On Some Fourier Methods for Inference

Andrey Feuerverger; Philip McDunnough

Journal of the American Statistical Association, Vol. 76, No. 374 (Jun., 1981), 379-387.

Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28198106%2976%3A374%3C379%3AOSFMFI%3E2.0.CO%3B2-O

On Some Fourier Methods for Inference

ANDREY FEUERVERGER and PHILIP McDUNNOUGH*

Common statistical procedures such as maximum likelihood and M-estimation admit generalized representations in the Fourier domain. The Fourier domain provides fertile ground for approaching a number of difficult problems in inference. In particular, the empirical characteristic function and its extension for stationary time series are shown to be fundamental tools which support numerically simple inference procedures having arbitrarily high asymptotic efficiency and certain robustness features as well. A numerical illustration involving the symmetric stable laws is given.

KEY WORDS: Fourier transforms; Maximum likelihood; Empirical characteristic functions; Asymptotic efficiency; Fisher information; Robustness.

1. INTRODUCTION

A number of difficult problems in statistical inference may be amenable to solution by transform methods. In this article we explore the applicability of some new procedures based on Fourier methods and empirical characteristic functions. These procedures are shown to provide a natural basis for asymptotically efficient inference, and certain natural tradeoffs between the efficiency of the procedures and their robustness are indicated. The range of potential applications emerges as being fairly broad. A numerical example is given for the symmetric stable laws.

The basic ideas may be introduced in the context of efficient inference. Thus suppose $X_1, X_2, \ldots, X_n$ to be iid variates with density in $\{f_\theta(x)\}$ and θ assumed (for simplicity here) to be a real univariate parameter. Now maximum likelihood leads to a likelihood equation which we may write as

$$\int \frac{\partial \log f_\theta(x)}{\partial \theta} dF_n(x) = 0 , \qquad (1.1)$$

where $F_n(x)$ is the empirical cumulative distribution function (cdf), or in the suggestive alternate form

$$\int \frac{\partial \log f_\theta(x)}{\partial \theta} d[F_n(x) - F_\theta(x)] = 0 , \qquad (1.2)$$

where F_θ is the cdf of f_θ, and the term introduced is just a "zero."

This last equation may be transformed. Using a general form of the Parseval Theorem (see Sec. 6) we obtain the following result, namely a Fourier domain version of the likelihood equation:

$$\int w_\theta(t)[c_n(t) - c_\theta(t)]\, dt = 0 . \qquad (1.3)$$

Here $c_\theta(t) = \int e^{itx} dF_\theta(x)$ is the characteristic function and

$$c_n(t) = \int e^{itx} dF_n(x) = \frac{1}{n} \sum_{j=1}^{n} e^{itX_j}$$

is the empirical characteristic function (ecf). The weight function

$$w_\theta(t) = \frac{1}{2\pi} \int \frac{\partial \log f_\theta(x)}{\partial \theta} e^{-itx}\, dx \qquad (1.4)$$

is given as the inverse Fourier transform of the score. In general, of course, the score function is not integrable and the proper interpretation of these equations is in terms of generalized functions (see Sec. 6).

Now one approach to equation (1.3) involves recognizing $\int w_\theta(t) c_\theta(t) \equiv 0$ as just the zero term, and working with the likelihood equation:

$$\int w_\theta(t) c_n(t)\, dt = 0 . \qquad (1.5)$$

This approach, however, can have merit only if the form of $w_\theta(t)$ is tractable, while that of $f_\theta(x)$ is not.

An alternative approach to (1.3) is to regard the weight function as known and then to solve the moment-type equation

$$\int w(t)[c_n(t) - c_\theta(t)]\, dt = 0 . \qquad (1.6)$$

We shall see (Sec. 2) that for appropriate $w(t)$ this provides an asymptotically efficient procedure. Heuristically this can be understood by comparing the behavior of the two components of (1.3) for θ in the neighborhood of the true value θ_0. Thus

$$\frac{\partial}{\partial\theta} \int w_\theta(t) c_n(t)\, dt \Big|_{\theta_0}$$
$$= \frac{1}{2\pi} \iint \frac{\partial^2 \log f_\theta(x)}{\partial\theta^2} \Big|_{\theta_0} c_n(t) e^{-itx}\, dx\, dt$$
$$= \int \frac{\partial^2 \log f_\theta(x)}{\partial\theta^2} \Big|_{\theta_0} dF_n(x)$$

and this converges at the usual rate to $-I(\theta_0)$ where $I(\theta)$

* Andrey Feuerverger is Associate Professor, and Philip McDunnough is Assistant Professor, both at the Department of Statistics, University of Toronto, Toronto, Ontario, Canada M5S 1A1. The research of both authors is supported by NSERC, Canada. The authors are grateful to C.L. Mallows for valuable comments on an earlier manuscript.

© Journal of the American Statistical Association
June 1981, Volume 76, Number 374
Theory and Methods Section

is the Fisher information. But for the other component we have

$$\frac{\partial}{\partial\theta_2}\int w_{\theta_1}(t)c_{\theta_2}(t)\,dt\,\Bigg|_{\theta_1=\theta_2=\theta_0} = I(\theta_0)\ .$$

Therefore if $w(t)$ in (1.6) is taken as $w_{\hat\theta}(t)$ where $\hat\theta$ is any consistent estimator of θ_0, then the compensation resulting from the asymptotic equality of the derivatives of the two components of (1.3) suggests that (1.6) will be asymptotically equivalent to maximum likelihood. (It is of interest that a parallel argument may be applied to (1.2).) We remark that if (1.6) were applied iteratively and θ_1 in w_{θ_1} and θ_2 in c_{θ_2} converged to the same value, then that value will, of course, satisfy the likelihood equation.

To illustrate some basic manipulations consider an arbitrary scale-location family

$$f(x) = \frac{1}{\sigma}g\left[\frac{x-\mu}{\sigma}\right]$$

where $g(x)$ is a symmetric density. The corresponding score functions are

$$\frac{\partial\log f(x)}{\partial\mu} = -\frac{1}{\sigma}g'\left[\frac{x-\mu}{\sigma}\right]\Big/ g\left[\frac{x-\mu}{\sigma}\right]$$

and

$$\frac{\partial\log f(x)}{\partial\sigma} = -\frac{1}{\sigma}-\frac{x-\mu}{\sigma^2}g'\left[\frac{x-\mu}{\sigma}\right]\Big/ g\left[\frac{x-\mu}{\sigma}\right]\ .$$

In describing the transforms of these functions we conveniently may drop constant factors and delta functions without affecting the procedure (1.6). (Delta functions at zero can be dropped because $c(0) = c_n(0) = 1$.) Thus, at $\mu = 0$, $\sigma = 1$ for example, we may take

$$w_\mu(t) = \int\frac{g'(x)}{g(x)}e^{-itx}\,dx$$

and

$$w_\sigma(t) = \int\frac{xg'(x)}{g(x)}e^{-itx}\,dx\ .$$

Note that (except for an inconsequential constant factor)

$$w_\sigma(t) = \frac{dw_\mu(t)}{dt}\ .$$

For symmetric g, w_σ is symmetric and thus acts only on $Re\,c(t)$, while w_μ is antisymmetric and acts only on $Im\,c(t)$.

