ONE-PARAMETER EXPONENTIAL FAMILIES GENERATED BY TRANSFORMATION GROUPS

By R. Borges and J. Pfanzagl[1]

University of Cologne

0. Summary. We consider a one-parameter exponential family generated by an arbitrary group of transformations of an abstract sample space. Topological assumptions about the group are not required. It is shown that such a family has densities either of the type of the normal distribution or of the type of the gamma distribution with respect to an invariant measure. This is a generalization of results of Dynkin (1951), Lindley (1958) and Ferguson (1962 and 1963).

1. Introduction. The important role of exponential distributions in statistical theory has been emphasized by Lehmann (1959). In [3] it was shown that, under the condition of mutual absolute continuity of the distributions, one-sided UMP tests for arbitrary sample sizes and arbitrary levels of significance exist only for one-parameter exponential families. This special role of exponential families in statistical theory suggests a closer examination of the properties of this family. Furthermore, much attention has been devoted to families of distributions which are generated from a single member of this family by a group of transformations of the sample space onto itself. We shall study more deeply in this paper properties of one-parameter exponential families which are generated by a group of transformations of the sample space (see Definition (2.1) and (2.4) below).

The first remarkable result in this direction was obtained by Dynkin (1951) in connection with the question of the existence of a set of sufficient statistics. The following special case of Dynkin's result is of interest here: If $f(x - \vartheta)$, $\vartheta \varepsilon R^1$ is a continuous density of a one-parameter exponential type then $f(x)$ is either a normal density or can be transformed into a density of a gamma distribution taking $x = \log y$ and $\vartheta = \log \sigma$.

Furthermore, for the comparison of the "fiducial" and "Bayesian" method the same problem is of interest. Richter (1954) proved [Theorem 15, p. 338] under relatively weak continuity assumptions on the distribution function that a fiducial distribution is an "a posteriori" distribution with respect to an "a priori" distribution if and only if the parameter ϑ and the sufficient statistic a for ϑ can be transformed into $\tau = \tau(\vartheta)$ and $t = T(a)$ such that (a) the distribution function depends only on $t - \tau$; (b) the "a priori" distribution for τ is the uniform distribution. (We have put quotation-marks since Richter distinguishes logically between the degree of credibility, that is "a posteriori" and "a priori" distributions, and the probability of random or chance experiments.) Lindley (1958)

Received 4 June 1963; revised 29 August 1964.

[1] This work was done while this author was visiting professor at the Mathematics Research Center, United States Army, Madison, under Contract No. DA-11-022-ORD-2059.

found essentially the same theorem under more restrictive differentiability assumptions and, in addition, that "if, for a random sample of any size from the distribution for x, there exists a single sufficient statistic for ϑ then the fiducial argument is inconsistent unless" conditions (a) and (b) above obtain.

In this connection Lindley (1958) proved under differentiability conditions again that the normal and gamma distributions are the only one-parameter exponential distributions generated by transformation groups. Ferguson (1962 and 1963) reconsidered the problem for location and scale parameter exponential families in R^1 without any regularity assumptions and obtained exactly the same result as Dynkin and Lindley.

In connection with the problem of the existence of optimal confidence procedures, Neyman (1938) (see also for example [11]) has proved the existence of type A_1 confidence procedures with confidence coefficient β for one-parameter exponential families possessing densities with respect to the n-dimensional Lebesgue measure satisfying certain differentiability assumptions. These confidence procedures are unbiased if the class of confidence procedures with respect to which they are most accurate (shortest, most selective) contains an unbiased one. On the other hand, in [2] the existence of subjective most accurate confidence procedures has been established. These are a.e. invariant and unbiased (Theorem 3.19 of [2]) if the family of distributions under consideration is generated by a group of transformations and the degree of credibility (*a priori* distribution) is a left invariant (Haar) measure. If a one-parameter exponential family generated by a group is given, it can easily be seen that Neyman's type A_1 confidence procedures are identical with the subjective most accurate ones with respect to a left invariant (Haar) measure. Naturally the question arises, as to whether the normal and gamma distributions are the only examples of one-parameter exponential families generated by groups.

The statistical problems mentioned above suggest investigations of one-parameter exponential families generated by a transformation group without restrictive assumptions such as: the group is isomorphic to the additive group of reals, only distributions in R^1 are considered and the distributions are absolutely continuous with respect to the Lebesgue measure. Our inquiries resulted in the Theorem given in Section 2 below.

Notations: $f \mid D$ denotes the function f defined on the domain D. If a is defined as b we shall write $a := b$. Furthermore R^1 denotes the set of real numbers. The exponential function is always denoted by exp, while e is used for the identity transformation $ex = x$.

