The Annals of Statistics
1973, Vol. 1, No. 5, 948–957

ON THE ASYMPTOTIC NORMALITY OF THE MAXIMUM-LIKELIHOOD ESTIMATE WHEN SAMPLING FROM A STABLE DISTRIBUTION[1]

BY WILLIAM H. DUMOUCHEL[2]

University of California, Berkeley

The large-sample distributions of the maximum-likelihood estimates for the index, skewness, scale, and location parameters (respectively α, β, c, and δ) of a stable distribution are studied. It is shown that if both α and δ are unknown, then the likelihood function L will have no maximum within $0 < \alpha \leqq 2$, $-\infty < \delta < \infty$, but that $L(\alpha, \delta) \to \infty$ as $(\alpha, \delta) \to (0, x_k)$ where x_k is any one of the n observed sample values. However, it is shown that the centroid of L is little affected by this behavior and, if the estimate $\hat{\alpha}$ is restricted to $\hat{\alpha} \geqq \varepsilon > 0$, then the maximum-likelihood estimates are consistent and $n^{\frac{1}{2}}(\hat{\alpha} - \alpha, \hat{\beta} - \beta, \hat{c} - c, \hat{\delta} - \delta)$ has a limiting normal distribution with mean $(0, 0, 0, 0)$ and covariance matrix I^{-1}, where I is the Fisher information matrix. There are some exceptional values of α and β for which the argument presented does not hold. The argument consists in showing that the family of stable distributions satisfies conditions given in the literature and in doing so it is proven that certain asymptotic expansions for stable densities can be differentiated arbitrarily with respect to the parameters.

1. Introduction. Independent, identically distributed variables $X, X_1, X_2, \cdots$ are said to have a *stable* distribution if for every positive integer n there exist constants $a_n > 0$, b_n such that $X_1 + \cdots + X_n$ has the same distribution as $a_n X + b_n$. Standard references for the theory of stable distributions are Gnedenko and Kolmogorov (1954; Chapter VII) and Feller (1966; Chapters VI, IX, and XVII). The coefficients a_n above are necessarily of the form $a_n = n^{1/\alpha}$, $0 < \alpha \leqq 2$, and the value α is called the *characteristic exponent* or *index* of the stable distribution. A parameterization of all stable distributions in terms of their characteristic functions is well known (see the above references). It may be written as (for all $-\infty < t < \infty$)

$$
\begin{aligned}
\phi(t; \alpha, \beta, c, \delta) &= E[e^{iXt} \mid \alpha, \beta, c, \delta] \\
&= \exp\left\{-|ct|^{\alpha} \exp\left[-i\frac{\pi}{2}\beta \operatorname{sgn} t\right] + i\delta t\right\}, \qquad \alpha \neq 1; \qquad (1.1)\\
&= \exp\{-|ct| - i(2\beta/\pi)ct \log|ct| + i\delta t\}, \qquad \alpha = 1; \\
&0 < \alpha \leqq 2,\ |\beta| \leqq \min(\alpha, 2-\alpha),\ c > 0,\ -\infty < \delta < \infty.
\end{aligned}
$$

Received July 1971; revised January 1973.

[1] This investigation was supported by PHS Research Grant No. GM-10525-08, National Institute of Health, Public Health Service.

[2] Presently at Department of Statistics, University of Michigan.

AMS 1970 *subject classifications*. Primary 62E20, 62F10; Secondary 41A60, 60E05.

Key words and phrases. Stable distributions, maximum-likelihood estimate, asymptotic normality.

(When $\alpha \neq 1$, another parameterization, in which the term $\exp[-i(\pi/2)\beta \operatorname{sgn} t]$ in (1.1) is replaced by $(1 + i\beta \operatorname{sgn} t \tan(\pi\alpha/2)$, is also commonly used. Feller (1966, page 548) and DuMouchel (1971b, page 12–14) discuss relationships between these and similar parameterizations of stable characteristic functions.) Besides α, there are three parameters. δ is a location parameter and c (sometimes $\gamma = c^\alpha$ is used) is a scale parameter. The real parameter β is an index of skewness: for $\beta = 0$ the distribution is symmetric about $x = \delta$, and the distributions having skewness parameter $-\beta$ are the mirror images about δ of those with skewness parameter $+\beta$. That is if $s_{\alpha,\beta}(x)$ is the density function (all stable distributions are absolutely continuous) of a stable distribution having $c = 1$ and $\delta = 0$, then $s_{\alpha,\beta}(x) = s_{\alpha,-\beta}(-x)$.