Now for the Gaussian case with $g(x) \equiv N(0, 1)$, w_μ, w_σ are proportional to the first and second derivatives of delta functions. For the Cauchy case with $g(x) = [\pi(1 + x^2)]^{-1}$, we have

$$\frac{g'(x)}{g(x)} = \frac{x}{1+x^2}\,,\quad \frac{xg'(x)}{g(x)} = \frac{1}{1+x^2}\,,$$

giving

$$w_\mu(t) = \mathrm{sgn}(t)e^{-|t|},\quad w_\sigma(t) = e^{-|t|}\ .$$

(Here again, we are ignoring constant factors and delta functions at zero.) Of course, the approach implicit here leads necessarily to maximum likelihood; other methods will be given below which do not require evaluation of the weight functions but yet have arbitrarily high asymptotic efficiency.

In Section 2 we present our main results concerning ecf procedures and in Section 3 discuss a robustness property of the ecf. In Section 4 we discuss the problem of inference for the stable laws and give a numerical example of our methods for the symmetric case. In Section 5 we present an extension for stationary time series, the polycharacteristic function (pcf), and outline an application to a specific Markov process. A number of mathematical technicalities which arise are omitted from the main text. These are collected in Section 6. Some concluding remarks are given in Section 7.

2. CONCERNING ECF PROCEDURES

A review of the properties of the ecf is given in Feuerverger and McDunnough (1981), hereafter referred to as FM. Here we shall note only that

$$Y(t) = \sqrt{n}\,[c_n(t) - c(t)]$$

is asymptotically normal at finite numbers of points, has zero mean, and covariance structure determined (both for finite n and the asymptotic case as well) by

$$\begin{aligned}\mathrm{cov}[Y(s), Y(t)] &= EY(s)\overline{Y(t)}\\ &= c(s-t) - c(s)c(-t)\ .\end{aligned} \tag{2.1}$$

In particular we obtain

$$\begin{aligned}&\mathrm{cov}[\mathrm{Re}\,Y(s), \mathrm{Re}\,Y(t)]\\ &\quad= \tfrac12[\mathrm{Re}\,c(s-t) + \mathrm{Re}\,c(s+t)] - \mathrm{Re}\,c(s)\,\mathrm{Re}\,c(t)\\ &\mathrm{cov}[\mathrm{Re}\,Y(s), \mathrm{Im}\,Y(t)]\\ &\quad= \tfrac12[\mathrm{Im}\,c(s-t) + \mathrm{Im}\,c(s+t)] - \mathrm{Re}\,c(s)\,\mathrm{Im}\,c(t)\\ &\mathrm{cov}[\mathrm{Im}\,Y(s), \mathrm{Im}\,Y(t)]\\ &\quad= \tfrac12[\mathrm{Re}\,c(s-t) - \mathrm{Re}\,c\,(s+t)] - \mathrm{Im}\,c(s)\,\mathrm{Im}\,c(t)\ .\end{aligned} \tag{2.2}$$

We summarize our main results in the following theorem. The technical conditions we require are quite mild and for convenience are collected in Section 6.

Theorem 2.1. Suppose $X_1, X_2, \ldots, X_n$ are iid with cf $c_\theta(x)$ where $\theta = (\theta_1, \ldots, \theta_l)$ has unknown true value $\theta_0 \in \Theta$. Let $0 < t_1 < \cdots < t_k$ be a fixed grid; define $\mathbf{z}'_\theta = [\mathrm{Re}\,c(t_1), \ldots, \mathrm{Re}\,c(t_k), \mathrm{Im}\,c(t_1), \ldots, \mathrm{Im}\,c(t_k)]$ and let $\mathbf{z}_n$ be its empirical counterpart. Let $n^{-1}\boldsymbol{\Sigma}$ be the covariance matrix of $\mathbf{z}_n$: the entries of $\boldsymbol{\Sigma}$ are given by (2.2). Then the estimation procedures I through IV given here yield consistent estimators having the same asymptotic normal distribution. Further, the asymptotic variances of these procedures can be made arbitrarily close to the Cramér-Rao (CR) bound by selecting the grid $\{t_j\}$ to be sufficiently fine and extended.

I (*the k-L procedure*). Estimate θ by maximizing the asymptotic normal form of the likelihood for $\mathbf{z}_n$.

II (*moment estimator*). Choose θ to solve the equation

$$\mathbf{D}'\mathbf{z}_n = \mathbf{D}'\mathbf{z}_\theta$$

where $\mathbf{D}$ is a consistent estimator of

$$\mathbf{D}_0 = \Sigma^{-1}\left[\frac{\partial z_\theta}{\partial\theta_1}, \ldots, \frac{\partial z_\theta}{\partial\theta_l}\right]$$

evaluated at θ_0.

III (*min Q procedure*). Choose θ to minimize the quadratic form

$$(\mathbf{z}_n - \mathbf{z}_\theta)' Q(\mathbf{z}_n - \mathbf{z}_\theta)$$

where Q is a consistent estimator of any matrix Q_0 with the property that $Q_0 \Sigma D_0 = D_0 F$ for some nonsingular matrix F.

IV (*harmonic regression*). Choose θ by fitting z_θ to z_n using nonlinear least squares and any consistent estimate of the asymptotically optimal weights.

We remark that the asymptotic log-likelihood referred to in I may be taken either as

$$-\tfrac{1}{2}\log\det\Sigma - \frac{n}{2}(\mathbf{z}_n - \mathbf{z}_\theta)'\,\Sigma^{-1}(\mathbf{z}_n - \mathbf{z}_\theta) \quad (2.3)$$

or as just the second term of this expression and that procedure II is motivated by differentiating this term. The procedures I and II given here are straightforward extensions of the results of FM to the multiparameter case and the results III and IV follow along similar lines. The condition in III is just that for the efficiency of a weight in least squares; the case $Q = D_0 D'_0$ leads to procedure II while $Q = \Sigma^{-1}$ gives procedure IV. Of course IV is just the regression formulation and can be obtained from (2.3) by replacing Σ by any consistent estimate.

