On the Efficiency of Empirical Characteristic Function Procedures

Andrey Feuerverger; Philip McDunnough

Journal of the Royal Statistical Society. Series B (Methodological), Vol. 43, No. 1 (1981), 20-27.

Stable URL:
http://links.jstor.org/sici?sici=0035-9246%281981%2943%3A1%3C20%3AOTEOEC%3E2.0.CO%3B2-3

J. R. Statist. Soc. B (1981),
43, *No.* 1, *pp.* 20–27

On the Efficiency of Empirical Characteristic Function Procedures

By Andrey Feuerverger and Philip McDunnough
University of Toronto

[Received November 1978. Final revision July 1979]

Summary

The asymptotic normality and arbitrarily high efficiency of some new statistical procedures based on the empirical characteristic function is established under general conditions.

Keywords: EMPIRICAL CHARACTERISTIC FUNCTION; ASYMPTOTIC EFFICIENCY; ASYMPTOTIC NORMALITY; MAXIMUM LIKELIHOOD ESTIMATION; FISHER INFORMATION

1. Introduction

Many distributions of statistical interest arise as sums of independent random variables. A celebrated example (aside from the classical Gaussian distribution) is given by the stable laws. Another, which arises in statistical physics (cf. McDunnough, 1979a), is the convolution of a binomial with a Poisson. Characteristic of such distributions is the generally unmanageable form of the probability density function (p.d.f.). Indeed, for the stable laws, no closed form for the densities is available except in a few special cases. This lack of a tractable form for the p.d.f. makes estimation via maximum likelihood of the parameters of these distributions extremely difficult. Consequently there is a need for alternative methods which effectively deal with inference problems involving such distributions.

Let $X_1, X_2, \ldots$ be a sample from a distribution with c.d.f. $F_\theta(x)$ and characteristic function $c_\theta(t) = \int e^{itx}\, dF_\theta(x)$ where θ is an unknown parameter. Note that the characteristic function of a sum of independent variables will, in general, be relatively simple. Now denote by $F_n(x)$ the empirical distribution function. Then the empirical characteristic function (e.c.f.) is defined as

$$c_n(t) = \int e^{itx}\, dF_n(x) = n^{-1} \sum_1^n \exp(itX_j).$$

Numerous authors have proposed the use of e.c.f. procedures in various statistical contexts. Most of these applications have been to problems that could not be treated conveniently by alternative methods. However e.c.f.-based tests for symmetry and goodness of fit have also appeared. See Heathcote (1972, 1977), Press (1972) and Paulson, Halcomb and Leitch (1975), for example. The e.c.f. is of interest also in connection with probability density estimation. See, for example, Parzen (1962). More recently, Feuerverger and Mureika (1977) have initiated the systematic study of e.c.f. properties with a view towards statistical applications.

A careful study of results such as appear in DuMouchel (1975) and Feigin and Heathcote (1976) in addition to the papers cited above reveals that e.c.f. procedures can have substantial efficiency. A further motivating example is afforded by the problem of estimating the scale θ in a centred Cauchy distribution having cf. $c(t) = e^{-\theta|t|}$. Let us consider an approach via maximum likelihood based on the asymptotic normal distribution (see Section 2) of the e.c.f. $c_n(t)$ at k equispaced points $t = \tau j, j = 1, 2, \ldots, k$. Thus suppose $\hat{\mathbf{c}}_k$ and $\mathbf{c}_k = E(\hat{\mathbf{c}}_k)$ are column vectors whose entries are $\operatorname{Re} c_n(\tau j)$ and $c(\tau j)$ respectively. Then the asymptotic Fisher information per unit observation is given by

$$I = n^{-1}\operatorname{var}\left\{\frac{\partial}{\partial\theta}[-\tfrac{1}{2}(\hat{\mathbf{c}}_k-\mathbf{c}_k)'\mathbf{\Sigma}_k^{-1}(\hat{\mathbf{c}}_k-\mathbf{c}_k)]\right\}$$
$$=\boldsymbol{\xi}_k'\mathbf{\Sigma}_k^{-1}\boldsymbol{\xi}_k,$$

