LOCATION AND SCALE PARAMETERS IN EXPONENTIAL FAMILIES OF DISTRIBUTIONS

By Thomas S. Ferguson

University of California, Los Angeles

1. Introduction and summary. Location and scale parameters, on the one hand, and distributions admitting sufficient statistics for the parameters, on the other, have played a large role in the development of modern statistics. This paper deals with the problem of finding those distributions involved in the intersection of these two domains.

In Sections 2 through 4 the preliminary definitions and lemmas are given. The main results found in Theorems 1 through 4 may be considered as a strengthening of the results of Dynkin [3] and Lindley [8]. Theorem 1 discovers the only possible forms assumed by the density of an exponential family of distributions having a location parameter. These forms were discovered by Dynkin under the superfluous assumptions that a density with respect to Lebesgue measure exist and have piecewise continuous derivatives of order one. Theorem 2 consists of the specialization of Theorem 1 to one-parameter exponential families of distributions. The resulting distributions, as found by Lindley, are either (1), the distributions of $(1/\gamma) \log X$, where X has a gamma distribution and $\gamma \neq 0$, or (2), corresponding to the case $\gamma = 0$, normal distributions. In Theorem 3, the result analogous to Theorem 2 for scale parameters is stated. In Theorem 4, those k-parameter exponential families of distributions which contain both location and scale parameters are found. If the parameters of a two-parameter exponential family of distributions may be taken to be location and scale parameters, then the distributions must be normal.

The final section contains a discussion of the family of distributions obtained from the distributions of Theorem 2 and their limits as $\gamma \to \pm\infty$. These limits are "non-regular" location parameter distributions admitting a complete sufficient statistic. This family of distributions is a main class of distributions to which Basu's theorem (on statistics independent of a complete sufficient statistic) applies. Furthermore, this family is seen to provide a natural setting in which to prove certain characterization theorems which have been proved separately for the normal and gamma distributions. Concluding the section is a theorem which, essentially, characterizes the gamma distribution by the maximum likelihood estimate of its scale parameter.

2. Definitions. Throughout this paper we shall be dealing with distributions on the real line only. The real line will be denoted by R, the Borel σ-field over R by $\mathcal{B}$, and a family of distributions over $(R, \mathcal{B})$, indexed by a parameter θ in a parameter space Θ, by P_θ, $\theta \varepsilon \Theta$. The corresponding cumulative distribution functions will be denoted by $F(x \mid \theta)$.

Received September 28, 1961; revised February 9, 1962.

DEFINITION 1. A real parameter, θ, is said to be a location parameter of a family of distributions over $(R, \mathcal{B})$ if $F(x \mid \theta)$ is a function only of $x - \theta$.

DEFINITION 2. A positive real parameter, θ, is said to be a scale parameter of a family of distributions over $(R, \mathcal{B})$ if $F(x \mid \theta)$ is a function only of $x\theta^{-1}$.

If a density, $f(x \mid \theta)$ with respect to Lebesgue measure exists, then θ is a location [scale] parameter if, and only if, $f(x \mid \theta) = g(x - \theta)[f(x \mid \theta) = \theta^{-1}g(x\theta^{-1})]$ almost everywhere for some function g. In terms of random variables, θ is a location [scale] parameter of the distribution of a random variable X if, and only if, the distribution of $X - \theta$ $[X\theta^{-1}]$ is independent of θ.

REMARK. If θ is a location parameter of the distribution of a random variable X, then e^{θ} is a scale parameter of the distribution of e^{X}. In fact, every scale parameter family of distributions can be constructed as follows. Choose random variables X and Y whose distributions both have θ as a location parameter, and choose three non-negative numbers p_1, p_2, p_3, whose sum is one. Then, e^{θ} is a scale parameter of the distribution of the random variable Z, defined to be equal to e^{X} with probability p_1, $-e^{Y}$ with probability p_2 and zero with probability p_3. Conversely, if θ is a scale parameter of the distribution of a random variable Z, then $\log \theta$ is a location parameter of the conditional distribution of $\log Z$ given $Z > 0$, and of the conditional distribution of $\log(-Z)$ given $Z < 0$. This remark will be useful in passing from problems involving location parameters to those involving scale parameters.

For simplicity, exponential families of distributions will be defined over $(R, \mathcal{B})$, although any measurable space would do as well. The notation $\nu \ll \mu$, where ν and μ are measured over $(R, \mathcal{B})$ is used to denote the statement "ν is absolutely continuous with respect to μ."

DEFINITION 3. A family P_θ, $\theta \in \Theta$, of probability measures over $(R, \mathcal{B})$ is said to constitute a k-parameter exponential family of distributions in θ, if

(1) there is a σ-finite measure, μ, on $(R, \mathcal{B})$ for which $P_\theta \ll \mu$ for all $\theta \in \Theta$,

(2) the densities, $p(x \mid \theta)$, of P_θ with respect to μ may be chosen of the form

$$p(x \mid \theta) = \exp\left\{T_0(x) + Q_0(\theta) + \sum_{j=1}^{k} T_j(x)Q_j(\theta)\right\} \tag{1}$$

where the $T_j(x)$ for $j = 0, 1, \cdots, k$ are measurable functions and where $T_0(x)$ may assume the value $-\infty$,

(3) the functions $\{1, T_1(x), \cdots, T_k(x)\}$ are almost surely $[\mu_0]$ linearly independent, where $d\mu_0 = e^{T_0(x)}\, d\mu$, and

(4) the functions $\{1, Q_1(\theta), \cdots, Q_k(\theta)\}$ are linearly independent.

Note that θ does not have to be a k-dimensional parameter. In this paper we shall deal only with parameter spaces, Θ, which are subsets of Euclidean spaces of one or two dimensions. In part (2) of this definition, $T_0(x)$ is allowed the value $-\infty$ in order that the density $p(x \mid \theta)$ may be zero on a set in R which is independent of θ. Parts (3) and (4) of this definition are required in order to insure that the value of k cannot be made smaller by a simple change of functions. As an example of the flavor of definition 3, the normal distributions with mean

θ and variance σ^2 would be considered as a one-parameter exponential family in θ (or in σ^2), as a two-parameter exponential family in (θ, σ^2), and as a two-parameter exponential family in α, if $\alpha = \theta = \sigma$.