2. Basic definitions and formulation of the Theorem. Let $(M, \mathcal{K})$ be a measurable space and $\mathcal{P}$ be a family of probability measure $P_\vartheta \mid \mathcal{K}$, $\vartheta \varepsilon \Theta$.

DEFINITION (2.1). A family $\mathcal{P}$ of probability measures $P_\vartheta \mid \mathcal{K}$, $\vartheta \varepsilon \Theta$, is said to be *generated by a group (of transformations)*, if

(a) the indexing set Θ of $\mathcal{P}$ is a group of $\mathcal{K}$-measurable transformations $\vartheta, \tau, \cdots$ with the product $(\vartheta\tau)x = \vartheta(\tau x)$ and the identity e;

(b) if $\vartheta K := \{\vartheta x : x \varepsilon K\}$ then $P_\vartheta(\vartheta K) = P_e(K)$ for all $\vartheta \varepsilon \Theta$ and all $K \varepsilon \mathcal{K}$. Conditions (a) and (b) are identical with Assumption GI of [2]. These conditions simplify the usual approach by identifying the indexing parameter of the probability measure with the transformation from the beginning. Clearly, in particular examples the usual indexing with parameters has to be replaced by one with transformations, see for example p. 58 and Section 4 of [2]. An example of families of distributions generated by a group is the distribution of n identically and independently distributed random variables $X_1, \cdots, X_n$ with the distribution $F((x - \alpha)/\sigma)$ where $\alpha \varepsilon R^1$ and $\sigma > 0$. The transformation is obviously $\tau_{\alpha,\sigma}(x_1, \cdots, x_n) = \sigma(x_1, \cdots, x_n) + (\alpha, \cdots, \alpha)$.

Definition (2.1) implies

$$(2.2) \qquad P_{\tau\vartheta}(K) = P_\vartheta(\tau^{-1}K).$$

This shows that Definition (2.1) is invariant against the choice of $P_e \varepsilon \mathcal{P}$.

To give still another insight into Definition (2.1) assume that the probability measures $P_\vartheta \mid \mathcal{K}$, $\vartheta \varepsilon \Theta$, possess densities $g(x, \vartheta)$ with respect to a measure $\mu \mid \mathcal{K}$ and that they are mutually absolutely continuous. Then another formulation of (2.2) is:

$$(2.3) \qquad g(x, \tau\vartheta)/g(x, \tau) = g(\tau^{-1}x, \vartheta)/g(\tau^{-1}x, e) \quad \text{a.e.}$$

For, an integral transformation yields

$$P_\vartheta(\tau^{-1}K) = \int_{\tau^{-1}K} \frac{g(x, \vartheta)}{g(x, e)} P_e(dx) = \int_K \frac{g(\tau^{-1}x, \vartheta)}{g(\tau^{-1}x, e)} P_\tau(dx).$$

On the other hand, obviously $P_{\tau\vartheta}(K) = \int_K [g(x, \tau\vartheta)/g(x, \tau)]P_\tau(dx)$. Hence (2.2) is equivalent to (2.3), q.e.d.

A family $\mathcal{P}$ of probability measures $P_\vartheta \mid \mathcal{K}$, $\vartheta \varepsilon \Theta$ is called a *one-parameter exponential family* [7] if each $P_\vartheta \mid \mathcal{K}$ possesses a density

$$(2.4) \qquad g(x, \vartheta) = h(x) \exp [u_\vartheta T(x) + v_\vartheta]$$

with respect to some measure $\mu \mid \mathcal{K}$. Thereby $h \mid M$ and $T \mid M$ are $\mathcal{K}$-measurable functions and u_ϑ, $v_\vartheta \varepsilon R^1$ are constants depending on ϑ. A measure $\lambda \mid \mathcal{K}$ is called *invariant*, if

$$(2.5) \qquad \lambda(\tau K) = \lambda(K) \quad \text{for all} \quad \tau \varepsilon \Theta \quad \text{and} \quad K \varepsilon \mathcal{K}.$$

THEOREM. *Let $\mathcal{P}$ be a one-parameter exponential family generated by a group of transformations and assume that $\mathcal{P}$ contains more than two probability measures. Then there exists an invariant measure $\lambda \mid \mathcal{K}$ depending on the family $\mathcal{P}$ such that the density of each $P_\vartheta \mid \mathcal{K}$ with respect to $\lambda \mid \mathcal{K}$ exists and is of one of the following two types*:

(a) $f(x, \vartheta) = \exp [-\frac{1}{2}(S(x) - w_\vartheta)^2]$ *with* $w_{\tau\vartheta} = w_\tau + a_\tau w_\vartheta$, *where* $a_\tau = \pm 1$ *and* $a_{\tau\vartheta} = a_\tau a_\vartheta$,

(b) $f(x, \vartheta) = (a_\vartheta S(x))^p \exp [-a_\vartheta S(x)]$ *with* $p > 0$, $a_{\tau\vartheta} = a_\tau a_\vartheta$ *and* $a_\vartheta > 0$.