A direct and more revealing proof of Theorem 2.1 is possible than that appearing in FM and we briefly indicate this here. Thus suppose $\mathbf{D}'(\mathbf{z}_n - \mathbf{z}_\theta) = 0$ where $\mathbf{D}$ is $2k \times 1$ (the multiparameter case would be treated similarly) or equivalently

$$\int [c_n(t) - c_\theta(t)]\, dW(t) = 0 \quad (2.4)$$

where $W(t) = \bar{W}(-t)$ is a step function. Then a standard differential argument (c.f. Section 6) yields that a consistent root $\hat{\theta}$ of (2.4) is asymptotically normal with asymptotic variance

$$n\,\text{var}(\hat{\theta}) = \frac{\iint c(s-t)\, dW(s)\, d\bar{W}(t) - \left[\int c(t)\, dW(t)\right]^2}{\left[\int \frac{\partial c(t)}{\partial\theta}\, dW(t)\right]^2} \le \frac{\iint c(s-t)\, dW(s)\, d\bar{W}(t)}{\left[\int \frac{\partial c(t)}{\partial\theta}\, dW(t)\right]^2}. \quad (2.5)$$

Now (2.5) is just the quotient of a Hermitian form and an associated squared linear form and is minimized when the jumps of $W(\cdot)$ are proportional to the entries of

$$[c(s-t)]^{-1}\left(\frac{\partial c(t)}{\partial\theta}\right). \quad (2.6)$$

Here we interpret $[c(s-t)]$ as being a matrix—which may be taken to have dimension $(2k+1)\times(2k+1)$ with s, t ranging over $-t_k, \ldots, -t_1, 0, t_1, \ldots, t_k$; and $[\partial c(t)/\partial\theta]$ as being a column vector. Taking $t_j = j\tau$, the inclusion of $t = 0$ (possible because $\partial c(0)/\partial\theta = 0$) makes $[c(s-t)]$ a Toeplitz matrix. Now substituting (2.6) in (2.5) we obtain

$$\left[\left(\frac{\partial c}{\partial\theta}\right)'[c(s-t)]^{-1}\left(\frac{\partial c}{\partial\theta}\right)\right]^{-1}. \quad (2.7)$$

We may now carry out a circulant approximation of the Toeplitz form. (This is standard in certain time series applications, dating from Whittle 1951.) Thus consider the spectral representation

$$[c(s-t)]^{-1} = \Sigma\lambda_j^{-1}\xi_j\xi'_j$$

where the $\{\lambda_j\}$, $\{\xi_j\}$ are eigenvalues, eigenvectors of $[c(s-t)]$. In the limit (as $k \to \infty$) these may be approximated by the eigenvalues, eigenvectors for the corresponding circulant (see Brillinger 1975, Sec. 3.7; Grenander and Szego 1958, Ch. 11). The eigenquantities for the circulant are given, for example, in Brillinger, Theorem 3.7.3, and substituting, we find that (2.7) can be made arbitrarily close to the CR bound.

The procedures of Theorem 2.1 are based on the ecf at only a finite number of points and thus are of *discrete type*. Analogous procedures of a *continuous type* require a more general context for proper treatment. In particular, concerning the interpretation of (1.4) when the score is not integrable, we remark that the Fourier integral will continue to exist when interpreted as a generalized function (see Sec. 6). In FM the problem of integrability of the score was treated by means of tapering. However, equations such as (1.3) and (1.4) have meaning more generally, provided these are viewed in the context of generalized function theory.

The procedures of Theorem 2.1 all have continuous analogs. We shall indicate these only heuristically. The moment procedure has the analog

$$\int \mathbf{w}(t)\,[c_n(t) - c_\theta(t)]\, dt = \mathbf{0}$$

with asymptotic covariance matrix

$$n\,\text{var}(\hat{\boldsymbol{\theta}}) = \left\{\int \mathbf{w}(t)\frac{\partial c(t)}{\partial\boldsymbol{\theta}}\, dt\right\}^{-1} \cdot \iint \mathbf{w}(s)\mathbf{w}'(t)K(s,t)\, ds\, dt \cdot \left\{\int \mathbf{w}(t)\frac{\partial c(t)}{\partial\boldsymbol{\theta}}\, dt\right\}^{-1} \quad (2.8)$$

where $\mathbf{w}$ and $\partial c/\partial\boldsymbol{\theta}$ here are $l \times 1$ and $1 \times l$, and

$$K(s,t) = c(s-t) - c(s)c(-t).$$

If $\mathbf{w}(t) = \mathbf{w}_{\theta_0}(t)$ in (2.8) we obtain the CR bound. This is also the case asymptotically if we use a consistent estimator of the efficient weight (i.e., $\mathbf{w}_{\hat\theta}$ where $\hat\theta$ is consistent for θ_0). Note also that (2.8) can be made arbitrarily close to the CR bound by a suitable selection (or estimate) of a continuous integrable $w(t)$, for example, through tapering as in FM.

The continuous analog for the min-Q procedure involves the Hermitian form

$$\iint [c_n(s) - c_\theta(s)]\,[\bar c_n(t) - \bar c_\theta(t)]\,A(s, t)\,ds\,dt \tag{2.9}$$

which must be minimized. In both the discrete and continuous cases, squaring and summing the moment equations leads to a rank l quadratic procedure which attains a zero minimum. Hence the kernel

$$A(s, t) = \sum_{j=1}^{l} w_{\theta_0}{}^{(j)}(s)\,\bar w_{\theta_0}{}^{(j)}(t) \tag{2.10}$$

leads to a CR-bound procedure. More generally, one may write down the asymptotic covariance for the procedure (2.9) and note that the CR bound is preserved if we take, formally,

$$A(s, t) = \sum_{j=1}^{l} w_{\theta_0}{}^{(j)}(s)\,\bar w_{\theta_0}{}^{(j)}(t) + \sum_{j=l+1}^{\infty} w_j(s)\,\bar w_j(t) \tag{2.11}$$

where the w_j, $j \geq l + 1$ are orthogonal to the $w_{\theta_0}{}^{(j)}$, $j \leq l$. Except for the difficulty that these are generalized functions, the theorems of Mercer and Schmidt (Riesz and Sz.Nagy, Sec. 97 and 98) suggest that such $A(s, t)$ provide most if not all solutions for asymptotically optimal quadratic weights.

Finally, we remark that procedures I and IV have continuous analogs. Because of results associated with the property of continuous dependence on the kernel (e.g., Courant and Hilbert 1953, p. 151), these may be equivalent to certain limits of discrete procedures. These continuous analogs have a real statistical interest, but we do not pursue this here.

3. A ROBUSTNESS PROPERTY OF THE ECF

A number of authors (Thornton and Paulson 1977; Heathcote 1977; FM 1981) have noted certain robustness properties for various procedures associated with the ecf. This robustness is related to the equivalence of a certain class of ecf based procedures with the M-estimators due to Huber. See for example Huber (1977).