The best-known stable distributions are the family of normal distributions corresponding to $\alpha = 2$. The nonnormal stable distributions have ordinarily been given very little attention in the field of statistical inference and indeed are surely not so important as the normal distributions. The reason for this is that the normal distributions are the only stable distributions which have a finite variance, and infinite variance does seem inappropriate in many statistical contexts. In fact, it seems to be widely felt that infinite variance is inappropriate in almost any context, because an empirical distribution with literally infinite variance seems almost a contradiction in terms. But this does not preclude the possibility of an infinite-variance distribution being involved in an inference situation. One can, for example, observe a truncated version of a variable known by theory to have such a distribution—a waiting time, for instance. Also a theoretical distribution that has infinite variance may be a good model for an empirical distribution which admittedly does not.

Mandelbrot (1960; 1962; 1963a; 1963b; 1967a; 1967b; 1969b; 1971) has pioneered in developing this idea, first in connection with certain economic phenomena and later in connection with other empirical phenomena. Take for example, the distribution of changes in stock market prices. Mandelbrot (1963b), Fama (1965), and others have shown that although these variables are certainly bounded, the probability of very large deviations is so great that many statistical techniques which depend for their validity on the asymptotic theory of finite variance distributions are inapplicable, even for rather large sample sizes. The sum of a large number of these variables is often dominated by one of the summands—a theoretical property of infinite variance distributions. In such a case, a mathematical model assuming such a distribution for the observations may be very useful.

Among infinite-variance distributions, the nonnormal stable distributions play an important role, not only because of their closure properties under convolution, but also because only a stable distribution can be the limiting distribution of sums of independent, identically distributed variables.

Under this impetus, some work in statistical inference in connection with stable distributions is beginning. The problem is greatly complicated by the absence, except for three special cases—the normal distribution, the Cauchy distribution, and the distribution of X^{-2}, where X has a (central) normal distribution—of a known closed form for the density or distribution function of stable random variables. The only "handles" available for most stable distributions are the characteristic function representation (1.1), the property of stability, and a few results on the behavior of $s_{\alpha,\beta}(x)$ as $x \to \infty$. Mandelbrot (1963b), and in more detail Fama (1965), proposed a graphical procedure for estimating the index α. Mandelbrot (1967b) proposed approximating the stable density function by a mixture of a uniform and a Pareto distribution, and then estimating α by the method of maximum likelihood. Fama and Roll (1968, 1971) proposed several easily computed estimators of α, c, and δ for the symmetric case where $\beta = 0$ is known, and measured their bias and mean squared error using computer simulations. DuMouchel (1971b) described the method of maximum likelihood for stable distributions and gave a table of the asymptotic standard deviations and correlations of the maximum likelihood estimators. This table may be used to compute the asymptotic relative efficiency of any other estimator whose asymptotic variance is known. DuMouchel (1971b) also gave tables for computing the loss of information incurred when various parts of the sample are censored, and presented a small-scale simulation experiment to compare the actual behavior of the maximum-likelihood estimator with the predictions of the asymptotic theory.

The main purpose of this paper is to demonstrate the validity of applying the well-known theory of maximum-likelihood estimation to stable distribution inference, at least if certain exceptional cases are taken into account. A secondary purpose is to show how "handles" like the characteristic function can sometimes be used to get a grip on the sampling theory of a distribution when an easy expression for the density function is not available.

Let $\theta = (\alpha, \beta, c, \delta)$ and let $\hat{\theta} = (\hat{\alpha}, \hat{\beta}, \hat{c}, \hat{\delta})$ be the maximum-likelihood estimate of θ based on $x = (x_1, \cdots, x_n)$, a sample of size n, where the likelihood function is defined by

$$L(\theta) = \prod_{k=1}^{n} s_{\alpha,\beta}\left(\frac{x_k - \delta}{c}\right)\Big/c\,. \tag{1.2}$$

2. The behavior of $L(\theta)$ when α is near 0.

PROPOSITION 1. *No matter what values $x_1, \cdots, x_n$ are observed, if both α and δ need to be estimated, $L(\alpha, \delta)$ has no maximum for $0 < \alpha \leqq 2$, $-\infty < \delta < \infty$. Instead, $L(\alpha, \delta) \to \infty$ as $(\alpha, \delta) \to (0, x_k)$; $k = 1, 2, \cdots, n$.*