It should be noted that our Definition (2.1) implies

$$f(x, \vartheta) = f(\vartheta^{-1}x, e) \quad \text{a.e.} \tag{2.6}$$

COROLLARY. *Under the assumptions of the Theorem there exists a sufficient statistic $S(x)$ such that the distribution functions $F_\vartheta(y)$ of $S(x)$ on the real line have densities*

(a) $\exp[-\frac{1}{2}(y - w_\vartheta)^2]$ *or*

(b) $(a_\vartheta y)^p \exp[-a_\vartheta y]$, $y > 0$, *with respect to some invariant measure over the real line involving the simple invariance to a group of translations and reflections in* (a) *and of multiplications in* (b). *Conversely, every choice of such an invariant measure yields an exponential family generated by a group.*

3. Properties of the exponential families and a useful convention. We collect a few well known [7] facts about one-parameter exponential families.

The function $c \mid R^1$ defined by

$$\exp[-c(\xi)] := \int_M h(x) \exp[\xi T(x)]\mu(dx) \tag{3.1}$$

is infinitely differentiable with

$$c''(\xi) < 0 \tag{3.2}$$

for all interior points ξ of the convex natural parameter space $\Xi := \{\xi : \exp c(\xi) > 0\}$. The exponential family of probability measures $Q_\xi \mid \mathcal{K}$, $\xi \varepsilon \Xi$, defined by

$$Q_\xi(K) := \int_K h(x) \exp[\xi T(x) + c(\xi)]\mu(dx), \tag{3.3}$$

contains $\mathcal{P}$, since

$$v_\vartheta = c(u_\vartheta). \tag{3.4}$$

Let $\mathcal{P}$ be a one-parameter exponential family generated by a group and assume that $\mathcal{P}$ contains more than two probability measures. Since $\mathcal{P}$ is by (2.2) invariant against the choice of $P_e \varepsilon \mathcal{P}$ and at least one u_ϑ is an interior point of Ξ, all u_ϑ are interior points of Ξ. Hence $E_\vartheta T$ and var T exist. Since T is only determined up to a linear transformation by (2.4) we can assume that $E_e T = 0$ and $\text{var}_e T = 1$. Moreover, for a given T the constants u_ϑ an v_ϑ are only unique up to an additive constant, hence we can choose $u_e = v_e = 0$. These assumptions imply $c(0) = c'(0) = 0$ and $c''(0) = -1$. By a change of the sign of T, if necessary, the finiteness of sup Ξ can always be achieved in the case that $\Xi \neq R^1$. Hence we make the following convention.

CONVENTION. Let $\mathcal{P}$ be a one-parameter exponential family generated by a group and assume that $\mathcal{P}$ contains at least three probability measures. The probability measure $P_e \varepsilon \mathcal{P}$ corresponding to the identity is chosen such that

$$u_e = c(0) = c'(0) = 0, \qquad c''(0) = -1 \tag{3.5}$$

and

$$\xi_0 := \sup \Xi < +\infty \quad \text{if} \quad \Xi \neq R^1. \tag{3.6}$$

In the example of the gamma distribution this convention implies $T(x) = x - 1$, $h(x) = x^{p-1} \exp -x$, $u_\sigma = 1 - 1/\sigma$ and $v_\sigma = -p \log \sigma + (1 - 1/\sigma)$, if μ is proportional to the Lebesgue measure. We remark that $\xi_0 = 1$. In the example of the normal distribution $N(\alpha,1)$, we have $T(x) = x$, $h(x) = \exp(-x^2/2)$, $u_\alpha = \alpha$ and $v_\alpha = -\alpha^2/2$, while $\Xi = R^1$.

If we deviate from the convention (3.5) we shall use other letters for T, u and c. Convention (3.5) implies

$$P_\vartheta(K) = \int_K \exp[u_\vartheta T(x) + v_\vartheta] P_e(dx). \tag{3.7}$$

4. A homorphism onto a matrix group. At first we study the transformations of the function $T(x)$ and the constants u_ϑ and v_ϑ induced by the group Θ of transformations of $x \varepsilon M$. It turns out that these transformations are linear and can be combined with u_ϑ and v_ϑ to matrices which form a group.

