The Annals of Statistics
1975, Vol. 3, No. 3, 747–753

CHARACTERIZING EXPONENTIAL FAMILY DISTRIBUTIONS BY MOMENT GENERATING FUNCTIONS[1]

BY ALLAN R. SAMPSON

Florida State University and Tel Aviv University

It is shown that if $\mathbf{T}$ has an unknown exponential family distribution with natural parameter $\boldsymbol{\theta}$, then $\mathbf{G}(\boldsymbol{\theta}) = E\mathbf{T}$ uniquely specifies the moment generating function. The converse is proved, namely, if $\{\mathbf{T}_{\boldsymbol{\theta}}\}$ is a family of random variables with moment generating functions of a certain form, then it must be an exponential family. Moreover, several necessary and sufficient conditions are given so that a function can be the mean value function of an exponential family distribution.

1. Introduction. In this paper we consider certain results characterizing exponential family distributions through their mean functions and moment generating functions. We also give necessary and sufficient conditions so that a function can be the mean value function of an exponential family distribution. Our work was motivated in part by Anderson [1]. She shows that the normal distribution is characterized by being an exponential family distribution having a mean function of a certain form. Bildikar and Patil [3] give a general discussion of multidimensional exponential family distributions and consider certain characterizations for this family. In particular they show that the ratio of the mean to the variance of the family's sufficient statistic characterizes the binomial, Poisson and negative binomial distributions.

Suppose $\{T_\theta, \theta \in \Theta\}$ is a family of scalar valued random variables where the following holds.

(1.1) There exists a σ-finite measure μ, such that the pdf $p(t, \theta)$ of T_θ with respect to μ is of the form

$$p(t, \theta) = \exp[\theta t + Q(\theta) + R(t)]$$

on some set S (independent of θ).

Thus T_θ is a random variable whose distribution is an exponential family distribution with a single natural parameter θ.

It is shown that if (1.1) holds and $g(\theta) = ET_\theta$ for $\theta \in \Theta$, then the moment generating function (mgf) of T_θ is uniquely determined and given by

$$m_{T,\theta}(s) = \exp[\textstyle\int_\theta^{\theta+s} g(w)\,dw]\,. \tag{1.2}$$

Received March 1972; revised May 1974.

[1] Part of this research was done while the author was a Visiting Lecturer at Tel Aviv University during 1971–72. The research at Florida State University was sponsored in part by ONR Grant No. N00014-67-A-0235-0006.

AMS 1970 *subject classifications.* Primary 62E10; Secondary 62H05.

Key words and phrases. Exponential family, characterization of distributions, normal, exponential, Wishart, moment generating function.

Additionally we prove the converse showing that if $\{T_\theta, \theta \in \Theta\}$ is a family of random variables whose mgf's are given by (1.2), then T_θ must have an exponential family distribution with respect to some σ-finite measure μ and ET_θ must be equal to $g(\theta)$. Also, we give necessary and sufficient conditions on functions $g(\theta)$ so that $g(\theta)$ can be equal to ET_θ, $\theta \in \Theta$, when $\{T_\theta\}$ is a family of random variables satisfying (1.1). Multivariate versions of these results are obtained.

In expressing our results, as will become apparent in the proofs, we could equivalently use mgf's or characteristic functions or bilateral Laplace Transforms (B.L.T.).

2. The scalar case. Because we are dealing with the natural parameter in (1.1), we consider Θ to be the natural parameter space. In order that the function $g(\theta)$ be differentiable everywhere on Θ, we make the following assumption.

ASSUMPTION. Let Θ be open.

Because additionally the natural parameter space is convex, Θ is alternatively written (a, b), which may possibly be a non-finite interval.

