SUFFICIENT STATISTICS IN THE CASE OF INDEPENDENT RANDOM VARIABLES[1]

BY L. BROWN

Cornell University

1. Introduction. In many statistical situations the information obtained from the observation of n-independent identically distributed real random variables $X_1, \cdots, X_n$ can be condensed into one "sufficient statistic", $\phi(x_1, \cdots, x_n)$. In a well known sense the statistic ϕ contains as much information about the distribution of $X_1, \cdots, X_n$ as do the observations $x_1, \cdots, x_n$ themselves [1].

The Neyman factorization theorem [6], [9] gives one characterization of the situations in which a sufficient statistic can be employed. Suppose the distribution of each X_i is a priori known to be one of the distributions in the set $\{P_\theta(\cdot): \theta \varepsilon \Theta\}$ where each $P_\theta(x)$ has density $p_\theta(x)$ with respect to a fixed σ finite measure μ. Neyman's theorem tells how the densities $\{p_\theta(\cdot)\}$ must be related to each other through any statistic which is sufficient for the problem.

A more definitive characterization valid under certain additional assumptions of the densities $p_\theta(\cdot)$ in terms of the sufficient statistic is given by Koopman [7], and Darmois [3]. A further related result was proved by Dynkin [4]. This characterization states exactly what the functional form of the possible densities must be—specifically, that each density must be a member of a certain exponential family of densities (sometimes called a Koopman-Darmois family). This family is determined by the sufficient statistic.

The assumptions in the theorems of [3] and [7] include significant limitations on the form of the densities and on the form of the sufficient statistics. Dynkin [4] states a theorem in which a very minimal assumption is made on the form of the sufficient statistic, but the form of the densities involved is significantly restricted.

In the first main theorem of this paper—Theorem 2.1—a different approach is used. Almost the entire burden of the assumptions is on the form of the statistics involved. The second main theorem—Theorems 8.1 and 8.1′—makes one assumption on the form of ϕ which is generally satisfied. The remainder of its hypotheses are very weak. The conclusion is of a local nature, as opposed to the global nature of the conclusion of Theorem 2.1. These results are a fairly complete characterization of the situation when the conclusion is valid that each density is a member of a certain exponential family of densities.

Only the case of a real sufficient statistic is considered in detail in this paper. Some analogous results are clearly true for n-dimensional or even more general sufficient statistics. I hope to pursue these questions in a later paper.

Received 30 December 1963.

[1] This research was supported in part by the Office of Naval Research contract number Nonr-266(04) (NR 047-005).

Section 2 of this paper contains definitions and a statement of the first main theorem—Theorem 2.1.

Section 3 contains examples which clarify the nature and importance of the assumptions in Theorem 2.1 and Theorem 8.1. Example 3.3 is of particular interest. First, it shows the falsity of a fairly natural conjecture. Second, it shows that a result of Dynkin [4] is false as stated. A possible corrected version of Dynkin's theorem is given at the end of Section 3.

The next three sections contain the proof of Theorem 2.1. Section 4 contains a point-set-topological result. The result of Section 5 is partly measure theoretic and partly topological. In Section 6 these results are used to complete the proof of Theorem 2.1.

The next section of the paper is devoted to corollaries and remarks which weaken the hypotheses of Theorem 2.1 regarding p and ϕ.

The second main theorem of this paper which applies to the case when the sufficient statistic is some type of mean is stated in Section 8. Its proof, which is sketched in that section, relies heavily on the methods of proof used in proving Theorem 2.1.

2. Definitions and statement of the main theorem. Let $\{p(x, \theta): \theta \varepsilon \Theta\}$ be a family of probability densities with respect to Lebesgue measure, denoted by μ, on the interval $I = (a, b)$, $-\infty \leqq a < b \leqq \infty$, of the real line, E^1.

Let $X_1, X_2, \cdots, X_n$ be n independent random variables each having the density $p(\cdot, \theta)$ for some $\theta \varepsilon \Theta$. Then [6], [9], the statistic $\phi(x_1, \cdots, x_n)$ is *sufficient* for θ (or for $\{p\}$) on the basis of $X_1, \cdots, X_n$ if and only if there exist functions v and w such that for all $\theta \varepsilon \Theta$

$$\prod_{i=1}^{n} p(x_i, \theta) = v(x_1, \cdots, x_n)w(\phi(x_1, \cdots, x_n), \theta) \tag{2.1}$$

for almost all $x_1, \cdots, x_n$ in $I^n = I \times I \times \cdots \times I$. We shall use the term *sufficient* (for $\{p\}$) when a factorization of the type 2.1 is satisfied even though the functions $p(x, \theta)$ may not be probability densities.

$q: I \times \Theta \to E^1$ is said to be an *n-parameter exponential family of functions* if there exist real valued functions C, Q_i, T_i, and h such that for all $\theta \varepsilon \Theta$

$$q(x, \theta) = C(\theta)h(x) \exp\left\{\sum_{i=1}^{n} Q_i(\theta)T_i(x)\right\} \qquad \text{a.e. } (\mu).$$

We state an assumption which we shall frequently refer to.

Assumption 2.1. Each density $p(x, \theta)$ is equivalent to Lebesgue measure on I, e.g. For each θ and $A \subset I$,

$$\int_A p(x, \theta)\, dx = 0 \quad \text{if and only if} \quad \int_A dx = 0.$$

When Assumption 2.1 is satisfied the Neyman factorization can be rewritten in the form

$$\prod_{i=1}^{n} p(x_i, \theta) = \left(\prod_{i=1}^{n} p(x_i, \theta_0)\right) w(\phi(x_1, \cdots, x_n), \theta) \qquad \text{a.e. } (\mu_n), \tag{2.2}$$

where θ_0 is any fixed parameter in Θ. This result is due to Bahadur [1]. (I will sometimes refer to (2.2) (or (2.3) below) as the N-B factorization.) In this case it is convenient to rewrite (2.2) as

$$\prod_{i=1}^{n} r(x_i, \theta) = w(\phi(x_1, \cdots, x_n), \theta) \qquad \text{a.e. } (\mu_n), \tag{2.3}$$

where

$$r(x, \theta) = p(x, \theta)/p(x, \theta_0). \tag{2.4}$$

As a consequence of Assumption 2.1, r is well defined.

Let $\mu_n(S)$ denote the Lebesgue measure of the set $S \subset E^n$. If $\phi: I^n \to E^1$ is measurable, $z \in I^{n-1}$, $B \in E^1$ is Borel measurable (i.e. $B \in \mathcal{B}$) and $A \in I$ is measurable, define

$$M_\phi(z, B, A) = \mu\{x: x \in A, \phi(x, z) \in B\}.$$

In words $M_\phi(z, B, A)$ is the cross sectional measure at $z \in I^{n-1}$ of the subset $\{(x, z): x \in A\} \cap \phi^{-1}(B)$ of I. The subscript ϕ will usually be omitted.

It is now possible to state the first main theorem of this paper.

THEOREM 2.1. *Let $\phi(x_1, \cdots, x_n)$ be sufficient for $\{p(x, \theta): \theta \in \Theta\}$ on the basis of n independent observations, $n \geqq 2$. Suppose that for each θ, $p(\cdot, \theta)$ satisfies Assumption 1. Suppose also that there exists a set $A \subset I$ with $\mu(A) > 0$ such that $\mu(B) = 0$ implies $M(z, B, A) = 0$ for all $z \in I^{n-1}$ and such that $\phi(\cdot, \cdot, \xi)$ is continuous on $A \times I$ for each $\xi \in I^{n-2}$. Then $\{p(\cdot, \theta)\}$ is a one-parameter exponential family.*

The following lemma is given so that the reader may better understand the condition imposed on ϕ by the hypotheses of Theorem 2.1. The hypotheses of the lemma are frequently satisfied in statistical problems. In a sense, Theorem 8.1 generalizes this lemma.

LEMMA 2.1. *Suppose $\phi(x_1, \cdots, x_n) = \prod_{i=1}^{n} \psi(x_i)$ or $\phi(x_1, \cdots, x_n) = \sum_{i=1}^{n} \psi(x_i)$ where $\psi(\cdot)$ is continuous and in the first case positive on I. Suppose there exists a measurable set $A \subset I$, $\mu(A) > 0$ on which ψ is absolutely continuous and $\psi'(x) > 0$ for all $x \in A$. Then the hypotheses of Theorem 2.1 are satisfied.*

PROOF. $\mu(B) = 0$ implies $\mu\{x: x \in A, \psi(x) \in B\} = 0$. The remainder of the proof is trivial.

3. Examples, and comments on a paper of Dynkin. These examples are designed to illustrate the nature of the hypotheses included in Theorem 2.1 and to show why these hypotheses are included in the statement of the theorem.

The first example illustrates the need for Assumption 2.1.