Consider a general class of M-estimators defined by the implicit equation $\Sigma\psi(X_j, \theta) = 0$, or equivalently

$$\int \psi(x, \theta)\,dF_n(x) = 0\,. \tag{3.1}$$

This equation has a representation in the Fourier domain: again by Parseval's theorem we may write

$$\int w_\theta(t) c_n(t)\,dt = 0 \tag{3.2}$$

where

$$w_\theta(t) = \frac{1}{2\pi}\int \psi(x, \theta) e^{-itx}\,dx\,. \tag{3.3}$$

For these procedures to be consistent we need at least the following restriction on $w_\theta(t)$ (i.e., on $\psi(x, \theta)$):

$$E\psi(X, \theta) = \int w_\theta(t) c_\theta(t)\,dt = 0 \tag{3.4}$$

and then (3.1) is equivalent to

$$\int w_\theta(t)[c_n(t) - c_\theta(t)]\,dt = 0\,. \tag{3.5}$$

It follows (Huber 1977, p. 14) that procedure (3.5) has influence curve proportional to

$$\psi(x, \theta) = \int w_\theta(t) e^{itx}\,dt\,. \tag{3.6}$$

For procedure (1.6), with *$w(t)$ considered fixed*, the influence curve may be calculated directly. Suppose that the function $\theta(\epsilon)$ is a differentiable root of the implicit equation

$$\int \hat w(y)\,d[(1 - \epsilon)\,F(y) + \epsilon H_x(y) - F_{\theta(\epsilon)}(y)] = 0 \tag{3.7}$$

where $\hat w$ is the Fourier transform of w and H_x gives unit mass at x. Then the influence curve (Huber 1977, p. 10) is just

$$IC(x, F) = \lim_{\epsilon\to 0} \frac{\theta(\epsilon) - \theta(0)}{\epsilon} = \left.\frac{d\theta(\epsilon)}{d\epsilon}\right|_{\epsilon=0}$$

and differentiating (3.7), we find

$$IC = \frac{\hat w(x) - \int \hat w(y)\,dF(y)}{\left.\frac{d}{d\theta}\int \hat w(y)\,dF_\theta(t)\right|_{\theta=\theta_0}}\,, \tag{3.8}$$

which, as a function of x, is just a linear function of the Fourier transform of $w(t)$.

These findings help to explain the robustness properties that have been noted. Thus, for example, for integrable $w(t)$, we obtain a bounded influence curve. On the other hand, $w(t)$ can have quite complex behavior and still preserve robustness. The case related to the M-estimator for location given by Huber (1964), which solves the Gaussian robustness problem, affords an interesting example. Here $\psi(x, \theta) = \psi(x - \theta)$ where, except for scaling,

$$\psi(x) = \begin{cases} x & |x| \leq 1 \\ \operatorname{sgn}(x) & |x| > 1 \end{cases}$$

while the inverse Fourier transform, obtained after some calculation, is

$$w_\psi(t) = \frac{\sin t}{i\pi t^2}\,.$$

We may note that efficiency and robustness are subject to trade. In FM the problem of integrability of the score was resolved by working with the truncated forms:

$$w_m(t) = \int_{-\infty}^{\infty} \frac{\partial \log f_\theta(x)}{\partial\theta} h_m(x) e^{-itx}\,dx \tag{3.9}$$

where

$$h_m(x) = \begin{cases} 1 & |x| \le m \\ 0 & |x| > m \end{cases}$$

(and the representation here serves to indicate the possibility of more general tapering). In the present context, however, two essential aspects may be clarified. First is that the influence curve for a tapered form is (a linear function of)

$$\frac{\partial \log f_\theta(x)}{\partial \theta} h_m(x) ,$$

which presumably is bounded. Second, not integrability, but rather boundedness, is seen to be essential for robustness, which suggests replacing the "vertical tapering" of (3.9) by some type of "horizontal tapering," which leads to

$$w_M(t) = \int_{-\infty}^{\infty} \left[\frac{\partial \log f_\theta(x)}{\partial \theta}\right]_M e^{-itx}\, dx$$

(interpreted as a generalized function) where the symbol

$$[x]_M = \begin{cases} x \quad , & |x| \le M \\ M \operatorname{sgn}(x), & |x| > M \end{cases} .$$

Trade-off between robustness and efficiency is accomplished in the selection of M. For Gaussian location, this approach leads of course to a type of optimal tradeoff. And in accordance with the general character of the results of Huber (1977), one may expect that useful tradeoff will generally occur.

Finally, we remark that the procedures based on finite grids $\{t_j\}$ are easily seen to have bounded influence. The reason for this is clear intuitively, for the behavior about $t = 0$ will not be accessible to discrete procedures, and outlying values appear as trigonometrically reduced. Note, however, that the bound on the influence can be very high should the grid pass very near the origin.

4. A NUMERICAL EXAMPLE: INFERENCE FOR THE STABLE LAWS

The methods of Section 2 lead to new inference procedures for the stable laws which have arbitrarily high asymptotic efficiency but are considerably simpler to execute than maximum likelihood. Essentially we propose to fit $c_\theta(t)$ to $c_n(t)$ by nonlinear weighted least squares. Here, as before, $c_n(t)$ is the ecf, and $c_\theta(t)$ is the cf with parametrization $\theta = (\mu, \sigma, \alpha)$:

$$c_\theta(t) = e^{i\mu t} \cdot e^{-|\sigma t|^\alpha} . \qquad (4.1)$$

We have indicated here only the symmetric case, but the method is in fact quite general.

The least squares procedure is based on a grid $0 < t_1 < \cdots < t_k$ and involves minimizing a quadratic form in the $2k$ variates $\mathrm{Re}[c_n(t_j) - c_\theta(t_j)]$, $\mathrm{Im}[c_n(t_j) - c_\theta(t_j)]$, $j = 1, 2, \ldots, k$. The covariance structure of these terms is given in (2.2) and may be estimated. To preserve the asymptotic properties a single iteration starting from any consistent estimates will suffice. The updating equation is found from a first order calculation:

$$\hat{\boldsymbol\theta}_{\text{new}} = \hat{\boldsymbol\theta}_{\text{old}} + (\mathbf{G}'\boldsymbol\Sigma^{-1}\mathbf{G})^{-1}\,\mathbf{G}'\boldsymbol\Sigma^{-1}\boldsymbol\phi_{\theta_{\text{old}}} . \qquad (4.2)$$

Here the $2k$ column vector $\boldsymbol\phi_\theta$ has entries consisting of the real and imaginary parts of $(c_n - c_\theta)$ at the t_j, and $\mathbf{G}$ is a $2k \times 3$ matrix

$$\mathbf{G} = \left[\frac{\partial\boldsymbol\phi}{\partial\mu}, \frac{\partial\boldsymbol\phi}{\partial\sigma}, \frac{\partial\boldsymbol\phi}{\partial\alpha}\right]$$

evaluated at $\hat{\boldsymbol\theta}_{\text{old}}$. The covariance matrix $\boldsymbol\Sigma$ of $\boldsymbol\phi$ is obtained from (2.2). It is of some practical value to note that if we work with centered variates such as $\hat{X}_j = (X_j - \hat\mu)/\hat\sigma$ where $\hat\mu$ and $\hat\sigma$ are estimated, then $\boldsymbol\Sigma$ will become block diagonal and the structure of $\mathbf{G}$ will match conveniently to allow (4.2) to separate into two parts: one involving only α, σ and the real part of the ecf based on the $\{\hat{X}_j\}$; the other involving only μ and the imaginary part. In particular, the $2k \times 2k$ inversion may then be replaced by two $k \times k$ inversions. This will not happen for nonsymmetric families.