THEOREM 1. *Let $\{T_\theta, \theta \in (a, b)\}$ be a family of random variables such that* (1.1) *holds and $ET_\theta = g(\theta)$. Then for $\theta \in (a, b)$ the* mgf *of T_θ exists and is given by* (1.2) *for $s \in (a - \theta, b - \theta)$.*

PROOF. Note $s \in (a - \theta, b - \theta)$ iff $\theta + s \in (a, b) = \Theta$. Because (1.1) holds and $s \in (a - \theta, b - \theta)$,

$$\begin{aligned} m_{T,\theta}(s) &= \textstyle\int_S \exp[(\theta + s)t + Q(\theta) + R(t)]\, d\mu \\ &= \exp[Q(\theta) - Q(\theta + s)] . \end{aligned}$$

Differentiate $m_{T,\theta}(s)$ by s and set $s = 0$ to obtain that $Q(\theta)$ satisfies the differential equation $g(\theta) = -dQ(\theta)/d\theta$. □

Theorem 1 is implicitly contained in [3], though not in the actual form given here.

A much more interesting result is the converse of Theorem 1 given in the theorem below.

THEOREM 2. *Suppose $\{T_\theta, \theta \in (a, b)\}$ is a family of random variables. For all θ, let $m_{T,\theta}(s)$, the* mgf *of T_θ, be given by* (1.2) *for $s \in (a - \theta, b - \theta)$. Then for all θ, $ET_\theta = g(\theta)$ and* (1.1) *holds.*

PROOF. That $ET_\theta = g(\theta)$ follows by evaluating $dm_{T,\theta}(s)/ds$ at $s = 0$.

To show (1.1), first fix $\theta = \theta_0$ for any $\theta_0 \in (a, b)$. Because $m_{T,\theta_0}(s) = \exp[\int_{\theta_0}^{\theta_0+s} g(w)\, dw]$ is a moment generating function, there exists a probability measure μ_{θ_0} such that

$$m_{T,\theta_0}(s) = \textstyle\int_{-\infty}^{\infty} e^{st}\, d\mu_{\theta_0} ,$$

for $s \in (a - \theta_0, b - \theta_0)$. But for any $\theta \in (a, b)$

$$\begin{aligned} m_{T,\theta}(s) &= \exp[\textstyle\int_\theta^{\theta+s} g(w)\,dw] \\ &= \exp[\textstyle\int_{\theta_0}^{\theta_0+s+\theta-\theta_0} g(w)\,dw + \int_\theta^{\theta_0} g(w)\,dw] \\ &= \exp[\textstyle\int_\theta^{\theta_0} g(w)\,dw] m_{T,\theta_0}(s + \theta - \theta_0)\,, \end{aligned}$$

because $s + \theta - \theta_0 \in (a - \theta_0, b - \theta_0)$ iff $s \in (a - \theta, b - \theta)$. Therefore,

$$m_{T,\theta}(s) = \textstyle\int_{-\infty}^{\infty} e^{st} p^*(t, \theta)\, d\mu_{\theta_0}\,,$$

where

$$p^*(t, \theta) = \exp[t\theta + \textstyle\int_\theta^{\theta_0} g(w)\,dw - t\theta_0]\,.$$

Observe that $p^*(t, \theta) > 0$ and $\int p^*(t, \theta)\, d\mu_{\theta_0} = 1$ so that p^* is a pdf with respect to μ_{θ_0}. In this case S is the support of μ_{θ_0} which does not depend on θ. The result now follows from the uniqueness theorem for moment generating functions. □

Note 2.1. The proof of Theorem 2 demonstrates that a set of functions of the form (1.2) for $s \in (a - \theta, b - \theta)$ and $\theta \in (a, b)$ is a set of mgf's if at least one of the functions is an mgf.

The question arises as to what are necessary and sufficient conditions on a function g, so that if $\{T_\theta\}$ is a family of random variables satisfying (1.1), ET_θ can be equal to $g(\theta)$ for $\theta \in \Theta$. Loosely speaking, the question is which functions can arise as expectations of exponential family random variables?

DEFINITION. $M = \{\gamma(\cdot)$: there exists a family $\{T_\theta, \theta \in (a, b)\}$ satisfying (1.1) with $\gamma(\theta) = ET_\theta\}$.

THEOREM 3. *In order that a function* $g \in M$, *it is necessary and sufficient that for some* $\theta_0 \in (a, b)$, $\exp[\int_{\theta_0}^{\theta_0+s} g(w)\,dw]$, $s \in (a - \theta_0, b - \theta_0)$, *be an* mgf.