EXAMPLE 3.1. Let $\Theta = E^2$ and let

$$
\begin{aligned}
p(x, (a, b)) &= C(a, b)x^b, && 1 \leqq x \leqq 2 \\
&= C(a, b)\beta^a x^b, && 4 \leqq x \leqq 5 \\
&= 0, && \text{otherwise}
\end{aligned} \tag{3.1}
$$

where β is a fixed positive constant $\beta \neq 1$. The statistic x_1x_2 is sufficient for $\{p\}$ on the basis of two independent observations X_1 and X_2 since

$$
p(x_1, (a, b))p(x_2, (a, b)) = C^2(a, b)h(x_1, x_2)\omega(x_1x_2, (a, b)) \tag{3.2}
$$

where

$$
\begin{aligned}
\omega(y, (a, b)) &= y^b, && 1 \leqq y \leqq 4 \\
&= \beta^a y^b, && 4 \leqq y \leqq 10 \\
&= \beta^{2a} y^b, && 16 \leqq y \leqq 25;
\end{aligned} \tag{3.3}
$$

$$
\begin{aligned}
h(x_1, x_2) &= 1 && 1 \leqq x_1 \leqq 2 \quad \text{or} \quad 4 \leqq x_1 \leqq 5, \\
& && \text{and } 1 \leqq x_2 \leqq 2 \quad \text{or} \quad 4 \leqq x_2 \leqq 5, \\
&= 0 && \text{otherwise.}
\end{aligned}
$$

In spite of the fact that there is a real sufficient statistic for θ, it is easy to see that the conclusion of Theorem 1 is not valid. $\{p\}$ is, in fact, a two parameter exponential family, rather than a one-parameter family. This does not provide a counterexample to Theorem 1 since each p does not satisfy Assumption 2.1. The facts that each p is positive on *two* disjoint intervals and that $\{p\}$ is a *two* parameter family are related. See Corollary 7.2.

The necessity for some assumption concerning the continuity of ϕ is shown by a similar example.

EXAMPLE 3.2. Let $\Theta = E^2$ and let

$$
\begin{aligned}
p(x, (a, b)) &= C(a, b)x^b, && 1 \leqq x \leqq 2 \\
&= C(a, b)\beta^a (x + 2)^b, && 2 \leqq x \leqq 3 \\
&= 0, && \text{otherwise}
\end{aligned} \tag{3.4}
$$

where C is chosen so that p is a probability density. Let $\phi(x_1, x_2) = \psi(x_1)\psi(x_2)$ where

$$
\begin{aligned}
\psi(x) &= x && 1 \leqq x \leqq 2 \\
&= x + 2 && 2 \leqq x \leqq 3 \\
&= x && \text{otherwise.}
\end{aligned} \tag{3.5}
$$

Then ϕ is sufficient for $\{p\}$ on the basis of two independent observations X_1 and X_2 because of the factorization

$$
p(x_1(a, b))p(x_2, (a, b)) = C(a, b)h(x_1, x_2)\omega(\phi(x_1, x_2), (a, b)), \tag{3.6}
$$

where ω is defined by (3.3) and h is appropriately chosen.

In this example $\{p\}$ is not a one-dimensional exponential family but it satisfies Assumption 2.1 and has a real sufficient statistic. This example does not provide a counterexample to Theorem 2.1 since ϕ is not continuous.

In the following example all the assumptions of Theorem 1 are satisfied except that ϕ does not satisfy the assumption that $\mu(B) = 0$ implies $M(y, B, A) = 0$. This example also provides a counterexample to Theorem 2 of [4].

EXAMPLE 3.3. Let $S = \{s_{k,l}\} l = 1, 2, \cdots, k = 1, 3, 5, \cdots, 2l - 1$ be a countable set of real numbers satisfying the following conditions:

(3.7)
(a) $s_{m,n} < s_{k,l}$ for $m/2^n < k/2^l$;
(b) $0 < s_{k,l} < 1$;
(c) S is dense in the interval [0, 1];
(d) $s_{k_1,l_1} \cdot s_{k_2,l_2} = s_{m_1,n_1} \cdot s_{m_2,n_2}$ implies $(k_1, l_1) = (m_1, n_1)$ or $(k_1, l_1) = (m_2, n_2)$.

Let $c(x)$ be the usual Cantor function defined on [0, 1] ([5], p. 83). $c(x)$ takes the values $k/2^l$ with k odd, $k \leqq 2^l - 1$ almost everywhere in [0, 1], say on the set $C \subset [0, 1]$. Let d be the unique real continuous function defined on [0, 1] satisfying $d(x) = s_{k,l}$ if $c(x) = k/2^l$. (3.7) (a), (b), and (c) guarantee that d exists.

Let $\phi(x_1, x_2) = d(x_1)\, d(x_2)$. If $x_1 \varepsilon C$ and $x_2 \varepsilon C$ then using (3.7) (d), $c(x_1)$ and $c(x_2)$ are determined uniquely (up to transposition of x_1 with x_2) by the value $d(x_1)\, d(x_2)$. Thus there exists a function ω such that

$$c^{\alpha}(x_1)c^{\alpha}(x_2) = \omega^{\alpha}(\phi(x_1, x_2)) \qquad \text{on } C \times C \text{ for } \alpha \,\varepsilon\, E'. \tag{3.8}$$

Also

$$d^{\alpha}(x_1)\, d^{\alpha}(x_2) = \phi^{\alpha}(x_1, x_2). \tag{3.9}$$

These two equations show that ϕ is sufficient for the family $\mathfrak{F}$ of densities on [0, 1] consisting of all $K_1(\alpha)c^{\alpha}(x)$ and of all $K_2(\alpha)\, d^{\alpha}(x)$. As a consequence of (3.7) (d), $\mathfrak{F}$ cannot be written as a one-parameter family of densities. This example is not a counterexample to Theorem 1 for it can be shown that there does not exist an A with $\mu(A) > 0$ such that for any B, $\mu(B) = 0$ implies $M_{\phi}(y, B, A) = 0$.

Example 3.3 provides a counterexample to Theorem 2 of [4]. (Actually, in order to satisfy all the hypotheses of [4], $\phi(x_1, x_2)$ should be chosen as $d(x_1)\, d(x_2)$ on $C \times C$ and as (x_1, x_2) or some statistic which is equivalent to (x_1, x_2), if $x_1, x_2 \not\varepsilon C \times C$.) In order to correct Theorem 2 of [4] it is enough (using the notation of [4]) to assume that the densities in the family π are continuously differentiable in Δ, rather than "piece-wise smooth" in Δ. This condition is sufficient to insure that the statement in the next to last sentence of the proof of Theorem 2 of [4] is correct.

The following theorem—Theorem A—which combines a corrected version of

Theorem 2 and Theorem 3 of [3] is given for comparison with Theorem 2.1 of this paper.

A statistic ϕ defined on a topological measure space G is called *trivial* if there exists an open subspace $\tilde{G} \subset G$ such that ϕ is equivalent to the identity statistic on $\tilde{G}$: that is, such that there exists a measurable function ω for which

$$x = \omega(\phi(x)) \qquad \text{for almost all } x \in \tilde{G}. \tag{3.10}$$

THEOREM A. *Let $\{p(x, \theta)\}$ be a family of probability densities on an interval I such that for each θ, $p(x, \theta)$ is continuous on I, is bounded away from 0 on I, and is continuously differentiable on I. Suppose there is a non-trivial, sufficient statistic ϕ for θ on the basis of n independent observations. Then $\{p\}$ is a ρ-parameter exponential family where $n > \rho$.*

PROOF OF THEOREM A. The hypotheses of Theorem A are sufficient to make valid the conclusion of the next to last sentence of the proof of Theorem 2 in [4], and hence to make valid the conclusion of that theorem. Then using the notation of [3] the rank ρ of the family π is less than n and $\{\prod_{i=1}^{n} r_k(x_i, \theta')\}$, $k = 1, 2, \cdots \rho$ is a sufficient statistic if the $r_k(\cdot, \theta')$ are linearly independent and not constant. Then, using Theorem 3 of [4], $\{p\}$ is a ρ parameter exponential family. This completes the proof of Theorem A.

By taking cross-sections, as is done in the proof of Theorem 2.1 (see 6.7), it can often be concluded from an examination of ϕ that ρ is in fact a specific value considerably less than $n - 1$.

4. Proof of Theorem 2.1—Part 1. The hypotheses of the Theorem 4.1 of this section are chosen so that this theorem can be easily applied in the proof of Theorem 2.1. Theorem 4.2, which follows easily from Theorem 4.1 has a more natural set of hypotheses. Theorem 4.2 should be compared to Theorem A and to Theorem 2.1, for it has the same conclusion as these two theorems, but a slightly different set of hypotheses.