PROOF. Use the series representation for $s_{\alpha,\beta}(x)$ derived by Bergström (1952) and by Feller (1966, page 549). It is absolutely convergent for $0 < \alpha < 1$, and $x > 0$:

$$s_{\alpha,\beta}(x) = \frac{-1}{\pi x} \sum_{j=1}^{\infty} \frac{\Gamma(1 + j\alpha)}{j!} (-x^{-\alpha})^j \sin\left[\frac{k\pi}{2}(\alpha + \beta)\right]. \tag{2.1}$$

Now as $\alpha \to 0$ and for fixed $x > 0$ it is shown in DuMouchel (1971a) that the right side of (2.1) is

$$s_{\alpha,\beta}(x) = \alpha(1 + \beta/\alpha)/2ex\{1 + O(\alpha)\}\,. \tag{2.2}$$

(Assume that, as $\alpha \to 0$, β varies so that the requirement $|\beta| \leqq \alpha$ of (1.1) is maintained.) For $x = 0$ and all $0 < \alpha < 1$, $s_{\alpha,\beta}(0)$ is known (see Bergström (1952)):

$$s_{\alpha,\beta}(0) = \frac{1}{\pi}\,\Gamma(1 + 1/\alpha)\sin\left[\frac{\pi}{2}(1 + \beta/\alpha)\right]. \tag{2.3}$$

Now substitute (2.2) and (2.3) into (1.2) and let $\delta \rightleftarrows x_1$, say, (for convenience, let $c = 1$):

$$L(\theta) = K\Gamma(1 + 1/\alpha)\prod_{k=2}^{n}\frac{\alpha}{|x_k - x_1|}\{1 + O(\alpha)\}\,.$$

The constant K depends only on x and β/α and is bounded away from 0 if $|\beta/\alpha|$ is bounded away from 1. If $|\beta/\alpha|$ is allowed to be arbitrarily close to 1, an adjustment of δ from $\delta = x_1$ to $\delta = x_1 - m(\alpha, \beta)$ is necessary, where $m(\alpha, \beta)$ is the mode of $s_{\alpha,\beta}(x)$. In either case, for x and $|\beta/\alpha|$ fixed, one has as $\alpha \to 0$

$$L(\theta) \propto \Gamma(1 + 1/\alpha)\alpha^{n-1}$$

which approaches ∞. If we use the Stirling approximation $\Gamma(1 + t) \sim (2\pi t)^{\frac{1}{2}}t^te^{-t}$, then $L(\theta)$ is proportional to $\alpha^{(2\alpha n-3\alpha-2)/2\alpha}e^{-1/\alpha}$ so that $\alpha^{-1} \gg n$ is necessary before $L(\theta)$ becomes large.

If δ were known and did not have to be estimated, this phenomenon would not occur, since there is 0 probability that any x-value would be exactly δ. In practice, of course, x-values are only measured with some finite precision, so that the stable likelihood (1.2) should be replaced by a multinomial model, thereby avoiding the present difficulty completely.

Although the stable likelihood is unbounded near $\alpha = 0$, estimates similar to the maximum likelihood estimate, such as the centroid of the likelihood function, or the maximum probability estimators of Weiss and Wolfowitz (1967), will be little affected by this behavior. Put another way, if the prior distribution of θ is relatively well-behaved as $\alpha \to 0$, then the posterior distribution of α will be well-behaved in the neighborhood of $\alpha = 0$. The following statement will be proved:

PROPOSITION 2. *With probability one, the quantity*

$$P(\varepsilon) = \iint_{\substack{\alpha<\varepsilon\\ -\infty<\delta<\infty}} L(\alpha, \delta)\,d\delta\,d\alpha$$

becomes proportional to ε^n *as* $\varepsilon \to 0$.