Inserting (2.4) and (3.5) into (2.3) we obtain the following basic transformation equation,

$$u_\vartheta T(\tau^{-1}x) + v_\vartheta = (u_{\tau\vartheta} - u_\tau)T(x) + (v_{\tau\vartheta} - v_\tau) \quad \text{a.e.} \tag{4.1}$$

This will be simplified in the following way. Choose a $\vartheta_0 \varepsilon \Theta$ such that $u_{\vartheta_0} \neq 0$. (The existence of such a $u_{\vartheta_0} \neq 0$ is guaranteed by the convention above.) We define

$$a_\tau := (u_{\tau\vartheta_0} - u_\tau)/u_{\vartheta_0} \tag{4.2}$$

and

$$b_\tau := (v_{\tau\vartheta_0} - v_\tau - v_{\vartheta_0})/u_{\vartheta_0} \tag{4.3}$$

for each $\tau \varepsilon \Theta$. Then (4.1) implies

$$T(\tau^{-1}x) = a_\tau T(x) + b_\tau \quad \text{a.e.} \tag{4.4}$$

We define now a 3×3 matrix

$$A_\tau := \begin{pmatrix} a_\tau & u_\tau & 0 \\ 0 & 1 & 0 \\ b_\tau & v_\tau & 1 \end{pmatrix}. \tag{4.5}$$

It is easily seen that

$$A_\tau A_\vartheta = A_{\tau\vartheta}. \tag{4.6}$$

For, from (4.4) and $(\tau\vartheta)^{-1}x = \vartheta^{-1}(\tau^{-1}x)$ we obtain that $a_{\tau\vartheta} = a_\tau a_\vartheta$ and $b_{\tau\vartheta} = a_\vartheta b_\tau + b_\vartheta$. Inserting (4.4) into (4.1) the relations $u_{\tau\vartheta} = u_\tau + a_\tau u_\vartheta$ and $v_{\tau\vartheta} = v_\tau + v_\vartheta + b_\tau u_\vartheta$ follow immediately.

We know from (3.4) that $v_\tau = c(u_\tau)$. Thus the element v_τ of A_τ is a function of u_τ. But even more is true. The elements a_τ and b_τ are functions of u_τ and sign a_τ. In order to see this we show at first that

$$c(a_\tau \xi + u_\tau) = b_\tau \xi + c(\xi) + c(u_\tau) \tag{4.7}$$

holds true for all $\xi \varepsilon \Xi$ and all $\tau \varepsilon \Theta$.

This follows from the standard transformation used in the proof of (2.3). We obtain from (4.4) and (3.7)

$$
\begin{aligned}
1 &= \int_{\tau^{-1}M} \exp [\xi T(x) + c(\xi)] P_e(dx) \\
&= \int_M \exp [\xi T(\tau^{-1}x) + c(\xi)] P_\tau(dx) \\
&= \int_M \exp [(a_\tau \xi + u_\tau) T(x) + b_\tau \xi + c(\xi) + c(u_\tau)] P_e(dx).
\end{aligned}
$$

Hence $a_\tau \xi + u_\tau \,\varepsilon\, \Xi$, and (4.7) is thus a consequence of (3.4).

Differentiating (4.7) twice and setting $\xi = u_e = 0$ afterwards we obtain from the second derivative, by (3.2) and by (3.5), that

$$(4.8) \qquad a_\tau^2 = (-c''(u_\tau))^{-1} > 0$$

and, utilizing also the first derivative of (4.7),

$$(4.9) \qquad b_\tau = (-c''(u_\tau))^{-\frac{1}{2}} c'(u_\tau) \operatorname{sign} a_\tau .$$

By (3.4), (4.8), and (4.9) the matrices A_ϑ with $a_\vartheta > 0$ are one-to-one functions of u_ϑ. But then the class of all A_ϑ with $a_\vartheta > 0$ forms a group by $a_{\tau\vartheta} = a_\tau a_\vartheta$. Hence we establish:

Lemma (4.10). *The group of all matrices A_ϑ with $a_\vartheta > 0$ is commutative.*

Proof. (1) Suppose $a_\vartheta = 1$ for all ϑ with $a_\vartheta > 0$. Then by (4.6) $u_{\tau\vartheta} = u_\tau + u_\vartheta = u_{\vartheta\tau}$. Since A_τ is a one-to-one function of u_τ for $a_\tau > 0$, this implies $A_{\tau\vartheta} = A_{\vartheta\tau}$ and hence by (4.6) $A_\tau A_\vartheta = A_\vartheta A_\tau$.