The symmetric stable family was used for numerical confirmation of the practicability of the methods proposed. Table 1 provides the asymptotic values of $n \cdot$ var for the parameters μ, α, σ evaluated at $\sigma = 1$ and $\alpha = 1.0, 1.1(.2)1.9$. These values are provided for procedures based on $k = 2, 3, 4, 6, 10, 20, 40, \infty$ equally spaced points $\tau, 2\tau, \ldots, k\tau$ on the ecf (a real and complex value being taken at each point). Below each value of the asymptotic variance we give the value of τ, the optimal pair being reported in each case. The cases $k = \infty$ were determined from Tables 1 and 2 of DuMouchel (1975).

The values given in Table 1 were obtained from the asymptotic covariance matrix $(\mathbf{G}'\,\boldsymbol\Sigma^{-1}\,\mathbf{G})^{-1}$ of the estimators using the block structure so that two $k \times k$ inversions were required. We found the subroutine MINV of SSP (1962) adequate for this purpose. (The case $k = 40$, $\alpha = 1.9$, however, did not pass our numerical checks and is not reported.)

The results in Table 1 are encouraging, especially when judged against the rigidity of uniform spacing. According to the results of DuMouchel (1975) stable samples have substantial information for the parameter α in the extreme order statistics. This suggests that for smaller α uniform spacing will not be nearly optimal for α. Spacings with higher concentrations at the origin seem of interest, but the ensuing lines of enquiry are extensive and we do not pursue these here. Table 1 does not show how the efficiencies vary with τ. These more detailed results may be obtained from the authors. We mention here however that in all cases the efficiencies varied slowly enough so that the selection of a compromise τ (allowing high efficiency for all three parameters) presents no difficulties.

A program to carry out k-point inference was developed based on a straightforward weighted nonlinear regression approach, which involved fitting via (4.2) (a Taylor expansion of) $\boldsymbol\phi$ to its estimate at $2k$ (real plus

Table 1. Asymptotic Variances N·VAR and Optimal Uniform Spacing Intervals for Estimating Symmetric Stable Parameters Using k ecf Points

	k	2	3	4	6	10	20	40	∞
$\alpha = 1.0$	μ	2.48	2.29	2.20	2.11	2.06	2.02	2.01	2.00
		.60	.50	.42	.34	.24	.15	.09	
	α	6.43	3.82	2.96	2.28	1.82	1.52	1.37	1.21
		.40	.34	.29	.23	.17	.11	.06	
	σ	2.54	2.35	2.27	2.19	2.15	2.12	2.10	2.10
		.74	.62	.53	.42	.30	.18	.10	
$\alpha = 1.1$	μ	2.52	2.37	2.30	2.24	2.19	2.17	2.16	2.16
		.54	.44	.37	.29	.21	.13	.08	
	α	6.12	3.80	3.03	2.41	2.00	1.73	1.60	1.49
		.41	.34	.29	.23	.17	.10	.06	
	σ	2.12	2.00	1.95	1.90	1.88	1.85	1.84	1.85
		.75	.62	.53	.40	.27	.15	.08	
$\alpha = 1.3$	μ	2.50	2.42	2.38	2.35	2.33	2.32	2.32	2.32
		.45	.36	.30	.23	.16	.10	.05	
	α	5.40	3.67	3.08	2.62	2.32	2.14	2.06	2.02
		.43	.34	.29	.22	.15	.09	.05	
	σ	1.57	1.54	1.51	1.49	1.47	1.46	1.46	1.46
		.76	.58	.45	.31	.20	.11	.06	
$\alpha = 1.5$	μ	2.41	2.37	2.36	2.34	2.34	2.34	2.34	2.34
		.38	.30	.25	.19	.12	.06	.03	
	α	4.50	3.35	2.96	2.67	2.50	2.41	2.38	2.37
		.43	.34	.28	.21	.14	.08	.04	
	σ	1.25	1.22	1.20	1.19	1.18	1.17	1.17	1.17
		.90	.44	.35	.24	.16	.09	.04	
$\alpha = 1.7$	μ	2.28	2.27	2.26	2.26	2.26	2.26	2.25	2.26
		.32	.25	.20	.14	.12	.07	.04	
	α	3.28	2.68	2.49	2.36	2.30	2.27	2.26	2.25
		.43	.32	.26	.19	.12	.07	.04	
	σ	.98	.95	.94	.94	.93	.93	.93	.93
		.48	.35	.28	.19	.12	.08	.04	
$\alpha = 1.9$	μ	2.12	2.12	2.12	2.12	2.12	2.11		2.11
		.24	.31	.25	.17	.15	.06		
	α	1.53	1.37	1.33	1.31	1.28	1.27		1.28
		.39	.28	.22	.15	.15	.10		
	σ	.72	.71	.71	.70	.70	.70		.70
		.38	.27	.33	.23	.17	.10		

imaginary) points starting from consistent estimates. The value for τ may be chosen from the initial estimates to minimize the asymptotic variances. A key feature of the program is its essential simplicity; we omit further descriptions here. Tests that we carried out using Cauchy data generated from Gaussian ratios showed the procedure to be well behaved. Computations were done on the University of Toronto IBM 360/165. The two (FORTRAN) programs described herein may be obtained from the authors.

5. AN EXTENSION FOR STATIONARY PROCESSES

Let $\{X_j\}$ be a strictly stationary ergodic time series, and $\mathbf{Y}_j^{(k)} = (X_j, X_{j-1}, \ldots, X_{j-k})'$. We define the poly-cf (pcf) functions

$$c^{(k)}(\mathbf{t}) = E \exp i\mathbf{t}'\mathbf{Y}_1^{(k)} \tag{5.1}$$

and the corresponding empirical (epcf) quantities

$$c_n^{(k)}(\mathbf{t}) = \frac{1}{n}\sum_{j=1}^{n} e^{i\mathbf{t}'\mathbf{Y}_j^{(k)}}. \tag{5.2}$$

Note that although $Ec_n^{(k)}(\mathbf{t}) = c^{(k)}(\mathbf{t})$, $c_n^{(k)}(\mathbf{t})$ may not be the best estimator for $c^{(k)}(\mathbf{t})$, just as $\bar{X}$ is not usually the best estimator for the process mean. Like $\bar{X}$, however, the epcf will have important asymptotic optimalities as an estimator of the pcf.