where $\xi_k' = (\partial/\partial\theta)\mathbf{c}_k$ has entries $-\tau\beta^j$ and $\mathbf{\Sigma}_k$ has (i,j)th entry $\beta^{|i-j|}-\beta^{i+j}$, where $\beta = e^{-\theta\tau}$. Now, for $\mathbf{\Sigma}_k$ we may obtain an explicit inverse by noting that it is a sum of two matrices: the first being the Toeplitz form associated with an autoregressive time series of order one whose inverse is known; and the second being a matrix of unit rank. Hence using the identity

$$(\mathbf{A}+\mathbf{BB}')^{-1} = \mathbf{A}^{-1}-(1+\mathbf{B}'\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{A}^{-1}\mathbf{BB}'\mathbf{A}^{-1},$$

where $\mathbf{A}$ is a matrix and $\mathbf{B}$ a vector, we obtain

$$\mathbf{\Sigma}_k^{-1} = \frac{2}{1-\beta^2}\begin{bmatrix} 1+\beta^2 & -\beta & & \\ -\beta & 1+\beta^2 & \ddots & \\ & \ddots & 1+\beta^2 & -\beta \\ & & -\beta & 1 \end{bmatrix}$$

It therefore follows, after some algebra, that

$$I \to \frac{2\tau^2\beta^2}{(1-\beta^2)^2} \quad \text{as } k\to\infty$$
$$\to 1/2\theta^2 \qquad \text{as } \tau\to 0$$

thus indicating the asymptotic efficiency of the procedure.

As a matter of convenience, our discussion throughout is confined to the case where θ is a single real parameter whose true value θ_0 lies in some open interval. We also assume that the c.d.f's $F_\theta(x)$ are absolutely continuous with densities $f_\theta(x)$.

Some properties of the e.c.f. are briefly reviewed in Section 2. The new procedures are given in Sections 3 and 4 and their efficiencies examined in Section 5. Some concluding remarks appear in Section 6.

2. SOME PROPERTIES OF THE E.C.F.

If $c_n(t)$ is the e.c.f. corresponding to a characteristic function $c(t)$ then we have, for fixed $T<\infty$, the convergence

$$\sup_{|t|\leq T}|c_n(t)-c(t)|\to 0$$

almost surely as $n\to\infty$. This result appears in Feuerverger and Mureika (1977) (hereafter FM) where it is shown that the uniform convergence cannot in general take place on the whole line. It is also shown that the convergence does hold on the whole line in the purely discrete case and in certain cases where the $c_n(t)$ are suitably tapered, or equivalently, corresponds to special density estimators. It is true also that this convergence result holds in general for certain sequences $T = T_n\to\infty$. See FM and Csorgo (1979). Such results naturally lead one to expect that the e.c.f. has valuable statistical applications.

Suppose that X corresponds to $c(t)$ and consider e^{itX} as a complex random process in t. Then $E(e^{itX}) = c(t)$ and $E(e^{isX}e^{itX}) = c(s+t)$ fully determine the first and second moment structure of this process. Now if $c_n(t) = n^{-1}\Sigma\exp(itX_j)$ (summing over j from 1 to n) is the associated e.c.f.

process then $c_n(t)$ is seen to be an average of n independent processes of the first type. It follows that $Y_n(t) = n^{\frac{1}{2}}(c_n(t)-c(t))$ has mean zero and the covariance structure:

$$\left.\begin{aligned}\operatorname{cov}(\operatorname{Re} Y_n(s), \operatorname{Re} Y_n(t)) &= \tfrac{1}{2}[\operatorname{Re} c(s+t)+\operatorname{Re} c(s-t)]-\operatorname{Re} c(s)\operatorname{Re} c(t),\\ \operatorname{cov}(\operatorname{Re} Y_n(s), \operatorname{Im} Y_n(t)) &= \tfrac{1}{2}[\operatorname{Im} c(s+t)+\operatorname{Im} c(s-t)]-\operatorname{Re} c(s)\operatorname{Im} c(t),\\ \operatorname{cov}(\operatorname{Im} Y_n(s), \operatorname{Im} Y_n(t)) &= \tfrac{1}{2}[-\operatorname{Re} c(s+t)+\operatorname{Re} c(s-t)]-\operatorname{Im} c(s)\operatorname{Im} c(t).\end{aligned}\right\} \quad (2.1)$$