We collect here a few facts about exponential families of distributions to be used later. For proofs the reader may refer to the excellent book of Lehmann [7].

a. If $T_0(x), T_1(x), \cdots, T_k(x)$ are the functions of x in equation (1), the set of points $\pi = (\pi_1, \cdots, \pi_k)$, for which

$$e^{-Q_0(\pi)} = \int \exp\left\{T_0(x) + \sum_{j=1}^{k} \pi_j T_j(x)\right\} d\mu(x) \tag{2}$$

is finite and positive, is called the natural parameter space associated with the exponential family.The natural parameter space, Π, is convex and, thanks to parts (3) and (4) of definition 3, contains an open set in k-dimensions.

b. The functions, $Q_0(\pi)$, and

$$E(\phi \mid \pi) = \int \phi(x) \exp\{T_0(x) + Q_0(\pi) + \sum \pi_j T_j(x)\} d\mu \tag{3}$$

are analytic functions of $\pi_1, \cdots, \pi_k$, at all interior points of Π, for all bounded measurable functions, ϕ.

3. Dominated location parameter families. In this section we shall prove a few lemmas concerning location parameter families of distributions which are dominated by a single σ-finite measure. In Lemma 2, it will be seen that the dominating measure may as well be taken to be Lebesgue measure.

Let us denote Lebesgue measure by l, and let P and μ denote arbitrary measures on $(R, \mathcal{B})$. Let θ denote a real number, and let A denote an arbitrary set of $\mathcal{B}$.

LEMMA 1. *If μ is σ-finite, and if $l(A) = 0$, then $\mu(A + \theta) = 0$ for almost all θ (with respect to l).*

PROOF. We may write $\mu = \sum_{i=1}^{\infty} \mu_i$, where the μ_i are finite measures. Let $\nu_i = \mu_i * l$, where $*$ denotes convolution. Then $\nu_i(A) = 0$ since $\nu_i \ll l$. But since $\nu_i(A) = \int \mu_i(A - \theta)\, d\theta = 0$, we have that $\mu_i(A - \theta) = 0$ for almost all θ, and all i. Hence $\mu(A - \theta) = 0$ for almost all θ.

LEMMA 2. *If μ is σ-finite, if $P_\theta(A)$ is defined to be $P(A - \theta)$ for all θ, and if $P_\theta \ll \mu$ for all θ, then $P \ll l$.*

PROOF. Suppose not; then there exists a set A such that $l(A) = 0$, and $P(A) > 0$. Define $A_\theta = A + \theta$. Then $P_\theta(A_\theta) > 0$ for all θ. Since $P_\theta \ll \mu$, this implies that $\mu(A_\theta) > 0$ for all θ. Thus $\mu(A + \theta) > 0$ for all θ and $l(A) = 0$, contradicting Lemma 1.

The implication of Lemma 2 is obvious. If P_θ is an exponential family of distributions in a location parameter, θ, then, since we have assumed that our exponential families are dominated by a σ-finite measure, μ, we may as well assume that μ is Lebesgue measure. Henceforth, in dealing with location parameters in exponential families, we shall take μ to be Lebesgue measure. In dealing with scale parameters in exponential families, the remark of Section 2 shows that the distributions may have a mass at the origin independent of θ. However, except for this mass at the origin, these distributions are absolutely continuous with respect to Lebesgue measure also.

Location parameter families of distributions are always stochastically continuous in the parameter. In the following lemma, we obtain a much stronger type of continuity in a location parameter when the family is dominated by a σ-finite measure.

LEMMA 3. *If $P_\theta(A)$ is defined to be $P(A - \theta)$, and if $P \ll l$, then for every measurable set A, $P_\theta(A)$ is a continuous function of θ.*

PROOF. If $\theta \to \theta_0$, then

$$|P_\theta(A) - P_{\theta_0}(A)| = |P(A - \theta) - P(A - \theta_0)|$$
$$\leqq P((A - \theta)\Delta(A - \theta_0)) \ll l((A - \theta)\Delta(A - \theta_0)) \to 0.$$

(See Halmos [5] pg. 266).

4. Location parameters in exponential families. We shall now combine the hypotheses about a family, P_θ, $\theta \varepsilon \Theta$, of distributions, that θ be a location parameter, $\Theta = R$, and that P_θ be a k-parameter exponential family in θ. The resulting distributions are described in Theorem 1, which has been proved by Dynkin [3] under certain regularity conditions. The following lemma shows that these regularity conditions are always satisfied.

LEMMA 4. *If P_θ is a k-parameter exponential family in a location parameter, θ, then a density with respect to Lebesgue measure exists and may be written in the form of formula (1). In addition, all derivatives of the functions $Q_i(\theta)$, $i = 0, 1, \cdots, k$, with respect to θ exist, and the functions $T_i(x)$, $i = 0, 1, \cdots, k$, may be chosen so that all derivatives with respect to x exist.*

PROOF. That the densities $p(x \mid \theta)$ of formula (1) may be taken with respect to Lebesgue measure follows from Lemma 2. Since, θ is a location parameter we may write

$$F(x - \theta) = \int_{-\infty}^{x} \exp\left\{T_0(y) + Q_0(\theta) + \sum_{j=1}^{k} T_j(y)Q_j(\theta)\right\} dy, \tag{4}$$

for all x and all θ, for some function F. Since the proof is rather long, we shall break it into several pieces.

1°. *Continuity of $Q_j(\theta)$.* Suppose that $\theta_n \to \theta_0$. We must show that

$$Q_j(\theta_n) \to Q_j(\theta_0) \quad \text{for} \quad j = 0, 1, \cdots, k.$$

In order to simplify the notation, we shall assume, as we may by changing the functions in the exponent of equation (4), that $Q_j(\theta_0) = 0$, for $j = 0, 1, \cdots, k$. With such an assumption, $\mu_0(A) = \int_A \exp\{T_0(y)\}\, dy$ becomes a probability measure. Since $\{1, T_1(x), \cdots, T_k(x)\}$ are almost certainly $[\mu_0]$ linearly independent, there exist $k + 1$ points $(t_1^{(i)}, t_2^{(i)}, \cdots, t_k^{(i)})$ $i = 1, \cdots, k + 1$, in R^k, not all lying in one $(k - 1)$-dimensional hyperplane, such that for all $\epsilon > 0$ and all i, the sets $A_i = \{x: \sum_{j=1}^{k} (T_j(x) - t_j^{(i)})^2 \leqq \epsilon\}$ have positive μ_0 measure. But, using the mean value theorem for integrals,

$$(5)\qquad \begin{aligned} P_{\theta_n}(A_i) - P_{\theta_0}(A_i) &= \int_{A_i} [\exp \{Q_0(\theta_n) + \textstyle\sum T_j(y)Q_j(\theta_n)\} - 1]\, d\mu_0(y) \\ &= [\exp \{Q_0(\theta_n) + \textstyle\sum \hat{t}_j^{(i)}Q_j(\theta_n)\} - 1]\, \mu_0(A_i) \end{aligned}$$

where $\sum_{j=1}^{k} (\hat{t}_j^{(i)} - t_j^{(i)})^2 \leqq \epsilon$ ($\hat{t}_j^{(i)}$ may depend on n). This converges to zero by Lemma 3, yielding the fact that as $n \to \infty$

$$(6)\qquad Q_0(\theta_n) + \sum_{j=1}^{k} t_j^{(i)} Q_j(\theta_n) \to 0, \qquad \text{for } i = 1, \cdots, k+1$$