The proof of this Theorem follows immediately from Theorems 1 and 2 and Note 2.1.

We now discuss some ways of verifying whether $g \in M$, or equivalently if a function $\exp[\int_{\theta_0}^{\theta_0+s} g(w)\,dw]$, $s \in (a - \theta_0, b - \theta_0)$, is an mgf.

LEMMA 1. *In order that* $\exp[\int_{\theta_0}^{\theta_0+s} g(w)\,dw]$, $s \in (a - \theta_0, b - \theta_0)$, *be an* mgf, *it is necessary and sufficient that*

(2.1) $\exp[\int_{\theta_0}^{u+v} g(w)\,dw]$ *is nonnegative definite for* $(2u, 2v) \in \Theta \times \Theta$, *where* $\Theta = (a, b)$.

PROOF. Note that $\exp[\int_{\theta_0}^{\theta_0+s} g(w)\,dw]$, $s \in (a - \theta_0, b - \theta_0)$, is an mgf iff

(2.2) $$\exp[\textstyle\int_{\theta_0}^{\theta_0-s} g(w)\,dw]\,, \quad s \in (\theta_0 - b, \theta_0 - a)$$

is a B.L.T. However, by Theorem 21 of [9] a necessary and sufficient condition that (2.2) be a B.L.T. is that $\exp[\int_{\theta_0}^{\theta_0-(u_1+v_1)} g(w)\,dw]$ be nonnegative definite for $\theta_0 - b < 2u_1, 2v_1 < \theta_0 - a$. Set $u = \theta_0/2 - u_1$ and $v = \theta_0/2 - v_1$ to see that this condition is equivalent to (2.1). □

Lemma 2. *Let $g(\theta) = P_n(\theta)$, where $P_n(\cdot)$ is a polynomial of degree n. If $n \geqq 2$, then $g \notin M$.*

Proof. If $g(\theta) = P_n(\theta)$, $n \geqq 2$, then by Theorem 1, $m_{T,\theta}(s) = \exp[Q_{n+1}(s)]$ where $Q_{n+1}(s)$ is a polynomial of degree $n + 1 > 2$. But by the Theorem of Marcinkiewicz (e.g., page 213 of Lukacs [7]), $m_{T,\theta}(s)$ cannot then be an mgf for any $\theta \in (a, b)$. The result now follows from Theorem 3. □

As noted earlier we could express our results in terms of the characteristic function $\phi_{T,\theta}(s)$ of T_θ. It can be shown that if $\{T_\theta, \theta \in \Theta\}$ satisfies (1.1), then

$$\phi_{T,\theta}(s) = \exp[G(\theta + is) - G(\theta)]\,, \tag{2.3}$$

where $G(z)$ is the analytic extension to the complex plane of $\int g(w)\,dw$. The proof follows the same lines as that of Theorem 1 with the added observation that

$$b(\theta) \equiv \int \exp[\theta t + R(t)]\,d\mu$$

considered as a function of the complex variable θ is analytical in the strip: real part of θ in Θ (e.g., Theorem 9, page 52 of [6]). Hence, equivalent to (1.2) being an mgf is that (2.3) is a characteristic function.

We can obtain additional results about $\{T_\theta\}$ by verifying other properties about $g(\theta)$. For example, suppose $\{T_\theta\}$ is a family of nonnegative random variables satisfying (1.1) for $\theta \in (-\infty, c)$ and $ET_\theta = g(\theta)$. Then a necessary and sufficient condition that for a given θ_0, T_{θ_0} have an infinitely divisible distribution is that for $t > 0$

$$g^{(k)}(\theta_0 - t) > 0 \qquad \text{for } k = 0, 1, \cdots,$$

where $g^{(k)}(x) \equiv d^k g(x)/dx^k$. A proof of this result can be obtained using Theorem 1, page 450 of Feller [4].

We now consider some simple examples of Theorem 1.

Example 1. If $g(\theta) = \theta$, then $m_T(s) = \exp[s\theta + s^2/2]$ which is the mgf of a normal random variable with mean θ and variance 1.