THEOREM 4.1. *Let $A \subset I$ be a set of points such that for some point, say $x' \in A$ there exists a sequence $\{x_i\}$ with $x_i \in A$, $i = 1, 2, \cdots$, and $\lim x_i = x'$. Suppose for each $\theta \in \Theta$, $r(\cdot, \theta)$ is a continuous function from $I = (a, b)$ into E^{1*} (the one point compactification of E^1), and, for each $\theta \in \Theta$, $r(\cdot, \theta)$ satisfies Assumption 1. Suppose there exists a continuous function $\phi: A \times I \to E^1$ and a function ω such that*

$$r(x_1, \theta)r(x_2, \theta) = \omega(\phi(x_1, x_2), \theta) \tag{4.1}$$

for all $(x_1, x_2) \in A \times I$ and all $\theta \in \Theta$. Then there exists an interval $K \subset I$ such that $K \cap A$ is not empty and such that for a fixed value $\theta_0 \in \Theta$, there exist functions C and Q such that

$$r(x, \theta) = C(\theta)\{r(x, \theta_0)\}^{Q(\theta)} \tag{4.2}$$

for all $x \in K \cap A$ and all $\theta \in \Theta$.

PROOF. Several preliminary results are needed. These will be stated and proved as lemmas.

Throughout this proof the hypotheses of Theorem 4.1 are assumed to be satisfied; and $\theta' \varepsilon \Theta$ will denote any fixed parameter such that $r(\cdot, \theta')$ is not a constant on all of I. (If no such value of θ' exists then $r(x, \theta) = C(\theta)$ for all x and θ, and the conclusion of the theorem is valid.)

LEMMA 4.1. *Suppose $r(\cdot, \theta')$ is constant on some interval $J \subset I$. Then either $\phi(x', y)$ is constant for $y \varepsilon J$, or there exists an interval $K \subset I$ such that $x' \varepsilon K$ and $r(x, \theta) = C(\theta)$ for $x \varepsilon K \cap A$ and all $\theta \varepsilon \Theta$.*

PROOF. Suppose $r(\cdot, \theta')$ is constant on an open interval J but $\phi(x', y)$ is not constant for $y \varepsilon J$. Since ϕ is continuous there is an open interval

$$D \subset \{z: \exists y \varepsilon J, \phi(x', y) = z\}. \tag{4.3}$$

There exists an $\epsilon > 0$ (not depending on θ') such that $|x - x'| < \epsilon, x \varepsilon A$, implies there exists a $y \varepsilon J$ such that $\phi(x, y) \varepsilon D$. (If this were not true, ϕ would not be continuous at any point x', y such that $y \varepsilon J, \phi(x', y) \varepsilon D$.) Thus there is a $y \varepsilon J$ such that

$$r(x, \theta')r(y, \theta') = \omega(\phi(x, y), \theta') = r(x', \theta')r(y, \theta'). \tag{4.4}$$

Hence $r(x, \theta') = r(x', \theta')$ for all $x \varepsilon A$ such that $|x - x'| < \epsilon$. Let $K = \{x: |x - x'| < \epsilon\}$. Then $r(x, \theta) = C(\theta)$ for $x \varepsilon K \cap A$. This completes the proof of the lemma.

It is only necessary for the remainder of the proof to deal with the case where $r(\cdot, \theta')$ constant on an interval J implies $\phi(x', \cdot)$ is constant on J; for if this condition is not satisfied then according to Lemma 4.1 the conclusion of Theorem 4.1 is valid. In the following lemmas we make that assumption.

LEMMA 4.2. *Suppose $r(\cdot, \theta')$ constant on the interval $J \subset I$ implies $\phi(x', \cdot)$ is constant on J. Let $L \subset I$ be an interval (not necessarily open) of positive length. Suppose x_1 is an extreme value of $r(\cdot, \theta')$ on L (*i.e.* x_1 is a minimum or maximum of r on L). Then either $\phi(x', x_1)$ is an extreme value of $\phi(x', \cdot)$ on L, or there exists an interval $K \subset I$ such that $x' \varepsilon K$ and $r(x, \theta) = C(\theta)$ for all $x \varepsilon K \cap A$ and all $\theta \varepsilon \Theta$.*

PROOF. Suppose $\phi(x', x_1)$ is not an extreme value of ϕ on L and suppose $r(\cdot, \theta')$ assumes its minimum on L at x_1. Then there exists an open interval D satisfying (4.3), and satisfying $\phi(x', x_1) \varepsilon D$. Using continuity there is an $\epsilon > 0$ such that $|x - x'| < \epsilon$, $x \varepsilon A$ implies there exists a $y \varepsilon J$ such that $\phi(x, y) = \phi(x', x_1)$, and implies that $\phi(x, x_1) \varepsilon D$. For such an x

$$r(x, \theta')r(y, \theta') = \omega(\phi(x, y), \theta') = \inf_{\phi \varepsilon D} \omega(\phi, \theta') = r(x', \theta')r(x_1, \theta'). \tag{4.5}$$

According to (4.5),

$$r(y, \theta') = \inf\{r(z, \theta'): \phi(x, z) \varepsilon D\}. \tag{4.6}$$

Since $x_1 \varepsilon \{z: \phi(x, z) \varepsilon D\}$, $r(y, \theta') \leqq r(x_1, \theta')$. By assumption, $r(x_1, \theta') \leqq r(y, \theta')$. Hence $r(y, \theta') = r(x_1, \theta')$. Then, using (4.5), $r(x, \theta') = r(x', \theta')$. Thus there is an open K, $x' \varepsilon K$, such that $x \varepsilon K \cap A$ implies $r(x, \theta) = C(\theta)$.

The procedure if $r(x, \theta')$ is a maximum is entirely analagous. This completes the proof of Lemma 4.2.

LEMMA 4.3. *For any interval $J \subset I$ assume $r(\cdot, \theta')$ constant on J implies $\phi(x', \cdot)$*

constant on J and $r(x_1, \theta')$ an extreme value of $r(\cdot, \theta')$ on J implies $\phi(x', x_1)$ an extreme value of $\phi(x', \cdot)$ on J. Then $r(y_1, \theta') = r(y_2, \theta')$ (if and) only if for any $x \varepsilon A$, $\phi(x', y_1) = \phi(x', y_2)$.

PROOF. Suppose $r(y_1, \theta') = r(y_2, \theta')$ but $\phi(x', y_1) \neq \phi(x', y_2)$. Assume $y_1 < y_2$. Let

$$
\begin{aligned}
S_1 &= \{x : \phi(x', x) = \phi(x', y_1), x \varepsilon I\}, \\
S_2 &= \{x : \phi(x', x) = \phi(x', y_2), x \varepsilon I\}.
\end{aligned} \tag{4.7}
$$

S_1 and S_2 are closed disjoint sets. Let y be a point such that $y_1 < y < y_2$ and $y \not\varepsilon S_1 \cup S_2$. Let

$$
\begin{aligned}
y_3 &= \sup\{x : x \varepsilon S_1, x < y\}, \\
y_4 &= \inf\{x : x \varepsilon S_2, x > y\}.
\end{aligned} \tag{4.8}
$$

Then $y_3 \varepsilon S_1$, $y_4 \varepsilon S_2$. The function $\phi(x', \cdot)$ assumes its extreme values on $[y_3, y_4]$ at y_3 and y_4. Hence y_3 and y_4 must be the extreme values of $r(\cdot, \theta')$ on $[y_3, y_4]$. Hence $r(y, \theta') = r(y_3, \theta') = r(y_4, \theta')$ for all $y \varepsilon [y_3, y_4]$; in contradiction to the hypotheses of the lemma. This completes the proof of the lemma.

For the remainder of the proof of the theorem, the special hypotheses of Lemma 4.3 are assumed true; for if they are not true Lemmas 4.1 and 4.2 show the conclusion of the theorem to be valid.

Suppose there are parameters θ', $\theta'' \varepsilon \Theta$ such that there does not exist a real C and k and an interval K for which $r(x, \theta'') = C(\theta')r^k(x, \theta')$ for $x \varepsilon A \cap K$; i.e., that $\{r(\cdot, \theta) : \theta \varepsilon \Theta\}$ is not of the form (4.2) on $A \cap K$. Then on $A \times I$

$$
\frac{r(x_1, \theta'')}{r^k(x_1, \theta')} \frac{r(x_2, \theta'')}{r^k(x_2, \theta')} = \frac{\omega(\phi(x_1, x_2), \theta'')}{\omega^k(\phi(x_1, x_2), \theta')} = \tilde{\omega}(\phi(x_1, x_2); \theta', \theta'', k). \tag{4.9}
$$

From these facts it is easily checked that for any k, the function $r(x, \theta'')/r^k(x, \theta')$ satisfies the hypotheses of Lemma 4.3 (including the hypotheses of Theorem 4.1). There must exist a k, x_1, x_2 such that $r(x_1, \theta') \neq r(x_2, \theta')$ (and hence $\phi(x', x_1) \neq \phi(x', x_2)$) but $r(x_1, \theta'')/r^k(x_1, \theta') = r(x_2, \theta'')/r^k(x_2, \theta')$. It follows that $\tilde{\omega}$ in (4.9) is not a 1-1 function, but this contradicts Lemma 4.3. This completes the proof of Theorem 4.1.

