Z. Wahrscheinlichkeitstheorie 2, 111—117 (1963)

A Characterization of the one Parameter Exponential Family of Distributions by Monotonicity of Likelihood Ratios

By

RUDOLF BORGES and JOHANN PFANZAGL*

Summary

Any one parameter exponential family of distributions has monotone likelihood ratios. As the product probabilities of n identical distributions of an exponential family form again an exponential family, it has monotone likelihood ratios for arbitrary n. Furthermore, the members of an exponential family are mutually absolutely continuous. In Part 1, we show that these properties uniquely characterize the exponential family. The application of this result to the theory of testing hypotheses (Part 2) shows that if a family of mutually absolutely continuous distributions has uniformly most powerful tests for arbitrary levels of significance, and arbitrary sample sizes, then it is necessarily an exponential family.

1. The main theorem

Let $\mathfrak{P} = \{P_\theta, \theta \in \Theta\}$ be a family of probability measures on a measurable space $(\mathfrak{X}, \mathscr{A})$. The family $\mathfrak{P}$ is assumed to be dominated by a σ-finite measure** $\mu|\mathscr{A}$. The density of P_θ with respect to μ is denoted by p_θ.

We shall say that the family $\mathfrak{P}$ has *monotone likelihood ratios* with respect to a probability measure $P_0|\mathscr{A}$ if there exists an $\mathscr{A}$-measurable function $T|\mathfrak{X}$ independent of $\theta \in \Theta$ and to each $\theta \in \Theta$ a nondecreasing function H_θ, such that

$$p_\theta(x)/p_0(x) = H_\theta(T(X)) \qquad (P_0 + P_\theta - \text{a.e.}) \tag{1}$$

The family $\mathfrak{P}$ is called an *exponential family* if there exist $\mathscr{A}$-measurable functions $g|\mathfrak{X}$ and $h|\mathfrak{X}$ and two functions $a|\Theta$ and $c|\Theta$, such that

$$p_\theta(x) = c(\theta)\, h(x) \exp[a(\theta)\, g(x)] \qquad (\mu - \text{a.e.})\,. \tag{2}$$

The product probability of n identical probability measures $P_\theta \in \mathfrak{P}$ will be denoted by $P_\theta^n = P_\theta \times \cdots \times P_\theta$, the family by $\mathfrak{P}^n = \{P_\theta^n : \theta \in \Theta\}$.

If $\mathfrak{P}$ is an exponential family, the distributions are mutually absolutely continuous and have monotone likelihood ratios for all $\theta, \tau \in \Theta$ such that $a(\theta) > a(\tau)$.

As the product family $\mathfrak{P}^n$ is again an exponential family, $\mathfrak{P}^n$ has these properties for arbitrary n. In the following it will be shown that this characterizes

* The research was done while this author was a Visiting Professor in the Department of Statistics at the University of Chicago. It was supported by Research Grants Nos. NSF-G10368 and NSF-G21058 from the Division of Mathematical, Physical and Engineering Sciences of the National Science Foundation.

** $f|D$ denotes a function f defined in the domain D.

the exponential family in a unique way. In fact, less is sufficient to show that $\mathfrak{P}$ is an exponential family.

Theorem 1. *Let $\mathfrak{P}$ be a family of probability measures P_θ, and $P_0 \notin \mathfrak{P}$. Let P_0 and each P_θ be mutually absolutely continuous. If the family $\mathfrak{P}^n$ of n-fold product probability measures of $\mathfrak{P}$ has monotone likelihood ratios with respect to the product probability measure P_0^n for each n, then $\mathfrak{P}$ is an exponential family with $a(\theta) \geqq 0$.*

At first we prove two lemmas.