(2) Suppose there is a τ with $0 < a_\tau < 1$ given. Then by (4.8) and (3.5) $u_\tau \neq 0$. Furthermore (4.6) implies that $u_{\tau^n} = [u_\tau/(1 - a_\tau)](1 - a_\tau^n)$. Hence

$$(*) \qquad \lim_{n\to\infty} u_{\tau^n} = u_\tau/(1 - a_\tau)$$

is finite, while $\lim_{n\to\infty} u_{\tau^{-n}}$ is not. This means that the convex natural parameter space has at most one finite boundary point.

On the other hand we obtain from (4.8) and (4.6) $\lim_{n\to\infty} c''(u_{\tau^n}) = -\lim_{n\to\infty} a_\tau^{-2n} = -\infty$. Since $c''(\xi)$ is a continuous function for all interior points of Ξ it follows that $\lim_{n\to\infty} u_{\tau^n} \not\varepsilon\, \Xi$. Therefore Ξ is bounded from one side by this limit and (3.6) entails finally $\lim_{n\to\infty} u_{\tau^n} = \sup \Xi = \xi_0 < +\infty$ and $0 = u_e \,\varepsilon\, \Xi$ yields

$$(**) \qquad 0 < \xi_0 = \sup_{\vartheta \varepsilon \Theta} u_\vartheta < +\infty .$$

We obtain from (*) at first for all ϑ with $0 < a_\vartheta < 1$ that

$$(***) \qquad u_\vartheta = \xi_0(1 - a_\vartheta).$$

For $a_\vartheta > 1$ (***) follows similarly by interchanging n with $-n$ above. For $a_\vartheta = 1$ we have $u_{\vartheta^n} = n u_\vartheta$. Since $\Xi \neq R^1$ this implies $u_\vartheta = 0$ and hence (***) holds.

Since (***) holds if $a_\vartheta > 0$, it follows that $u_{\tau\vartheta} = u_{\vartheta\tau}$ which in turn implies as under (1) that $A_\vartheta A_\tau = A_\tau A_\vartheta$, q.e.d.

Corollary (4.11). *If there exists a $\tau \,\varepsilon\, \Theta$ such that $|a_\tau| \neq 1$ then*

(a) $0 < \xi_0 = \sup_{\vartheta \varepsilon \Theta} u_\vartheta < +\infty$,
(b) $a_\vartheta = 1 - u_\vartheta/\xi_0$ *for all* $\vartheta \varepsilon \Theta$,
(c) *the group of matrices* A_ϑ, $\vartheta \varepsilon \Theta$, *is commutative.*

PROOF. By assumption there is a τ with $|a_\tau| \neq 1$. Then $|a_{\tau^2}| \neq 1$ and by (4.8) $a_{\tau^2} > 0$. Hence assertion (a) holds by (**) of Part (2) of the proof of (4.10). This yields $\sup_{n=0,\pm1,\ldots} u_{\tau^n} = \sup_{n=0,\pm1,\ldots} u_\tau(1 - a_\tau^n)/(1 - a_\tau) \leqq \xi_0 < +\infty$, which implies $a_\tau > 0$. Assertion (b) follows now from (***) and assertion (c) from the Lemma, q.e.d.

Finally, we show that there exists a constant k such that

$$(4.12) \qquad b_\tau = -ku_\tau \qquad \text{for } a_\tau > 0.$$

For, since $\mathcal{P}$ contains more than two different probability measures and $a_{\vartheta\tau} = a_\vartheta a_\tau$ there is a ϑ_1 with $u_{\vartheta_1} \neq 0$ and $a_{\vartheta_1} > 0$. By (4.6) and Lemma (4.10) it follows that $b_\tau = (v_{\tau\vartheta_1} - v_\tau - v_{\vartheta_1})/u_{\vartheta_1} = (v_{\vartheta_1\tau} - v_{\vartheta_1} - v_\tau)/u_{\vartheta_1} = (b_{\vartheta_1}/u_{\vartheta_1})u_\tau$. Setting $k = -b_{\vartheta_1}/u_{\vartheta_1}$ we obtain (4.12), q.e.d.

5. The normal distributions. In this section we consider the case that $u_{\tau\vartheta} = u_\tau + a_\tau u_\vartheta$ with $|a_\tau| = 1$, that is the case $\Xi = R^1$.