We shall consider only the case of Markov processes having a single parameter θ; the generality of the method, however, will be evident. Thus set $\mathbf{Y}_j = \mathbf{Y}_j^{(1)} = (X_j, X_{j-1})'$, $c(\mathbf{t}) = c^{(1)}(\mathbf{t})$, and

$$c_n(\mathbf{t}) = c_n^{(1)}(\mathbf{t}) = \frac{1}{n}\sum_{j=1}^{n} \exp(i\mathbf{t}'\mathbf{Y}_j).$$

Then $Ec_n(\mathbf{t}) = c(\mathbf{t})$ and

$$\begin{aligned} \lim n \cdot \operatorname{cov}[c_n(\mathbf{s}), c_n(\mathbf{t})] &= c(\mathbf{s} - \mathbf{t}) - c(\mathbf{s})\,\bar{c}(\mathbf{t}) \\ &\quad + 2\sum_{r=1}^{\infty} \operatorname{cov}[e^{i\mathbf{s}'\mathbf{Y}_j}, e^{i\mathbf{t}'\mathbf{Y}_{j+r}}]. \end{aligned} \tag{5.3}$$

A mixing condition we now assume is the convergence of the infinite sum. For example, by the maximal property of the Gaussian correlation (e.g., Rényi 1959 and his reference to Gebelein) or otherwise, we may show this holds for stationary Gaussian processes with absolutely summable correlations (and more generally for stationary processes whose maximal correlations are absolutely summable). Now consider the moment equation

$$\int w(\mathbf{t})[c_n(\mathbf{t}) - c_\theta(\mathbf{t})]\, d\mathbf{t} = 0 \tag{5.4}$$

involving the weight function $w(t_1, t_2)$. A formal differential argument yields

$$n \operatorname{var}(\hat{\theta}) = \frac{\displaystyle\iint w(\mathbf{s})w(\mathbf{t})\, \boldsymbol{\Sigma}(\mathbf{s}, \mathbf{t})\, d\mathbf{s}\, d\mathbf{t}}{\left[\displaystyle\int w(\mathbf{t}) \frac{\partial c(\mathbf{t})}{\partial \theta}\, d\mathbf{t}\right]^2} \tag{5.5}$$

where $\hat{\theta}$ denotes a consistent root of (5.4) and $\boldsymbol{\Sigma}(\mathbf{s}, \mathbf{t})$ is given by (5.3). Taking

$$\begin{aligned} w(\mathbf{t}) &= w_{\theta_0}(\mathbf{t}) \\ &= \left[\frac{1}{2\pi}\right]^2 \int e^{-i\mathbf{t}'\mathbf{y}_j} \left.\frac{\partial \log f(x_j \mid x_{j-1})}{\partial \theta}\right|_{\theta_0} dx_{j-1}\, dx_j, \end{aligned} \tag{5.6}$$

we can show that

$$\begin{aligned} \int w_{\theta_0}(\mathbf{s})\, \boldsymbol{\Sigma}(\mathbf{s}, \mathbf{t})\, d\mathbf{s} &= E\left\{\frac{\partial \log f(X_2 \mid X_1)}{\partial \theta} e^{-i\mathbf{t}'\mathbf{Y}_2}\right\} \\ &\quad + 2\sum_{r=1}^{\infty} E\left\{\frac{\partial \log f(X_2 \mid X_1)}{\partial \theta} e^{-i\mathbf{t}'\mathbf{Y}_{2+r}}\right\}. \end{aligned} \tag{5.7}$$

Substituting this now in (5.5), we may show that the numerator reduces to

$$E_{\theta_0}\left[\frac{\partial \log f(X_j \mid X_{j-1})}{\partial \theta_0}\right]^2 \tag{5.8}$$

while the denominator reduces to the square of this term.

Hence, under the Markov assumption, (5.5) becomes the asymptotic Fisher information per observation. As before, the generalized function $w_{\theta_0}(\mathbf{t})$ may be consistently estimated and can be approximated by bounded integrable functions (by tapering of the score, say) to give robust procedures. The arbitrarily high efficiency of a k-L type procedure may now also be established as before.

The convergence of the infinite sum in (5.3) ensures the convergence in probability of $c_n(\mathbf{t})$ to $c_\theta(\mathbf{t})$ but (owing to the dependence) does not imply the asymptotic normality of $c_n(\mathbf{t})$. To prove this one may proceed by using the cumulant mixing conditions of Brillinger (1975, p. 26). To see this, consider the asymptotic distribution, for example, of $\sum_{j=1}^n \cos tX_j$. Let $p_l(X_j)$ be the degree l polynomial in the Taylor series of $\cos tX_j$; set $a = E(\cos tX_j)$, $a_l = E[p_l(X_j)]$ and

$$e_n^{\,l} = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} [(\cos tX_j - a) - (p_l(X_j) - a_l)].$$

If the mixing conditions hold, then

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} (p_l(X_j) - a_l) \xrightarrow{d} N(0, \sigma_l^2)$$

where

$$\sigma_l^2 = \lim_{n\to\infty} \operatorname{var} \frac{1}{\sqrt{n}} \left\{ \sum_{j=1}^{n} [p_l(X_j) - a_l] \right\}.$$

Let

$$\sigma^2 = \lim_{n\to\infty} \operatorname{var} \frac{1}{\sqrt{n}} \left\{ \sum_{j=1}^{n} (\cos tX_j - a) \right\}.$$

Then if (a) $\lim_{l\to\infty} \sigma_l^2 = \sigma^2 > 0$ and (b) $\lim_{l\to\infty} \lim_{n\to\infty} \operatorname{var}(e_n^{\,l}) = 0$ the asymptotic normality of $\sum_{j=1}^n \cos tX_j$ will follow from the lemma of Bernstein (c.f. Hannan 1970, p. 242). We note that the conditions (a) and (b) are satisfied by stationary ergodic Gaussian processes.