Because $c_n(t)$ is an average of bounded processes it follows also, by means of the multidimensional central limit theorem, that $Y_n(t_1), Y_n(t_2), \ldots, Y_n(t_m)$ converge in distribution to $Y(t_1), Y(t_2), \ldots, Y(t_m)$ where $Y(t) = \overline{Y(-t)}$ is a complex Gaussian process with covariance structure identical to Y_n: $EY(t) = 0$, and $EY(s)\,Y(t) = c(s+t)-c(s)\,c(t)$. FM prove the weak convergence of $Y_n(t)$ to $Y(t)$ in any finite interval provided that $E\,|\,X\,|^{1+\delta} < \infty$. However, the moment restriction is not easily removed and the entire question is one of considerable depth. Some clarification of this is given by Csorgo (1979).

For our purposes, we shall need only the following:

Lemma 2.1. Let $W(t)$ be a function of bounded variation with $dW(t) = \overline{dW(-t)}$. Then

$$U_n = \int_{-\infty}^{\infty} \sqrt{n}(c_n(t)-c(t))\,dW(t) \quad (2.2)$$

has mean zero and variance $\iint K(s,t)\,dW(s)\,dW(t)$ where $K(s,t) = c(s+t)-c(s)\,c(t)$, and is asymptotically normal with the same mean and variance.

This result follows on applying the Central Limit Theorem to the variables $\int \exp(itX_j)\,dW(t)$, $j = 1, 2, \ldots, n$.

3. THE $k-L$ PROCEDURE

The term "$k-L$ procedure" will refer to maximum likelihood estimation based on the asymptotic distribution at k points of the e.c.f., where k and the points $t_1, t_2, \ldots, t_k$ are fixed and do not vary with the sample size n.

Thus suppose $c_n(t_1), c_n(t_2), \ldots, c_n(t_k)$ to be the values of the e.c.f. at the sampled points. From (2.1) we obtain (to within a constant) the approximate log-likelihood

$$-k \log(\det \boldsymbol{\Sigma}) - n(\mathbf{z}_n-\mathbf{z})'\boldsymbol{\Sigma}^{-1}(\mathbf{z}_n-\mathbf{z}),$$

where $\mathbf{z}' = (\operatorname{Re} c(t_1), \ldots, \operatorname{Re} c(t_k), \operatorname{Im} c(t_1), \ldots, \operatorname{Im} c(t_k))$, $\mathbf{z}_n$ is its empirical counterpart, and $\boldsymbol{\Sigma} = \operatorname{var}(\mathbf{z}_n)$. This suggests estimation by minimizing the criterion function

$$l_n(\theta) = (\mathbf{z}_n-\mathbf{z})'\boldsymbol{\Sigma}^{-1}(\mathbf{z}_n-\mathbf{z}).$$

Note, incidentally, that $l_n(\theta)$ involves only the characteristic function $c_\theta(t)$ and not, directly, the p.d.f.

We thus consider the approximate likelihood equations

$$F(\theta, \mathbf{z}_n) = 0, \quad (3.1)$$

where $F(\theta, \mathbf{z}_n) = \partial l_n(\theta)/\partial\theta$. This is a random implicit equation which depends only on a fixed finite number of statistics. The asymptotic behaviour of such estimation equations is considered in McDunnough (1979b). Briefly, in order that there exist a statistic $\hat{\theta}$ which is an asymptotic random root of (3.1), and consistent for θ_0, we require that $\mathbf{z}_n$ converge almost surely to some $\boldsymbol{\lambda}(\theta_0)$ satisfying (3.1) and that $F(\cdot,\cdot)$ be continuously differentiable with

$$\frac{\partial F(\theta_0, \boldsymbol{\lambda}(\theta_0))}{\partial\theta} \neq 0.$$

These requirements, of course, are met here. In particular

$$\frac{\partial F(\theta, \mathbf{z})}{\partial \theta} = 2\frac{\partial \mathbf{z}'}{\partial \theta}\mathbf{\Sigma}^{-1}\frac{\partial \mathbf{z}}{\partial \theta} \neq 0.$$

In fact, a stronger result may be obtained. Assume θ defined in $[a, b]$ but known to lie in (a, b). Suppose θ to be identifiable from $\boldsymbol{\lambda}(\theta) = \mathbf{z}$ in $[a, b]$ in the sense that different θ's yield different $\mathbf{z}$'s, and suppose $\mathbf{\Sigma}^{-1}$ exists for all θ. Let $\hat{\theta}$ minimize $l_n(\theta)$ over $[a, b]$. Then necessarily $\hat{\theta} \to \theta_0$ almost surely. For if $\hat{\theta} \to \theta' \neq \theta_0$ almost surely then $l_n(\hat{\theta}) \not\to 0$ which clearly is impossible. If $\hat{\theta}$ does not converge to any value, the argument applies to any convergent subsequence.