When ϵ is sufficiently small, equation (6) will certainly be invertible for the $Q_j(\theta_n)$, showing that $Q_j(\theta_n) \to 0$ as $n \to \infty$, for $j = 0, 1, \cdots, k$, thus proving continuity.

2°. *Differentiability of* $Q_j(\theta)$. The derivative $F'(x) = f(x)$ exists almost everywhere, let us say except for $x \in N$ where $l(N) = 0$. We shall express the distribution function as a function of the natural parameters $F(x - \theta) = G(x \mid Q(\theta))$, where $Q(\theta)$ represents the vector $(Q_1(\theta), \cdots, Q_k(\theta))$. The function G, as we mentioned in the remarks after definition 3, is an analytic function in

$$(Q_1, \cdots, Q_k).$$

Suppose that $\theta_n \to \theta_0$. We may write, using the mean value theorem for derivatives,

$$(7)\qquad \frac{F(x - \theta_n) - F(x - \theta_0)}{\theta_n - \theta_0} = \sum_{j=1}^{k} \frac{\partial}{\partial Q_j} G(x \mid Q_n^*) \frac{(Q_j(\theta_n) - Q_j(Q_0))}{\theta_n - \theta_0}$$

where $Q_n^* = \alpha Q(\theta_n) + (1 - \alpha)Q(\theta_0)$ for some $0 \leqq \alpha \leqq 1$. As $n \to \infty$, the continuity of $Q(\theta)$ may be applied to prove that for all j,

$$(8)\qquad (\partial/\partial Q_j)G(x \mid Q_n^*) \to (\partial/\partial Q_j)G(x \mid Q(\theta_0))$$

for all x. The left side of equation (7) will converge (to $-f(x - \theta_0)$) for all $x \not\in N - \theta_0$. There exist $x_1, \cdots, x_k, x_j \not\in N - \theta_0$ for all j, for which the matrix whose (i, j)th element is

$$(9)\qquad (\partial/\partial Q_j)G(x_i \mid Q(\theta_0))$$

has non-vanishing determinant. (This may be verified by induction. If the determinant were zero for almost all $x_1, \cdots, x_k$, then, expanding the determinant in its first row, we would obtain for fixed $x_2, \cdots, x_k$,

$$(10)\qquad c_1(\partial/\partial Q_1)G(x \mid Q(\theta_0)) + \cdots + c_k(\partial/\partial Q_k)G(x \mid Q(\theta_0)) = 0$$

for almost all, hence all, x. By induction, $x_2, \cdots, x_k$, may be chosen so that $c_1 \neq 0$. But equation (10) is equivalent to

$$(11)\qquad e^{Q_0(Q)} \int_{-\infty}^{x} \left\{\sum_{j=1}^{k} c_j \left(T_j(y) + \frac{\partial Q_0}{\partial Q_j}\right)\right\} \exp \{\textstyle\sum T_j(y)Q_j\}\, d\mu_0 = 0$$

for all x. This would contradict the assumption that $\{1, T_1(y), \cdots, T_k(y)\}$ were almost surely $[\mu_0]$ linearly independent.) Thus, when n becomes sufficiently large, the k equations, obtained from equation (7) by replacing x alternately by $x_1, x_2, \cdots, x_k$, will be invertible for the ratios $(Q_j(\theta_n) - Q_j(\theta_0))/(\theta_n - \theta_0)$ which will then converge proving differentiability of $Q_j(\theta)$ for $j = 1, \cdots, k$. Differentiability of $Q_0(\theta)$ will follow from this, since Q_0 is analytic in $Q_1, \cdots, Q_k$.

3°. *Essential continuity of* $T_j(x)$. The derivative with respect to θ of the right side of equation (4) exists everywhere thanks to the differentiability of the $Q_j(\theta)$, and is a continuous function of x. Since this derivative is equal to $-f(x - \theta)$, we find that this density is a continuous function. Letting $h(x) = \log f(x)$, we find from equation (4) that

$$h(x - \theta) = T_0(x) + Q_0(\theta) + \textstyle\sum_{j=1}^{k} T_j(x)Q_j(\theta) \tag{12}$$

for all θ, and almost all x, say for $x \not\in N_\theta$ where $l(N_\theta) = 0$. By taking $\theta_1, \theta_2, \cdots, \theta_{k+1}$ so that the matrix whose (i, j)th element is a_{ij} where $a_{ij} = Q_j(\theta_i)$ for $1 \leqq j \leqq k$ and $i = 1, \cdots, k + 1$, and where $a_{ij} = 1$ for $j = k + 1$ and $i = 1, \cdots, k + 1$, has a non-vanishing determinant, which may be accomplished since $\{1, Q_1(\theta), \cdots, Q_k(\theta)\}$ are linearly independent, we may solve for the $T_j(x)$

$$T_j(x) = \textstyle\sum_{j=1}^{k+1} a^{ij}(h(x - \theta_i) - Q_0(\theta_i)), \tag{13}$$

for $j = 1, 2, \cdots, k + 1$, (where $T_{k+1}(x)$ is a notation for $T_0(x)$) where the a^{ij} are the elements of the inverse matrix of the a_{ij}. Thus the $T_j(x)$ are equal almost everywhere to continuous functions. We may therefore choose the $T_j(x)$ to be continuous. Equation (12) is now valid for all θ and all x.

4°. *Infinite differentiability*. The function $h(x - \theta)$ in equation (12) is differentiable with respect to θ, and therefore with respect to x. From equation (13), each $T_j(x)$ is therefore differentiable. Hence, h is twice differentiable. Since we can solve equation (12) for the $Q_j(\theta)$ as we solved it for the $T_j(x)$ in equation (13), we may continue inductively to show that h and each Q_j and T_j have infinitely many derivatives. This completes the proof of the lemma.