Example 2. If $g(\theta) = -\theta^{-1}$, then $m_T(s) = (1 + s/\theta)^{-1}$ which is the mgf of an exponential distribution with parameter $-\theta$.

Example 3. If $g(\theta) = e^\theta$, then $m_T(s) = \exp[e^\theta(e^s - 1)]$ which is the mgf of a Poisson random variable with parameter e^θ.

Example 4. If $g(\theta) = (1 + e^{-\theta})^{-1}$, then $m_T(s) = (1 + e^{s+\theta})/(1 + e^\theta)$ which is the mgf of a Bernoulli random variable with probability of success $(1 + e^{-\theta})^{-1}$.

Up to reparametrization by a constant, Example 1 is essentially Theorem 1 of [1]; Examples 3 and 4 overlap Theorem 5.1 of [3].

3. The vector case. In this section we discuss exponential families with more than one sufficient statistic.

Suppose $\{\mathbf{T}_\theta, \boldsymbol{\theta} \in \Theta\}$ is a family of k-dimensional random vectors satisfying:

(3.1) There exists a σ-finite measure μ, such that the pdf $p(\mathbf{t}, \boldsymbol{\theta})$ of $\mathbf{T}_\theta$ with respect to μ is of the form
$$p(\mathbf{t}, \boldsymbol{\theta}) = \exp[\boldsymbol{\theta}'\mathbf{t} + Q(\boldsymbol{\theta}) + R(\mathbf{t})]$$
on some set S (independent of $\boldsymbol{\theta}$).

Without loss of generality, assume Θ is an open k-dimensional rectangle. (It is supposed that Θ is actually k-dimensional, so that convexity of the natural parameter space implies Θ contains such a rectangle which may be used in place of Θ.)

Because the vector results are the analogues of the corresponding ones in the previous section, we delete all proofs.

Analogous to $(a - \theta, b - \theta)$, define the following set
$$A(\boldsymbol{\theta}) = \{\mathbf{s} : \boldsymbol{\theta} + \mathbf{s} \in \Theta\}.$$
Note $A(\boldsymbol{\theta})$ is an open rectangle containing the origin. For any function h of $\boldsymbol{\theta}$, define $\partial h(\boldsymbol{\theta})/\partial \boldsymbol{\theta} \equiv (\partial h(\boldsymbol{\theta})/\partial \theta_1, \cdots, \partial h(\boldsymbol{\theta})/\partial \theta_k)'$.

THEOREM 4. *Let* $\{\mathbf{T}_\theta, \boldsymbol{\theta} \in \Theta\}$ *be a family of random vectors such that* (3.1) *holds and* $E\mathbf{T}_\theta = \mathbf{G}(\boldsymbol{\theta})$. *Then the differential equation*
$$-\mathbf{G}(\boldsymbol{\theta}) = \partial \Gamma(\boldsymbol{\theta})/\partial \boldsymbol{\theta} \tag{3.2}$$
admits a solution and the mgf *of* $\mathbf{T}_\theta$ *exists and is given for* $\mathbf{s} \in A(\boldsymbol{\theta})$ *by*
$$m_{\mathbf{T},\theta}(\mathbf{s}) = \exp[\Gamma(\boldsymbol{\theta}) - \Gamma(\boldsymbol{\theta} + \mathbf{s})], \tag{3.3}$$
where $\Gamma(\boldsymbol{\theta})$ *is a solution of* (3.2).

Note that the condition that there exists a solution to (3.2) is equivalent to the integrability of $g(\theta)$ in the univariate case.

THEOREM 5. *Let* $\{\mathbf{T}_\theta, \boldsymbol{\theta} \in \Theta\}$ *be a family of random vectors, such that for all* $\boldsymbol{\theta}$, $m_{\mathbf{T},\theta}(\mathbf{s})$ *is given by* (3.3) *and* (3.2). *Then* $E\mathbf{T}_\theta = G(\boldsymbol{\theta})$ *and* (3.1) *holds.*

THEOREM 6. *In order that a function* $\mathbf{G}$ *be the mean value function of a family given by* (3.1), *it is necessary and sufficient that for some* $\boldsymbol{\theta}_0 \in \Theta$, $\exp[\Gamma(\boldsymbol{\theta}_0) - \Gamma(\boldsymbol{\theta}_0 + \mathbf{s})]$, *where* Γ *is given by* (3.3), *is an* mgf *for* $\mathbf{s} \in A(\boldsymbol{\theta}_0)$.

