cap (A_{2k} \times A_{2k})$ at $x$ has probability one, $\psi_{2k}(f)(x)$ is the origin. Therefore, $\varphi$ satisfies the hypotheses of Lemma 2 if we set $B = \psi_{2k}(f)(A^*)$ for an appropriate $A^* \subset A_{2k}$ for which $P_{2k}^i(A^*) = 1$ (we may choose $A^* = A_{2k} \cap A')$. This implies $P_{2k}^i(\psi_k(g) = \varphi \circ \psi_k(f)) = 1$ where $\varphi$ is a real linear map, and taking into account the translation of $f$, we obtain the lemma.

Since for fixed $t_0 \in T$ and $\varepsilon > 0$ each Borel function $g$ is bounded on a set having probability at least $1 - \varepsilon$, it is elementary to obtain

**LEMMA 4.** The conclusion of Lemma 3 holds if assumption (i) is omitted.

**LEMMA 5.** Let $g$ and $f_1, \ldots, f_m$ be real-valued Borel functions on $x, f = (f_1, \ldots, f_m)$, satisfying (i) for some integer $k$, fixed $t_1 \in T$ and $\alpha \in (0, 1)$, if $A \in \mathfrak{B}$ and $P^i_k(A)$ $\geq \alpha$, then the inner Lebesgue measure of $\psi_k(f)(A)$ is positive; (ii) for $n \geq k$ there is a locally Borel function $\varphi_n: M_n \rightarrow R$ so that $\psi_n(g) = \varphi_n \circ \psi_n(f)$ on $A_n \in \mathfrak{B}$ with $P^i_k(A_n) = 1$. Then the conclusion of Lemma 3 holds.

**Proof:** Condition (i) implies that there is $C \in \mathfrak{B}$ so that $P^i_k(C) is positive and $D \in \mathfrak{B}$ and $P^i_k(C \cap D)$ positive implies the inner measure of $\psi_k(f)(D)$ is positive. Now if $E$ and $F$ are nonvoid sets and each point of $E$ is a density point of $E$, then each point of $E + F$ is a density point of $E + F$. Therefore, defining $\gamma = C \times \mathfrak{B} \subset \mathfrak{B}^{+1}$, it follows that $f_1, \ldots, f_m$ and $g$ satisfy the assumptions of Lemma 3 on $\gamma$, and this fact leads directly to the proof.

In the same way we obtain

**LEMMA 6.** Replace (i) in Lemma 5 by (i') for each $\psi_n(f)(A)$ is a Lebesgue set for each $A \in \mathfrak{B}$, and for some integer $k$, if $A \in \mathfrak{B}$ and $P^i_k(A) = 1$, then the Lebesgue measure of $\psi_k(f)(A)$ is positive. Then the conclusion of Lemma 3 holds.
I am indebted to William Gustin for suggesting the proof of the next lemma.

**Lemma 7.** Let $f_1, \ldots, f_m$ be real-valued continuous functions defined on a connected compact topological space $Y$, $f = (f_1, \ldots, f_m)$, so that $a_0 + \sum_{j=1}^{m} a_j f_j = 0$ implies all the real constants $a_j$ are zero. If $n \geq m$, then $\psi_n(f)(y)$ has a nonempty interior.

**Proof:** We may assume each $f_j(x_0) = 0$ for $x_0 \in Y$. The set $\psi_n(f)(y)$ is the algebraic sum $\sum_{\lambda=1}^{n} Q_{\lambda}$ where $Q_{\lambda} = f(y)$. The $Q_{\lambda}$ satisfy the hypotheses of the theorem of Gustin and Green\(^2\), and this completes the proof.

Let $E = \{a_0, \ldots, a_n\} \subset Q$, the rationals, where $a_0 = 0$. For sufficiently large $m$, if $\varphi$ is a real function on the algebraic sum of $E$ $2m$ times satisfying $\varphi(\sum_{i=1}^{2m} \tau_i) = \varphi(\sum_{i=1}^{m} \tau_i) + \varphi(\sum_{i=m+1}^{2m} \tau_i)$, then $\varphi(a_i) = a_i \varphi(a_i)/a_i$. This fact and an approximation by sets of probability $1 - \epsilon$ gives

**Lemma 8.** Let $g$ and $f_1, \ldots, f_m$ be real-valued Borel functions on $X$ so that with probability one $f_j(X) \subset Q$ for each $j$. If for $n \geq n_0$ there is a locally Borel function $\varphi_n:M_n \to R$ so that $\psi_n(g) = \varphi_n \circ \psi_n(f)$ on $A_n \subset \mathbb{R}^n$ with $P_n(A_n) = 1$, then the conclusion of Lemma 3 holds.

Each closed additive subgroup $G \subset R^m$ is isomorphic to $R^p \times Z^q$ ($0 \leq p + q \leq m$) and the isomorphism preserves Borel sets, null sets with respect to the Haar measures on $G$ and $R^p \times Z^q$, and Lebesgue sets. This fact and Lemmas 5, 6, and 8 lead to the following