Lemma 1. *Let $Q|\mathscr{B}$ be a probability measure and $f_\theta(y)$ with $\theta \in \Theta$ be a family of $\mathscr{B}$-measurable nonnegative functions. Assume that the class $\mathscr{C}$ of all sets $\{y : f_\theta(y) \leqq c\}$ with $\theta \in \Theta$ and $c \geqq 0$ is linearly ordered by the relation of inclusion.*

Then a) *there exists a $\mathscr{B}$-measurable nonnegative function S such that for $S(y) > 0$.*

$$S(y) = Q\text{-ess sup}\{S(z) : S(z) \leqq S(y)\} \tag{3}$$

and b) *there exists a left continuous nondecreasing function G_θ such that $G_\theta(0) = 0$ and*

$$G_\theta(S(y)) = f_\theta(y) \qquad Q - \text{a.e.} \tag{4}$$

If $Q(C_0) = Q\,(\bigcap_{i=1}^{\infty} C_i)$ for $C_i \in \mathscr{C}$, $i = 0, 1, \ldots$, implies $C_0 = \bigcap_{i=1}^{\infty} C_i$, then Equation (4) *holds everywhere.*

Remark. This lemma is a modification of Lemmas 1 and 2 in [*4*], where the class $\mathscr{C}$ is linearly ordered by the relation of inclusion up to sets of probability zero. Hence we can restrict ourselves to a short outline of the proof.

Proof. a) We choose one $D \in \mathscr{C}$ for each value $Q(C)$ on $\mathscr{C}$ and denote the class of all D by $\mathscr{D}$.

Hence the linear order of $\mathscr{C}$ implies that $\mathscr{D}$ is linearly ordered. As each equation $Q(D) = d$ has at most one solution $D \in \mathscr{D}$, this implies that the intersection and the union of any subclass of $\mathscr{D}$ can be written as the intersection or the union of a countable subclass of $\mathscr{D}$. Hence the intersection and the union of each subclass of $\mathscr{D}$ are $\mathscr{B}$-measurable.

We define

$$\begin{cases} D_y := \bigcap\{D : y \in D \in \mathscr{D}\} \\ S(y) := Q(D_y). \end{cases} \tag{5}$$

First we observe that S is $\mathscr{B}$-measurable. For, (5) implies

$$\begin{aligned}\{y : S(y) < c\} &= \{y : y \in D \text{ for some } D \in \mathscr{D} \text{ with } Q(D) < c\} \\ &= \bigcup\{D : Q(D) < c \text{ and } D \in \mathscr{D}\}.\end{aligned}$$

Furthermore we obtain from (5) that up to a Q-null set

$$B = \{z : S(z) \leqq Q(B)\} \tag{6}$$

if $B \in \mathscr{C}$ or $B = D_y$.

For, first let $B = D \in \mathscr{D}$. Then $y \in D$ implies $D_y \subset D$ and hence $S(y) \leqq Q(D)$. As the class of all D_y is obviously linearly ordered and contains at most two different D_{y_1}, D_{y_2} with $Q(D_{y_1}) = Q(D_{y_2})$, this implies by (5) up to a Q-null set

$$\bigcup\{D_y : S(y) \leqq Q(D)\} \subset D.$$

From $y \in D_y$ we have

$$\{y : S(y) \leqq Q(D)\} \subset \bigcup \{D_y : S(y) \leqq Q(D)\}.$$

Therefore (6) holds up to a Q-null set for $B = D$. Since by definition of $\mathscr{D}$ each $C \in \mathscr{C}$ is equal to some $D \in \mathscr{D}$ up to a Q-null set, (6) holds for all $B \in \mathscr{C}$. Since each D_y can be written as a countable intersection, (6) follows for $B = D_y$ by the Monotone Limit Theorem.

Setting $B = D_y$ in Equation (6), we obtain from (5) that

$$S(y) = Q\{z : S(z) \leqq S(y)\}. \tag{7}$$

From (7) we obtain

$$\begin{aligned} Q\{z : S(y) - \varepsilon < S(z) \leqq S(y)\} \\ = S(y) - \sup\{S(z) : S(z) \leqq S(y) - \varepsilon\} \geqq \varepsilon > 0. \end{aligned} \tag{8}$$

This yields (3).