We now outline very briefly one application of the epcf to the stationary Markov emigration-immigration process (e.g., McDunnough 1979a). Briefly, this process $\{X_t,\ t = 0, \pm 1, \ldots\}$ takes on only positive integer values and has been used to model systems of infinitely many randomly moving particles. The process is determined by its bivariate probability generating function

$$E[z_1^{X_1} z_2^{X_2}] = \exp\{\nu[(z_1 - 1) + (z_2 - 1) + \rho(z_1 - 1)(z_2 - 1)]\}$$

where $\nu = E(X_1) = \operatorname{var}(X_1)$ and $\rho = \operatorname{corr}(X_1, X_2)$. The bivariate cf is obtained by setting $z_1 = e^{it_1}$, $z_2 = e^{it_2}$, and because of the discreteness we may take $-\pi \le t_1, t_2 \le \pi$. Now for ν we have the estimator $\bar{X}$, but the nature of the probability function complicates estimation of the parameter ρ, which is related to Avogadro's number. This process, however, satisfies all of the assumptions made here. Hence a k-L approach based on the epcf leads in a straightforward way to procedures having arbitrarily high asymptotic efficiency.

6. TECHNICAL NOTES

Let $\theta \in \Theta$ be an $l \times 1$ parameter and $\mathbf{T}_n$ be a $k \times 1$ statistic where $k \ge l$. The estimation procedures we have proposed follow one of two general types:

I Solve the random implicit equation

$$F(\theta, \mathbf{T}_n) = \mathbf{0}, \quad F: R^{l\times k} \to R^l.$$

II Minimize the criterion function

$$G(\theta, \mathbf{T}_n), \quad G: R^{l\times k} \to R.$$

We shall assume that G is continuous and that F is continuously differentiable. The fact that k is fixed and not dependent on n simplifies the asymptotic properties for our procedures, and in fact the following three-part result is adequate for our needs.

Theorem 6.1. Let θ_0 denote the actual θ, and let Θ be an open rectangle. Then

(a) If $\mathbf{T}_n \xrightarrow{\text{a.s.}} \boldsymbol{\lambda}(\theta_0)$, $F(\theta_0, \boldsymbol{\lambda}(\theta_0)) = \mathbf{0}$, and $\partial F(\theta_0, \boldsymbol{\lambda}(\theta_0))/\partial\theta$ is invertible, then there exists a statistic $\hat{\theta} \xrightarrow{\text{a.s.}} \theta_0$ that is an asymptotic random root for procedure I.

(b) If $\mathbf{T}_n$ is asymptotically $N(\boldsymbol{\lambda}(\theta_0), \boldsymbol{\Sigma}/n)$, F is as in (*a*), and $\hat{\theta} \xrightarrow{p} \theta_0$ is a root of I, then asymptotically $\hat{\theta}$ has a normal distribution with mean θ_0 and covariance matrix

$$\left[\frac{\partial F(\theta_0, \lambda(\theta_0))}{\partial\theta}\right]^{\prime -1} \frac{\boldsymbol{\Sigma}}{n} \left[\frac{\partial F(\theta_0, \boldsymbol{\lambda}(\theta_0))}{\partial\theta}\right]^{-1}.$$

(c) If $T_n \xrightarrow{\text{a.s.}} \boldsymbol{\lambda}(\theta_0)$, if $G(\theta, \lambda(\theta_0)) = 0$ when $\theta = \theta_0$ but not otherwise, and if $\hat{\theta}$ is a solution for II in the closure of Θ, here assumed bounded, then $\hat{\theta} \xrightarrow{\text{a.s.}} \theta_0$.

Part (a) of the theorem is proved as in Section 2 of McDunnough (1979b), whereas (b) follows on expanding F about $[\theta_0, \boldsymbol{\lambda}(\theta_0)]$ and using a differential argument (c.f. Rao 1973, p. 385). The proof for (c) is given in FM. We remark that for (b) the differentiability of F is required only at $(\theta_0, \lambda(\theta_0))$.

We next give the regularity conditions, which we require for Theorem 2.1. Following through in multiparameter form the arguments given in FM, we see that the following mild conditions are sufficient.

I Θ is an open rectangle.
II The covariance $\boldsymbol{\Sigma}$ of $\mathbf{z}_n$ is invertible at $\theta = \theta_0$.
III $(\partial\mathbf{z}/\partial\theta)' \,\boldsymbol{\Sigma}^{-1} (\partial\mathbf{z}/\partial\theta)$ is invertible at $\theta = \theta_0$.
IV $c_\theta(t)$ is continuously differentiable (in θ) at $\theta = \theta_0$.

We remark that conditions II and III hold very generally because of linear independence properties of the functions e^{it_jx}.

The efficiency statement in the theorem requires in addition the following conditions.

V $(\partial \ln f_\theta(x)/\partial\theta_i)\,(\partial \ln f_\theta(x)/\partial\theta_j)\, f_\theta(x)$ is integrable in x for $i, j = 1, 2, \ldots, l$.
VI $\partial \ln f_\theta(x)/\partial\theta_i$ is integrable over bounded intervals in x.

VII $(\partial \ln f_\theta(x)/\partial\theta_i)^2$ is integrable over bounded intervals in x.

VIII $c_\theta(t)$ and $\partial c_\theta(t)/\partial\theta_i$ are integrable in t at $\theta = \theta_0$. (For lattice distributions we require only that the derivative be integrable over a single period.)

IX $\int c_\theta(t)e^{-itx}\,dt$ can be differentiated (with respect to θ) through the integral sign.

Finally, we remark that the Parseval theorem is justified in the context of the generalized functions appearing in (1.3) and (3.2). For example, the existence of the weights $w_{\theta_i}(t)$ as generalized functions follows under the very mild requirement that the score functions are bounded by polynomials. It can then be readily established that the expressions (1.3) and (3.2) may be given consistent sequential interpretations under which the indicated Parseval equalities are immediately justified. The sequential approach to generalized functions is given in Antosik, Mikusinksi, and Sikorski (1973). For an alternative treatment, see Lighthill (1970).

7. SOME CONCLUDING REMARKS

In this section we summarize a number of issues that remain to be resolved: the range of these mirrors, to some extent, the range of potential applications of the methods proposed here, and certain generalizations of these methods. Some of these will be explored in a subsequent work.

For stable laws a number of questions remain open. What is the optimal spacing of points in the frequency domain? In conjunction with this, what properties do the procedures have for finite sample sizes? On another front, what is the form of the weight functions (1.4) corresponding to the parameters of the stable laws? (An *analytical* solution of this problem would indeed represent a fundamental contribution.) How much can be done by way of tractable approximation schemes to either the weight or the score functions?

The Fourier domain supports a natural characterization of independence. It is evident that in some appropriate sense, our efficiency results will carry over to such testing contexts. One feels heuristically that in physically realistic models, a few points on the multivariate ecf will generally be more informative than a few points on the multivariate cdf for testing independence. This matter requires investigation. Similar questions arise in the context of testing for goodness of fit.