PROOF. Let $A(\alpha) = \int_{-\infty}^{+\infty}\prod_1^n s_\alpha(x_k - \delta)d\delta$. (Assume for convenience that the parameters c and β/α are fixed.) With probability 1 no two x's will be exactly identical. Suppose they are labeled so that $-\infty = x_0 < x_1 < \cdots < x_n < x_{n+1} = \infty$. Then, as $\alpha \to 0$, using (2.2) one has

$$A(\alpha) \sim \alpha^{n-1} \sum_{k=1}^{n} c_k \int_{(x_{k-1}+x_k)/2}^{(x_k+x_{k+1})/2} s_\alpha(x_k - \delta)d\delta \,, \tag{2.4}$$

where $c_k = \prod_{j \neq k} [(1 + \beta/\alpha)/2e|x_j - x_k|]$.

The coefficient of α^{n-1} in (2.4) is bounded by $\sum C_k$, and it can be shown (see DuMouchel (1971a)) that as $\alpha \to 0$ it tends to

$$C = e^{-1} \sum_1^n C_k + \tfrac{1}{2}(1 - e^{-1})[C_1(1 - \beta/\alpha) + C_n(1 + \beta/\alpha)] \,.$$

The proof is completed by noting that $P(\varepsilon) = \int_0^\varepsilon A(\alpha)d\alpha$, so that $P(\varepsilon) \sim C\varepsilon^n/n$ as $\varepsilon \to 0$.

3. The asymptotic distribution of the restricted maximum-likelihood estimate. If the maximum-likelihood estimate $\hat{\theta}$ is defined to be the value of θ at which $L(\theta)$ is a maximum, subject to the restriction $\alpha \geqq \varepsilon > 0$, ε arbitrarily small, then $\hat{\theta}$ is consistent and asymptotically normal. The proof of this statement consists of showing that the family of probability densities whose characteristic functions are given by (1.1) satisfies conditions given in the literature for the consistency and asymptotic normality of the maximum likelihood estimate. Several sets of such conditions are available, including those of Cramér (1946), LeCam (1952) and Wald (1941; 1949). Within the qualifications of the following theorem, the family of stable distributions satisfies the conditions given in each of the above references; only one such set of conditions will be verified here. The following lemma is a restatement of LeCam (1952; pages 76–78) adapted somewhat to the present case.

LEMMA. *Denote the permissible values of $\theta = (\alpha, \beta, c, \delta)$ by Θ, let C be an open subset of Θ, and assume $\bar{C} \subset \Theta$. Then, if Conditions* 1—6 *of the next section are satisfied, and if $\theta_0 \in C$ is the true value of θ*; (1) *$\hat{\theta}_n$, the maximum-likelihood estimate based on the first n observations, is consistent and asymptotically normal, and the limiting covariance matrix of $n^{\frac{1}{2}}(\hat{\theta}_n - \theta_0)$ is $I^{-1}(\theta_0)$, where I is the Fisher information matrix defined below*; *and* (2) *there exists a positive number ρ, independent θ_0, having the following property*: *with probability* 1 *the sequence $(x_1, x_2, \cdots)$ is such that, for n sufficiently large, $\hat{\theta}_n$ is the unique solution of the likelihood equation $\partial L(\hat{\theta})/\partial\theta = 0$ also satisfying $|\hat{\theta}_n - \theta_0| < \rho$.*

Using this lemma, the verification of Conditions 1—6 of the next section proves the following theorem.

THEOREM. *When sampling from a stable distribution, $\hat{\theta}_n$, the maximum likelihood estimate for $\theta = (\alpha, \beta, c, \delta)$ based on the first n observations, restricted so that $\hat{\alpha}_n$, the estimate for α, satisfies $\hat{\alpha}_n > \varepsilon$, ε arbitrarily small and positive, is consistent and asymptotically normal as long as θ_0, the true value of θ, is in the interior of the parameter space* (*that is, the cases $\alpha_0 \leqq \varepsilon$, $\alpha_0 = 2$, and $\beta_0 = \pm \min(\alpha_0, 2 - \alpha_0)$ are excluded*) *and the additional case* ($\alpha_0 = 1$, $\beta_0 \neq 0$) *is excluded.*

The permissible set of values of θ is

$$\Theta = \{(\alpha, \beta, c, \delta) : 0 < \varepsilon \leqq \alpha < 1 \quad \text{or} \quad 1 < \alpha < 2, \\ |\beta| < \min(\alpha, 2 - \alpha), 0 < c < \infty, -\infty < \delta < \infty\} .$$

Then C is an arbitrary open subset of Θ whose closure, $\bar{C}$, is also contained in Θ.