We define a left continuous nondecreasing function G_θ by

$$G_\theta(s) := \inf\{c : Q\{y : f_\theta(y) \leqq c\} \geqq s \;\&\; c \geqq 0\}. \tag{9}$$

Hence the following inequalities are equivalent

$$G_\theta(s) \leqq c \quad \text{and} \quad s \leqq Q\{y : f_\theta(y) \leqq c\}. \tag{10}$$

Since equation (6) holds for $B \in \mathscr{C}$ up to a Q-null set we obtain (4) almost everywhere. If $Q(C_0) = Q(\bigcap_{i=1}^{\infty} C_i)$ with $C_i \in \mathscr{C}$, $i = 0, 1, \ldots,$ implies $C_0 = \bigcap_{i=1}^{\infty} C_i$ then $Q(D_y) \leqq Q(D)$ implies $D_y \subset D$ without restriction. Hence in this case, all relations in the proof of (6) hold everywhere and hence (6) and (4) too, q.e.d.

Lemma 2. *Let $f_1 | \mathfrak{X}$ and $f_2 | \mathfrak{X}$ be two real functions with the following properties*

1. *If for some $x_i, y_i \in \mathfrak{X}$ $(i = 1, \ldots, r)$*

$$\sum_{i=1}^{r} f_1(x_i) < \sum_{i=1}^{r} f_1(y_i) \tag{11}$$

then

$$\sum_{i=1}^{r} f_2(x_i) \leqq \sum_{i=1}^{r} f_2(y_i). \tag{12}$$

2. *There exist $x_1, x_2, \in \mathfrak{X}$ such that*

$$f_1(x_1) < f_1(x_2). \tag{13}$$

Then there exists a function $\lambda(x)$ and constants $a_k \geqq 0$ and b_k such that for $k = 1, 2$

$$f_k(x) = a_k \lambda(x) + b_k. \tag{14}$$

Remark. This lemma could also be proved by regarding the sums $\sum f_1(x_i)$ as elements of a semi-group of real numbers. Using the order relations one can extablish Cauchy's functional equation.

Proof. By Assumption 2. for every $x \in \mathfrak{X}$ and every positive integer n there exists an integer $m_n(x)$ such that

$$\frac{m_n(x)}{n}(f_1(x_2) - f_1(x_1)) < (f_1(x) - f_1(x_1)) < \frac{m_n(x) + 2}{n}(f_1(x_2) - f_1(x_1)). \tag{15}$$

Both inequalities can easily be transformed into an inequality of the form (11) where the x_i's and y_i's are equal to x, x_1 or x_2. By Assumption 1. this inequality implies (12), which in turn yields

$$(16)\quad \frac{m_n(x)}{n}\left(f_2(x_2)-f_2(x_1)\right) \leqq \left(f_2(x)-f_2(x_1)\right) \leqq \frac{m_n(x)+2}{n}\left(f_2(x_2)-f_2(x_1)\right).$$

Hence the strong inequalities (15) yield the weak inequalities (16).

We define

$$(17)\quad \lambda(x) := \lim_{n\to\infty} \frac{m_n(x)}{n}$$

and obtain

$$(18)\quad f_k(x) = \lambda(x)\left[f_k(x_2)-f_k(x_1)\right] + f_k(x_1).$$

Letting $a_k = [f_k(x_2) - f_k(x_1)]$ and $b_k = f_k(x_1)$ for $k = 1, 2$, the desired result follows, since $a_k \geqq 0$ by (13) and Assumption 1.

Proof of Theorem 1. 1. By assumption the likelihood ratios $\prod_{i=1}^{n} p_\theta(x_i)/p_0(x_i)$ are monotonic for each n. Hence for every n there exist a measurable function T_n, independent of θ, and nondecreasing functions $H_\theta^{(n)}$ depending on θ such that

$$(19)\quad H_\theta^{(n)}(T_n(x_1,\ldots,x_n)) = \prod_{i=1}^{n} \frac{p_\theta(x_i)}{p_0(x_i)} \qquad P_0^n - \text{a.e.}$$

This yields

$$(20)\quad H_\theta^{(n)}(T_n(x_1,\ldots,x_n)) = \prod_{i=1}^{n} H_\theta(T(x_i)) \quad P_0^n - \text{a.e.},$$

where we drop the index 1 at H and T for $n = 1$.

We shall replace (20) by a similar set of equations which hold everywhere. These equations will yield the exponentiality by Lemma 2.