The multivariate form of (2.1) becomes:
$$\exp[\Gamma(\boldsymbol{\theta}_0) - \Gamma(\mathbf{u} + \boldsymbol{v})] \quad \text{is nonnegative definite for} \quad (2\mathbf{u}, 2\boldsymbol{v}) \in \Theta \times \Theta,$$
where Γ is again given by (3.2).

EXAMPLE 5. If $\mathbf{G}(\boldsymbol{\theta}) = \Sigma\boldsymbol{\theta}$, Σ positive definite, then (3.2) becomes $-\Sigma\boldsymbol{\theta} = \partial\Gamma(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ which admits only solutions of the form $\Gamma(\boldsymbol{\theta}) = -\frac{1}{2}\boldsymbol{\theta}'\Sigma\boldsymbol{\theta} + c$ (e.g., Graybill [5], page 262). Hence, $m_{\mathbf{T},\theta}(\mathbf{s}) = \exp[-\frac{1}{2}\boldsymbol{\theta}'\Sigma\boldsymbol{\theta} + \frac{1}{2}(\boldsymbol{\theta} + \mathbf{s})'\Sigma(\boldsymbol{\theta} + \mathbf{s})] = \exp[\mathbf{s}'\Sigma\boldsymbol{\theta} + \frac{1}{2}\mathbf{s}'\Sigma\mathbf{s}]$ which is the mgf of a normal random vector with mean $\Sigma\boldsymbol{\theta}$ and covariance matrix Σ.

The difference between the format of this example and Theorem 2 of [1] is a result of our parametrizing by the natural parameter space.

EXAMPLE 6. If $\mathbf{G}(\boldsymbol{\theta}) = (-\theta_1^{-1}, \cdots, -\theta_k^{-1})'$, then $\Gamma(\boldsymbol{\theta}) = -\sum_{i=1}^k \log \theta_i + c$, so that $m_{\mathbf{T},\boldsymbol{\theta}}(\mathbf{s}) = \prod_{i=1}^k (1 + s_i/\theta_i)^{-1}$ which is the mgf of a vector of independent exponential random variables with natural parameters $-\theta_1, \cdots, -\theta_k$.

Example 6 has an interesting interpretation in the formulation of possible multivariate generalizations of the exponential distribution. It implies that if we seek a multivariate exponential distribution of the form

$$p(\mathbf{t}, \boldsymbol{\theta}) = c(\boldsymbol{\theta})h(\mathbf{t}) \exp(\boldsymbol{\theta}'\mathbf{t})$$

requiring that

$$EX_i = -\theta_i^{-1} \qquad 1 \leqq i \leqq k, \tag{3.4}$$

then $p(\mathbf{t}, \boldsymbol{\theta})$ must be the trivial independent distribution. Note that (3.4) is much less restrictive than the usual condition of exponential marginals. For a review of multivariate exponential distributions see Marshall and Olkin [8].

More generally we have the following lemma about multivariate generalizations of arbitrary univariate exponential family distributions, the proof being obvious.

LEMMA 3. *Let* $\{T_\theta : \theta \in \Theta\}$ *be a family of random variables satisfying* (1.1) *and having* $ET_\theta = g(\theta)$. *Let* $\{\mathbf{Y}_{\boldsymbol{\theta}} \equiv (Y_1, \cdots, Y_k)', \boldsymbol{\theta} \equiv (\theta_1, \cdots, \theta_k) \in \times_{i=1}^k \Theta\}$ *be a family of random variables satisfying* (3.1). *If* $E\mathbf{Y}_{\boldsymbol{\theta}} = (g(\theta_1), \cdots, g(\theta_k))'$, *then* $Y_1, \cdots, Y_k$ *are* i.i.d. *with the same* pdf *as* T_θ.