The next theorem may be of some independent interest, as it is another theorem belonging to the same class of theorems as Theorem 2.1.

THEOREM 4.2. *Suppose for each θ, $p(x, \theta)$ is a continuous function from $I = (a, b)$ into E^{1*} (the point compactification of E^1), and for each θ, $p(x, \theta)$ satisfies Assumption 1. Suppose there exists a real-valued function ϕ which is continuous as a function from I^n into E^1 and which is sufficient for θ on the basis of n-independent observations $X_1, X_2, \cdots, X_n$, $n \geqq 2$. Assume the Neyman-Bahadur factorization (2.2) holds everywhere in I^n,* i.e.

$$
\prod_{i=1}^{n} p(x_i, \theta) = \left(\prod_{i=1}^{n} p(x_i, \theta_0)\right) \omega(\phi(x_1, \cdots, x_n)) \tag{4.10}
$$

$$
\text{for} \quad (x_1, \cdots, x_n) \varepsilon I^n.
$$

Then $\{p\}$ is a one-parameter exponential family of densities.

Note: It is not necessary in Theorem 4.2 to assume that the functions $p(x, \theta)$ are probability densities.

PROOF. Let $r(x, \theta) = p(x, \theta)/p(x, \theta_0)$ for some fixed $\theta_0 \varepsilon \Theta$. As remarked previously $\prod_{i=1}^{n} r(x_i, \theta) = \omega(\phi(x_1, x_2, \cdots, x_n), \theta)$. If $n > 2$ fix $y_3, \cdots, y_n \varepsilon I^{n-2}$. Then

$$(4.11) \qquad \begin{aligned} r(x_1, \theta)r(x_2, \theta) &= \omega(\phi(x_1, x_2; y_3, \cdots, y_n), \theta)/\prod_{i=3}^{n} r(y_i, \theta) \\ &= \tilde{\omega}(\tilde{\phi}(x_1, x_2), \theta). \end{aligned}$$

It is easy to finish checking that $\{r\}$ satisfies the hypotheses of Theorem 4.1 where $A \subset I$ can be any set. In particular, Theorem 4.1 then states that every $x \varepsilon I$ has a neighborhood such that $\{r\}$ is an exponential family on that neighborhood. It is easily checked that this can be true only if $\{r\}$ is an exponential family on I. Since $p(x, \theta) = p(x, \theta_0)r(x, \theta)$, $\{p\}$ is also an exponential family. This completes the proof of Theorem 4.2.

5. Proof of Theorem 2.1—Part 2. The important result of this section, so far as the proof of Theorem 2.1 is concerned, is that for each θ, r is a continuous function from I into E^{1*}. This result is contained in Theorem 5.1.

LEMMA 5.1. *Let $\phi: I \times I \to E^1$ be continuous. Let A be a measurable subset of $I = (a, b)$ such that for each fixed $y \varepsilon [c, d](-\infty < a < c < d < b < \infty)$, and any subset $B \varepsilon \mathfrak{L}$, (the measurable subsets of E^1) $\mu(B) = 0$ implies $M(y, B, A) = 0$. Then for any fixed $B \varepsilon \mathfrak{L}$, $M(\cdot, B, A)$ is continuous as a function on $[c, d]$.*

PROOF. Each of the following assertions will be proved as they are stated.

(1) B an open interval, and A a closed set implies $M(\cdot, B, A)$ is continuous: Let $B = (a, b)$, $B_k = (a_k, b_k)$ with $a_k \searrow a$ and $b_k \nearrow b$ (strictly decreasing and increasing to). Then $M(y, B_k, A) \nearrow M(y, B, A)$ for each $y \varepsilon [c, d]$. Using the uniform continuity of ϕ on $A \times [c, d]$, for every given $y_0 \varepsilon [c, d]$ and $\epsilon > 0$ there exists a neighborhood N_k of y_0 such that $y, z \varepsilon N_k$ imply

$$(5.1) \qquad M(y, B_k, A) \leqq M(z, B_k, A) + \epsilon^k \leqq M(z, B, A) + \epsilon^k.$$

Let $y_k \to y_0$ such that $\lambda = \lim_{k\to\infty} M(y_k, B, A) = \liminf_{y\to y_0} M(y, B, A)$. Then $y \varepsilon N_k$ implies $M(y, B_k, A) \leqq \lambda + \epsilon^k$ so that

$$(5.2) \quad M(y_0, B, A) = \lim_{k\to\infty} M(y_0, B_k, A) \leqq \lambda = \liminf_{y\to y_0} M(y, B, A).$$

Using an analogous procedure,

$$(5.3) \qquad M(y_0, B, A) \geqq \limsup_{y\to y_0} M(y, B, A),$$

(5.3) and (5.2) together establish the truth of Assertion 1.

(2) If B is open and A is closed, M is lower semi-continuous, l.s.c., (i.e., satisfies (5.2)). If B is closed and A is closed, M is upper semi-continuous, u.s.c., (satisfies (5.3)).

If B is open, it is the union of a countable number of disjoint open intervals β_k, $B = \bigcup_{k=1}^{\infty} \beta_k$. Using Assertion 1, $M(\cdot, \bigcup_{k=1}^{n} \beta_k, A)$ is continuous. For any fixed $y \varepsilon [c, d]$, $M(y, \cdot, A)$ is an absolutely continuous set function on $(\mathfrak{L}, \mu)$ by the hypothesis of the lemma. Hence $M(y, \bigcup_1^n \beta_k, A) \to M(y, B, A)$ as $n \to \infty$, and $M(y, \bigcup_1^n \beta_k, A)$ is an increasing sequence (in n). Thus $M(y, B, A)$ is the pointwise limit of an increasing sequence of continuous functions, which implies it is l.s.c. If B is a closed set, $M(y, B, A) = \mu(A) - M(y, B^c, A)$ where $B^c = E^1 - B$. B^c is open, so that $M(y, B^c, A)$ is l.s.c., and $M(y, B, A)$ is u.s.c.

(3) $M(y, \cdot, A)$ is locally uniformly (in y) absolutely continuous as a set function on $(\mathfrak{L}, \mu)$: Suppose this is not so. Then there exists an $\epsilon > 0$, a decreasing sequence of nested sets γ_i, and a sequence y_i such that $M(y_i, \gamma_i, A) > \epsilon$ but $\mu(\gamma_i) \to 0$. Since Lebesgue measure is regular we may assume the γ_i are closed nested sets without any loss of generality. Since $[c, d]$ is a closed interval the sequence y_i has an accumulation point, say $y_0 \varepsilon [c, d]$. Using (2), $M(y_0, \bigcap \gamma_i, A) \geqq \lim \sup_{y_i \to y_0} M(y_i, \gamma_i, A) \geqq \epsilon > 0$. Since $\mu(\bigcap \gamma_i) = 0$, this contradicts the hypothesis of the lemma, proving (3).

(4) If B is an open set or a closed set and A is a closed set, then $M(\cdot, B, A)$ is continuous: If B is open $B = \bigcup_{k=1}^\infty \beta_k$. $M(y, \bigcup_{k=1}^n \beta_k, A) \to M(y, B, A)$ uniformly in y since $\mu(\bigcup_{k=n+1}^\infty \beta_k) \to 0$ and $M(y, \bigcup_{k=n+1}^\infty \beta_k, A) = M(y, B, A) - M(y, \bigcup_{k=1}^n \beta_k, A)$ is uniformly (in y) absolutely continuous according to (3), $M(y, \bigcup_{k=1}^n \beta_k, A)$ is continuous, hence so also is $M(y, B, A)$. If B is closed $M(y, B, A) = \mu(J) - M(y, B^c, A)$, so that $M(y, B, A)$ is continuous.

(5) For any $B \varepsilon \mathfrak{L}$ and closed A, $M(\cdot, B, A)$ is continuous: There exists a $B' \subseteq B$ such that $\mu(B - B') = 0$ and $B' = \bigcup_{k=1}^\infty b_k$ where the b_k are nested-increasing closed sets. Reasoning as in (4), $M(\cdot, B', A)$ is continuous, and using the hypotheses of the theorem $M(y, B - B', A) = 0$ for all $y \varepsilon [c, d]$. Hence $M(\cdot, B, A)$ is continuous.