2. At first we put $Q|\mathscr{B} = P_0|\mathscr{A}$ in Lemma 1 and $f_\theta(x) = H_\theta(T(x))$. Let $\mathscr{C}$ be the class of all sets $\{x : H_\theta(T(x)) \leqq c\}$ with $\theta \in \Theta$ and $c \geqq 0$. Since the elements of $\mathscr{C}$ are of the form $\{x : T(x) < k\}$ or $\{x : T(x) \leqq k\}$, the class $\mathscr{C}$ is ordered by the relation of inclusion. Hence by Lemma 1 there exists a measurable nonnegative function $S(x)$ such that for $S(x) > 0$

$$(21)\quad S(x) = P_0 - \operatorname{ess\,sup}\{S(y) : S(y) \leqq S(x)\}.$$

Furthermore there exists a left continuous nondecreasing function G_θ depending on $\theta \in \Theta$ such that

$$(22)\quad G_\theta(S(x)) = H_\theta(T(x)) \qquad P_0 - \text{a.e.}$$

Inserting this into (20) we obtain

$$(23)\quad H_\theta^{(n)}(T_n(x_1,\ldots,x_n)) = \prod_{i=1}^{n} G_\theta(S(x_i)) \qquad P_0^n - \text{a.e.}$$

3. Secondly let $Q|\mathscr{B}$ of Lemma 1 be the product P_0^n of n identical probability measures $P_0|\mathscr{A}$. Furthermore put $f_\theta(y) = \prod_{i=1}^{n} G_\theta(S(x_i))$. Denote the class of all sets

$$(24)\quad C_\theta(c) := \{(x_1,\ldots,x_n) : \prod_{i=1}^{n} G_\theta(S(x_i)) \leqq c\}$$

with $\theta \in \Theta$ and $c \geqq 0$ by $\mathscr{C}_n$.

By (23) the sets $C_\theta(c)$ are of the form

$$\{(x_1, \ldots, x_n) : T_n(x_1, \ldots, x_n) < k\} \quad \text{or} \quad \{(x_1, \ldots, x_n) : T_n(x_1, \ldots, x_n) \leqq k\}$$

up to a P_0^n-null set.

Hence for $D := \bigcap_{i=1}^{\infty} C_{\vartheta_i}(c_i)$ the inequality $P_0^n(D) \leqq P_0^n(C_\tau(d))$ yields $D \subset C_\tau(d)$ up to a P_0^n-null set. Let $(y_1, \ldots, y_n) \in D$. Then the monotonicity of G_{ϑ_i} and definition (24) yield

$$\{(x_1, \ldots, x_n) : S(x_i) \leqq S(y_i), \quad i = 1, \ldots, n\} \subset C_{\vartheta_i}(c_i).$$

Hence we have

$$\{(x_1, \ldots, x_n) : S(x_i) \leqq S(y_i), \quad i = 1, \ldots, n\} \subset D \subset C_\tau(d) \quad P_0^n - \text{a.e.}$$

By the left continuity of G_τ and by (21) it follows that $(y_1, \ldots, y_n) \in C_\tau(d)$. Hence $P_0^n(D) \leqq P_0^n(C_\tau(d))$ implies $D \subset C_\tau(d)$ without restriction. Similarly $P_0^n(D) \geqq P_0^n(C_\tau(d))$ implies $D \supset C_\tau(d)$. Hence $P_0^n(\bigcap_{i=1}^{\infty} C_{\vartheta_i}(c_i)) = P_0^n(C_\tau(d))$ implies $\bigcap_{i=1}^{\infty} C_{\vartheta_i}(c_i) = C_\tau(d)$. Setting $\vartheta_i = \vartheta$ and $c_i = c$ for all $i = 1, 2, \ldots$ we obtain furthermore that $P_0^n(C_\vartheta(c)) \leqq P_0^n(C_\tau(d))$ implies $C_\vartheta(c) \subset C_\tau(d)$, that is $\mathscr{C}_n$ is linearly ordered by inclusion.

