Statistica Sinica 3(1993), 157-171

MAXIMUM LIKELIHOOD CHARACTERIZATIONS OF DISTRIBUTIONS

Albert W. Marshall and Ingram Olkin

University of British Columbia and Western Washington University, and Stanford University

Abstract: In a classic paper, Teicher (1961) showed that certain location and scale families of distributions are characterized by maximum likelihood estimators of their respective families. In particular, the distributions that are characterized in this manner are the exponential and normal distributions as scale parameter families, and the normal distribution as a location family. We have extended these results to the gamma distribution as a scale parameter family, and to the multivariate normal distribution as a location parameter family. Similar results are obtained for elliptically contoured families and Laplace distributions.

Key words and phrases: Gamma distribution, normal distribution, Laplace distribution, functional equations, location families, scale families, elliptically contoured families.

1. Introduction

Several results are known that characterize parametric families of distributions from properties of maximum likelihood estimators. Methods developed by Teicher (1961) in a classic paper on this subject are utilized here to obtain multivariate versions of known univariate results.

More specifically, the following cases are considered.

I. Location parameter families $\{F(x-\theta), -\infty < \theta < \infty\}$

a. Teicher (1961) showed that if, for sample sizes $n = 2, 3$, a maximum likelihood estimator (MLE) of a location parameter θ is the sample mean $\bar{x}$, then with some regularity conditions, it follows that F is a normal distribution. As indicated by Teicher (1961), this result has its origins in the work of Gauss.

In the multivariate case, where x and θ are vectors, we show in Section 2 that if for sample sizes $n = 2, 3$, the sample mean vector is a MLE of a location parameter θ, then with appropriate regularity conditions, F is a multivariate normal distribution with some covariance matrix.

b. Kagan, Linnik and Rao (1973) showed that if, for sample size $n = 4$, a maximum likelihood estimator of a location parameter θ is the sample median, then with regularity conditions, F is a Laplace distribution. This result has its origins in the work of Laplace (see Johnson and Kotz (1970, Chapter 23, p.22)). Rao and Ghosh (1971) showed that the result fails for sample sizes $n = 2, 3$.

In the multivariate case we show in Section 3 that if for sample size $n = 4$ the vector of medians (computed component by component) is a MLE of a location parameter θ, then F is the product of univariate Laplace marginal distributions. This fact came as a disappointment to us since we were originally motivated to study this case by the hope of finding a broader class of interesting multivariate Laplace distributions.

II. Scale parameter families $\{F(x/\sigma), \sigma > 0\}$

a. Teicher (1961) showed that if the second sample moment is a MLE of σ^2 where $\sigma > 0$ is a scale parameter, then with appropriate regularity conditions, F is a normal distribution with zero mean.

In the multivariate case, the scale parameter can be extended, in terms of densities, to families of the form $\{\phi(\boldsymbol{x}, \Lambda) = c(\Lambda) f(\boldsymbol{x}\Lambda\boldsymbol{x}'), \boldsymbol{x} \in R^p\}$. Families of this kind are just those with elliptically contoured densities. We show in Section 4 with appropriate regularity conditions that if the sample covariance matrix is a MLE for $\Sigma = \Lambda^{-1}$, then ϕ is a normal density function with expectation 0.

b. Teicher (1961) showed that if the first sample moment is a MLE of σ, then with appropriate regularity conditions, $F(0) = 0$ and F is an exponential distribution with mean 1.

This result is generalized in two directions in Section 5. There, the first sample moment is replaced by $\sum x_i/rn$ where $r > 0$; in this case the exponential distribution is replaced by the gamma distribution with shape parameter r.

In the multivariate case, with the family $\{F(x_1/\sigma_1, \dots, x_p/\sigma_p)\}$, suppose that a maximum likelihood estimator of $(\sigma_1, \dots, \sigma_p)$ is the vector with components of the form $\sum x_{i\alpha}/r_i n$. Then with the required regularity conditions, F is a product of univariate gamma distributions.

There is a closely related extension of Teicher's result described under II.a above, which applies to scale parameter families of the kind considered in the preceding paragraph. If the vector with components $(\sum x_{i\alpha}^2/r_i n)^{1/2}$ is a MLE for σ^2, where all the r_i are odd integers, then

$$f(\boldsymbol{x}) = \prod_{i=1}^{p} \left\{ c_i x_i^{r_i - 1} \exp(-x_i^2/2) \right\}, \quad -\infty < x_i < \infty.$$

The case $r_1 = p = 1$ is the result of Teicher (1961).

2. Characterization of the Multivariate Normal Distribution by the Sample Mean Vector as a MLE of the Location Parameter

The following theorem with $p = 1$ is due to Teicher (1961).

Theorem 2.1. *Suppose $\{F(\boldsymbol{x} - \boldsymbol{\mu}), \boldsymbol{\mu} \in R^p\}$ is a translation family of distributions on R^p, where F has a density f that is lower semi-continuous at $\boldsymbol{x} = \mathbf{0}$. If, for all samples of size 2 and 3, a MLE of $\boldsymbol{\mu}$ is the mean vector $\bar{\boldsymbol{x}}$, then F is a multivariate normal distribution with mean zero.*

Proof. For sample size $n + 1 = 2$ or 3, the hypothesis asserts that

$$\prod_1^{n+1} f(\boldsymbol{x}_i - \bar{\boldsymbol{x}}) \geq \prod_1^{n+1} f(\boldsymbol{x}_i - \boldsymbol{\mu}) \quad \text{for all} \quad \boldsymbol{\mu}.$$

or equivalently,

$$\prod_1^{n+1} f(\boldsymbol{u}_i) \geq \prod_1^{n+1} f(\boldsymbol{u}_i - \boldsymbol{\theta}) \quad \text{for all} \quad \boldsymbol{\theta}, \tag{2.1}$$

where $\boldsymbol{\theta} = \bar{\boldsymbol{x}} - \boldsymbol{\mu}$, $\boldsymbol{u}_i = \boldsymbol{x}_i - \bar{\boldsymbol{x}}$, so that $\sum_1^{n+1} \boldsymbol{u}_i = 0$; with $h = \log f$ this becomes

$$\begin{aligned} &\sum_1^n h(\boldsymbol{u}_i) + h(\boldsymbol{u}_{n+1}) \\ = \; &\sum_1^n h(\boldsymbol{u}_i) + h\Big(-\sum_1^n \boldsymbol{u}_i\Big) \geq \sum_1^n h(\boldsymbol{u}_i - \boldsymbol{\theta}) + h\Big(-\sum_1^n \boldsymbol{u}_i - \boldsymbol{\theta}\Big). \end{aligned} \tag{2.2}$$