So now, let $\hat{\theta}$ be any consistent root of (3.1). A standard differential argument shows that $\hat{\theta}$ is asymptotically normal with mean θ and asymptotic variance

$$\left(\frac{\partial \mathbf{z}'}{\partial \theta}\mathbf{\Sigma}^{-1}\frac{\partial \mathbf{z}}{\partial \theta}\right)^{-1}, \tag{3.2}$$

all quantities being evaluated at the true θ_0. (Specific reference to this is omitted for convenience of printing.)

For fixed k we have the design question for choice of the t_j. It is convenient to use equal spacing: $t_j = j\tau$, $j = 1, 2, ..., k$ and to set the sampling rate τ to minimize (3.2), or at least its estimated version. For the Cauchy scale family of Section 1 for example, the maximal asymptotic efficiency thus obtained is 61 per cent for $k = 1$, 81 per cent for $k = 2$ and just over 90 for $k = 3$. In fact, in Section 5 we show that by using a sufficiently extensive grid $\{t_j\}$, (3.2) can be made arbitrarily close to the Cramér–Rao bound, so that the $k-L$ procedure can attain arbitrarily high asymptotic efficiency.

We point out that the $k-L$ procedure lends itself conveniently to adaptive inference in the special location-scale case with unknown distributional form. Thus suppose that an unknown c.f. is given by

$$c(t) = e^{-i\mu t}\,\phi(\sigma t),$$

where μ, σ are location, scale parameters, and ϕ is unknown. Note that the meanings attributable to μ, σ are determined by the manner of "standardization" used for ϕ. For a discussion of this point, see Fraser (1976). The adaptive procedure may now be described as follows: first ϕ is estimated by $\hat{\phi}$, the e.c.f. of the standardized sample $(X_j - \hat{\mu})/\hat{\sigma}$, $j = 1, 2, ..., n$. Here $\hat{\mu}, \hat{\sigma}$ are consistent estimates in correspondence with the intended manner of standardization. We then apply the $k-L$ procedure at points $t_1, t_2, ..., t_k$ to estimate μ and σ using $\hat{\phi}$. The needed covariance structures may be estimated from $\hat{\phi}$. This procedure clearly should lead, under general conditions, to estimates having the same asymptotic properties as those derived from the $k-L$ procedure with distributional form known. However a rigorous proof of this assertion is not in our present scope.

Finally, we note that the idea of basing inference on several e.c.f. points occurs, for example, in Press (1972, 1975) and Heathcote and Feigin (1976). See also Quandt and Ramsay (1978). The idea of using maximum likelihood based on asymptotic forms in this context does however appear to be new.

4. A MOMENT ESTIMATOR

An alternative to the $k-L$ procedure may be obtained by generalizing the moment estimators of Press (1972, 1975). Thus consider the estimation equation

$$\mathbf{d}'(\mathbf{z}_n - \mathbf{z}) = 0, \tag{4.1}$$

where $\mathbf{z}_n$, $\mathbf{z}$ are as in Section 3, and $\mathbf{d}$ is a $2k \times 1$ vector of *known* constants. If $\hat{\theta}$ is a consistent root of (4.1) then to first order (with terms evaluated at θ_0)

$$\hat{\theta} = \theta_0 - \mathbf{d}'(\mathbf{z}_n - \mathbf{z}) \Big/ \left(\mathbf{d}' \frac{\partial \mathbf{z}}{\partial \theta}\right)$$

and we have, by Lemma 2.1, that $\hat{\theta}$ is asymptotically normal with asymptotic mean θ_0 and asymptotic variance

$$\sigma^2 = \mathbf{d}' \boldsymbol{\Sigma} \mathbf{d} \Big/ \left(\mathbf{d}' \frac{\partial \mathbf{z}}{\partial \theta}\right)^2 \tag{4.2}$$