(6) Finally, any measurable A can be written as $A \supset \bigcup_{k=1}^\infty \alpha_k$ where the α_k are nested-increasing closed sets and $\mu(A - \bigcup_{k=1}^\infty \alpha_k) = 0$. Hence $M(y, B, \bigcup_{k=1}^\infty \alpha_k) \to M(y, B, A)$ uniformly in y. Thus $M(y, B, A)$ is continuous. This completes the proof of Lemma 4.1.

I am indebted to H. Kesten for supplying parts (1) and (2) of the proof of this lemma and for further discussions concerning it.

THEOREM 5.1. *Suppose the hypotheses of Theorem* 2.1 *are satisfied, and* $n = 2$. (*In particular, let* $A \subset I$ *be such that* $\mu(A) > 0$ *and* $\mu(B) = 0$ *implies* $M_\phi(y, B, A) = 0$ *for all* $y \varepsilon I$). *Then for each* θ, $r(\cdot, \theta)$ *has a continuous version* (i.e.; *there exists a function* r' *such that* r' *is continuous and* $r'(\cdot) = r(\cdot, \theta)$ a.e.).

(The theorem is also true for all $n > 2$. This can easily be shown from the case $n = 2$ by the argument at (6.4).)

PROOF. If the theorem is not true then there exists a $\theta' \varepsilon \Theta$ and a point $x_0 \varepsilon I$ such that

$$p_1 = \text{ess lim inf}_{x \to x_0} r(x, \theta') \neq \text{ess lim sup}_{x \to x_0} r(x, \theta') = p_2. \tag{5.4}$$

Let

$$D_k = \{x: \sigma^k \leqq x < \sigma^{k+1}\} k = \cdots -2, -1, 0, 1, 2, \cdots, \qquad 1 < \sigma < p_2/p_1. \tag{5.5}$$

For some k, say k', $\mu(r^{-1}(D_{k'}) \cap A) > 0$. Using (5.5)

$$M_{r(\cdot)r(\cdot)}(y, B, A) = M_{\phi(\cdot,\cdot)}(y, \omega^{-1}(B), A) \tag{5.6}$$

for almost all $y \varepsilon I$. Let $B' = \{x: p_3\sigma^{k'} \leqq x \leqq p_2\sigma^{k'+1}\}$, where $p_1 < p_3 < p_2$. Then, using (5.6) and the inequality from (5.7), p_3 may be chosen so that $p_1\sigma^{k'+1} < p_3\sigma^{k'}$. Using (5.4)

$$\begin{aligned} &\text{ess lim inf}_{x\to x_0} M_{r\cdot r}(x, B', A) = 0, \\ &\text{ess lim sup}_{x\to x_0} M_{r\cdot r}(x, B', A) = \mu(D_{k'} \cap A) > 0. \end{aligned} \tag{5.7}$$

(5.7) together with (5.6) contradicts Lemma 5.1. This completes the proof of the theorem.

6. Proof of Theorem 2.1—Part 3. In this section the results of the previous two sections are combined in order to complete the proof of Theorem 2.1.

Throughout this section we shall assume that the hypotheses of Theorem 2.1 are satisfied. As in Theorem 4.2, it suffices to prove this theorem for the case $n = 2$, and to deal with the functions r. Throughout this section unless otherwise noted it is assumed that $n = 2$. This assumption will be removed at the end of the proof.

Using Theorem 5.1, it is no loss of generality to assume that each $r(\cdot, \theta)$ is a continuous function on I, and we shall do so throughout this section.

The line of proof is to establish that the Neyman-Bahadur factorization (see (2.2)) is an equality everywhere on $A \times I$ (for an appropriate A and ω). Then Theorem 4.1 can be applied and the proof of Theorem 2.1 can be completed. Several lemmas tending in this direction will be stated and proved.

Let $S \subset I$ be the set guaranteed by the hypotheses of the theorem, i.e. $\mu(B) = 0$ implies $M_\phi(y, B, S) = 0$ for all $y \varepsilon I$. Let T be a countable set of points such that T is dense in I, i.e. $\bar{T} = I$. Choose T such that for each $y \varepsilon T$ the N-B factorization ((2.3) or (6.2)) is valid for almost all $x \varepsilon I$. For any measurable set E let E^d denote the set of points of E which are points of density of E ([7], p. 285–95).

For any $y \varepsilon T$, define U_y by

$$U_y = \{x: x \varepsilon S^d, \phi(x, y) \varepsilon [\phi(S^d, y)]^d\}.$$

The facts that $\mu(S^d) = \mu(S)$ and $\mu\{\phi(x, S^d)\}^d = \mu\{\phi(x, S^d)\}$ imply $\mu(U_y) = \mu(S)$. Let $A = \bigcap_{y\varepsilon T} U_y$. Then $\mu(A) = \mu(S) > 0$. A satisfies the assumptions of Theorem 2.1 in place of S. Note that $A = A^d$.

Lemma 6.1. *There exists a continuous function ω such that*

$$r(x, \theta)r(y, \theta) = \omega(\phi(x, y), \theta) \qquad \textit{for all } (x, y) \varepsilon A \times I \textit{ and all } \theta \varepsilon \Theta. \tag{6.1}$$

Proof. By hypotheses

$$r(x, \theta)r(y, \theta) = w(\phi(x, y), \theta) \qquad \text{a.e. on } A \times T. \tag{6.2}$$

For each θ, and $\zeta \varepsilon \phi(A, T)$, let $\omega(\zeta, \theta) = w(\zeta, \theta)$ if there exists $(x_0, y_0) \varepsilon A \times T$

such that $\phi(x_0, y_0) = \zeta$ and equality holds in (6.1) for that (x_0, y_0). If there does not exist such a value $(x_0, y_0) \varepsilon A \times T$, then choose any $(x_0, y_0) \varepsilon A \times T$ such that $\phi(x_0, y_0) = \zeta$ and define $\omega(\zeta, \theta) = r(x_0, \theta)r(y_0, \theta)$. Note that for each $\theta \varepsilon \Theta$, $\omega(\zeta, \theta) = w(\zeta, \theta)$ for almost all $\zeta \varepsilon \phi(A, T)$.

Consider $(x_1, y_1) \varepsilon A \times T$ and any fixed value $\theta' \varepsilon \Theta$. Let $(x_0, y_0) \varepsilon A \times T$ be a point such that (6.1) is an equality and $\phi(x_0, y_0) = \phi(x_1, y_1)$. If X_i is any neighborhood of x_i, $\phi(x_i, y_i)$ is a point of density of the set $\phi(X_i \cap A, y_i)$, $i = 0, 1$. Since (6.2) and thus (6.1) is an equality at $\phi(x, y_0)$ for almost all x and since r is continuous at x_0 then for every $\epsilon > 0$, $\phi(x_0, y_0) = \phi(x_1, y_1)$ is a point of density of the set $\{\zeta: |\omega(\zeta) - r(x_0, \theta')r(y_0, \theta')| < \epsilon\} = Z_\epsilon$. Let $W_\epsilon = \phi(X_1 \cap A, y_1) \cap Z_\epsilon$. From the construction of A $\mu(W_\epsilon) > 0$ which in turn implies $\mu\{x: \phi(x, y_1) \varepsilon W_\epsilon, x \varepsilon A\} > 0$. Thus there must be points in every neighborhood X_1 of x_1 such that

$$(6.3) \qquad |r(x, \theta')r(y_1, \theta') - r(x_0, \theta')r(y_0, \theta')| < \epsilon.$$

Since (6.3) holds for every $\epsilon > 0$, $r(x_1, \theta')r(y_1, \theta') = r(x_0, \theta')r(y_0, \theta')$ and (6.1) is valid at (x_1, y_1). Thus for every $\theta \varepsilon \Theta$ and $(x, y) \varepsilon A \times T$, (6.1) is an equality.

Let $\zeta_i \to \zeta_0$, $\zeta_i \varepsilon \phi(A \times T)$ where $(x_i, y_i) \varepsilon A \times T$ such that $(x_i, y_i) \to (x, y)$ and $\phi(x_i, y_i) = \zeta_i \cdot \lim_{i\to\infty} \omega(\phi(x_i, y_i), \theta) = \lim_{i\to\infty} r(x_i, \theta)r(y_i, \theta) = r(x, \theta)r(y, \theta)$. Hence $\lim_{\zeta\to\zeta_0} \omega(\zeta, \theta)$ exists, $\zeta \varepsilon \phi(A \times T)$. If $\zeta_0 \not\varepsilon \phi(A \times T)$, define $\omega(\zeta_0, \theta)$ by $\omega(\zeta_0, \theta) = \lim_{\zeta_i\to\zeta_0} \omega(\zeta_i, \theta)$.