At first we consider the subgroup of all ϑ with $a_\vartheta = 1$. Then we obtain from (4.7) and (4.12) the functional equation

$$(5.1) \qquad c(u_\tau + u_\vartheta) = c(u_\tau) + c(u_\vartheta) - ku_\tau u_\vartheta .$$

This is equivalent to a functional equation solved by Sinzow (see Aczél [1] p. 63). Since $c(\xi)$ is continuous, the unique solution of (5.1) is given by

$$(5.2) \qquad c(u_\vartheta) = (-k/2)u_\vartheta^2 + lu_\vartheta .$$

(The reader can easily check that the proof given in [1] p. 63–64 does apply to any additive subgroup of R^1.)

To prove that (5.2) holds also for ϑ with $a_\vartheta = -1$, we show at first that

$$(5.3) \qquad b_\vartheta = ku_\vartheta - 2l \qquad \text{for } a_\vartheta = -1.$$

For, $a_\vartheta = a_\tau = -1$ yields $a_{\tau\vartheta} = a_{\vartheta\tau} = 1$. Hence we obtain from (4.7) and (5.2) $b_\vartheta u_\tau - b_\tau u_\vartheta = c(u_\vartheta - u_\tau) - c(u_\tau - u_\vartheta) = 2l(u_\vartheta - u_\tau)$, while (4.6) and (4.12) imply that $b_\vartheta - b_\tau = b_{\tau\vartheta} = -ku_{\tau\vartheta} = k(u_\vartheta - u_\tau)$. The solution of these two linear equations is (5.3). (Because $\mathcal{P}$ contains more than two probability measures, there exist τ, ϑ with $u_\vartheta - u_\tau \neq u_e = 0$.) q.e.d.

Since $a_\vartheta = -1$ implies $u_\vartheta = u_{\vartheta^{-1}}$, we obtain from (4.7), from (3.5) and from (5.3) $2c(u_\vartheta) = c(u_\vartheta) + c(u_{\vartheta^{-1}}) = c(0) - b_\vartheta u_\vartheta = -ku_\vartheta^2 + 2lu_\vartheta$, that is (5.2) holds also in the case $a_\vartheta = -1$.

We remark, that Convention (3.5) implies that $k = 1$ and $l = 0$ if the set of all u_ϑ, $\vartheta \varepsilon \Theta$, is dense in R^1. Otherwise there exist solutions of (4.7) with $k \neq 1$ and $l \neq 0$. An example is given at the end of this section. Clearly the constants k and l are uniquely determined by the family $\mathcal{P}$ itself. Since $c(\xi)$ is by (3.2) concave, it follows that $k > 0$.

To simplify the notation we define

(5.4) $$S(x) := k^{-\frac{1}{2}}(T(x) + l)$$

and

(5.5) $$w_\vartheta := k^{\frac{1}{2}} u_\vartheta .$$

Then from (3.7) and (5.2) we obtain that

(5.6) $$P_\vartheta(K) = \int_K \exp\left[-\tfrac{1}{2}(S(x) - w_\vartheta)^2 + \tfrac{1}{2}S^2(x)\right]P_e(dx).$$

The comparison with the normal distribution suggests the definition of a measure $\lambda \mid \mathcal{K}$ by

(5.7) $$\lambda(K) := \int_K \exp\left[\tfrac{1}{2}S^2(x)\right]P_e\, d(x).$$

We prove that λ is invariant, that is, that (2.5) holds. For this purpose we observe that (4.4), (4.12) and (5.3) for $a_\tau = 1$ and $a_\tau = -1$ resp. yield $T(\tau^{-1}x) + l = a_\tau T(x) + b_\tau + l = a_\tau(T(x) + l) - a_\tau k u_\tau$ a.e. Multiplication by $k^{-\frac{1}{2}}$ results in

(5.8) $$S(\tau^{-1}x) = a_\tau(S(x) - w_\tau) \quad \text{a.e.}$$

By the standard transformation used in the proof of (2.3) we obtain from (5.7), using (5.8) and (5.6) that

$$\begin{aligned}\lambda(\tau^{-1}K) &= \int_K \exp\left[\tfrac{1}{2}S^2(\tau^{-1}x)\right]P_\tau(dx)\\ &= \int_K \exp\left[\tfrac{1}{2}(S(x) - w_\tau)^2 - \tfrac{1}{2}(S(x) - w_\tau)^2 + \tfrac{1}{2}S^2(x)\right]P_e(dx)\\ &= \lambda(K) \quad \text{q.e.d.}\end{aligned}$$

Since the integrand in (5.7) is finite, λ is σ-finite. Therefore, by (5.6), $P_\vartheta \mid \mathcal{K}$ can be written as an integral with respect to an invariant measure:

(5.9) $$P_\vartheta(K) = \int_K \exp\left[-\tfrac{1}{2}(S(x) - w_\vartheta)^2\right]\lambda(dx).$$

This proves the first part of the Theorem.