The ecf may lend itself to adaptive inference more conveniently than density based approaches. For example, suppose an unknown cf is given by $c(t) = e^{i\mu t}\,\phi\,(\sigma t)$, where μ, σ are location and scale parameters and ϕ is unknown but "standardized" according to the meanings attributed to μ, σ. Then we may estimate ϕ by the ecf $\hat{\phi}$ of the "standardized" sample $(X_j - \hat{\mu})/\hat{\sigma}$ where $\hat{\mu}$, $\hat{\sigma}$ are any consistent estimates and then apply the k-L procedure to estimate μ and σ using $\hat{\phi}$ and the covariance structure as estimated from $\hat{\phi}$. It seems reasonable to expect that this approach will lead to estimates having the same asymptotic properties as those of the k-L procedure with ϕ known.

The pcf of Section 5 may emerge to be a useful new approach for dealing with certain problems in stationary time series analysis. For example, the wide class of linear processes involves convolutions whose densities are (except in Gaussian cases) rarely tractable; the pcf is, however, a quite natural tool here. Second, any likelihood approach that involves the approximation of a time series as an autoregressive process of order p can be transformed via a Parseval-type argument to an approach based on the p-variate epcf. A k-L procedure will then approximate arbitrarily well (asymptotically) the efficiency attainable with such a likelihood approach. Finally, the pcf lends itself in a natural way to approaching problems in nonlinear time series modeling. Such problems may quite generally be formulated by using the idea of time-invariant transformation of a process to independence, which we may characterize, in turn, by means of the pcf functions.

It is interesting to note that it is possible to define a stationary version of the ecf (the secf, say) by averaging not $\exp(itX_j)$, but rather $\exp(itX_j + \Phi_j)$, where the Φ_j are independent random variables with uniform distribution on $[-\pi, \pi]$ and independent of the X's. This is essentially the quantogram of Kendall (1974, 1977) and Kent (1975). One may easily see that the ecf and secf are equivalent in a certain analytic sense, and it is tempting to conjecture that the secf supports tractable asymptotically efficient inference procedures. We point out, however, that it is readily established that the Fisher information at k points of the secf increases with k, but essentially does not depend on n.

We end with a conjecture. It is natural to wonder what it is about Fourier transformation that leads to the arbitrarily high asymptotic efficiency of the discrete procedures and to what extent this useful result can be generalized. Monte Carlo studies by Quandt and Ramsay (1978) and Leslie and Khalique (1980) indicate that useful efficiencies are possible for procedures based on the empirical Laplace transform in certain cases. Numerical evaluations by Brockwell and Brown (1980) show high efficiencies for a procedure that estimates the so-called positive stable laws based on fitting certain negative moments EX^{-t} empirically. In a context unrelated to this, Brockwell and Brown also give a completeness result involving the positive laws and the functions x^{-t}. Our conjecture is that the efficiency results of the k-L type we have obtained will hold for a process $n^{-1}\sum_1^n g(t, X_j)$ essentially if and only if the class of functions $\{g(t, x)\}_{t\in T}$ is such that, for arbitrary $\boldsymbol{\theta}$ and for an arbitrary countable subset $\{t_j\}$ dense in T, the score functions may be approximated arbitrarily well in a certain sense by finite linear combinations in the class of random variables $\{g(t_j, X)\}$. We shall consider this matter in our subsequent work.

[*Received August 1979. Revised August 1980.*]

REFERENCES

ANTOSIK, P., MIKUSINSKI, J., and SIKORSKI, R. (1973), *Theory of Distributions: The Sequential Approach,* New York: Elsevier.

BRILLINGER, D.R. (1975), *Time Series: Data Analysis and Theory,* New York: Holt, Rinehart & Winston.

BROCKWELL, P.J., and BROWN, B.M. (1980), "Efficient Estimation for the Positive Stable Laws," unpublished manuscript.

COURANT, R., and HILBERT, D. (1953), *Methods of Mathematical Physics I,* New York: Interscience.

CSORGO, S. (1981), "Limit Behaviour of the Empirical Characteristic Function," *Annals of Probability,* in press.

DuMOUCHEL, W.H. (1975), "Stable Distributions in Statistical Inference: 2. Information From Stably Distributed Samples," *Journal of the American Statistical Association,* 70, 386–393.

FEUERVERGER, A., and McDUNNOUGH, P. (1981), "On the Efficiency of Empirical Characteristic Function Procedures," *Journal of the Royal Statistical Society,* Ser. B, 43, in press.

GRENANDER, U., and SZEGO, G. (1958), *Toeplitz Forms and Their Applications.* Berkeley: University of California Press.

HANNAN, E.J. (1970), *Multiple Time Series,* New York: Wiley.

HEATHCOTE, C.R. (1977), "The Integrated Squared Error Estimation of Parameters," *Biometrika,* 64, 255–264.

HUBER, P.J. (1964), "Robust Estimation of a Location Parameter," *Annals of Mathematical Statistics,* 35, 73–101.

——— (1977), *Robust Statistical Procedures,* Philadelphia: Society for Industrial and Applied Mathematics.

KENDALL, D.G. (1974), "Hunting Quanta," *Philosophical Transactions of the Royal Society of London,* Ser. A, 276, 231–266.

——— (1977), "Hunting Quanta " (2nd ed.), *Proceedings of the Symposium to Honour Jerzy Neyman, Warszawa, 1974,* Warsaw: Polish Scientific Publishers, 111–159.

KENT, T.J. (1975), "A Weak Convergence Theorem for the Empirical Characteristic Function," *Journal of Applied Probability,* 12, 515–523.

LESLIE, R.T., and KHALIQUE, A. (1980), "Parametric Estimation Via Laplace Transforms," unpublished manuscript.

LIGHTHILL, M.J. (1970), *Fourier Analysis and Generalized Functions,* New York: Cambridge University Press.

McDUNNOUGH, P. (1979a), "Estimating the Law of Randomly Moving Particles," *Journal of Applied Probability,* 16, 25–35.

——— (1979b), "The Estimation of an Autocorrelation Parameter of a Gaussian Vector Process," *Journal of the Royal Statistical Society,* Ser. B, 41, 203–209.

QUANDT, R.E., and RAMSAY, J.B. (1978), "Estimating Mixtures of Normal Distributions and Switching Regressions," *Journal of the American Statistical Association,* 73, 730–752.

RAO, C.R. (1973), *Linear Statistical Inference and Its Applications,* New York: John Wiley.

RÉNYI, A. (1959), "On Measures of Dependence," *Acta Mathematica Academiae Scientiarum Hungaricae* 10, 441–451.

RIESZ, F., and SZ-NAGY, B. (1955), *Functional Analysis,* New York: Unger.

SSP (1962), *Scientific Subroutine Package,* IBM, 360A-CM-07X.

THORNTON, J.C., and PAULSON, A.S. (1977), "Asymptotic Distributions of Characteristic Function-Based Estimators for the Stable Laws," *Sankhyā,* 39, 341–354.

WHITTLE, P. (1951), *Hypothesis Testing in Time Series Analysis,* Uppsala: Almqvist and Wiksell.