Since $\overline{A \times T} \supset A \times I$, ω is now defined on all of $\phi(A \times I)$. Furthermore, according to the previous paragraph, $\omega(\cdot, \theta)$ is a continuous function, and $r(x, \theta)r(y, \theta) = \omega(\phi(x, y), \theta)$ for all $x, y \varepsilon A \times I$ and all $\theta \varepsilon \Theta$. This completes the proof of Lemma 6.1.

Theorem 4.1 can now be applied to establish the existence of an interval K such that $K \cap A$ is non-empty and such that $r(x, \theta) = C(\theta)\{r(x, \theta_0)\}^{Q(\theta)}$ for all $x \varepsilon K \cap A$. The following lemma uses this hypotheses.

LEMMA 6.2. *Suppose* $x' \varepsilon K \cap A$, *and the equation*

$$(6.4) \qquad r(x, \theta) = C(\theta)\{r(x, \theta_0)\}^{Q(\theta)}$$

is valid for all $x \varepsilon K \cap A$ *and all* $\theta \varepsilon \Theta$. *Then* (6.4) *is valid for all* $x \varepsilon I$ *and all* $\theta \varepsilon \Theta$.

PROOF. Let N be a neighborhood of x' such that $y \varepsilon N$ implies $M(y, \phi(K \cap A, x'), K \cap A) > 0$. The existence of such an interval is guaranteed by Lemma 5.1. For any $\theta \varepsilon \Theta$ and any $y \varepsilon N$ there exist points $x_1, x_2 \varepsilon K \cap A$ such that $\phi(x_1, y) = \phi(x_2, x')$. Using (6.1)

$$(6.5) \qquad r(y, \theta)r(x_1, \theta) = r(x', \theta)r(x_2, \theta).$$

Using (6.4) and (6.5)

$$(6.6) \quad r(y, \theta) = C(\theta)[r(x', \theta_0)r(x_2, \theta_0)/r(x_1, \theta_0)]^{Q(\theta)} = C(\theta)\{r(y, \theta_0)\}^{Q(\theta)}$$

which is the desired equation on N.

Now, suppose $N' = (c', d')$ is a maximal interval containing N on which (6.4) is valid. Let N'' be a neighborhood of c' (or d') such that $M(y, \phi(K \cap A, c'), K \cap A) > 0$ (or $M(y, \phi(K \cap A, d'), K \cap A) > 0$). Proceed as in the preceding paragraph to show that (6.4) is valid on $N' \cup N''$. Therefore $N' = I$. This completes the proof of the lemma.

The proof of Theorem 2.1 can now be easily completed. Suppose $n \geqq 2$. Then there exists $y_3, y_4, \cdots, y_n \varepsilon I^{n-2}$ such that for every $\theta \varepsilon \Theta$,

$$
\begin{aligned}
r(x_1, \theta)r(x_2, \theta) &= \omega(\phi(x_1, x_2; y_3, \cdots, y_n), \theta)/\prod_{i=3}^{n} r(y_i, \theta) \\
&= \tilde{\omega}(\tilde{\phi}(x_1, x_2), \theta)
\end{aligned} \tag{6.7}
$$

for almost all $x_1, x_2 \varepsilon I \times I$.

Using the results of Sections 4 and 5 and the preceding results of this section $r(x, \theta) = C(\theta)\{r(x, \theta_0)\}^{Q(\theta)}$. Hence,

$$
p(x, \theta) = C(\theta)p(x, \theta_0) \exp\{Q(\theta) \ln r(x, \theta_0)\}, \tag{6.8}
$$

which is the desired factorization. This completes the proof of Theorem 2.1.

7. Corollaries to the theorem. In this section several corollaries are stated which weaken in some fashion the hypotheses of Theorem 2.1. The corollaries given here by no means exhaust the possibilities for results of the general type of Theorem 2.1 which have weaker hypotheses than that theorem. They should be sufficient, however, to guide the reader in search of other possible corollaries to Theorem 2.1. These corollaries also serve to amplify understanding of the situation when the presence of sufficient statistics implies the densities are of exponential type.

The proofs of these corollaries mainly consist of minor revisions in the proof of Theorem 2.1. These revisions will only be sketched, rather than given in full.

COROLLARY 7.1. *Let $\{p(x, \theta): \theta \varepsilon \Theta\}$ be a family of probability distributions on an interval I, which satisfy Assumption 2.1 on I. Suppose there exists a set $A \subset I$ such that $\mu(A) > 0$, and there exists a continuous function $\phi: A \times I \rightarrow E^1$ (where A is given the topology inherited from I) and a function ω such that for some $\theta_0 \varepsilon \Theta$,*

$$
p(x_1, \theta)p(x_2, \theta) = p(x_1, \theta_0)p(x_2, \theta_0)\omega(\phi(x_1, x_2), \theta) \quad \text{a.e. } (\mu_2(A \times I)) \tag{7.1}
$$

(i.e. *ϕ is sufficient for p on $A \times I$). Suppose $\mu(B) = 0$ implies $M(z, B, A) = 0$ for all $z \varepsilon I$. Then $\{p(x, \theta)\}$ is a one-parameter exponential family.*

PROOF. The reader may check that the proof of Theorem 2.1 uses only the hypotheses of this corollary, rather than the somewhat stronger (though more natural) hypotheses of Theorem 2.1.

Corollary 7.1 can also be applied in the case $n > 2$ if, for instance, ϕ is sufficient for $\prod_{i=1}^{n} p(x_i, \theta)$ on $A \times I \times S$, $S \subset I^{n-2}$, $\mu_{n-2}(S) > 0$, for an appropriate set A. For then as in the proof of Theorem 2.1 (6.7) there is a $\pi \varepsilon S$ such that $\phi(\cdot, \cdot, \pi)$ is sufficient on $A \times I$ for $p(x_1, \theta)p(x_2, \theta)$.

It may be that the hypotheses of Theorem 2.1 (or Corollary 7.1) do not hold (in the case $n = 2$) for all of $I \times I$, but do hold on $I_k \times I_k$ where the I_k are intervals. (Sometimes it is also desirable that $I = \cup \bar{I}_k[\bar{I}_k$ denotes the closure of I_k].) Many corollaries of Theorem 2.1 can be derived which deal with this type of situation. Only a few of the simplest of such results will be given in Corollary 7.2. To simplify matters, with no great loss of generality, only the case $n = 2$ will be treated.

COROLLARY 7.2. *Let I_k, $k = 1, 2, \cdots, m$ be a set of disjoint intervals. Let $\{p(x, \theta)\}$ be a family of probability densities on $\cup I_k$, and let $\phi(x_1, x_2)$ be sufficient for θ on the basis of x_1, x_2. The following results are true:*

(1) *If $m < \infty$, and the hypotheses of Theorem 2.1 are valid on $I_k \times I_k$ $k = 1, 2, \cdots, m$, then $\{p\}$ is (at most) a $(2m - 1)$-parameter exponential family.*

(2) *In addition to the special hypotheses of (1) if $I = \cup \bar{I}_k$, if $p(\cdot, \theta)$ is continuous and if $0 < p(\cdot, \theta) < \infty$ for each θ then p is (at most) an m-parameter exponential family.*

(3) *If $m < \infty$ and if there exists a set $A \subset I_k$ for some $k = 1, 2, \cdots, m$ such that $\mu(A) > 0$; ϕ is continuous on $A \times I_k$ for each k; and $\mu(B) = 0$ implies $M(z, B, A) = 0$ for all $z \in \cup I_k$ (not necessarily $\overline{\cup I_k}$), then $\{p\}$ is (at most) an m parameter exponential family.*

(4) *If in addition to the special hypotheses of (3) $I = \cup \bar{I}_k, p(\cdot, \theta)$ is continuous, and $0 < p(\cdot, \theta) < \infty$ for each θ then $\{p\}$ is a 1-parameter exponential family.*

(5) *Suppose $m \leqq \infty$ and the special hypotheses of (3) are satisfied. If in addition for all $k = 2, 3, \cdots,$*

$$\mu\left[\phi(A, I_k) \cap \left(\bigcup_{l=1}^{k-1} \phi(A, I_l)\right)\right] > 0 \tag{7.2}$$

then $\{p\}$ is a one parameter exponential family.

PROOF.

(1) On each interval I_k, $p(x, \theta) = C(\theta)p(x, \theta_0)\ \{r(x, \theta_0)\}^{Q(\theta)}$, $x \in I_k$. If for example, the statistic ϕ is trivial (see (3.10)) in every neighborhood of $I_j \times I_l$, $j \neq l$, then the $C(\theta)$ and $Q(\theta)$ may be chosen arbitrarily (a total of $2m$ choices) on each I_k, except that the condition $\int p(x, \theta)\, dx = 1$ must be fulfilled. This leaves $2m - 1$ free parameter choices. It is then easily checked that $\{p\}$ is (at most) a $2m - 1$ parameter exponential family. (It is not necessary for this result that $\cup \bar{I}_k = I$.)