We shall now state the following theorem.

THEOREM 1. *An exponential family of distributions in a location parameter has a density with respect to Lebesgue measure of the form*

$$f(x) = \exp\left\{\sum_{i=1}^{m} e^{\alpha_i x} p_i(x)\right\}, \tag{14}$$

where $\alpha_1, \alpha_2, \cdots, \alpha_m$ *are complex numbers and* $p_1(x), \cdots, p_m(x)$ *are polynomials in* x *with complex coefficients.*

Naturally, if $f(x)$ is to be a probability density, the complex constants in (14) must be chosen so that the function $f(x)$ is real, positive, and satisfies the condition $\int_{-\infty}^{\infty} f(x)\, dx = 1$. Under these conditions, formula (14) is, conversely, the density of an exponential family of distributions in a location parameter.

The theorem above was proved by Dynkin [3], under the assumption that the density with respect to Lebesgue measure exist and have a piecewise continuous derivative of order one. Lemma 4 implies that a density with respect to Lebesgue measure exists and has continuous derivatives of all orders. Therefore, we refer the reader to Dynkin's paper for the completion of the proof of Theorem 1.

5. A one-parameter generalization of the normal distributions. The main result of this paper is the specialization of Theorem 1 to one-parameter exponential families of distributions. This result, found in Theorem 2 below, strengthens the result of Lindley [8], proved under the assumption of the existence of two derivatives of the density. We prefer, however, to state this result in a more compact form which will emphasize the fact that the distributions involved are stochastically continuous in the parameters. For this reason we shall develop some notation to describe the resulting family of distributions.

If a random variable, X, has a gamma distribution with density

$$f_X(x) = \begin{cases} [\Gamma(\alpha)\beta^{\alpha}]^{-1}e^{-(x/\beta)}x^{\alpha-1} & \text{if } x > 0 \\ 0 & \text{if } x \leqq 0 \end{cases} \tag{15}$$

where $\alpha > 0$ and $\beta > 0$, the density of the random variable $Y = (1/\gamma) \log X$, for $\gamma \neq 0$, is

$$f_Y(y) = [|\gamma|\alpha^{\alpha}/\Gamma(\alpha)] \exp\{-\alpha e^{\gamma(y-\theta)} + \alpha\gamma(y - \theta)\} \tag{16}$$

where θ is a location parameter related to β by the formula $\gamma\theta = \log \alpha\beta$. The mean and variance of Y are $\theta + (1/\gamma)[\psi(\alpha) - \log \alpha]$ and $(1/\gamma^2)\psi'(\alpha)$, respectively, where $\psi(\alpha)$ and $\psi'(\alpha)$ are the digamma and trigamma functions (the first and second derivatives of $\log \Gamma(\alpha)$). We shall use the notation $N(\theta, \sigma^2, \gamma)$ to denote the above distribution of Y, where σ^2 is the variance of Y and is related to α by the formula,

$$\sigma^2\gamma^2 = \psi'(\alpha). \tag{17}$$

The symbol $N(\theta, \sigma^2, \gamma)$ is defined for all θ, $\sigma^2 > 0$, and $\gamma \neq 0$. We will now show that $N(\theta, \sigma^2, \gamma)$ converges as $\gamma \to 0$ to a normal distribution with mean θ and variance σ^2.

With σ^2 fixed and $\gamma \to 0$, formula (17) implies that $\alpha \to \infty$; but $\alpha\psi'(\alpha) \to 1$ as $\alpha \to \infty$, so that $\alpha\sigma^2\gamma^2 \to 1$. From Stirling's formula, as $\alpha \to \infty$

$$\Gamma(\alpha)e^{\alpha}\alpha^{-\alpha}\alpha^{\frac{1}{2}} \to (2\pi)^{\frac{1}{2}},$$

so that

$$\begin{aligned} f_Y(y) &\sim [(|\gamma|\alpha^{\frac{1}{2}})/(2\pi)^{\frac{1}{2}}] \exp\{-\alpha[e^{\gamma(y-\theta)} - \gamma(y - \theta) - 1]\} \\ &\sim [(2\pi)^{\frac{1}{2}}\sigma]^{-1} \exp\{-(\sigma^2\gamma^2)^{-1}[e^{\gamma(y-\theta)} - \gamma(y - \theta) - 1]\} \\ &\to [(2\pi)^{\frac{1}{2}}\sigma]^{-1} \exp\{-(2\sigma^2)^{-1}(y - \theta)^2\}. \end{aligned} \tag{18}$$

Thus, we may define $N(\theta, \sigma^2, \gamma)$ at $\gamma = 0$ as the normal distribution with mean θ and variance σ^2, and the distributions will be stochastically continuous in the parameters.

Now we shall extend the definition of $N(\theta, \sigma^2, \gamma)$ to the case where γ is allowed to assume the values $+\infty$ and $-\infty$. As $\gamma \to \pm\infty$, formula (17) implies that $\alpha \to 0$; but $\alpha^2\psi'(\alpha) \to 1$ as $\alpha \to 0$, so that $\alpha^2\sigma^2\gamma^2 \to 1$. Since, as $\alpha \to 0$, both $\alpha\Gamma(\alpha) \to 1$ and $\alpha^\alpha \to 1$, we have, as $\gamma \to -\infty$,

$$
\begin{aligned}
f_Y(y) &\sim |\gamma|\alpha \exp\{-\alpha e^{\gamma(y-\theta)} + \alpha\gamma(y-\theta)\} \\
&\sim (1/\sigma) \exp\{-(\sigma|\gamma|)^{-1} e^{\gamma(y-\theta)} + (|\gamma|/\sigma\gamma)(y-\theta)\} \\
&\to \begin{cases} (1/\sigma) \exp\{-(1/\sigma)(y-\theta)\} & \text{if } y > \theta \\ 0 & \text{if } y \leqq \theta. \end{cases}
\end{aligned}
\tag{19}
$$

Similarly, as $\gamma \to +\infty$,

$$
f_Y(y) \to \begin{cases} (1/\sigma) \exp\{(1/\sigma)(y-\theta)\} & \text{if } y < \theta \\ 0 & \text{if } y \geqq \theta. \end{cases}
\tag{20}
$$

We now define $N(\theta, \sigma^2, -\infty)$ and $N(\theta, \sigma^2, +\infty)$ to be the limiting distributions of formulas (19) and (20) respectively.