From (2.2) with $\boldsymbol{u}_1 = \cdots = \boldsymbol{u}_n \equiv \boldsymbol{t}$,

$$nh(\boldsymbol{t}) + h(-n\boldsymbol{t}) \geq nh(\boldsymbol{t} - \boldsymbol{\theta}) + h(-n\boldsymbol{t} - \boldsymbol{\theta}),$$

which for $n = 1$ yields

$$h(\boldsymbol{t}) + h(-\boldsymbol{t}) \geq h(\boldsymbol{t} - \boldsymbol{\theta}) + h(-\boldsymbol{t} - \boldsymbol{\theta}). \tag{2.3}$$

An argument is required to show that $f(\boldsymbol{y})$ is finite everywhere, and that $f(\boldsymbol{y}) > 0$ for $\boldsymbol{y} = \mathbf{0}$. For $y \in R^1$, Teicher (1961, p.1216) provides such an argument which carries through word for word for $\boldsymbol{y} \in R^p$ based upon (2.1).

In (2.3) replace $\boldsymbol{t}$ by $-\boldsymbol{t}$ and $\boldsymbol{\theta}$ by $-\boldsymbol{\theta}$ to yield

$$h(\boldsymbol{t}) + h(-\boldsymbol{t}) \geq h(-\boldsymbol{t} + \boldsymbol{\theta}) + h(\boldsymbol{t} + \boldsymbol{\theta}). \tag{2.4}$$

Let $g(\boldsymbol{t}) = h(\boldsymbol{t}) + h(-\boldsymbol{t})$; the sum of (2.3) and (2.4) shows that g is mid-point concave and reaches a maximum at $\boldsymbol{\theta} = \mathbf{0}$. Furthermore, g satisfies (2.2). With $n = 2$ in (2.2), and from the symmetry of g, it follows that

$$g(\boldsymbol{u}_1 - \boldsymbol{\theta}) + g(\boldsymbol{u}_2 - \boldsymbol{\theta}) + g(\boldsymbol{u}_1 + \boldsymbol{u}_2 + \boldsymbol{\theta})$$

reaches a maximum at $\boldsymbol{\theta} = \mathbf{0}$, so

$$\nabla\Big[g(\boldsymbol{u}_1) + g(\boldsymbol{u}_2) - g(\boldsymbol{u}_1 + \boldsymbol{u}_2)\Big] = 0$$

for all $\boldsymbol{u}_1, \boldsymbol{u}_2$ for which the gradient ∇ exists. Thus

$$(g_1(\boldsymbol{x}_1), \ldots, g_p(\boldsymbol{x}_1)) + (g_1(\boldsymbol{x}_2), \ldots, g_p(\boldsymbol{x}_2)) = \Big(g_1(\boldsymbol{x}_1 + \boldsymbol{x}_2), \ldots, g_p(\boldsymbol{x}_1 + \boldsymbol{x}_2)\Big), \tag{2.5}$$

where $g_j(\boldsymbol{z}) \equiv \partial g(\boldsymbol{z})/\partial z_j$. But (2.5) is the well-known Cauchy equation, from which we conclude that

$$g_j(\boldsymbol{z}) = a_{j1}z_1 + \cdots + a_{jp}z_p, \quad j = 1, \ldots, p. \tag{2.6}$$

(See, e.g., Aczél (1966, Chapters 2 and 8).)

We assert that all solutions of the set of partial differential equations (2.6) have the form

$$g(\boldsymbol{z}) = -(z_1, \ldots, z_p)B(z_1, \ldots, z_p)' + \text{constant}, \tag{2.7}$$

where B is a positive semi-definite matrix.

From (2.6) note that $g_{jk}(\boldsymbol{z}) = \frac{\partial^2 g(\boldsymbol{z})}{\partial z_j \partial z_k} = a_{jk}$, so that $a_{jk} = a_{kj}$, $j \neq k$. The proof is by induction. The case $p = 1$ was resolved by Teicher (1961). Assume (2.7) holds for $p - 1$. First note that for some function q

$$g(\boldsymbol{z}) = \int g_1(\boldsymbol{z})dz_1 = \frac{1}{2}a_{11}z_1^2 + z_1(a_{12}z_2 + \cdots + a_{1p}z_p) + q(z_2, \ldots, z_p). \tag{2.8}$$

Differentiation with respect to z_j, $j \neq 1$, yields

$$g_j(\boldsymbol{z}) = a_{1j}z_1 + q_j(z_2, \ldots, z_p), \quad j = 2, \ldots, p, \tag{2.9}$$

which, by (2.6) is equal to

$$g_j(\boldsymbol{z}) = a_{j1}z_1 + \cdots + a_{jp}z_p, \quad j = 2, \ldots, p. \tag{2.10}$$

Equating (2.9) and (2.10) yields the differential equation

$$q_j(z_2, \ldots, z_p) = a_{j2}z_2 + \cdots + a_{jp}z_p, \quad j = 2, \ldots, p,$$

which is (2.6) in $p - 1$ variables. By the induction hypothesis $q(z_2, \ldots, z_p)$ is a quadratic form plus a constant, so that from (2.8) we obtain (2.7).

Thus

$$g(\boldsymbol{z}) = h(\boldsymbol{z}) + h(-\boldsymbol{z}) = -\boldsymbol{z}B\boldsymbol{z}' + \text{constant} \tag{2.11}$$

which implies that

$$h(\boldsymbol{z}) = -\frac{1}{2}\boldsymbol{z}B\boldsymbol{z}' + b(\boldsymbol{z}) + \text{constant}, \tag{2.12}$$

where $b(\boldsymbol{z}) = -b(-\boldsymbol{z})$ is an odd function.