Whether the invariant measures $\nu \mid \mathcal{K}$ are unique up to a multiplicative factor depends on the group Θ and the field $\mathcal{K}$. If $\mathcal{K}$ is the class of all subsets of the set of all integers and Θ the group of their translations $\nu \mid \mathcal{K}$ is proportional to the counting measure. In the case that $\mathcal{K}$ is equal to the Borel field $\mathcal{B}$ over R^1 and Θ a group which is dense in the group of all translations, $\nu \mid \mathcal{B}$ is unique and proportional to the Lebesgue measure. But if Θ is a discrete group of translations of R^1 which is not dense in itself and $\mathcal{K} = \mathcal{B}$, then $\nu \mid \mathcal{B}$ is not unique. For example, each integral over a non negative periodic function with respect to Lebesgue measure is invariant. In the last case $\lambda \mid \mathcal{K}$ is an invariant measure appropriate for our purposes. There are clearly certain connections between the multiplicity of solutions of the functional equation (4.7) and the existence of several invariant measures in case of groups for which $\{u_\vartheta : \vartheta \in \Theta\}$ is not dense in R^1.

Examples of the form (5.9) are the normal distribution and the distributions occurring in the discrete local limit theorems [6], possessing probabilities

(5.10) $$p(n) = C_{\alpha,\beta} \exp\left[-\tfrac{1}{2}(\beta n - \alpha)^2\right] \qquad \text{for } n = 0, \pm 1, \pm 2, \cdots .$$

To illustrate the necessity of the case $a_\tau = -1$ we give an example: Consider the densities

$$g(y, \alpha) = C \exp[-\tfrac{1}{2}(y - \alpha)^2] \quad \text{if } 2k \leqq y \leqq 2k + 1 \text{ for } k = 0, \pm 1, \cdots,$$
$$= 0 \quad \text{otherwise,}$$

with respect to the Lebesgue measure for $\alpha = 0, \pm 1, \cdots$. Obviously

$$g(y, \alpha) = g(y - \alpha, 0) \quad \text{for } \alpha = 0, \pm 2, \cdots$$
$$= g(-(y - \alpha), 0) \quad \text{for } \alpha = \pm 1, \pm 3, \cdots,$$

which shows that the family of densities $g(y, \alpha)$ can not be generated by translations only. Furthermore, we have in this example $E_e Y = lk^{-\frac{1}{2}} \neq 0$ and $\text{Var}_e\, Y = k^{-1} = 1 - l^2/k \neq 1$, where Y denotes a random variable with density $g(y, 0)$. For, we have $y = S(y) = (T(y) + l)/k^{\frac{1}{2}}$.

6. The gamma distributions. In this section we consider the case that there exists a τ with $|a_\tau| \neq 1$ and $a_\tau \neq 0$. According to Corollary (4.11.a) the natural parameter space Ξ is then not R^1.

By Corollary (4.11.b) we have $a_\vartheta > 0$ for all $\vartheta \in \Theta$ and

$$u_\vartheta = \xi_0(1 - a_\vartheta). \tag{6.1}$$

Since $a_{\tau\vartheta} = a_\tau a_\vartheta$ this formula suggests the study of the exponential family $\mathcal{P}$ generated by a group as function of a_ϑ instead of u_ϑ. At first we transform the functional equation (4.7) into one depending on a_ϑ. Inserting (6.1) and (4.12) into (4.7) we obtain $c(\xi_0(1 - a_\vartheta a_\tau)) = c(\xi_0(1 - a_\vartheta)) + c(\xi_0(1 - a_\tau)) - k\xi_0^2(1 - a_\vartheta)(1 - a_\tau)$. Setting, as [1] p. 63–64 suggests,

$$d(\eta) := c(\xi_0(1 - \eta)) - k\xi_0^2(1 - \eta), \tag{6.2}$$

one of Cauchy's functional equations [1] p. 47–48 results:

$$d(a_\vartheta a_\tau) = d(a_\vartheta) + d(a_\tau). \tag{6.3}$$

Since $c(\xi)$ and hence $d(\eta)$ is continuous and the set of all a_ϑ, $\vartheta \in \Theta$, is a multiplicative group of positive real numbers, there exists a constant p such that

$$d(a_\vartheta) = p \log a_\vartheta. \tag{6.4}$$

But from (6.2) and (3.2) we find $d''(\eta) = \xi_0^2 c''(\xi_0(1 - \eta)) < 0$, so that $d(\eta)$ is concave and hence $p > 0$.