Hence by Lemma 1 there exists a measurable function $S_n(x_1, \ldots, x_n)$ independent of $\theta \in \Theta$ and a monotone non-increasing function $G_\theta^{(n)}$ depending on $\theta \in \Theta$ such that everywhere

$$G_\theta^{(n)}(S_n(x_1, \ldots, x_n)) = \prod_{i=1}^{n} G_\theta(S(x_1)). \tag{25}$$

4. Now we shall apply Lemma 2. By (25) the inequality

$$\prod_{i=1}^{n} G_\tau(S(x_i)) < \prod_{i=1}^{n} G_\tau(S(y_i)) \tag{26}$$

for some $\tau \in \Theta$ yields $S_n(x_1, \ldots, x_n) < S_n(y_1, \ldots, y_n)$ and this in turn implies

$$\prod_{i=1}^{n} G_\theta(S(x_i)) \leqq \prod_{i=1}^{n} G_\theta(S(y_i)) \tag{27}$$

for all $\theta \in \Theta$. Since P_τ and P_0 are mutually absolutely continuous and not identical there exist $x_1, x_2 \in \mathfrak{X}$ such that

$$0 < G_\tau(S(x_1)) < G_\tau(S(x_2)) < +\infty. \tag{28}$$

With $f_1(x) = \log G_\tau(S(x))$ and $f_2(x) = \log G_\theta(S(x))$ we obtain by Lemma 2

$$G_\theta(S(x)) = c(\theta) \exp[a(\theta) g(x)] \tag{29}$$

with $a(\theta) \geqq 0$. From (1) and (22) we obtain the desired result

$$p_\theta(x) = c(\theta) p_0(x) \exp[a(\theta) g(x)] \qquad P_0 - \text{a.e.} \tag{30}$$

2. The statistical significance of the main theorem

In [4] (p. 170) it had been shown: If for each level of significance a (randomized) test exists for testing a hypothesis P_0, which test is most powerful against any member of a class $\mathfrak{P}$, then $\mathfrak{P}$ has nondecreasing likelihood ratios with respect to P_0. Using this result, Theorem 1 immediately implies:

Theorem 2. *Let $\mathfrak{P}$ be a family of probability measures P_θ and $P_0 \notin \mathfrak{P}$. Let P_0 and each P_θ be mutually absolutely continuous. If for any sample size n and any level of significance there exists a (randomized) test for the hypothesis P_0^n which is uniformly most powerful against the class of alternatives $\mathfrak{P}^n$, then $\mathfrak{P}$ is an exponential family.*

This theorem clearly shows that the concept of a "uniformly most powerful" test is far too restrictive to be generally applicable in testing theory. Under the assumption of mutual absolute continuity, its applicability is limited to a specific family, namely the exponential family. A treatment of the general case — abandoning the assumption of mutual absolute continuity — will be given in [*1*].

Theorem 1 is closely related to a well-known theorem stating that a class of mutual absolutely continuous probability measures, admitting a sufficient statistic, is an exponential family. The relation to Theorem 1 consists in the fact that — according to the factorization theorem — $T(x)$ is a sufficient statistic. The most recent version of the theorem on sufficient statistics and exponential families is contained in DYNKIN ([*2*] Section 2). He assumes that μ is the Lebesgue measure on the real line and the densities $p_\theta(x)$ are continuous and have piecewise continuous derivatives with respect to x. Earlier versions of this theorem (for references see e.g. LEHMANN [*3*] p. 51) use even more restrictive regularity conditions concerning the density function $p_\theta(x)$.

The conditions of Theorem 1 are more restrictive in a single point: We assume that the likelihood ratios are *monotone* functions of the sufficient statistic. This additional assumption is suggested by some questions originating in the theory of testing hypotheses as outlined above. This is sufficient to arrive at results which are independent of any further regularity conditions on the density functions and any further condition on the measure $\mu\,|\,\mathscr{A}$, but σ-additivity. Especially $\mu\,|\,\mathscr{A}$ can be the counting measure.