A few words about this family of distributions are in order. The symbol $N(\theta, \sigma^2, \gamma)$ is now defined for all finite θ, all positive finite σ^2, and all γ on the extended real line. This family of distributions is stochastically continuous in these three parameters, and is, furthermore, a one-parameter exponential family in θ, a one-parameter exponential family in σ^2, and a two-parameter exponential family in (θ, σ^2). The parameter σ^2 represents the variance, and θ is a location parameter. Given a sample of size n, $X_1, X_2, \cdots, X_n$, from the distribution $N(\theta, \sigma^2, \gamma)$, a sufficient statistic for θ may be found in the mean of order γ of $X_1, \cdots, X_n$, to be denoted by $M_\gamma(X_1, \cdots, X_n)$. More precisely, for fixed σ^2 and γ the statistic

$$
M_\gamma(X_1, \cdots, X_n) = \begin{cases} (1/\gamma) \log\left\{(1/n) \sum_{i=1}^{n} e^{\gamma X_i}\right\} & \text{if } \gamma \neq 0 \text{ or } \pm\infty \\ \bar{X} & \text{if } \gamma = 0 \\ \max_i X_i & \text{if } \gamma = +\infty \\ \min_i X_i & \text{if } \gamma = -\infty \end{cases}
\tag{21}
$$

is sufficient for θ, and is, in fact, the maximum likelihood estimate of θ. For fixed $X_1, \cdots, X_n$, $M_\gamma(X_1, \cdots, X_n)$ is a continuous strictly increasing function of γ between $-\infty$ and $+\infty$ inclusive, except in the trivial case where all the X_i are equal, when $M_\gamma(X_1, \cdots, X_n)$ is a constant.

Several other notations are possible to describe this family. One notation worth considering is to let $N^*(\theta, \sigma^2, \gamma)$ be defined as $N(\theta, \sigma^2, \gamma/\sigma)$. For the distributions $N^*(\theta, \sigma^2, \gamma)$, which are still continuous in the parameters, the coefficient of skewness is a function of γ alone, and takes values from -2 to $+2$. Furthermore, (θ, σ^2) are location and scale parameters for all values of γ, and σ^2 is still the variance of the distribution. However, with N^* one loses the property of being a two-parameter exponential family in (θ, σ^2). Another advantage of $N(\theta, \sigma^2, \gamma)$, exploited in Theorem 5 below, is that the maximum likelihood estimate (21) of θ is independent of σ^2.

THEOREM 2. *The parameter of a one-parameter exponential family of distributions can be taken to be a location parameter if, and only if, the family of distributions is* $N(\theta, \sigma^2, \gamma)$ *for some fixed positive* σ^2 *and finite* γ.

Lemma 4 above implies that the density with respect to Lebesgue measure is infinitely differentiable. The reader is therefore referred to Lindley's paper for the completion of the proof. Of course, this theorem is also an immediate consequence of Theorem 1.

The author conjectures that Theorem 2 can be strengthened and extended to allow infinite values of γ; namely, that the distributions $N(\theta, \sigma^2, \gamma)$ are the only distributions, dominated by a σ-finite measure μ, independent of the location parameter, θ, which have a complete sufficient statistic for θ for any sample size $\geqq 2$. One can easily prove a theorem parallel to Theorem 2 to the effect that the only "non-regular" distributions of Pitman [10] with a one-dimensional sufficient statistic for a location parameter, θ, are the distributions $N(\theta, \sigma^2, \pm\infty)$. In fact, Theorem 8 of Dynkin's paper [3] contains such a theorem. Whether or not there exist distributions other than $N(\theta, \sigma^2, \gamma)$ having a complete sufficient for a location parameter θ for a sample size $\geqq 2$ in the dominated case, is an open problem. In the non-dominated case, the author has been able to find only one other such example, that being the geometric distribution with a probability mass function on the points $\theta, \theta + 1, \theta + 2, \cdots$

$$P(X = x \mid \theta) = (1 - p)p^{x-\theta}, \qquad x = \theta, \theta + 1, \cdots$$

For a sample of size n from this distribution, min X_i is a complete sufficient statistic for the location parameter θ.

6. Scale parameters in one-parameter exponential families. To describe the result for scale parameters corresponding to Theorem 2, we introduce some more notation. We shall use $L(\theta, \sigma^2, \gamma)$ to denote the distribution of the random variable $X = e^Y$ where Y has the distribution $N(\log \theta, \sigma^2, \gamma)$. The parameter θ is a scale parameter of the distribution $L(\theta, \sigma^2, \gamma)$. This class of distributions contains certain well known distributions. Of course, $L(\theta, \sigma^2, 1)$ are just the gamma distributions, and $L(\theta, \sigma^2, 0)$ are the lognormal distributions. In addition,

$$L(\theta, \psi'(\tfrac{1}{2}), -1)$$

are completely asymmetric stable distributions with characteristic exponent $\frac{1}{2}$, $L(\theta, (1/\gamma^2)\psi'(1/\gamma), \gamma)$ are the Weibull distributions when $\gamma > 0$, $L(\theta, 1, +\infty)$ is the uniform distribution over the interval from zero to θ, and $L(\theta, \frac{1}{4}\psi'(\frac{1}{2}), 2)$ are half-normal distributions.

If X has a distribution $L(\theta, \sigma^2, \gamma)$, the distribution of $-X$ will be denoted by $-L(\theta, \sigma^2, \gamma)$. The distribution degenerate at zero will be denoted by $D(0)$. Finally, we shall denote mixtures of distributions by the sums of the corresponding symbols. For example $pL(\theta, \sigma^2, \gamma) + (1 - p)D(0)$ represents the mixture of a distribution $L(\theta, \sigma^2, \gamma)$ at proportion p, with a distribution of mass $(1 - p)$ at the origin, $0 \leqq p \leqq 1$.