Assume without loss of generality that B is symmetric. To show that B is positive semi-definite, recall that (2.3) and (2.4) imply

$$g(\boldsymbol{z}) + g(-\boldsymbol{z}) \geq g(\boldsymbol{z} - \boldsymbol{\theta}) + g(-\boldsymbol{z} - \boldsymbol{\theta}),$$

which, from (2.11) yields

$$\boldsymbol{\theta}B\boldsymbol{\theta}' \geq 0. \text{ for all } \boldsymbol{\theta}. \tag{2.13}$$

To show that $b(\boldsymbol{z}) = 0$, substitute (2.12) in (2.3) to yield

$$b(\boldsymbol{z} - \boldsymbol{\theta}) - b(\boldsymbol{z} + \boldsymbol{\theta}) \leq \boldsymbol{\theta}B\boldsymbol{\theta}',$$

which for $\boldsymbol{\theta}$ replaced by $-\boldsymbol{\theta}$ yields

$$b(\boldsymbol{z} + \boldsymbol{\theta}) - b(\boldsymbol{z} - \boldsymbol{\theta}) \leq \boldsymbol{\theta}B\boldsymbol{\theta}',$$

so that

$$|b(\boldsymbol{z} + \boldsymbol{\theta}) - b(\boldsymbol{z} - \boldsymbol{\theta})| \leq \boldsymbol{\theta}B\boldsymbol{\theta}' \text{ for all } \boldsymbol{z}, \boldsymbol{\theta}. \tag{2.14}$$

With $\boldsymbol{\theta} = (0, \ldots, 0, \theta_j, 0, \ldots, 0)$, (2.14) becomes

$$\left|\frac{b(z_1, \ldots, z_j + \theta_j, \ldots, z_p) - b(z_1, \ldots, z_j - \theta_j, \ldots, z_p)}{2\theta_j}\right| \leq \frac{1}{2}b_{jj}\theta_j. \tag{2.15}$$

Taking the limit in (2.15) as $\theta_j \to 0$, $j = 1, \ldots, p$, yields $\dot{b}(\boldsymbol{z}) = 0$, which implies that $b(\boldsymbol{z})$ is a constant. Because $b(\boldsymbol{z})$ is an odd function, the constant must be zero.

In summary, from (2.12) we obtain that

$$f(\boldsymbol{z}) = \text{constant} \exp(-\frac{1}{2}\boldsymbol{z}B\boldsymbol{z}'), \quad -\infty < z_j < \infty, \ i = 1, \ldots, p,$$

where B is a positive semi-definite matrix. For B positive definite, the constant $(2\pi)^{-\frac{p}{2}}|B|^{\frac{1}{2}}$, is determined so that f is a probability density function.

3. Characterization of a Multivariate Laplace Distribution by the Sample Median as a MLE for Location

The following univariate result was obtained by Kagan, Linnik and Rao (1973).

Theorem 3.1. *Let $\{F(x-\theta), \theta \in R^1\}$ be a translation family of distributions, and suppose that F has a density f lower semicontinuous at $x = 0$. If the sample median is a MLE of θ for samples of size $n = 4$, then for some $a > 0$,*

$$f(x) = \frac{1}{2} a \exp(-a|x|), \quad -\infty < x < \infty.$$

Rao and Ghosh (1971) show by a conterexample that the theorem is false if the sample size is reduced to 2 or 3.

The proof of the multivariate version given below follows the univariate proof, but a differential equation generated in the univariate proof becomes a set of partial differential equations in p variables. It is interesting to note that the condition $n = 4$ does not change in the multivariate case.

Theorem 3.2. *Let $\{F(\boldsymbol{x}-\boldsymbol{\theta}), \boldsymbol{\theta} \in R^p\}$ be a translation family of p-dimensional distributions, and suppose that F has a density f lower semicontinuous at $\boldsymbol{x} = \boldsymbol{0}$. If the vector of sample medians is a MLE of $\boldsymbol{\theta}$ for samples of size $n = 4$, then for some $\boldsymbol{a} > \boldsymbol{0}$,*

$$f(\boldsymbol{x}) = \Big(\prod_1^p a_i/2\Big) \exp\Big(-\sum_1^p a_i|x_i|\Big), \quad \boldsymbol{x} \in R^p.$$

Proof. The proof for general p requires some cumbersome notation, and can be made more transparent if given in detail for $p = 2$. We adopt this course, and then describe the modifications required for the general case.

Write $\boldsymbol{\theta} = (\mu, \nu)$, so that by hypothesis,

$$\prod_{i=1}^n f(x_i, y_i) \geq \prod_{i=1}^n f(x_i - \mu, y_i - \nu) \tag{3.1}$$

for all (x_i, y_i) with median $\boldsymbol{0}$ and all μ, ν. Inequality (3.1) with $n = 4$ and points $(-x, -y)$, $(-x, -y)$, (x, y), (x, y) is equivalent to (2.1) with $n = 1$, from which it followed that $f(\boldsymbol{0}) > 0$ and f is everywhere finite.

Suppose that $n = 4$ and the ordered x_i (ordered y_i) are respectively

$$-x_1 \leq -x \leq 0 \leq x \leq x_2 \;\; (-y_1 \leq -y \leq 0 \leq y \leq y_2). \tag{3.2}$$

These observations can be paired in a number of ways to generate a sample, but for purposes of this proof, only two possible samples need be considered:

$$(-x, -\varepsilon y_1), (-x, -\varepsilon y), (x, \varepsilon y), (x_2, \varepsilon y_2), \;\; \varepsilon = \pm 1.$$

Then (3.1) becomes

$$\begin{aligned} & f(-x_1,-\varepsilon y_1)f(-x,-\varepsilon y)f(x,\varepsilon y)f(x_2,\varepsilon y_2) \\ \geq \;& f(-x_1-\mu,-\varepsilon y_1-\nu)f(-x-\mu,-\varepsilon y-\nu)f(x-\mu,\varepsilon y-\nu)f(x_2-\mu,\varepsilon y_2-\nu) \end{aligned} \tag{3.3}$$

whenever (3.2) holds.