Theorem 2 refers to the general case which assumes no relationship between the hypothesis and the class of alternatives. Applications, however, often suggest to think of a family with a naturally ordered parameter space. If in such a case a family of tests exists for any hypothesis $\vartheta_0 \in \Theta$, these tests being uniformly most powerful against the class of alternatives $\{\vartheta : \vartheta > \vartheta_0\}$, then the immediate application of Theorem 2 yields that to any ϑ_0 there exists a measurable function $g(x;\vartheta_0)$ and functions $c(\vartheta;\vartheta_0)$ and $a(\vartheta;\vartheta_0)$ of $\vartheta > \vartheta_0$ such that P_ϑ — a.e. for $\vartheta > \vartheta_0$

$$p_\vartheta(x) = p_{\vartheta_0}(x)\, c(\vartheta;\vartheta_0) \exp[a(\vartheta;\vartheta_0)\, g(x;\vartheta_0)] \tag{31}$$

Consider three parameters $\vartheta_1, \vartheta_2, \vartheta_3$ such that $\vartheta_3 > \vartheta_2 > \vartheta_1$ and Equation (31) for $(\vartheta, \vartheta_0) = (\vartheta_3, \vartheta_2), (\vartheta_3, \vartheta_1)$ and $(\vartheta_2, \vartheta_1)$. Since we have assumed that

$$\{x : p_\vartheta(x) > 0\}$$

is independent of ϑ up to null sets, comparison of these three equations yields

$$[a(\vartheta_3;\vartheta_1) - a(\vartheta_2;\vartheta_1)]\, g(x;\vartheta_1) - a(\vartheta_3;\vartheta_2)\, g(x;\vartheta_2) = \log \frac{c(\vartheta_3;\vartheta_2)\, c(\vartheta_2;\vartheta_1)}{c(\vartheta_3;\vartheta_1)} \tag{32}$$

Hence the functions $g(x;\vartheta)$ with $\vartheta \in \Theta$ can be chosen such that they are mutually linear dependent, say

$$g(x;\vartheta) = b_0(\vartheta)\, g(x;\vartheta_0) + d_0(\vartheta) \tag{33}$$

for any $\vartheta \in \Theta$ and a fixed $\vartheta_0 \in \Theta$. With ϑ_0 fixed we define

$$c(\vartheta) := \begin{cases} c(\vartheta;\vartheta_0) & \text{if} \quad \vartheta > \vartheta_0 \\ 1 & \text{if} \quad \vartheta = \vartheta_0 \\ (c(\vartheta_0;\vartheta)\exp[a(\vartheta_0;\vartheta)\, d_0(\vartheta)])^{-1} & \text{if} \quad \vartheta < \vartheta_0 \end{cases} \tag{34}$$

and

$$a(\vartheta) := \begin{cases} a(\vartheta;\vartheta_0) & \text{if} \quad \vartheta > \vartheta_0 \\ 0 & \text{if} \quad \vartheta = \vartheta_0 \\ -a(\vartheta_0;\vartheta)\, b_0(\vartheta) & \text{if} \quad \vartheta < \vartheta_0 . \end{cases} \tag{35}$$

Furthermore, let $h(x) := p_{\vartheta_0}(x)$ and $g(x) := g(x;\vartheta_0)$. Then Equation (2) follows from (31) and (33)—(35).

Thus we have shown the following: Let $\mathfrak{P}$ be a family of mutually absolutely continuous probability measures with an ordered parameter space Θ. Assume that for each level of significance α, each sample size n and each hypothesis $\vartheta_0 \in \Theta$ there exists a test which is uniformly most powerful against all $\vartheta > \vartheta_0$. Then the family of probability measures is exponential in the sense of Equation (2).

References

[1] Borges, R.: On the existence of uniformly most powerful tests for arbitrary sample sizes (to be published).

[2] Dynkin, E. B.: Necessary and sufficient statistics for a family of probability distributions. Uspechi Mat. Nauk (N.S.) **6** no. 1 (41), 68—90 (1951). Englisch translation: Selected translations in mathematical statistics and probability, **1**, 17—40 (1961).

[3] Lehmann, E. L.: Testing Statistical Hypotheses. New York: Wiley 1959.

[4] Pfanzagl, J.: Über die Existenz überall trennscharfer Tests. Metrika **3**, 169—176 (1960) and **4**, 105—106 (1961).

Universität zu Köln
Seminar für Wirtschafts- und Sozialstatistik
5 Köln-Lindenthal
Albertus-Magnus-Platz 1

(Received November 14, 1962)