THEOREM 3. *A scale parameter family P_θ of distributions is also a one-parameter exponential family of distributions in θ, if and only if, P_θ is, for fixed θ ($\theta = 1$, say), the mixture $p_1L(\theta_1, \sigma^2, \gamma) + p_2(-L(\theta_2, \sigma^2, \gamma)) + p_3D(0)$, for some finite γ, $\sigma^2 > 0$, $\theta_1 > 0$ and $\theta_2 > 0$, where p_1, p_2, p_3 are non-negative constants whose sum is one.*

PROOF. The "if" part of the theorem is easy to check, so we can restrict our attention to the "only if" part. We know from a previous discussion that in a scale parameter family there may be a mass, p_3, $0 \leqq p_3 \leqq 1$, at the origin provided it is independent of the parameter. Subtracting this mass from the distributions will make them absolutely continuous with respect to Lebesgue measure. The conditional distribution of $Y = \log X$, given $X > 0$, is a one-parameter exponential family in a location parameter $\mu = \log \theta$, where X has the distribution P_θ. Hence, the conditional distribution of Y given $X > 0$ must be for fixed μ, $N(\mu_1, \sigma_1^2, \gamma_1)$ for some μ_1, $\sigma_1^2 > 0$ and finite γ_1, so that the conditional distribution of X, given $X > 0$, must be for fixed θ, $L(\theta_1, \sigma_1^2, \gamma_1)$ where $\mu_1 = \log \theta_1$. Similarly, the distribution of X given $X < 0$ must be for fixed θ, $-L(\theta_2, \sigma_2^2, \gamma_2)$ for some $\theta_2 > 0$, $\sigma_2^2 > 0$ and finite γ_2. Therefore, the logarithm of the density, $h(x)$, has the following form: for $x > 0$,

$$
\begin{aligned}
&\text{either } h(x) = ax^\gamma + b \log x + c, \qquad \gamma \neq 0, \\
&\text{or } h(x) = a(\log x)^2 + b \log x + c,
\end{aligned} \tag{22}
$$

and for $x < 0$

$$
\begin{aligned}
&\text{either } h(x) = a'|x|^{\gamma'} + b' \log |x| + c' \qquad \gamma' \neq 0, \\
&\text{or } h(x) = a'(\log |x|)^2 + b' \log |x| + c'.
\end{aligned} \tag{23}
$$

Now we shall combine these two parts. If $h(x) = ax^\gamma + b(\log x) + c$ for $x > 0$, then, computing $h(x/\theta) - \log \theta$, we find that $Q_1(\theta) = c_1\theta^{-\gamma}$, and

$$Q_0(\theta) = -(b + 1) \log \theta.$$

Since the Q_i must be the same whether x is positive or negative, we may deduce that for $x < 0$, $h(x) = a'|x|^\gamma + b \log |x| + c'$. If, on the other hand,

$$h(x) = a(\log x)^2 + b(\log x) + c \quad \text{for} \quad x > 0,$$

then $Q_1(\theta) = c_1(\log \theta) + c_2$ and $Q_0(\theta) = a(\log \theta)^2 + c_3(\log \theta) + c_4$. From this, we may deduce that for $x < 0$, $h(x) = a(\log |x|)^2 + b'(\log |x|) + c'$. Thus, either

$$h(x) = \begin{cases} a\, x^\gamma + b \log x + c & \text{for } x > 0 \\ a'|x|^\gamma + b \log |x| + c' & \text{for } x < 0 \end{cases} \tag{24}$$

or

$$h(x) = \begin{cases} a(\log x)^2 + b \log x + c & \text{for } x > 0 \\ a(\log |x|)^2 + b' \log |x| + c' & \text{for } x < 0. \end{cases} \tag{25}$$

It is clear that the same value of γ must be used for the negative and positive parts of the density. The fact that the coefficients of $\log |x|$ must be the same in equation (24), and the coefficients of $(\log |x|)^2$ must be the same in equation (25), is equivalent to the fact that the same value of the parameter σ^2 must be used for the negative and positive parts of the density. This completes the proof of the theorem.

This class of distributions contains the normal distribution with mean zero and variance θ^2:

$$\tfrac{1}{2}L(\theta, \tfrac{1}{4}\psi'(\tfrac{1}{2}), 2) + \tfrac{1}{2}(-L(\theta, \tfrac{1}{4}\psi'(\tfrac{1}{2}), 2));$$

and the Laplace distribution centered at the origin:

$$\tfrac{1}{2}L(\theta, \psi'(1), 1) + \tfrac{1}{2}(-L(\theta, \psi'(1), 1)).$$

A generalization of Theorem 3, allowing infinite values of γ would be to characterize $p_1L(\theta_1\theta, \sigma^2, \gamma) + p_2(-L(\theta_2\theta, \sigma^2, \gamma)) + p_3D(0)$ as that class of distributions for which there exists a complete sufficient statistic for a scale parameter, θ, in the dominated case, irrespective of the sample size. However, as in the corresponding statement involving location parameters, it is not known if this generalization is true without added regularity conditions.

7. Location-scale parameters; a characterization of the normal distribution. To complete our study, we present a theorem describing those k-parameter exponential families which contain both location and scale parameters. We shall use the following definition of parameters which are jointly location and scale parameters.

DEFINITION 4. A two-dimensional parameter (μ, σ), with $\sigma > 0$, is said to be a location-scale parameter of a family of distributions, $F(x \mid \mu, \sigma)$, if $F(x \mid \mu, \sigma)$ is a function only of $(x - \mu)/\sigma$.

If a density, $f(x \mid \mu, \sigma)$, with respect to Lebesgue measure exists, then (μ, σ) is a location-scale parameter if, and only if, $f(x \mid \mu, \sigma) = (1/\sigma)g((x - \mu)/\sigma)$ for some function, g. In terms of random variables, (μ, σ) is a location-scale parameter of the distribution of X if, and only if, the distribution of $(X - \mu)/\sigma$ is independent of μ and σ, when μ and σ are the true values of the parameters.

It is important to notice that if (μ, σ) is a location-scale parameter, then μ is a true location parameter (i.e., the distribution of $X - \mu$ does not depend on μ, for each fixed $\sigma > 0$). However, σ does not satisfy our definition of a scale parameter of the distribution of X unless $\mu = 0$.

The following theorem also has been proved by Dynkin [3] under the assumption that the density has a piecewise continuous derivative. Since Lemma 4 implies that this assumption is automatically satisfied in the present context, we present this theorem without proof, referring the reader to Dynkin's paper.