If we now define

$$S(x) := \xi_0 T(x) + k\xi_0^2 \tag{6.5}$$

and insert this expression into (3.7) and (4.4) resp. we obtain

$$P_\vartheta(K) = \int_K a_\vartheta^p \exp[(1 - a_\vartheta)S(x)]P_e(dx) \tag{6.6}$$

and

$$S(\tau^{-1}x) = a_\tau S(x). \tag{6.7}$$

We shall now show that

$$S(x) > 0 \quad \text{a.e.} \tag{6.8}$$

holds. This is seen by the following argument. According to the assumption made at the beginning of this section there exists a ϑ with $a_\vartheta > 1$ by Corollary (4.11). Since the set $\{x : S(x) \leqq 0\}$ is invariant we obtain from (6.6)

$$a_\tau^{-p} P_e\{S(x) \leqq 0\} = a_\tau^{-p} P_\tau\{S(x) \leqq 0\} = P_e\{S(x) = 0\} + \int_{S(x)<0} \exp[(1 - a_\tau)S(x)] P_e(dx).$$

Setting $\tau = \vartheta^n$ and letting $n \to +\infty$ it follows by the Monotone Limit Theorem that $0 = P_e\{S(x) = 0\} + \infty P_e\{S(x) < 0\}$ and hence $P_e\{S(x) \leqq 0\} = 0$, q.e.d.

Knowing that $S(x) > 0$ a.e. we can define a σ-finite measure $\lambda \mid \mathcal{K}$ by

$$\lambda(K) = \int_K S(x)^{-p} \exp S(x) P_e(dx). \tag{6.9}$$

The invariance property now follows readily from (6.6) and (6.7). We conclude that $P_\vartheta \mid \mathcal{K}$ is in this case also expressible as an integral with respect to an invariant measure. This proves the second part of the Theorem in Section 2:

$$P_\vartheta(K) = \int_K (a_\vartheta S(x))^p \exp[-a_\vartheta S(x)] \lambda(dx). \tag{6.10}$$

If the group $\{a_\vartheta : \vartheta \in \Theta\}$ is dense in the multiplicative group of positive real numbers and $\mathcal{K}$ the Borel field of R^1, then λ is the Haar measure of the later group, that is $\lambda(B) = \kappa \int_B y^{-1}\, dy$. Hence the gamma distributions

$$P_\sigma(B) = [1/\Gamma(p)] \int_B (y/\sigma)^p \exp(-y/\sigma) y^{-1}\, dy$$

are examples.

Acknowledgment. We wish to thank Mr. K. Daniel for several improvements of the English and the referee for the suggestion to formulate the Corollary of the Theorem.

REFERENCES

[1] Aczél, J. (1961) *Vorlesungen über Funktionalgleichungen und ihre Anwendungen.* Birkhäuser Verlag, Basel.

[2] Borges, R. (1962) Subjektivtrennscharfe Konfidenzbereiche, *Z. Wahrscheinlichkeitstheorie* **1** 47–69.

[3] Borges, R. and Pfanzagl, J. (1963) A characterization of the one-parameter exponential family of distributions by monotonicity of likelihood ratios. *Z. Wahrscheinlichkeitstheorie* **2** 111–117.

[4] Dynkin, E. B. (1951) Necessary and sufficient statistics for a family of probability distributions. *Uspehi Mat. Nauk.* (N.S.) **6** No. 1 (41) 68–90 (Also, *Select. Transl. Math. Statist. and Prob.* **1** 17–40).

[5] Ferguson, T. S. (1962 and 1963) Location and scale parameters in exponential families of distributions. *Ann. Math. Statist.* **33** 986–1001 and **34** 1603.

[6] Gnedenko, B. W. (1962) *Lehrbuch der Wahrscheinlichkeitsrechnung.* 3rd ed. Akademieverlag, Berlin.

[7] LEHMANN, E. L. (1959) *Testing Statistical Hypotheses*. Wiley, New York.

[8] LINDLEY, D. V. (1958) Fiducial distributions and Bayes' Theorem, *J. Roy. Statist. Soc. Ser. B* **20** 102–107.

[9] NEYMAN, J. (1938) L'estimation statistique traitée comme un problème classique de probabilité. *Actualités Sci. Ind.* **739** 26–57.

[10] RICHTER, H. (1954) Zur Grundlegung der Wahrscheinlichkeitstheorie, Teil V. Indirekte Wahrscheinlichkeitstheorie, *Math. Ann.* **128** 305–339.

[11] SCHMETTERER, L. (1956) *Einführung in die mathematische Statistik*. Springer-Verlag, Wien, first ed.