In (3.3), interchange (x_1, y_1) and (x_2, y_2), (μ, ν) and $(-\mu, -\nu)$ to obtain

$$\begin{aligned} & f(x_1,\varepsilon y_1)f(-x,-\varepsilon y)f(x,\varepsilon y)f(-x_2,-\varepsilon y_2) \\ \geq \;& f(x_1+\mu,\varepsilon y_1+\nu)f(-x+\mu,-\varepsilon y+\nu)f(x+\mu,\varepsilon y+\nu)f(-x_2+\mu,-\varepsilon y_2+\nu), \end{aligned} \tag{3.4}$$

With $g(x,y) = f(x,y)f(-x,-y)$, multiplication of (3.3) and (3.4) yields

$$\begin{aligned} & g(x_1,\varepsilon y_1)g^2(x,\varepsilon y)g(x_2,\varepsilon y_2) \\ \geq \;& g(x_1+\mu,\varepsilon y_1+\nu)g(x+\mu,\varepsilon y+\nu)g(-x+\mu,-\varepsilon y+\nu)g(-x_2+\mu,-\varepsilon y_2+\nu). \end{aligned} \tag{3.5}$$

With $h = \log g$, (3.5) can be rewritten as

$$\begin{aligned} & h(x_1,\varepsilon y_1) + 2h(x,\varepsilon y) + h(x_2,\varepsilon y_2) \\ \geq \;& h(x_1+\mu,\varepsilon y_1+\nu) + h(x+\mu,\varepsilon y+\nu) + h(-x+\mu,-\varepsilon y+\nu) \\ & \quad + h(-x_2+\mu,-\varepsilon y_2+\nu). \end{aligned} \tag{3.6}$$

With $x_1 = x$, $y_1 = y$ in (3.6), if follows that

$$\begin{aligned} 2h(x,\varepsilon y) &\geq h(x+\mu,\varepsilon y+\nu) + h(-x+\mu,-\varepsilon y+\nu) \\ &= h(x+\mu,\varepsilon y+\nu) + h(x-\mu,\varepsilon y-\nu) \end{aligned} \tag{3.7}$$

so that h is midpoint (Jensen) concave.

Since f is lower semicontinuous at $\mathbf{0}$ and $f(\mathbf{0}) > \mathbf{0}$ for some $\delta > 0$ it follows that $f(x,y) > \delta$ whenever the norm $\|(x,y)\|$ is sufficiently small. Suppose $g(u,v) = 0$ for some $(u,v) \neq \mathbf{0}$, and let $c = \inf\{\|(u,v)\| : g(u,v) = 0\}$. Then $c > 0$. It is possible to choose (x_1,y_1), (x,y) and (x_2,y_2) and (μ,ν) with sufficiently small norms such that $g(x_1,\varepsilon y_1) = 0$, $\|(x_1+\mu,\varepsilon y_1+\nu)\| < c$, $\|(x+\mu,\varepsilon y+\nu)\| < c$, $\|(x-\mu,\varepsilon y-v)\| < c$, $\|(x_2+\mu,\varepsilon y_2-\nu)\| < c$. For such a choice, (3.6) fails to hold for g in place of h. This contradiction leads to the conclusion that $g(x,y) > 0$ for all x, y and it follows from (3.6) that h is continuous and convex. (See, e.g., Roberts and Varberg (1973).)

According to (3.6), the right-hand side of (3.6) as a function at μ and ν reaches a maximum at $\mu = \nu = 0$, and differentiation with respect to these variables yields

$$h_j(x_1,\varepsilon y_1) + h_j(x,\varepsilon y) + h_j(-x,-\varepsilon y) + h_j(-x_2,-\varepsilon y_2) = 0, \quad j = 1,2. \tag{3.8}$$

Set $x_1 = x_2 = x$, $y_1 = y_2 = y$ in (3.8); it follows that

$$h_j(x, \varepsilon y) + h_j(-x, -\varepsilon y) = 0, \quad j = 1, 2, \quad x, y > 0. \tag{3.9}$$

Combined with (3.8), (3.9) implies

$$h_j(x_1, \varepsilon y_1) + h_j(-x_2, -\varepsilon y_2) = 0, \quad j = 1, 2, \quad x, y > 0. \tag{3.10}$$

A consequence of (3.10) is that $h_j(x, \varepsilon y)$ is a constant for $x, y > 0$, and continuity yields this for $x, y \geq 0$. For some functions q_1 and q_2 integration yields

$$h(x, \varepsilon y) = b_1 - a_1 x + q_1(\varepsilon y) = b_2 - a_2 \varepsilon y + q_2(x),$$

which with $y = 0$ yields

$$b_1 - a_1 x + q_1(0) = b_2 + q_2(x).$$

Thus

$$h(x, \varepsilon y) = b - a_1 x - b_1 \varepsilon y, \quad x, y \geq 0. \tag{3.11}$$

In (3.3), set $x_1 = x$, $y_1 = y$ and take logarithms to obtain

$$\begin{aligned} h(x,y) &\geq \log f(x-\mu, \varepsilon y - \nu) + \log f(-x-\mu, -\varepsilon y - \nu) \qquad (3.12)\\ &= h(x+\mu, \varepsilon y + \nu) + \log f(x-\mu, \varepsilon y - \nu) - \log f(x+\mu, \varepsilon y - \nu), \quad x, y \geq 0. \end{aligned}$$

With $x > 0$, $y > 0$ and μ, ν chosen so that $x - |\mu| > 0$, $\varepsilon y - |\nu| > 0$, (3.12) becomes

$$\begin{aligned} &\log f(x-\mu, \varepsilon y - \nu) - \log f(x+\mu, \varepsilon y + \nu)\\ \leq\ & -(a_1 x + a_2 \varepsilon y) + [a_1(x+\mu) + b_1(\varepsilon y + \nu)]\\ =\ & a_1 \mu + b_1 \nu. \end{aligned} \tag{3.13}$$

In (3.13) replace (μ, ν) by $-(\mu, \nu)$; this reverses inequality (3.13) and shows that (3.13) holds with equality. Taking derivatives in (3.13) with respect to x and y yields

$$\frac{\partial}{\partial x} \log f(x, \varepsilon y) = -a_1, \quad \frac{\partial}{\partial y} \log f(x, \varepsilon y) = -b_1. \tag{3.14}$$

Consequently,

$$\log f(x, \varepsilon y) = -a_1 x + t_1(\varepsilon y) = -b_1 y + t_2(x\varepsilon).$$

Set $y = 0$ to conclude that $t_2(x) = -a_1 x + t_1(0)$, so that

$$f(x, \varepsilon y) = c \exp(-a_1 x - b_1 \varepsilon y), \quad x, y \geq 0, \tag{3.15}$$

where the constant c does not depend upon the choice $\varepsilon = 1$ or $\varepsilon = -1$ because of the lower semi-continuity of f at zero.

The starting point for the above development was (3.3), which does not change if (x, y) and $(-x, -y)$, $(-x_1, y_1)$ and $(x_1, -y_1)$ are interchanged. Thus, (3.14) again holds. Because of continuity the constant c is unchanged.

Because f is everywhere a product of its marginals, and because the marginals satisfy the conditions of Theorem 3.1, the proof for $p = 2$ is complete.