with $\boldsymbol{\Sigma}$ as before. This quantity is minimized by choosing (evaluated at $\theta = \theta_0$)

$$\mathbf{d} = \mathbf{d}_0 = \boldsymbol{\Sigma}^{-1} (\partial \mathbf{z} / \partial \theta)$$

and in this case (4.2) reduces to (3.2). Thus the $k-L$ procedure, and (4.1) with $\mathbf{d} = \mathbf{d}_0$ give asymptotically equivalent procedures. Of course $\mathbf{d}_0$ is not known in practice (a "deficiency", we note, that is not shared by the $k-L$ procedure) but following a standard argument, use of *any* consistent estimate of $\mathbf{d}_0$ will yield a procedure having the same asymptotic properties as the $k-L$ procedure. Note that $\mathbf{d}_0$ involves only the characteristic function (and its derivative) evaluated at a finite number of points. Therefore these procedures, like those of the previous section could be easily used in estimation problems involving, for example, the stable laws.

5. EFFICIENCY OF THE PROCEDURES

For our present purpose, it will be convenient to think of the estimation equation (4.1) in a form such as

$$\mathbf{w}'(\mathbf{c}_n - \mathbf{c}) = 0,$$

where $\mathbf{c}' = (c(t_1), \ldots, c(t_k), \overline{c(t_1)}, \ldots, \overline{c(t_k)})$ and $\mathbf{c}_n$ is its empirical version. Here $\mathbf{w}'$ has the form $\mathbf{w}' = (w_1, \ldots, w_k, \overline{w_1}, \ldots, \overline{w_k})$, and again, some θ-subscripts are suppressed. It is also convenient to think in terms of a uniform spacing for the $\{t_j\}$.

We then note that (4.1) may be considered as the form of a discrete approximation to the continuous-type estimation equation

$$\int_{-\infty}^{\infty} w(t)(c_n(t) - c(t))\, dt = 0, \tag{5.1}$$

where $w(t) = \overline{w(-t)}$ is an integrable function. We then have that a consistent root $\hat{\theta}$ is asymptotically normal with asymptotic variance

$$n \operatorname{var}(\hat{\theta}) = \iint w(s)\, w(t)\, K(s,t)\, ds\, dt \Big/ \left(\int w(s) \frac{\partial c(s)}{\partial \theta} ds \right)^2, \tag{5.2}$$

where

$$K(s,t) = c(s+t) - c(s)\, c(t). \tag{5.3}$$

Note that a corresponding approximation of (5.2) will have the form (4.2). To minimize $\operatorname{var}(\hat{\theta})$ we use a variational method.

Now, since (5.2) is invariant under rescaling of w, our minimum will in fact be that of

$$\iint w(s)\, w(t)\, K(s,t)\, ds\, dt \tag{5.4}$$

subject to

$$\int w(t) \frac{\partial c(t)}{\partial \theta} dt = 1. \tag{5.5}$$

Suppose $w(t)$ is already our solution but now is replaced by $w(t)+\delta w(t)$ in (5.4) to give

$$\iint [w(s)\,w(t)+2w(s)\,\delta w(t)+\delta w(s)\,\delta w(t)]\; K(s,t)\,ds\,dt. \tag{5.6}$$

Due to (5.5) our variant must satisfy

$$\int \delta w(t)\frac{\partial c(t)}{\partial\theta}\,dt = 0. \tag{5.7}$$

Following the usual variational argument, the last term of (5.6) being of second order may be neglected. Then because the first term is assumed to be at a minimum the middle term must be always zero. Thus for all $\delta w(t)$ satisfying (5.7),

$$\int \delta w(t)\left(\int w(s)\,K(s,t)\,ds\right)dt = 0.$$

This means that the function $\int w(s)\,K(s,t)\,ds$ is orthogonal to every function which is orthogonal to $\partial c(t)/\partial\theta$. This suggests that

$$\int w(s)\,K(s,t)\,ds = \lambda\frac{\partial c(t)}{\partial\theta}. \tag{5.8}$$