THEOREM 4. *If* (μ, σ) *is a location-scale parameter for a k-parameter exponential family of distributions in* (μ, σ), *then k is even, and the logarithm of the density is a polynomial of effective degree k.*

COROLLARY. *The only 2-parameter exponential family of distributions in a location-scale parameter is the family of normal distributions.*

The generalization of this result to distributions having a complete sufficient statistic for a location-scale parameter, will involve, in addition to the normal distributions, two other distinct types of "non-regular" distributions. The extreme type, involving sufficient statistics due to non-regularity at both ends of the distribution, is the uniform distribution over the interval

$$(\mu - \sqrt{3}\sigma, \mu + \sqrt{3}\sigma);$$

having, for a sample, $X_1, \cdots, X_n$, a sufficient statistic of the form

$$(\min_i X_i, \max_i X_i)$$

for the location-scale parameter (μ, σ). The other type is mixed, having one of the sufficient statistics due to "non-regularity" and the other not. This type consists of the exponential distributions $N(\theta, \sigma^2, -\infty)$ and $N(\theta, \sigma^2, +\infty)$. Whether there are distributions other than $N(\theta, \sigma^2, 0)$, $N(\theta, \sigma^2, -\infty)$, $N(\theta, \sigma^2, +\infty)$, and the uniform distributions, having a complete sufficient statistic for a location-scale parameter in the dominated case is an open problem.

8. Applications. The author's motivation for investigating location and scale parameters in exponential families was to obtain an idea of the generality of Basu's theorem on statistics independent of a sufficient statistic, [1]. This theorem states that if T is a (boundedly) complete sufficient statistic for a parameter θ, and if a statistic U has a distribution which does not depend on θ, then the statistics T and U are stochastically independent. The main classes of distributions for which a k-dimensional sufficient statistic exists for a sample size greater than k, in the so-called "regular" case, are the k-parameter exponential families. (For a proof and a statement of the regularity conditions see Dynkin [3].) Furthermore, when the parameter space has the same dimensionality as the sufficient statistic, the sufficient statistic is complete (see [7] p. 132). On the other hand, the main families of univariate distributions for which it is easy to find statistics whose distributions are independent of the parameter(s) are the location and/or scale parameter families. Given a sample $X_1, \cdots, X_n$ from a distribution with a location parameter θ, the joint distribution of the differences, $X_1 - X_2, \cdots, X_{n-1} - X_n$, does not involve θ. Thus, if there is a complete sufficient statistic, T, for θ, T will be stochastically independent of any function of the differences of the X_i. Similarly, the joint distribution of the ratios of the X_i does not depend on any scale parameter, and the joint distribution of the ratios of the differences does not depend on any location-scale parameter. Thus, the main, "regular," univariate applications of Basu's theorem involve location and/or scale parameters in exponential families of distributions, in which the dimensionality of the parameter space is equal to the dimensionality of the sufficient statistic. The resulting distributions are found in Theorems 2 and 3.

It should be pointed out that there are, of course, other applications of Basu's theorem. First, there is the "non-regular" case of sufficient statistic, as, for example, $N(\theta, \sigma^2, \pm\infty)$ and $L(\theta, \sigma^2, \pm\infty)$. Second, there are applications which, strictly speaking, do not involve location or scale parameters. For example, in a sample from the lognormal distribution $L(1, \sigma^2, 0)$, $\sum (\log X_i)^2$ is independent of any function $g(X_1, \cdots, X_n)$ for which $g(X_1, \cdots, X_n) = g(X_1^\lambda, \cdots, X_n^\lambda)$ for all λ. Actually, this is just a restatement of the fact that σ is a scale parameter for the $N(0, \sigma^2, 0)$ distribution. Third, R. G. Laha [6] mentions that for the $N(\theta, 1, 0)$ distribution, he has found statistics independent of $\bar{X}$ which are not location invariant.

As an example of the use of these results, consider the two theorems proved by Laha in [6]; these state that when $X_1, \cdots, X_n$ is a sample from a normal [resp. gamma] distribution, then $\bar{X}$ and $g(X_1, \cdots, X_n)$ are stochastically independent if, and only if, for every λ [resp. $\lambda > 0$] $g(X_1, \cdots, X_n)$ and

$$g(\lambda + X_1, \cdots, \lambda + X_n) [\text{resp. } g(\lambda X_1, \cdots, \lambda X_n)]$$

are identically distributed. These theorems follow immediately from Basu's theorem and its converse, [2]. But these theorems may be generalized to samples from any of the distributions of Theorems 2 and 3. Furthermore, these two theorems may be combined into a single theorem as follows: when $X_1, \cdots, X_n$ is a sample from $N(\theta, \sigma^2, \gamma)$, then $M_\gamma(X_1, \cdots, X_n)$ (see formula (21)) and

$$g(X_1, \cdots, X_n)$$

are stochastically independent if and only if $g(X_1, \cdots, X_n)$ and

$$g(\lambda + X_1, \cdots, \lambda + X_n)$$

are identically distributed for every λ. This contains Laha's result for the normal distribution when $\gamma = 0$, and his result for the gamma distribution when $\gamma = 1$. The importance of this extension lies not so much in the fact that a larger class of distributions is involved. Indeed, the whole of this theorem, except for

$$\gamma = \pm\infty,$$

follows immediately from Laha's theorems. It is rather that this larger class of distributions (1) constitutes a natural setting in which to prove many theorems as one, and (2) provides a maximal class to which the extension can be made.

Two other examples of the way theorems involving normal and gamma distributions may be unified by the class of distributions $N(\theta, \sigma^2, \gamma)$ will be mentioned. The first example deals with the characterization of the normal and gamma distributions by independence of certain statistics. The well-known Kac-Bernstein-Gnedenko result, see [9], is as follows. If random variables X and Y are independent, and if $X + Y$ is independent of $X - Y$, then X and Y have normal distributions with identical variances. A corresponding result on gamma distributions stated by Lukacs [9] under the assumption that X and Y are positive random variables is: if X and Y are independent, and if $X + Y$ is independent of X/Y then X and Y have gamma distributions. These results can be combined under one roof as follows. If X and Y are independent, and if for some finite γ, $M_\gamma(X, Y)$ is independent of $X - Y$, then X and Y have $N(\theta, \sigma^2, \gamma)$ distributions. Again, this theorem could easily be proved from the above results of Lukacs, and Kac-Bernstein-Gnedenko. The above generalization is not true in the case $\gamma = \pm\infty$, a counterexample being the geometric distribution, as is easily seen by applying Basu's theorem. The author intends to treat these problems in another paper.

The final example is that of characterizing distributions involving location parameters by the functional form of the maximum likelihood estimate. A theorem exists, due originally to Gauss [4], which states that under the assumption that the density has one continuous deviative, if $X_1, \cdots, X_n$ is a sample from a distribution with location parameter, θ, and if for some $n \geqq 3$, $\bar{X}$ is the maximum likelihood estimate of θ, then the distribution is normal. A proof under very weak conditions, provided that $\bar{X}$ is the maximum likelihood estimate for $n = 2$ and $n = 3$, may be found in Teicher [11]. This theorem can be extended to the class $N(\theta, \sigma^2, \gamma)$ for γ finite, as follows.