For general p, it is notationally convenient to replace (x_1, y_1) by $\boldsymbol{x}^{(1)}$, (x, y) by $\boldsymbol{x}$ and (x_2, y_2) by $\boldsymbol{x}^{(2)}$. For p-dimensional vectors $\boldsymbol{z} = (z_1, \ldots, z_p)$ and $\boldsymbol{a} = (a_1, \ldots, a_p)$, write $\boldsymbol{a} \circ \boldsymbol{z} = (a_1 z_1, \ldots, a_p z_p)$. Then (3.2) becomes

$$-x_i^{(1)} \le -x_i \le 0 \le x_i \le x_i^{(2)}, \quad i = 1, \ldots, p$$

and (3.3) becomes

$$\begin{aligned} & f\Big(-\varepsilon \circ \boldsymbol{x}^{(1)}\Big) f(-\varepsilon \circ \boldsymbol{x}) f(\varepsilon \circ \boldsymbol{x}) f\Big(\varepsilon \circ \boldsymbol{x}^{(2)}\Big) \\ \le\ & f\Big(-\varepsilon \circ \boldsymbol{x}^{(1)} - \boldsymbol{\theta}\Big) f\Big(-\varepsilon \circ \boldsymbol{x}^{(1)} - \boldsymbol{\theta}\Big) f(\varepsilon \circ \boldsymbol{x} - \boldsymbol{\theta}) f\Big(\varepsilon \circ \boldsymbol{x}^{(2)} - \boldsymbol{\theta}\Big). \end{aligned}$$

Then the proof goes through in essentially the same way as with $p = 2$.

4. Characterization of the Multivariate Normal Distribution by the Sample Covariance Matrix as a MLE of the Population Covariance Matrix

Theorem 4.1. *Let $\{\phi : \phi(\boldsymbol{x}; \Lambda) = c(\Lambda) f(\boldsymbol{x}\Lambda\boldsymbol{x}'), \boldsymbol{x} \in R^p\}$ be a family of density functions where the parameter Λ is a positive definite matrix, c is a normalizing constant depending on Λ, and f is a function continuous on $(-\infty, \infty)$ such that $f(z) > 0$ for $z \in (0, \delta)$ for some $\delta > 0$, and*

$$\lim_{z \to 0} \frac{f(z\lambda)}{f(z)} = a(\lambda), \quad \text{for all } \lambda > 0. \tag{4.1}$$

If, for all samples $\boldsymbol{x}_\alpha = (x_{1\alpha}, \ldots, x_{p\alpha})$, $\alpha = 1, \ldots, n$, of size n, a MLE of $\Sigma = \Lambda^{-1}$ is $\hat{\Sigma} = \sum_1^n \boldsymbol{x}_\alpha' \boldsymbol{x}_\alpha / n$, then

$$\phi(\boldsymbol{x}; \Lambda) = c(\Lambda) \exp\Big(-\frac{1}{2}\boldsymbol{x}\Lambda\boldsymbol{x}'\Big)$$

is a normal density function with expectation 0.

Proof. Suppose $n > p$. Then $\hat{\Sigma}$ is nonsingular with probability 1. (See Das Gupta (1971), Eaton and Perlman (1973).)

Let $\hat{\Lambda} = \hat{\Sigma}^{-1}$. By hypothesis

$$|\hat{\Lambda}|^{\frac{1}{2}} \prod_1^n f\left(\boldsymbol{x}_i \hat{\Lambda} \boldsymbol{x}_i'\right) \geq |\Lambda|^{\frac{1}{2}} \prod_1^n f\left(\boldsymbol{x}_i \Lambda \boldsymbol{x}_i'\right). \tag{4.2}$$

Let $\boldsymbol{y}_i = \boldsymbol{x}_i \hat{\Lambda}^{\frac{1}{2}}$, $\Psi = \hat{\Lambda}^{-\frac{1}{2}} \Lambda \hat{\Lambda}^{-\frac{1}{2}}$, so that

$$\prod_1^n f(\boldsymbol{y}_i \boldsymbol{y}_i') \geq |\Psi|^{\frac{1}{2}} \prod_1^n f(\boldsymbol{y}_i \Psi \boldsymbol{y}_i'), \sum \boldsymbol{y}_i' \boldsymbol{y}_i / n = I_p, \tag{4.3}$$

for all Ψ positive definite, where I_p is the identity matrix of order p.

Let $\Psi = \lambda I_p$, $\lambda > 0$, $\boldsymbol{y}_i \boldsymbol{y}_i' = z_i$. Then $\sum z_i = \operatorname{tr} \sum \boldsymbol{y}_i' \boldsymbol{y}_i = pn$, and (4.3) becomes

$$\prod_1^n f(z_i) \geq \lambda^{pn/2} \prod_1^n f(\lambda z_i). \tag{4.4}$$

With $g(z) = f^{2/p}(z)$, (4.4) becomes

$$\prod_1^n g(z_i) \geq \lambda^n \prod_1^n g(\lambda z_i), \quad \sum_1^n z_i = pn. \tag{4.5}$$

But this is exactly inequality (5.2) with p and r replace by 1 and p, which as we will see in Section 5 implies

$$g(z) = b_1 x^{A'(1)} \exp\Big(- [A'(1) + 1] z/p \Big),$$

where $A(\lambda) = a^{2/p}(\lambda)$. Consequently

$$f(z) = b z^{A'(1)p/2} \exp\Big(- [A'(1) + 1] z/2 \Big).$$

As in Section 2, $A(\lambda) = \lambda^t$ for some $t \geq 0$, and for $\sum z_i / n$ to be a MLE of λ, it is necessary that $t = 0$, in which case

$$f(z) = b \exp(-z/2).$$

5. Characterization of a Multivariate Gamma Distribution by the Mean Vector

In this section we extend the result of Teicher (1961) in two directions. One extension is from the exponential distribution to the gamma distribution; the other is from univariate to multivariate. We wish to acknowledge our debt to Teicher in that the extension borrows heavily from his proof.