To solve this integral equation, we first abandon the constraint (5.5) in favour of $\lambda = 1$ and apply the inverse Fourier transformation to (5.8) recalling that $K(s,t)$ is given by (5.3). For the right-hand side of (5.8) this gives

$$\frac{1}{2\pi}\int \frac{\partial c(t)}{\partial\theta}e^{-itx}\,dt = \frac{\partial f_\theta(x)}{\partial\theta}$$

assuming interchangeability of the derivative and integral. The left-hand side, on the other hand, becomes

$$\frac{1}{2\pi}\iint w(s)\,c(s+t)\,e^{-itx}\,ds\,dt - \left(\int w(s)\,c(s)\,ds\right) f(x)$$

and the first term reduces to

$$\begin{aligned}\frac{1}{2\pi}\iint w(s)\,c(s+t)\,e^{-i(s+t)x}\,e^{isx}\,ds\,dt &\\ = \int w(s)\,f(x)\,e^{isx}\,ds &\\ = f(x)\int w(s)\,e^{isx}\,ds.&\end{aligned}$$

The equation resulting is therefore

$$\int w(s)\,e^{isx}\,ds = \int w(s)\,c(s)\,ds + \frac{1}{f(x)}\frac{\partial f(x)}{\partial\theta}.$$

If we ignore the constant term on the right, which leads only to a Dirac function at the origin, we obtain, after an additional inverse Fourier transformation, the following result:

$$w(t) = w_0(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \left.\frac{\partial \log f_\theta(x)}{\partial\theta}\right|_{\theta_0} e^{-itx}\,dx. \tag{5.9}$$

We remark first that because $(c_n(t)-c_\theta(t))$ is uninformative at the origin and continuous around the origin it is clearly reasonable to expect that removal of a Dirac function at the origin should obtain without cost. Of more immediate concern is the fact that the score function $\partial \log f_\theta(x)/\partial\theta\,|_{\theta_0}$ need not be integrable.

First, let us show formally that (5.9) leads to an asymptotically efficient solution. Now,

$$\begin{aligned}\int w(s)\,K(s,t)\,ds &= \frac{1}{2\pi}\iint e^{-i(s+t)x}\,e^{itx}\,\frac{\partial \log f(x)}{\partial\theta}\,c(s+t)\,dx\,ds\\ &= \frac{1}{2\pi}\iint e^{-isx}\,\frac{\partial \log f(x)}{\partial\theta}\,c(s)\,c(t)\,dx\,ds\\ &= \int e^{itx}\,\frac{\partial \log f(x)}{\partial\theta}\,f(x)\,dx - c(t)\int\frac{\partial \log f(x)}{\partial\theta}\,f(x)\,dx.\end{aligned}$$

Then, provided that the derivative may be moved outside the integral in each case, the second term is zero and the first is $\partial c(t)/\partial\theta$. It follows that (5.9) is a solution of (5.8) with $\lambda = 1$. Substituting first (5.8) into (5.2) and then using (5.9) gives

$$\begin{aligned}(n\,var\,(\hat\theta))^{-1} &= \int w(t)\,\frac{\partial c(t)}{\partial\theta}\,dt\\ &= \frac{\partial}{\partial\theta}\iint w(t)\,f(x)\,e^{itx}\,dx\,dt\\ &= \frac{\partial}{\partial\theta}\int\left.\frac{\partial \log f(x)}{\partial\theta}\right|_{\theta_0} f(x)\,dx\\ &= E\left(\left.\frac{\partial \log f_\theta(x)}{\partial\theta}\right|_{\theta_0}\right)^2 = I(\theta_0),\end{aligned}$$

where $I(\theta_0)$ in fact is the Fisher information.

To resolve the question of integrability of the score, we work with

$$w_m(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\partial \log f_\theta(x)}{\partial\theta}\,h_m(x)\,e^{-itx}\,dx$$

where

$$h_m(x) = \begin{cases}1, & \text{for } |x| \leqslant m,\\ 0, & \text{otherwise.}\end{cases}$$

(This representation for the truncated integral serves to stress the possibility of more general "tapering" functions $h_m(x)$.) Note that $w_m(t)$ is continuous and integrable.