THEOREM 5. *Let $X_1, X_2, \cdots, X_n$ be a sample from a distribution with unknown location parameter, whose density admits one continuous derivative. If for every n the maximum likelihood estimate of the location parameter exists and is equal to $M_\gamma(X_1, \cdots, X_n)$ for some $\gamma \neq \pm\infty$, then the distribution is $N(\theta, \sigma^2, \gamma)$ for some $\sigma^2 > 0$. For $\gamma = 0$, one value of $n = 3$ suffices for this result.*

PROOF. The likelihood function is $L = \prod_{i=1}^n f(x_i - \theta)$, where f admits one continuous derivative. The maximum likelihood estimate, $\hat{\theta}$, satisfies the equation $\sum_{i=1}^n g(x_i - \hat{\theta}) = 0$, where $g(x) = f'(x)/f(x)$, for x for which $f(x) \neq 0$, and is undefined otherwise. For $\gamma \neq 0$, we have that

$$(26) \qquad \sum_{i=1}^n g\left[x_i - (1/\gamma)\log(1/n)\sum_{j=1}^n e^{\gamma x_j}\right] = 0$$

for all n and all $x_1, \cdots, x_n$. If we let $h(x)$ denote $g((1/\gamma)\log x)$ for x positive, equation (26) is equivalent to the equation

$$(27) \qquad \sum_{i=1}^n h(z_i) = 0$$

for all n, and all positive $z_1, \cdots, z_n$ for which $\sum_1^n z_i = n$. Noting that

$$h(z_1 + \epsilon) + h(z_2 - \epsilon) + \sum_3^n h(z_i) = 0$$

provided in addition that $-z_1 < \epsilon < z_2$, and subtracting (27), we find that

$$(28) \qquad h(z_1 + \epsilon) - h(z_1) = h(z_2) - h(z_2 - \epsilon).$$

When $n \geqq 3$, this implies that $h(z + \epsilon) - h(z)$ is a function of ϵ only, provided $0 < z < n$ and $0 < z + \epsilon < n$. This, continuity of h, and the fact that $h(1) = 0$, imply that $h(z) = a(z - 1)$ for $0 < z < n$. Since true for all n, this equation is true for all z, and

$$\log f(y) = a((1/\gamma)e^{\gamma y} - y) + c, \tag{29}$$

for all y. This with $a = 1/\sigma^2\gamma$ is the logarithm of the density of $N(0, \sigma^2, \gamma)$.

For $\gamma = 0$, the same method will arrive at the equation, $\sum_1^n g(z_i) = 0$, for all z_i for which $\sum_1^n z_i = 0$, similar to equation (31), without the restriction that the z_i be positive. Continuing as before, we will conclude that $g(x) = ax$ for all x, implying normality, and completing the proof.

This theorem can as easily be stated in terms of scale parameters for distributions on the positive real line. The resulting set of distributions are, naturally, $L(\theta, \sigma^2, \gamma)$ for finite γ. In such a form, Theorem 2 of Teicher's paper [11] is concerned with the case $\gamma = 1$, under a different set of regularity conditions, one of which rules out values of σ^2 other than $\sigma^2 = \psi'(1)$, thus yielding a characterization of the exponential distribution, when, in fact, $\bar{X}$ is the maximum likelihood estimate of θ for all the gamma distributions $L(\theta, \sigma^2, 1)$.

We see from the proof of Theorem 5 for the case $\gamma \neq 0$, that if the hypotheses are valid for just one $n \geqq 3$, then the logarithm of the density has the form (29) at least for those y for which $\gamma y < \log n$. That it is not necessary that it have this form for all y is seen by considering the $N(\theta, \sigma^2, 1)$ distribution truncated at $B + \theta$. Let $Y_1, Y_2, \cdots, Y_n$ be a sample from a distribution with density

$$f(y \mid \theta) = \begin{cases} c(\sigma) \exp\{-(1/\sigma)^2 e^{(y-\theta)} + (1/\sigma)^2(y - \theta)\} & \text{if } y - \theta < B \\ 0 & \text{if } y - \theta > B. \end{cases} \tag{30}$$

Then, θ is a location parameter whose maximum likelihood estimate is

$$\hat{\theta} = \max\{M_1(Y_1, \cdots, Y_n), \max_i Y_i - B\}. \tag{31}$$

But whenever $B > \log n$, $\hat{\theta} = M_1(Y_1, \cdots, Y_n)$ irrespective of the values of the Y_i. Although the density (30) does not satisfy the differentiality assumption of Theorem 5, it can be patched up easily, by letting the density come down to zero after $B + \theta$, smoothly and quickly enough.

Finally, we show by a counterexample, why nothing similar to Theorem 5 can be proved for the distributions $N(\theta, \sigma^2, \pm\infty)$. Let $X_1, \cdots, X_n$ be a sample from a distribution with density of the form

$$f(x \mid \theta) = \begin{cases} g(x - \theta) & \text{if } x < \theta \\ 0 & \text{if } x > \theta \end{cases} \tag{32}$$

where $g(x)$ is a strictly increasing function of x on $(-\infty, 0)$ (as for formula (30) with $B = 0$). Then, the maximum likelihood estimate of the location parameter, θ, is always $\max_i X_i = M_{+\infty}(X_1, \cdots, X_n)$.

REFERENCES

[1] BASU, D. (1955). On statistics independent of a complete sufficient statistic. *Sankhya* **15** 377–380.

[2] BASU, D. (1958). On statistics independent of sufficient statistics. *Sankhya* **20** 223–226.

[3] 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, *Selected Transl. Math. Stat. Prob.* pp. 17–40).

[4] GAUSS, C. F. (1809). *Theoria motus corporum coelestium in sectionibus conicus solem ambientium.* Hamburg.

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

[6] LAHA, R. G. (1956). On some properties of the normal and gamma distributions. *Proc. Amer. Math. Soc.* **7** 172–174.

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

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

[9] LUKACS, E. (1956). Characterization of populations by properties of suitable statistics, 195–214. *Proc. Third Berkeley Symp. Math. Stat. Prob.* Univ. of California Press.

[10] PITMAN, E. J. G. (1936). Sufficient statistics and intrinsic accuracy. *Proc. Cambridge Philos. Soc.* **32** 567–579.

[11] TEICHER, H. (1961). Maximum likelihood characterization of distributions. *Ann. Math. Statist.* **32** 1214–1222.