Some notation will simplify many expressions. Retain the notation $\boldsymbol{a} \circ \boldsymbol{z} = (a_1 z_1, \ldots, a_p z_p)$, and write $\boldsymbol{a}^{-1} \circ \boldsymbol{z} = (\frac{z_1}{a_1}, \ldots, \frac{z_p}{a_p})$ when $a_i \neq 0$, $i = 1, \ldots, p$. By $\boldsymbol{z} > \boldsymbol{0}$ (or $\boldsymbol{z} \geq \boldsymbol{0}$) we mean $z_i > 0 (z_i \geq 0)$, $i = 1, \ldots, p$. Denote the vectors $\boldsymbol{e} = (1, \ldots, 1)$, $\boldsymbol{z}^{(j)} = (z_1, \ldots, z_{j-1}, z_{j+1}, \ldots, z_p)$, $\boldsymbol{z}^{[j]} = (1, \ldots, 1, z_j, 1, \ldots, 1)$.

Theorem 5.1. *Let* $\{F(\frac{x_1}{\sigma_1}, \ldots, \frac{x_p}{\sigma_p}), \boldsymbol{\sigma} > \boldsymbol{0}\}$ *be a scale parameter family of distributions such that*

(i) *For some* $\boldsymbol{\delta} > \boldsymbol{0}$, $f(\boldsymbol{z}) > \boldsymbol{0}$ *for all* $\boldsymbol{z}$ *such that* $\boldsymbol{0} < \boldsymbol{z} < \boldsymbol{\delta}$, *and*

$$\lim_{\boldsymbol{z} \to 0} \frac{f(\boldsymbol{\lambda} \circ \boldsymbol{z})}{f(\boldsymbol{z})} = B(\boldsymbol{\lambda})$$

exists and is finite in some neighborhood $\lambda_0^{-1} < \lambda_j < \lambda_0$, $\lambda_0 > 1$, $j = 1, \ldots, p$.

(ii) *for all samples* $\boldsymbol{x}_\alpha = (x_{1\alpha}, \ldots, x_{p\alpha})$, $\alpha = 1, \ldots, n$ *of size* n, $\hat{\sigma}_i = \sum_\alpha^n x_{i\alpha}/r_i n$, $r_i > 0$, *is a maximum likelihood estimate of* σ_i, $i = 1, \ldots, p$.

Then f *is the density of a multivariate gamma distribution, where*

$$f(\boldsymbol{x}) = b^{-1}\Big(\prod_1^p x_i^{r_i - 1}\Big) \exp\Big(-\frac{1}{2}\sum_1^p x_i\Big).$$

Proof. Notice first that (ii) implies $F(\boldsymbol{z}) = 0$ for any $z_j < 0$, $j = 1, \ldots, p$, so it is sufficient to consider only $\boldsymbol{z} > \boldsymbol{0}$. From (ii) it follows that for all $\boldsymbol{x}_\alpha > 0$, $\sigma_i > 0$, $\alpha = 1, \ldots, n$ and $i = 1, \ldots, p$,

$$\Big(\prod_1^p \hat{\sigma}_i^{-n}\Big) \prod_1^n f\Big(\frac{x_{1\alpha}}{\hat{\sigma}_1}, \ldots, \frac{x_{p\alpha}}{\hat{\sigma}_p}\Big) \geq \Big(\prod_1^p \sigma_i^{-n}\Big) \prod_1^n f\Big(\frac{x_{1\alpha}}{\hat{\sigma}_1}, \ldots, \frac{x_{p\alpha}}{\hat{\sigma}_p}\Big). \tag{5.1}$$

Let $y_{i\alpha} = x_{i\alpha}/\hat{\sigma}_i$, $\boldsymbol{y}_\alpha = (y_{1\alpha}, \ldots, y_{p\alpha})$, $\lambda_i = \hat{\sigma}_i/\sigma_i$, $\sigma_i > 0$, $\alpha = 1, \ldots, n$ and $i = 1, \ldots, p$, so that (5.1) becomes

$$\prod_1^n f(\boldsymbol{y}_\alpha) \geq \Big(\prod_1^p \lambda_i^n\Big) \prod_1^n f(\boldsymbol{\lambda} \circ \boldsymbol{y}_\alpha), \tag{5.2}$$

for all $\lambda_i > 0$, $y_{i\alpha} > 0$, $\sum_{\alpha=1}^n y_{i\alpha} = r_i n$, $i = 1, \ldots, p$ and $\alpha = 1, \ldots, n$. Let $y_{i\alpha} = k_i/m$, $\alpha = 1, \ldots, m$; $y_{i\alpha} = (r_i n - k_i)/(n - m)$, $\alpha = m + 1, \ldots, n$, where $0 < k_i < r_i n$, $m < n$. Then $\sum_{\alpha=1}^n y_{i\alpha} = r_i n$ and (5.2) becomes

$$\begin{aligned} &f^m\Big(\frac{k_1}{m}, \ldots, \frac{k_p}{m}\Big) f^{n-m}\Big(\frac{r_1 n - k_1}{n - m}, \ldots, \frac{r_p n - k_p}{n - m}\Big) \\ \geq\ &\Big(\prod_1^p \lambda_i^n\Big) f^m\Big(\lambda_1 \frac{k_1}{m}, \ldots, \lambda_p \frac{k_p}{m}\Big) f^{n-m}\Big(\lambda_1 \frac{r_1 n - k_1}{n - m}, \ldots, \lambda_p \frac{r_p n - k_p}{n - m}\Big). \end{aligned} \tag{5.3}$$

Let $m \to \infty$, $n \to \infty$, $k_i \to \infty$ in such a way that $k_i/m \to a_i$, $m/n \to c$, $0 < c < 1$, where $0 < a_i < r_i/c$, $i = 1, \ldots, p$. Because f is continuous, it follows from (5.3) that

$$\begin{aligned} & f(\boldsymbol{a}) f^{\bar{c}/c}\Big(\frac{r_1 - a_1 c}{\bar{c}}, \ldots, \frac{r_p - a_p c}{\bar{c}}\Big) \\ \geq\ & \Big(\prod_1^p \lambda_i^{1/c}\Big) f(\boldsymbol{\lambda} \circ \boldsymbol{a}) f^{\bar{c}/c}\Big(\lambda_1 \frac{r_1 - a_1 c}{\bar{c}}, \ldots, \lambda_p \frac{r_p - a_p c}{\bar{c}}\Big), \end{aligned} \tag{5.4}$$

where $\bar{c} = 1 - c$.