We substitute $w_m(t)$ directly into (5.2). First as to the denominator, we note

$$\begin{aligned}\int_{-\infty}^{\infty} w_m(t)\,\frac{\partial c(t)}{\partial\theta}\,dt &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-m}^{m}\frac{\partial \log f(x)}{\partial\theta}\,e^{-itx}\,\frac{\partial c(t)}{\partial\theta}\,dx\,dt\\ &= \int_{-m}^{m}\frac{\partial \log f(x)}{\partial\theta}\,\frac{\partial f(x)}{\partial\theta}\,dx \to I(\theta_0).\end{aligned}$$

Similarly,

$$\int_{-\infty}^{\infty} w_m(t)\, c(t)\, dt = \int_{-m}^{m} \frac{\partial \log f(x)}{\partial \theta} f(x)\, dx \to 0$$

so that the numerator effectively is

$$\begin{aligned}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} & w_m(s)\, w_m(t)\, c(s+t)\, ds\, dt \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-m}^{m} w_m(t) \frac{\partial \log f(x)}{\partial \theta} e^{-i(s+t)x} e^{itx}\, c(s+t)\, dx\, ds\, dt \\ &= \int_{-\infty}^{\infty}\int_{-m}^{m} w_m(t) \frac{\partial \log f(x)}{\partial \theta} f(x)\, e^{itx}\, dx\, dt \\ &= \int_{-m}^{m} \left(\frac{\partial \log f(x)}{\partial \theta}\right) f(x)\, dx \to I(\theta_0).\end{aligned}$$

It follows that (5.2) can be made arbitrarily close to $I^{-1}(\theta_0)$ by a proper selection of $w(t)$. Now any $w(t)$ can be approximated by a step function which implies that (4.2) and hence (3.2) can be made arbitrarily close to $I^{-1}(\theta_0)$. This establishes the arbitrarily high efficiency of the $k-L$ procedure (as well as the moment method).

6. CONCLUSION

This article has proposed two highly efficient estimation procedures based on the empirical characteristic function. The primary motivation for these approaches has been to handle problems where maximum likelihood estimation is difficult. The methods have some generality. And they lend themselves to adaptive inference perhaps more conveniently than density based approaches.

Finally, we note that the procedures here considered should generally be robust in that outlying values will be trigonometrically reduced. On this point, see also Heathcote (1977).

Added in proof: For further work the readers is referred to a paper submitted by the authors to J. Amer. Statist. Ass.

ACKNOWLEDGEMENT

This research was supported by NSERC, Canada.

REFERENCES

CSÖRGÖ, S. (1979). Limit behaviour of the empirical characteristic function. Unpublished.

DUMOUCHEL, W. H. (1975). Stable distributions in statistical inference: 2. Information from stably distributed samples. *J. Amer. Statist. Ass.*, **70**, 386–393.

FEIGIN, P. D. and HEATHCOTE, C. R. (1976). The empirical characteristic function and the Cramer–Von Mises statistic. *Sankhyā* A, **38**, 309–325.

FEUERVERGER, A. and MUREIKA, R. A. (1977). The empirical characteristic function and its applications. *Ann. Statist.*, **55**, 88–97.

FRASER, D. A. S. (1976). Necessary analysis and adaptive inference. *J. Amer. Statist. Ass.*, **71**, 99–110.

HEATHCOTE, C. R. (1972). A test of goodness of fit for symmetric random variables. *Austral. J. Statist.*, **14**, 172–181.

—— (1977). The integrated squared error estimation of parameters. *Biometrika*, **64**, 255–264.

MCDUNNOUGH, P. (1979a). Estimating the law of randomly moving particles. *J. Appl. Prob.*, **16**, 25–35.

—— (1979b). The estimation of an autocorrelation parameter of a Gaussian vector process. *J. R. Statist. Soc.* B, **41**, 203–209.

PARZEN, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.*, **33**, 1065–1076.

PAULSON, A. S., HALCOMB, E. W. and LEITCH, R. A. (1975). The estimation of the parameters of the stable laws. *Biometrika*, **62**, 163–170.

PRESS, S. J. (1972). Estimation in univariate and multivariate stable distributions. *J. Amer. Statist. Assoc.*, **67**, 842–846.

—— (1975). Stable distributions: probability inference and applications in finance—a survey, and a review of recent results. In *Statistical Distributions in Scientific Work*, Vol. I (G. P. Patil, ed.). Dordrecht: Reidel.

QUANDT, R. E. and RAMSAY, J. B. (1979). Estimating mixtures of normal distributions and switching regressions. *J. Amer. Statist. Ass.*, **73**, 730–752.