4. Statement and verification of the conditions.

CONDITION 1. For every x, $-\infty < x < \infty$, $f_\theta(x)$, defined as $s_{\alpha,\beta}((x - \delta)/c)/c$, is a continuous function of θ for every $\theta \in \Theta$, and admits continuous partial derivatives of first and second order with respect to θ for every $\theta \in \bar{C}$.

PROOF. Represent $f_\theta(x)$ using the Fourier inversion formula, as

$$f_\theta(x) = \int_{-\infty}^{\infty} e^{-ixt}\phi(t; \alpha, \beta, c, \delta)\, dt \tag{4.1}$$

where ϕ is defined by (1.1). It is readily seen that differentiations with respect to the parameters through the integral sign in (4.1) are permitted. Assumption 1 holds, then, except for the discontinuity of $f_\theta(x)$ with respect to α at $\alpha = 1$ and $\beta \neq 0$. If α must be estimated, the points θ having $\alpha = 1$, $\beta \neq 0$ must be omitted from θ for the proof to apply. See DuMouchel (1971b; pages 11–14) for a discussion of a slightly different parameterization of stable distributions in which $f_\theta(x)$ is continuous with respect to α at $\alpha = 1$. The $\alpha = 1$ case will not be considered further here.

CONDITION 2. For some fixed integer K sufficiently large, let $g_\theta(x_1, \cdots, x_K) = \prod_1^K f_\theta(x_K)$. Then for all $\theta_0 \in \bar{C}$, $E_{\theta_0}[\sup_{\theta \in \Theta - C} \log(g_\theta/g_{\theta_0})] < \infty$.

PROOF. If Θ is allowed to include points where α becomes arbitrarily close to 0, this condition is violated, as was seen in Section 2. If α is restricted away from 0, then $f_\theta(x) = s_{\alpha,\beta}((x - \delta)/c)/c$ is bounded uniformly for all x, α, β, c, and δ except for $x \to \delta$ and either $c \to 0$ or $(\alpha, \beta) \to (1,1)$. In these cases the distribution associated with θ tends toward a 1-point distribution concentrated at δ. If $x \neq \delta$, $c \to 0$, by (2.1)

$$s\left(\frac{x - \delta}{c}\right)\Big/c \propto \left|\frac{x - \delta}{c}\right|^{-\alpha-1}\Big/c \propto c^\alpha ,$$

while $s(0)/c \propto c^{-1}$. Thus if $K > 1 + (1/\alpha)$ then as $c \to 0$

$$g_\theta(x) = \prod_1^K s\left(\frac{x_k - \delta}{c}\right)\Big/c \to 0 .$$

A similar proof holds for the case $(\alpha, \beta) \to (1,1)$, completing the proof of condition 2.

CONDITION 3. Let $A(x; \theta)$ be the vector of first derivatives $[(\partial \log f_\theta(x))/\partial\theta]$ and let $B(x; \theta)$ be the matrix of second derivatives $(\partial^2 \log f_\theta(x))/\partial\theta^2$. Then there exists a function $C(x)$ such that, for all $\theta \in \bar{C}$, $\int C(x) f_\theta(x) dx \leqq M < \infty$, and each element of $B(x; \theta)$ is bounded in absolute value by $C(x)$.

PROOF. For x fixed and θ in the closed set $\bar{C}$, the elements of $|B(x, \theta)|$ will have a maximum $C(x)$, which is itself bounded in any closed x-interval. It is only necessary to study the behavior of $C(x)$ as $x \to \pm\infty$. Consider the infinite series representation of $s_{\alpha,\beta}(x)$ given by (2.1). It is an absolutely convergent series for $0 < \alpha < 1$, and Bergström (1952) further showed that, even if $1 < \alpha < 2$, it is an asymptotic expansion of $s_{\alpha,\beta}(x)$ as $x \to \infty$. That is, if one approximates $s_{\alpha,\beta}(x)$ by the first J terms of (2.1), then the error will be of the order of the first neglected term, as $x \to \infty$. In the next section it will be shown that the asymptotic expansion (2.1) may be formally differentiated an arbitrary number of times with respect to α, β, or x, to yield an asymptotic expansion for the corresponding derivative of $s_{\alpha,\beta}(x)$. Using this result, simple but rather tedious calculations show that $C(x) = O(\log|x|)^2$ as $x \to \pm\infty$, finishing the proof of Assumption 3. Notice that for $\alpha = 2$ or $\beta = -\alpha$ $(\alpha < 1)$ or $\beta = 2 - \alpha$ $(\alpha > 1)$ every term in the expansion (2.1) becomes 0. This reflects the fact that the right tails of these distributions decrease faster than any power of $x^{-\alpha}$. Assumption 3 fails in this case and so, in the above lemma, $\bar{C}$ must be restricted so as not to include the end points of the parameter space, namely $\alpha = 2$ or $|\beta| = \min(\alpha, 2 - \alpha)$.