To show that $f(\boldsymbol{z}) > \mathbf{0}$ for all $\boldsymbol{z} > \mathbf{0}$, suppose that for some $\boldsymbol{z} > \mathbf{0}$, $f(\boldsymbol{z}) = \mathbf{0}$. Choose $\bar{c} \in (0, \min_j r_j/z_j)$ and let $a_j = (r_j - \bar{c} z_j)/c$; because of the constraints on $\bar{c}$, it must be that $0 < a_j < r_j/c$, and moreover, $z_j = (r_j - a_j c)/\bar{c}$, $j = 1, \ldots, p$. It follows by assumption that the left hand side of (5.4) is zero. Now take λ_j sufficiently small so that $\lambda_j a_j < \delta_j$, $\lambda_j z_j < \delta_j$, $j = 1, \ldots, p$. By (i), the right hand side of (5.4) is positive, a contradiction. Hence $f(\boldsymbol{z}) > \mathbf{0}$ for all $\boldsymbol{z} > \mathbf{0}$.

Consequently, from (5.4) it follows that

$$\begin{aligned} & f^{\bar{c}/c}\Big(\frac{r_1 - a_1 c}{\bar{c}}, \ldots, \frac{r_p - a_p c}{\bar{c}}\Big) \\ \geq\ & \Big(\prod_1^p \lambda_i^{1/c}\Big) \frac{f(\boldsymbol{\lambda} \circ \boldsymbol{a})}{f(\boldsymbol{a})} f^{\bar{c}/c}\Big(\lambda_1 \frac{r_1 - a_1 c}{\bar{c}}, \ldots, \lambda_p \frac{r_p - a_p c}{\bar{c}}\Big), \end{aligned}$$

which, together with the continuity of f and (i) yields

$$f^{\bar{c}/c}\Big(\frac{r_1}{\bar{c}}, \ldots, \frac{r_p}{\bar{c}}\Big) \geq \Big(\prod_1^p \lambda_i^{1/c}\Big) B(\boldsymbol{\lambda}) f^{\bar{c}/c}\Big(\lambda_1 \frac{r_1}{\bar{c}}, \ldots, \lambda_p \frac{r_p}{\bar{c}}\Big), \quad \text{for all } \boldsymbol{\lambda} > \mathbf{0}. \tag{5.5}$$

Let $r_i/\bar{c} = z_i$ (so that $\boldsymbol{z} > \boldsymbol{r}$) and rewrite (5.5) as

$$f(\boldsymbol{z}) \geq \Big(\prod_1^p \lambda_i^{1/\bar{c}}\Big) [B(\boldsymbol{\lambda})]^{c/\bar{c}} f(\boldsymbol{\lambda} \circ \boldsymbol{z}), \quad \boldsymbol{z} \geq \boldsymbol{r}, \quad \boldsymbol{\lambda} > \mathbf{0}, \tag{5.6}$$

where the validity of (5.6) for $z_i = r_i$ $(i = 1, \ldots, p)$ is a consequence of continuity.

From the definition of $B(\boldsymbol{\lambda})$ in (i) it follows that

$$B(\boldsymbol{\lambda}) B\big(\boldsymbol{\lambda}^{-1}\big) = 1. \tag{5.7}$$

Thus $B(\boldsymbol{\lambda}) > 0$, $\boldsymbol{\lambda} > \mathbf{0}$.

Let $h = \log f$. Because the factors in (5.6) are positive, (5.6) is equivalent to

$$h(\boldsymbol{z}) - h(\boldsymbol{\lambda} \circ \boldsymbol{z}) \geq \frac{1}{\bar{c}} \sum_1^p \log \lambda_i + \frac{c}{\bar{c}} \log B(\boldsymbol{\lambda}), \quad \boldsymbol{z} \geq \boldsymbol{r}, \quad \boldsymbol{\lambda} > \mathbf{0}. \tag{5.8}$$

With $\boldsymbol{\lambda}$ replaced by $\boldsymbol{\lambda}^{-1}$ together with (5.7), it follows that

$$h(\boldsymbol{z}) - h\left(\boldsymbol{\lambda}^{-1} \circ \boldsymbol{z}\right) \geq -\frac{1}{\bar{c}} \sum_{1}^{p} \log \lambda_i - \frac{c}{\bar{c}} \log B(\boldsymbol{\lambda}). \tag{5.9}$$

The addition of (5.8) and (5.9) yields

$$h(\boldsymbol{\lambda} \circ \boldsymbol{z}) + h\left(\boldsymbol{\lambda}^{-1} \circ \boldsymbol{z}\right) \leq 2h(\boldsymbol{z}), \quad \boldsymbol{z} \geq \boldsymbol{r}, \quad \boldsymbol{\lambda} > \boldsymbol{0}. \tag{5.10}$$

Let $\lambda_j = e^{b_j}$, $z_j = e^{y_j}, j = 1, \ldots, p$, and $H(\boldsymbol{y}) = h(e^{y_1}, \ldots, e^{y_p})$. Then (5.10) becomes $H(\boldsymbol{b} + \boldsymbol{y}) + H(-\boldsymbol{b} + \boldsymbol{y}) \leq 2H(\boldsymbol{y})$, which together with the continuity of f says that $H(\boldsymbol{y})$ is concave in $\boldsymbol{y} > (\log r_1, \ldots, \log r_p)$. Consequently, h is differentiable in $\times_{j=1}^{p}(r_j, \infty)$, except possibly for a countable subset $\mathcal{D}$ of $\times_{j=1}^{p}(r_j, \infty)$.

If $\lambda_j < 1$, then from (5.8)

$$\frac{h(\boldsymbol{z}) - h(\boldsymbol{\lambda} \circ \boldsymbol{z})}{z_j(1 - \lambda_j)} \geq \frac{\sum_1^p \log \lambda_i}{\bar{c} z_j (1 - \lambda_j)} + \frac{c \log B(\boldsymbol{\lambda})}{\bar{c} z_j (1 - \lambda_j)}. \tag{5.11}$$

From (5.7) $B(\boldsymbol{e}) = 1$, so that with $\lambda_i = 1$, $i \neq j$,

$$h_j(\boldsymbol{z}) \geq -\frac{1}{\bar{c} z_j} - \frac{c}{\bar{c} z_j} \lim_{\lambda_j \uparrow 1} \frac{\log B(\boldsymbol{e}) - \log B(\boldsymbol{\lambda}^{[j]})}{1 - \lambda_j}. \tag{5.12}$$