4. Comments. While the case when T is a symmetric matrix random variable could be considered as part of the vector case, it is more convenient to handle separately. We only sketch such an approach. Replace in (3.1) the density by

$$p(t, \Lambda) = \exp[-\mathrm{tr}\,(t\Lambda) + Q(\Lambda) + R(t)], \tag{4.1}$$

where t, Λ are positive definite $p \times p$-matrices. Define the operator $\partial^*/\partial^*\lambda_{ij} = (\frac{1}{2})^{\Delta(i,j)}\partial/\partial\lambda_{ij}$, where $\Delta(i, i) = 0$ and for $i \neq j$, $\Delta(i, j) = 1$. Further define for any real-valued function $F(\Lambda)$, $\partial^*F(\Lambda)/\partial^*\Lambda = \{\partial^*F(\Lambda)/\partial^*\lambda_{ij}\}$, where $\Lambda = \{\lambda_{ij}\}$.

Let $\{T_\Lambda\}$ be a family of random variables each having density of the form (4.1) and satisfying $ET_\Lambda = G(\Lambda)$. Then the mgf of T_Λ is

$$\exp[\Gamma(\Lambda) - \Gamma(\Lambda - S)], \tag{4.2}$$

where $\Lambda - S \in \Theta$ and Γ is a solution of

$$G(\Lambda) = \partial^*\Gamma(\Lambda)/\partial^*\Lambda. \tag{4.3}$$

EXAMPLE 7. If $\Theta = \{\Lambda : \Lambda \text{ is positive definite}\}$ and $G(\Lambda) = (n/2)\Lambda^{-1}$ then (4.3) becomes $(n/2)\Lambda^{-1} = \partial^*\Gamma(\Lambda)/\partial^*\Lambda$, which has as a solution $\Gamma(\Lambda) = (n/2) \log |\Lambda| + c$ (e.g., page 267, [5]). Then (4.2) becomes $\exp[(n/2) \log |\Lambda| - (n/2) \log |\Lambda - S|] = |\Lambda|^{n/2}/|\Lambda - S|^{n/2}$, for $S \in I(\Lambda) = \{S : \Lambda - S \in \Theta\}$. This is the mgf of $V_{11}, \cdots, V_{pp}, 2V_{12}, \cdots, 2V_{p-1,p}$, where $V = \frac{1}{2}W$ and W has the Wishart distribution with parameter Λ^{-1} on n degrees of freedom. For a derivation of the characteristic function of W and thus its mgf, see Anderson [2], page 160.

Finally note that in our theorems we could use other central moments of T_θ in place of ET_θ. The difference is that the simple first-order linear differential equation relating $g(\theta)$ to $Q(\theta)$ becomes a much more complicated nonlinear differential equation of higher order. Along these lines there are presentations in [1] and [3], respectively, of characterizations of normal distributions by variances and of characterizations of binomial, Poisson and negative binomial random variables by Var T_θ/ET_θ.

Acknowledgment. I wish to thank Professors G. Kimeldorf and J. Sethuraman for helpful discussions.

REFERENCES

[1] ANDERSON, M. R. (1971). A characterization of the multivariate normal distribution. *Ann. Math. Statist.* **42** 824–827.

[2] ANDERSON, T. W. (1958). *An Introduction to Multivariate Statistical Analysis.* Wiley, New York.

[3] BILDIKAR, S. and PATIL, G. P. (1968). Multivariate exponential-type distributions. *Ann. Math. Statist.* **39** 1316–1326.

[4] FELLER, W. (1971). *An Introduction to Probability Theory and its Applications.* **2** (2nd ed.). Wiley, New York.

[5] GRAYBILL, F. A. (1969). *Introduction to Matrices with Applications in Statistics.* Wadsworth, Belmont, California.

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

[7] LUKACS, E. (1970). *Characteristic Functions.* (2nd ed.). Hafner, New York.

[8] MARSHALL, A. W. and OLKIN, I. O. (1967). A multivariate exponential distribution. *J. Amer. Statist. Assoc.* **62** 30–44.

[9] WIDDER, D. V. (1946). *The Laplace Transform.* Princeton Univ. Press.

DEPARTMENT OF STATISTICS
FLORIDA STATE UNIVERSITY
TALLAHASSEE, FLORIDA 32306