(2) The condition of continuity (and $\cup \bar{I}_k = I$) imposes $m - 1$ additional restrictions on the choice of parameters in 1 when $0 < p(x, \theta) < \infty$ at the end-points of the intervals I_k (which can easily be ascertained from ϕ on $I_k \times I_k$). This leaves (at most) an m-parameter exponential family.

(3) Although the proof of Theorem 2.1 is not constructed specifically for the situation in this corollary, it is not hard to check that in this situation,

$$r(x, \theta) = C(k, \theta)(r(x, \theta))^{Q(\theta)}, x \in I_k. \tag{7.3}$$

The constants $C(\cdot, \theta)$ can be arbitrarily chosen, except that $\int p(x, \theta)\, dx = 1$.

This leaves $(m - 1)$-parameter choices, which together with the choice of $Q(\theta)$ make $\{p\}$ an m-parameter exponential family.

(4) The proof of this is just like the proof of (2).

(5) The condition (7.2) implies that on each interval I_k, $\{p\}$ is a 1-parameter exponential family. (7.2) also implies that the choice of $C(\theta)$ and $Q(\theta)$ for the interval I_1 determines the choice for I_2, then for I_3, and so on inductively. The choice of $C(\theta)$ in turn must be dictated by the condition $\int p(x, \theta)\, dx \neq 0$. Hence $\{p\}$ is a 1-parameter exponential family. This completes the proof of the corollary.

It should be noted before concluding these considerations that even if $\phi(x, y)$ does not satisfy any of the preceding hypotheses, a transformation of the range of ϕ; and/or of its domain, I, may yield a new problem in which ϕ does satisfy the desired hypotheses. To be more precise there may exist appropriate functions α, β such that $\tilde{\phi} = \alpha(\phi(\beta(x), \beta(y)))$ is continuous as a function of $\beta(x)$, $\beta(y)$. If α and β are appropriately chosen—α one to one and β almost 1-1 and measure preserving are sufficient conditions—then ϕ sufficient for $\{p(x, \theta)\}$ implies $\tilde{\phi}$ sufficient for $\{\tilde{p}(\beta, \theta)\}$ (in the Neyman-Bahadur factorization). If β is almost 1-1 then β^{-1} exists a.e., and $\{\tilde{p}(\beta, \theta)\}$ is an exponential family satisfying (6.4) if and only if $\{p(x, \theta)\}$ is a one-parameter exponential family.

8. An important special case: $\phi(x_1, \cdots, x_n) = \sum_{i=1}^n \psi(x_i)$. This section will begin with an example of perhaps the most useful transformation of the type discussed at the end of Section 7. Suppose $\phi(x_1, \cdots, x_n) = \sum_{i=1}^n \psi(x_i)$ as is often the case in many statistical problems. It will be shown using a change of variables that there is always a problem equivalent in the sense of the preceding section in which $\psi(x)$ is monotone non-decreasing. This transformation is the first step in the proof of a "local" theorem regarding sufficient statistics of the type $\phi = \sum \psi$ in which continuity of ψ is not hypothesized.

For $x \varepsilon I$ (assume I is bounded) let

$$(8.1) \quad t(x) = \mu\{\xi: \xi \varepsilon I, \psi(\xi) < \psi(x)\} + \mu\{\xi: \xi \varepsilon I, \psi(\xi) = \psi(x), \xi \leqq x\}.$$

Let

$$(8.2) \qquad \tilde{\psi}(y) = \psi(t^{-1}(y)), \qquad \tilde{r}(y, \theta) = r(t^{-1}(y), \theta)$$

(where r is defined as before by (2.4)). It is necessary to prove

LEMMA 8.1. *If $\phi = \sum \psi$ is sufficient for p on the basis of n independent observations, then $\tilde{\psi}$ and $\tilde{r}$ are well defined by* (8.2) *and* (8.1) *almost everywhere $\mu(I^n)$, $\tilde{\psi}$ is sufficient for $\{\tilde{r}\}$, and $\{\tilde{r}\}$ is a one-parameter family satisfying* (6.4) *if and only if the family $\{r\}$ is also.*

PROOF. The only problem with the definition of $\tilde{\psi}$ occurs if $t(x_1) = t(x_2)$ but $\psi(x_1) \neq \psi(x_2)$. Consider the collection of equivalence classes $\mathcal{E}(x) = \{\xi: t(\xi) = t(x)\}$. If $\psi(\mathcal{E}_0)$ is not a point, then there is an interval i of positive length (not necessarily open) such that $\psi(\mathcal{E}_0) \subset i$ and $x \not\varepsilon \mathcal{E}_0$ implies $\psi(x) \not\varepsilon i$. Hence $\{\mathcal{E}: \psi(\mathcal{E})$ is not a point$\}$ is countable. This establishes that (8.2) defines $\tilde{\psi}$ uniquely except perhaps for a countable set of points (which will be discontinuities of $\tilde{\psi}$). For each θ there exists a $z \varepsilon I^{n-1}$ such that

$$(8.3) \qquad r(x, \theta) = (1/k)\omega(\lambda + \psi(x), \theta), \qquad \text{a.e. } \mu(I)$$

where $k = \prod_{i=2}^{n} r(z_i, \theta)$, $\lambda = \sum_{i=2}^{n} \psi(z_i)$. It may even be assumed that a version of $r(\cdot, \theta)$ has been chosen such that (8.3) is valid everywhere. Then

$$(8.4) \qquad \tilde{r}(y, \theta) = r(t^{-1}(y), \theta) = (1/k)\omega(\lambda + \psi(t^{-1}(y)), \theta)$$

is uniquely defined except perhaps for a countable number of points.

It can easily be checked that the transformation t is measure preserving in the sense that for any Borel set B, $\mu(t^{-1}(B)) = \mu(t(t^{-1}(B)))$. It follows that

$$(8.5) \qquad \begin{aligned} \prod \tilde{r}(x_i, \theta) &= \prod r(t^{-1}(x_i), \theta) \\ &= \omega(\sum \psi(t^{-1}(x_i), \theta) = \omega(\sum \tilde{\psi}(x_i), \theta), \qquad \text{a.e.} \end{aligned}$$

So $\sum \tilde{\psi}$ is sufficient for $\{\tilde{r}\}$. Similarly $\{\tilde{r}\}$ is a one parameter exponential family if and only if $\{r\}$ is. This completes the proof of the lemma.

It should be clear that in the preceding considerations we have nowhere used the fact that I is an interval, only that it is bounded. If I is unbounded, a slightly different definition of t may be used to yield a transformation with the desired properties.

It is clear from Corollary 7.2 that it cannot be expected for general ϕ that r is a one parameter exponential family on all of I. However under very weak conditions it can be asserted that r is locally a one-parameter exponential family. The following theorem contains this result.

The hypotheses of Theorem 8.1 may seem at first a bit strange. The difficulty is that Theorem 8.1 is stated without the assumption that $\psi(x)$ be monotone non-decreasing. As was shown in the first part of this section there is no loss of generality in assuming that $\psi(x)$ is monotone. Theorem 8.1′ consists of the statement of Theorem 8.1 specialized to the case where $\psi(x)$ is monotone. The first step of the proof of Theorem 8.1 is to show by using the transformation t (8.1) that the conditions of Theorem 8.1 imply there exists an equivalent problem satisfying the conditions of Theorem 8.1′.

THEOREM 8.1. *Let $\{p(x, \theta)\}$ be a family of probability densities on a measurable subset J of E^1 satisfying Assumption 2.1 on J. Suppose $\phi(x_1, \cdots, x_n) = \sum_{i=1}^{n} \psi(x_i)$ is sufficient for $\{p\}$. Suppose there is a subset $A \subset J$ with $\mu(A) > 0$ such that for any $B \subset \psi(A)$, $\mu(B) = 0$ implies $\mu(\psi^{-1}(B)) = 0$. Let $x_0 \varepsilon J$ be any point such that for all $\epsilon > 0$:*

$$(8.6) \qquad \mu\{\psi^{-1}\{x\colon x \varepsilon \psi(J),\ 0 < \psi(x) - \psi(x_0) < \epsilon\}\} > 0$$

and

$$\mu\{\psi^{-1}\{x\colon x \varepsilon \psi(J), 0 < \psi(x_0) - \psi(x) < \psi(x) < \epsilon\}\} > 0.$$

Then there exists a neighborhood Q of $\psi(x_0)$ such that p is a one parameter exponential family on $\psi^{-1}(Q \cap \psi(J))$.