From (5.8) with λ^{-1} in place of λ, it follows that

$$h(\boldsymbol{z}) - h\left(\boldsymbol{\lambda}^{-1} \circ \boldsymbol{z}\right) \geq -\frac{\sum_1^p \log \lambda_i}{\bar{c}} + \frac{c}{\bar{c}} \log B\left(\boldsymbol{\lambda}^{-1}\right), \quad \boldsymbol{z} \geq \boldsymbol{r}, \quad \boldsymbol{\lambda} > \boldsymbol{0};$$

and with $z_i/\lambda_i = t_i$, $i = 1, \ldots, p$,

$$h(\boldsymbol{\lambda} \circ \boldsymbol{t}) - h(\boldsymbol{t}) \geq -\frac{\sum_1^p \log \lambda_i}{\bar{c}} - \frac{c}{\bar{c}} \log B(\boldsymbol{\lambda}), \quad \boldsymbol{t} \geq \boldsymbol{\lambda}^{-1} \circ \boldsymbol{r}, \quad \boldsymbol{\lambda} > \boldsymbol{0};$$

so that if $\lambda_j < 1$,

$$\frac{h(\boldsymbol{z}) - h(\boldsymbol{\lambda} \circ \boldsymbol{z})}{z_j(1 - \lambda_j)} \leq \frac{\sum_1^p \log \lambda_j}{\bar{c} z_j (1 - \lambda_j)} + \frac{c \log B(\boldsymbol{\lambda})}{\bar{c} z_j (1 - \lambda_j)}, \quad \boldsymbol{z} \geq \boldsymbol{\lambda}^{-1} \circ \boldsymbol{r}, \quad \boldsymbol{\lambda} > \boldsymbol{0}$$

which is the reversal of (5.9), and leads to a reversal of (5.10).

Because B is differentiable at $\boldsymbol{e}$, we obtain

$$h_j(\boldsymbol{z}) = -\frac{1}{\bar{c} z_j} - \frac{c B_j(\boldsymbol{\lambda}^{[j]})}{\bar{c} z_j}. \tag{5.13}$$

Recall that $r_j/\bar{c} = z_j$, so that (5.13) becomes

$$\begin{aligned} h_j(z) &= -\frac{1}{r_j} - \frac{(1 - r_j/z_j)}{r_j} B_j\left(\boldsymbol{\lambda}^{[j]}\right) \\ &= -\frac{(1 + B_j(\boldsymbol{\lambda}^{[j]}))}{r_j} + \frac{B_j(\boldsymbol{\lambda}^{[j]})}{z_j}, \quad j = 1, \ldots, p. \end{aligned} \tag{5.14}$$

With the continuity of h, it follows that

$$h(z) = -\frac{(1 + B_j(\boldsymbol{\lambda}^{[j]}))}{r_j} z_j + B_j\left(\boldsymbol{\lambda}^{[j]}\right) \log z_j + q\left(z^{(j)}\right), \quad j = 1, \ldots, p. \tag{5.15}$$

Because (5.15) holds for each $j = 1, \cdots, p$, it follows (e.g., by an iterative argument) that

$$h(z) = -\sum_{j=1}^{p} \frac{[1 + B_j(\boldsymbol{\lambda}^{[j]})] z_j}{r_j} + \sum_{j=1}^{p} B_j\left(\boldsymbol{\lambda}^{[j]}\right) \log z_j + \text{constant},$$

whereby

$$f(z) = b \prod_{1}^{p} z_j^{a_j - 1} \exp\left(-\sum_{1}^{p} \xi_j z_j\right), \tag{5.16}$$

where $a_j = 1 + B_j(\boldsymbol{\lambda}^{[j]})$ and $\xi_j = a_j/r_j$.

A direct computation of $\lim_{z \to 0}[f(\boldsymbol{\lambda} \circ z)/f(z)]$ shows that $B(\boldsymbol{\lambda}) = \prod_1^p \lambda_j^{a_j - 1}$, so that $B(\boldsymbol{\lambda}^{[j]}) = \lambda_j^{a_j - 1}$. But in order for $\sum_{\alpha=1}^{n} x_{i\alpha}/r_i n$ to be a maximum likelihood estimate of σ_i it is necessary that $a_i = r_i$, in which case

$$f(z) = b^{-1} \prod_{1}^{p} z_j^{r_j - 1} \exp\left(-\sum_{1}^{p} \xi_j z_j\right),$$

where $b = \prod_1^p [\xi_j \Gamma(r_j)]$. But note that $\hat{\sigma}$ is a MLE for this distribution only if $\xi_j = 1$, $j = 1, \ldots, n$.

The following theorem is a minor variant of Theorem 5.1.

Theorem 5.2. *Suppose that all of the conditions of Theorem 5.1 are satisfied except that $\hat{\sigma}_i = (\sum_{\alpha=1}^{n} x_{i\alpha}^2/r_i n)^{1/2}$, where r_i is an odd integer, $i = 1, 2, \ldots, p$. Then the density of F has the form*

$$f(\boldsymbol{x}) = (\text{constant}) \prod_{i=1}^{p} x_i^{r_i - 1} \exp\left(-\frac{1}{2} x_i^2\right).$$

The proof of Theorem 5.2 mimics that of Theorem 5.1 and is omitted.

With $r_i = 1$, $i = 1, \ldots, p$ in Theorem 5.2, f is just a product of standard univariate normal densities. In case $p = 1$, this is the result of Teicher (1961) described under II.a in the Introduction.

References

Aczél, J. (1966). *Lectures on Functional Equations and their Application.* Academic Press, New York.

Das Gupta, S. (1971). Nonsingularity of the sample covariance matrix. *Sankhyā Ser.A* **33**, 475–478.

Eaton, M. L. and Perlman, M. D. (1973). The non-singularity of generalized sample covariance matrices. *Ann. Statist.* **1**, 710–717.

Johnson, Norman L. and Kotz, Samuel (1970). *Continuous univariate distributions -2.* Houghton-Mifflin, Boston.

Kagan, A. M., Linnik, Y. V. and Rao, C. R. (1973). *Characterization Problems in Mathematical Statistics.* John Wiley, New York.

Rao, C. R. and Ghosh, J. K. (1971). A note on some translation-parameter families of densities of which the median is the M. L. E. *Sankhyā Ser.A* **33**, 91–92.

Roberts, A. Wayne and Varberg, Dale E. (1973). *Convex Functions.* Academic Press, New York.

Teicher, H. (1961). Maximum likelihood characterization of distributions. *Ann. Math. Statist.* **32**, 1214–1222.

University of British Columbia and Western Washington University, U.S.A.
Department of Statistics, Stanford University, Sequoia Hall, Stanford, CA 94305, U.S.A.

(Received February 1992; accepted August 1992)