CONDITION 4. For every $\theta \in \bar{C}$, $E_\theta(A(x, \theta)) = 0$ and $E_\theta(A(x, \theta)A^T(x, \theta)) = -E_\theta(B(x, \theta)) = I(\theta)$, say, where $I(\theta)$ is the Fisher information matrix at the point θ.

PROOF. It is sufficient to show that $\int_{-\infty}^{+\infty} \dot{f}_\theta(x)dx = 0$, where $\dot{f}_\theta(x)$ is a derivative of $f_\theta(x)$ with respect to θ. Using the asymptotic expansions referred to above, it is easy to show that $|\dot{f}_\theta|$ is integrable. This in turn implies that $\int_{-\infty}^{+\infty} \dot{f}_\theta(x)dx = \dot{\phi}(0; \theta)$ where $\dot{\phi}$ is the corresponding derivative of (1.1). Since $\phi(0; \theta) \equiv 1$, then $\dot{\phi}(0; \theta) \equiv 0$.

CONDITION 5. For all $\theta \in \Theta$, and for all $\theta_0 \in \bar{C}$, $\theta_0 \neq \theta$, $\int |f_\theta(x) - f_{\theta_0}(x)|dx > 0$.

PROOF. This is an obvious consequence of the fact that the family of characteristic functions (1.1) are all distinct.

CONDITION 6. For every $\theta \in \bar{C}$, the Fisher information matrix $I(\theta) = \int_{+\infty}^{-\infty} -B(x; \theta)f_\theta(x)dx$ is nonsingular.

PROOF. If θ were a 1-dimensional parameter, this assumption is equivalent to assuming that $\int_{-\infty}^{+\infty} ((\partial f_\theta)/\partial\theta)^2 1/f_\theta dx > 0$ or, equivalently, that $\partial f_\theta/\partial\theta$ is not an identically 0 function of x. In the present case it is equivalent to prove:

CONDITION 6′. For every $\theta \in \bar{C}$, and for every $a = (a_1, a_2, a_3, a_4)$, then $g(a, x)$ is identically 0 for all x only if $a_1 = a_2 = a_3 = a_4 = 0$, where

$$g(a, x) = a_1 \frac{\partial f_\theta}{\partial \alpha} + a_2 \frac{\partial f_\theta}{\partial \beta} + a_3 \frac{\partial f_\theta}{\partial c} + a_4 \frac{\partial f_\theta}{\partial \delta}.$$

Now $g(a, x)$ is of the form $\int_{-\infty}^{+\infty} e^{-ixt}\psi(a, t)dt$, where ψ is a linear combination of derivatives of the stable characteristic function (1.1), and $g(a, x) = 0$ for all x iff $\psi(a, t) = 0$ for all t. The derivatives of (1.1) (evaluated at $c = 1$ and $\delta = 0$) are

$$\phi_1 = \frac{\partial\phi}{\partial\alpha} = -\exp\left[-i\frac{\pi}{2}\beta \operatorname{sgn} t\right]|t|^\alpha \log|t|\phi ,$$

$$\phi_2 = \frac{\partial\phi}{\partial\beta} = i\frac{\pi}{2}\operatorname{sgn} t \exp\left[-i\frac{\pi}{2}\beta \operatorname{sgn} t\right]|t|^\alpha\phi ,$$

$$\phi_3 = \frac{\partial\phi}{\partial c} = -\exp\left[-i\frac{\pi}{2}\beta \operatorname{sgn} t\right]\alpha|t|^\alpha\phi ,$$

$$\phi_4 = \frac{\partial\phi}{\partial\delta} = it\phi .$$

Since the ϕ_k, $k = 1, \cdots, 4$ are separate multiples of $|t|^\alpha \log|t|$, t^α, and t, it is impossible for $\sum a_k\phi_k \equiv 0$, as long as $\alpha \neq 1$, except, perhaps, for a sum of the form $a_2\phi_2 + a_3\phi_3$, since ϕ_2 and ϕ_3 are multiples of $\exp[-i(\pi/2)\beta \operatorname{sgn} t]|t|^\alpha\phi$. But that would imply $a_2(i(\pi/2)\operatorname{sgn}(t)) - a_3\alpha = 0$ for both positive and negative t, which is impossible unless $a_2 = a_3 = 0$.