A simpler (though at first glance less general) statement of the preceding is

THEOREM 8.1$'$. *Let $\{p(x, \theta)\}$ be a family of probability densities on an interval $I \subset E^1$, satisfying Assumption* 2.1 *on I. Suppose $\phi(x_1, \cdots, x_n) = \sum \psi(x_i)$ is sufficient for $\{p\}$ where ψ is monotone non-decreasing. Suppose $\psi'(x) > 0$ on a set of positive measure $A \subset I$. Let x_0 be a point such that ψ is continuous at x_0. Then there exists a neighborhood K of x_0 such that $\{p\}$ is a one-parameter exponential family on K, having the form* (8.10).

PROOF. By transforming the real line according to $t(x) = x/(1 + |x|)$ we may assume that the set J is bounded. Then the transformation t of (8.1) can be used. It is easily checked using Lemma 8.1 that the point x_0 of Theorem 8.1 becomes an x_0 of Theorem 8.1$'$, and similarly the set A [or rather $\psi(\psi^{-1}(A))$] transforms into an interval A appropriate for the hypotheses of Theorem 8.1$'$. The conclusion of Theorem 8.1$'$ is slightly stronger than the transformation by t of the conclusion of Theorem 8.1. (It may be stronger at points x such that $\mu(\psi^{-1}(\psi(x))) > 0$.)

The hypotheses of Theorem 8.1$'$ will be assumed throughout the remainder of this proof except where otherwise noted. It will also be assumed that $n = 2$, which is no loss of generality. The proof of Theorem 8.1$'$ follows approximately the outline of the proof of Theorem 2.1 with only a few major differences. Where it is possible, the proof of this theorem will consist of references to the proof of Theorem 2.1.

The first step in the proof of Theorem 8.1 is an analog to Theorem 5.1; namely that for each θ there is a version of $r(x, \theta)$ which is continuous at any continuity point of ψ. To show this, let y_i, $i = 1, 2, \cdots$, be any sequence in I such that $y_i \to y_0$ where ψ is continuous at y_0. Let σ be a continuous function defined on I such that $\sigma(y_i) = \psi(y_i)$ $i = 0, 1, 2, \cdots$.

Using Lemma 5.1, for any B,

$$\begin{aligned}\lim_{i\to\infty} M_{\Sigma\psi}(y_i, B, A) &= \lim_{i\to\infty} M_{\Sigma\sigma}(y_i, B, A) \\ &= M_{\Sigma\sigma}(y_0, B, A) = M_{\Sigma\psi}(y_0, B, A).\end{aligned} \tag{8.7}$$

Hence $M(\cdot, B, A)$ is continuous at y_0. Using the procedure of the proof of Theorem 5.1 it is then easy to show that r is continuous at y_0.

We now turn to the analog of Lemma 6.1. It will be shown there exist versions of ω, r, and ψ such that the N-B factorization is an equality on $A' \times J$, where A' and J are suitably chosen. The first main change from Lemma 6.1 lies in the choice of A and T. In this case let $A = \{x: x \varepsilon S^d, \psi(x) \varepsilon \{\psi(S^d)\}^d\}$ where S is any set with $\mu(S) > 0$ such that $B \subset S$. $\mu(B) = 0$ implies $\mu(\psi^{-1}(B)) = 0$. In particular, S can always be a set of the form $S = \{x: \epsilon < \psi'(x) < 1/\epsilon\}$ for some $\epsilon > 0$. Let T be the set of continuity points of ψ such that $y \varepsilon T$ implies the N-B factorization is valid for almost all $x \varepsilon I$. The proof that

$$r(x, \theta)r(y, \theta) = \omega(\phi(x, y), \theta), \quad (x, y) \varepsilon A \times T, \theta \varepsilon \Theta \tag{8.8}$$

is almost word for word the same as the proof of the analagous fact in Lemma 6.1.

Let x_0, $y_0 \varepsilon A \times T$ (as defined in the preceding paragraph). Since $A^d = A$ and

$T^d = T$, for any neighborhood Z of x_0, y_0,

$$\beta = \{\zeta : \exists (x, y) \in Z \cap (A \times T) : \psi(x) + \psi(y) = \zeta\}$$

contains an open interval of positive length about the point $\phi(x_0, y_0)$([5], p. 68). If Z is chosen small enough, $\zeta \in B$ implies $|\omega(\zeta, \theta,) - \omega(\phi(x_0, y_0), \theta)| < \epsilon$. Hence $\omega(\cdot, \theta)$ is continuous at $\zeta_0 = \phi(x_0, y_0)$.

Again using [5], p. 68 there exists a non-empty set $A' \subset A$ and an interval $J \subset I$ such that $(A')^d = A'(\mu(A') > 0)$ and such that $\phi(A', J) \subset K$ where K is a closed interval contained in the $\overline{\text{interior of}}$ $\phi(A \times T)$. ω is uniformly continuous on $\phi(A', J)$, hence if $x_1, y_1 \in (A' \times J)$, $\lim_{\zeta \to \phi(x_1, y_1)} \omega(\zeta, \theta)$ exists. If ψ is chosen so as to be continuous from the left (or right) it is then easily checked using the previous two paragraphs that $\lim_{y \to x^-} r(y, \theta)$ always exists; and if r is also chosen to be continuous from the left (or right) the N-B factorization (8.8) is valid everywhere in $A' \times J$. This analog of Lemma 6.1 is satisfactory for the proof of Theorem 8.1'.

It is a somewhat tedious matter to alter the statement and proof of Theorem 4.1 to fit the conditions of the theorem at hand. I will not do this in detail here.

The continuity conditions of ψ and r on A' and J and validity of the N-B factorization on $A' \times J$ which have been previously established are sufficient to prove analogs of Lemmas 4.1, 4.2, and 4.3, and from there to establish that there exists a sufficiently small open set O such that $O \cap A' \neq \phi$,

$$r(x, \theta) = c(\theta)[r(x, \theta_0)]^{Q(\theta)}, \qquad \text{a.e. } O \cap A'. \tag{8.9}$$

Since $M(z, B, A')$ is continuous at any continuity point of ψ, the method of Lemma 6.2 applied in a sufficiently small neighborhood N of z proves that r is an exponential family on that neighborhood. It must, in fact, be true that in such a neighborhood

$$r(x, \theta) = c(\theta)e^{Q(\theta)\psi(x)}, \qquad \text{a.e. } (N). \tag{8.10}$$

This completes the proof of Theorems 8.1 and 8.1'.

In particular, when ψ is monotone non-decreasing there will be a sequence of disjoint open intervals I_i such that $\{p\}$ where restricted to any given I_i is a one-parameter exponential family. Corollaries analagous to Corollary 7.2, 1–5, are possible. For instance, suppose

$$\mu\{\phi^{-1}(\phi(I_j, I_k)) \cap (I_j \times I_j)\} > 0 \quad \text{and} \quad \mu\{\phi^{-1}(\phi(I_j, I_j)) \cap (I_j \times I_k)\} > 0. \tag{8.11}$$

Then $\{p\}$ is a one-parameter exponential family on $I_i \cup I_j$.

The considerations of the preceding paragraph appropriately transformed by t^{-1} of course also apply if the hypotheses of Theorem 8.1 are satisfied.

Recall from Example 3.3 that some hypotheses such as that concerning the subset A in Theorem 8.1 is necessary. It can thus be seen that when $\phi = \sum \psi$ is sufficient Theorem 8.1 and its corollaries give a nearly complete characterization of the situation when it is possible to conclude that the densities are locally exponential families.

REFERENCES

[1] BAHADUR, R. R. (1954). Sufficiency and statistical decision functions. *Ann. Math. Statist.* **25** 423–463.

[2] BARANKIN, E. W. and MAITRA, A. P. (1963). Generalizations of the Fisher-Darmois-Koopman-Pitman theorem on sufficient statistics. *Sankhyā* (Series A) **25** 217–244.

[3] DARMOIS, G. (1935). Sur les lois de probabilite a estimation exhaustive. *C. R. Acad. Sci. Paris* **260** 1265–1266.

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

[5] HALMOS, P. R. (1950). *Measure Theory.* Van Nostrand, New York.

[6] HALMOS, P. R. and SAVAGE, L. J. (1949). Applications of the Radon-Nikodym theorem to the theory of sufficient statistics. *Ann. Math. Statist.* **20** 225–241.

[7] KOOPMAN, L. H. (1936). On distributions admitting a sufficient statistic. *Trans. Amer. Math. Soc.* **39** 399–409.

[8] MUNROE, M. E. (1953). *Introduction to Measure and Integration.* Wesley, Cambridge.

[9] NEYMAN, J. (1935). Sur un teorema concernente la cosidette statisticle sufficienti. *Giorn. Ist. Ital. Attuari* **6** 320–334.

[10] ZHURAVLEV, O. G. (1963). Minimal sufficient statistics for a sequence of independent random variables. *Theor. Probability Appl.* **7** 218–220.