**Theorem 1.** Let $f_1, \ldots, f_m$ be real-valued Borel functions on $X$, $f = (f_1, \ldots, f_m)$, satisfying either (i) for some integer $k$, fixed $t_1 \in T$ and $\alpha \in (0,1)$, if $\Lambda \subset \mathbb{R}^k$ and $P_k^\alpha(\Lambda) \geq \alpha$, then the inner Haar measure of $\psi_k(f)(A)$ obtained from the smallest closed additive group in $R^m$ containing $\bigcup (\psi_k(f)(\mathbb{R}^n))$ is positive; or (ii) for each integer $n \psi_n(f)(A)$ is a Lebesgue set when $\alpha \in \mathbb{R}^n$, and for some integer $k$, if $\Lambda \subset \mathbb{R}^k$ and $P_k^\alpha(\Lambda) = 1$, then the Haar measure of $\psi_k(f)(A)$ obtained from the smallest closed additive group in $R^m$ containing $\bigcup (\psi_k(f)(\mathbb{R}^n))$ is positive. If $g$ is a real-valued Borel function on $X$ so that for each $n \geq n_0$ there is a locally Borel $\varphi_n:M_n \to R$, $M_n \subset R^m$, so that $P_n^\alpha(\psi_n(g) = \varphi_n \circ \psi_n(f)) = 1$, then there are $m + 1$ real constants $a_0, \ldots, a_m$ so that $P_n(g = a_0 + \sum_{j=1}^{m} a_j f_j) = 1$.

**Theorem 2.** Let $f_1, \ldots, f_m$ be real-valued Borel functions on $X$, $f = (f_1, \ldots, f_m)$, which satisfy either (i) or (ii) of Theorem 1. If for each $n \geq n_0$ and $t \in T$ there are locally Borel $\varphi_{1,n}:M_{1,n} \to R, M_{1,n} \subset R^m$, satisfying

$$P_n^\alpha(\varphi_{1,n} \circ \psi_n(f) P_n^\alpha(d\xi_1 \ldots d\xi_n)$$

for $(t,A) \in T \times \mathbb{R}^n$, then (1) holds; each $\varphi_j$ in (1) is a linear combination of the $f_j$ with probability one, and thus $p \leq m$ for a version of $dP_j/dP_k$.

**Proof:** Following Dynkin\(^3\), define $\xi(t,x) = \ln(dP_j/dP_k(x))$ up to a null set. Let $\{g,s \in S\}$ be a basis for the smallest real linear space containing the functions of $x \xi(t,\cdot)$. Then for each $s \in S, g_s$ satisfies the hypotheses of the $g$ in Theorem 1 and this completes the proof.
Remark: It is clear that in case there are \( p \) integers \( 0 \leq p \leq m \) so that \( P_i(b_j f_j + d_j \in Q) = 1 \) for real \( d_j \) and \( b_j \neq 0, j = 1, \ldots, p \), and the remaining \( f_j \) satisfy the hypotheses of Theorem 2, then the same conclusion holds, because of Lemma 8. In particular, the hypotheses of Theorem 2 are satisfied for lattice-valued \( f_j \) whenever they are Borel functions. If \( \mathcal{X} \) is a connected locally Euclidean space, if the \( f_j \) are continuous, and if the \( f_j \{ P_i \} \), and Lebesgue measure are "compatible," then Lemma 7 will often permit a rapid answer as to whether the \( f_j \) satisfy the hypotheses of Theorem 2. Moreover, if \( \mathcal{X} \) is a Borel set in a Euclidean space and \( \{ P_i \} \) is equivalent to Lebesgue measure, then, for example, locally Lipschitzian \( f_j \) satisfy the hypotheses of Theorem 2. In particular, if \( \mathcal{X} \) is a real interval, then it is sufficient that real-valued continuous \( f \) satisfy Lusin's condition \((N)\) or be absolutely continuous —this latter fact was proved by L. Brown.\(^4\)

We mention counterexamples to some possible weakening of the hypotheses. Let \( \mathcal{X} \) be the positive integers and let \( P_i \) be the \( i \)th prime. If \( f(x) = \ln p_x \), then evidently \( \psi_n(g) = \varphi \circ \psi_n(f) \) on \( \mathcal{X} \) for each real-valued \( g \) on \( \mathcal{X} \). Let \( \mathcal{X} \subset R \) be a Hamel basis of positive outer measure.\(^5\) Since the algebraic sum of \( \mathcal{X} \) \( n \) times always has inner measure zero, it is not difficult to see that "inner measure" cannot be replaced by "outer measure" in Theorem 1. J. von Neumann\(^6\) constructs a strictly increasing \( f:(0, \infty) \rightarrow (0, \infty) \), so that \( f((0, \infty)) \) is an algebraically independent set and in particular is a subset of a Hamel basis. Since \( f(A) \) is an analytic set for each Borel \( A \subset (0, \infty) \), it is not difficult to prove that \( \psi_n(f)(A) \) is a Lebesgue null set for each \( A \subset \mathcal{X} \). Evidently, for each real-valued \( g \) on \( (0, \infty) \), \( \psi_n(g) = \varphi \circ \psi_n(f) \) everywhere \( \varphi \) is independent of \( n \). Finally, a minor modification of an example of L. Brown\(^4\) shows that both continuity and equation (2) being satisfied are not sufficient conditions on \( f \) to ensure that equation (1) holds. Indeed, it is not hard to construct a continuous Cantor function \( f \) on \([0,1] = \mathcal{X} \) so that the image by \( f \) of the complement of the Cantor set is a countable subset of a Hamel basis. Let \( P = \{ p: p: [0,1] \rightarrow [0, \infty), p(x) = 0 \text{ if and only if } x \text{ lies in the Cantor set, } p(x) = p(y) \text{ if } x \text{ and } y \text{ lie in the same open interval complement to the Cantor set, } \int_0^1 p(x)dx = 1 \} \). Then identifying \( P \) with \( \{ P_i \} \), it is easy to verify that while equation (2) holds, \( f \) never meets the requirements of Theorem 2, and that \( \{ P_i \} \) is not an exponential family.

---