This completes the proof that $\hat\theta_n$ converges to θ_0 with probability 1 and is asymptotically normal. For the exceptional cases involving the normal distribution ($\alpha_0 = 2$), or the maximally skewed distributions ($\beta_0 = \pm\min(\alpha_0, 2 - \alpha_0)$), the rate of convergence of $\hat\alpha_n$ or $\hat\beta_n$ respectively to the true value will be faster than $n^{-\frac{1}{2}}$, since it can be shown that the Fisher information about α or β approaches ∞ as $\alpha_0 \to 2$ or $\beta_0 \to \pm\min(\alpha_0, 2 - \alpha_0)$. Further investigation of the properties of $\hat\alpha_n$ and $\hat\beta_n$ is needed for these cases.

5. Asymptotic expansions for the derivatives of $s_{\alpha,\beta}(x)$. Consider (2.1) to be of the form

$$s_{\alpha,\beta}(x) = \sum_{j=1}^\infty a_j(\alpha, \beta, x) .$$

For arbitrary nonnegative integers m_1, m_2, m_3 let

$$\dot s_{\alpha,\beta}(x) = (\partial^{m_1+m_2+m_3}/\partial\alpha^{m_1}\partial\beta^{m_2}\partial x^{m_3})s_{\alpha,\beta}(x) ,$$

and let $\dot a_j(a, \beta, x)$ be the corresponding derivative of $a_j(\alpha, \beta, x)$. With the help of asymptotic formulas for the derivatives of $\Gamma(1 + k\alpha)$ it can be shown that for j sufficiently large (i.e., $j \geqq J(\varepsilon)$), $|\dot a_j| \leqq 2^{-j}$ for all α, β, x satisfying

$$0 < \alpha \leqq 1 - \varepsilon , \qquad -\alpha \leqq \beta \leqq \alpha , \qquad \text{and} \qquad x \geqq \varepsilon > 0 .$$

Therefore the series $\sum \dot a_j$ converges, uniformly with respect to the above values of α, β, and x, to the derivative $\dot s$.

Bergström (1952) shows that the series (2.1) is a divergent asymptotic expansion, as $x \to \infty$, for $s_{\alpha,\beta}(x)$, if $1 < \alpha < 2$. Similarly it is true that the series $\sum \dot a_j$ is a divergent asymptotic expansion, as $x \to \infty$, for $\dot s$, if $1 < \alpha < 2$. To prove this, notice that each $\dot{a}_j(\alpha, \beta, x)$ is

$$\dot{a}_j(\alpha, \beta, x) = x^{-j\alpha-1-m_3} \sum_{n=0}^{m_1} b_{jn}(\alpha, \beta)(\log x)^n ,$$

$$(5.1) \qquad b_{jn}(\alpha, \beta) = \frac{(-1)^{j+m_3-1}}{\pi j!} [\prod_{k=1}^{m_3} (j\alpha + k)] \binom{m_1}{n} \left(\frac{\partial^{m_1-n+m_2}}{\partial\alpha^{m_1-n}\partial\beta^{m_2}}\right) \times \Gamma(1 + j\alpha) \sin \frac{j\pi}{2} (\alpha + \beta) .$$

Since

$$\dot{s}_{\alpha,\beta}(x) = \int_{-\infty}^{+\infty} e^{-itx}(-it)^{m_3}\dot{\phi}(t, \alpha, \beta)\, dt ,$$

where $\dot{\phi} = (\partial^{m_1+m_2}/\partial\alpha^{m_1}\partial\beta^{m_2})\phi$, then, by the same method which Bergström (1952) used for s (see also DuMouchel (1971b) where the method is used to derive on expansion for $s_{1,\beta}(x)$), it can be shown that $\dot{s}$ will have an expansion in powers of $x^{-\alpha}$ and $\log x$ as $x \to \infty$. The coefficients of the expansion are tedious to compute for the high-order derivatives, but, since the coefficients of asymptotic expansions are unique, when $\alpha < 1$ they must coincide with the coefficients of the convergent series, namely the b_{jn} of (5.1). Now since the aforementioned tedious computation of the coefficients of the asymptotic expansion is identical for $\alpha > 1$ as for $\alpha < 1$, it follows that the b_{jn} defined by (5.1) work for $\alpha > 1$ also. Therefore the asymptotic expansion (2.1) may be differentiated arbitrarily with respect to α, β, and x.

Acknowledgment. I wish to thank Professor L. M. LeCam for the benefit of several discussions on this topic.

REFERENCES

[1] BERGSTRÖM, HARALD (1952). On some expansions of stable distributions. *Ark. Mat.* **2** 375–378.

[2] CRAMÉR, HARALD (1946). *Mathematical Methods of Statistics*, Princeton Univ. Press. 500–504.

[3] DUMOUCHEL, WILLIAM H. (1971a). On the limiting form of stable distributions when the characteristic exponent approaches 0. Submitted for publication.

[4] DUMOUCHEL, WILLIAM H. (1971b). Stable distributions in statistical inference. Ph. D. dissertation, Yale Univ.

[5] FAMA, EUGENE F. (1965). The behavior of stock market prices. *J. Bus. Univ. Chicago* **38** 34–105.

[6] FAMA, EUGENE F. and ROLL, RICHARD (1968). Some properties of symmetric stable distributions. *J. Amer. Statist. Assoc.* **63** 817–836.

[7] FAMA, EUGENE F. and ROLL, RICHARD (1971). Parameter estimates for symmetric stable distributions. *J. Amer. Statist. Assoc.* **66** 331–338.

[8] FELLER, WILLIAM (1966). *An Introduction to Probability Theory and its Applications*, **2**. Wiley, New York.

[9] GNEDENKO, V. B. and KOLMOGOROV, A. N. (1954). *Limit Distributions for Sums of Random Variables*, (transl. by K. L. Chung). Addison-Wesley, Reading.

[10] LECAM, LUCIEN M. (1952). On some asymptotic properties of maximum-likeihood estimates and related Bayes estimates. Ph. D. dissertation, Univ. of California, Berkeley.

[11] MANDELBROT, BENOIT (1960). The Pareto-Levy law and the distribution of income. *Internat. Econ. Rev.* **1** 79–106.

[12] MANDELBROT, BENOIT (1962). A class of long-tailed probability distributions and the

empirical distribution of city sizes. Research note, Thomas J. Watson Research Center, Yorktown Heights, New York.

[13] MANDELBROT, BENOIT (1963a). The stable Paretian income distribution when the apparent exponent is near two. *Internat. Econ. Rev.* **4** 111–114.

[14] MANDELBROT, BENOIT (1963b). The variation of certain speculative prices. *J. Bus. Univ. Chicago* **26** 394–419.

[15] MANDELBROT, BENOIT (1967a). How long is the coast of Britain? Statistical self-similarity and fractional dimension. *Science* **156** 636–638.

[16] MANDELBROT, BENOIT (1967b). The variation of some other speculative prices. *J. Bus. Univ. Chicago* **40** 393–413.

[17] MANDELBROT, BENOIT (1969a). Long-run linearity, locally Gaussian process, *H*-spectra and infinite variances. *Internat. Econ. Rev.* **10** 82–111.

[18] MANDELBROT, BENOIT (1969b). Efficient arbitrating, price variation, random walks, and martingale models. Presented to the Annual Winter Meeting of the Econometric Society, 28 December 1969.

[19] MANDELBROT, BENOIT (1971). The birth and mutation process: explicit distributions for the number of mutants in a culture of bacteria. Unpublished manuscript.

[20] WALD, ABRAHAM (1941). Asymptotically most powerful tests of hypotheses. *Ann. Math. Statist.* **12** 1–19.

[21] WALD, ABRAHAM (1949). Note on the consistency of the maximum-likelihood estimate. *Ann. Math. Statist.* **20** 595–601.

[22] WEISS, L. and WOLFOWITZ, J. (1967). Maximum probability estimators. *Ann. Inst. Statist. Math.* **19** 193–206.

DEPARTMENT OF STATISTICS
UNIVERSITY OF MICHIGAN
ANN ARBOR, MICHIGAN 48104