On Statistical Transform Methods and Their Efficiency
Author(s): Andrey Feuerverger and Philip McDunnough
Source: *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, Vol. 12, No. 4 (Dec., 1984), pp. 303-317
Published by: Statistical Society of Canada
Stable URL: http://www.jstor.org/stable/3314814
Accessed: 12/06/2009 15:42

The Canadian Journal of Statistics
Vol. 12, No. 4, 1984, Pages 303–317
La Revue Canadienne de Statistique

On statistical transform methods and their efficiency*

Andrey FEUERVERGER and Philip McDUNNOUGH

University of Toronto

Key words and phrases: k-L procedures, generalized moment procedures, transforms, empirical transforms, asymptotic efficiency, approximate Bahadur efficiency, Pitman efficiency, maximum likelihood, Fisher information, Cramer-Rao bound, the Hilbert spaces $L_2(F)$.
AMS 1980 *subject classifications*: Primary 62F99; Secondary 62M99.

ABSTRACT

General conditions for the asymptotic efficiency of certain new inference procedures based on empirical transform functions are developed. A number of important processes, such as the empirical characteristic function, the empirical moment generating function, and the empirical moments, are considered as special cases.

RÉSUMÉ

On développe des conditions générales garantissant l'efficacité asymptotique de certaines nouvelles procédures d'inférence fondées sur des transformations empiriques. Ces conditions s'appliquent à un ensemble de processus importants, y compris la fonction caractéristique empirique, la fonction génératrice des moments, ainsi que les moments empiriques eux-mêmes.

1. INTRODUCTION AND SUMMARY

This paper is concerned with the asymptotic efficiency of certain new inference procedures based on transforms. These can be viewed as being alternatives to maximum likelihood, but applicable to a large number of statistical problems where maximum likelihood is not feasible. As the methods involved are fairly broad, a number of alternative viewpoints are presented. In particular, the contexts of estimation and testing are both considered below.

Suppose x_j, $j = 1, 2, \ldots, n$, are independent and identically distributed with density $f_\theta(x)$, where $\theta \in \Theta$, but the true value $\theta = \theta_0$ is unknown. Let $F_\theta(x)$ and $F_n(x)$ denote the actual and empirical cumulative distribution functions. The transform procedures with which we are concerned are based in each case on a kernel $g_t(x)$ such that

$$\mathscr{E}_\theta g_t(x) \equiv G_\theta(t) \equiv \int g_t(x)\, dF_\theta(x) \tag{1.1}$$

exists and is finite for all $\theta \in \Theta$ and all $t \in T$. An empirical version of this transform (or expectation) may be defined as

$$\hat{\mathscr{E}} g_t(x) \equiv G_n(t) \equiv \int g_t(x)\, dF_n(x) = \frac{1}{n}\sum g_t(x_j). \tag{1.2}$$

*Based on an invited talk to the Canadian Statistical Society's annual meeting in Halifax, 1981.

It turns out that the kernel functions $g_t(x)$ potentially of interest in statistical contexts are numerous. However, the following examples are typical:

(1) Take $g_t(x) = 1$, $x \geq t$, and 0 otherwise. Then $\mathscr{E}_\theta g_t$ and $\hat{\mathscr{E}} g_t$ (equivalently G_θ G_n) coincide with F_θ, F_n, respectively.

(2) Take $g_t(x) = e^{tx}$. Then $\mathscr{E}_\theta g_t$ and $\hat{\mathscr{E}} g_t$ coincide with the moment generating function (mgf) $L_\theta(t) = \int e^{tx}\, dF_\theta(x)$, when it exists, and the empirical mgf $L_n(t) = \int e^{tx}\, dF_n(x) = (1/n) \sum e^{tx_j}$, respectively.

(3) Take $g_t(x) = e^{itx}$. Here $\mathscr{E}_\theta g_t$, $\hat{\mathscr{E}} g_t$ are complex-valued and coincide with the characteristic function (cf) $\phi_\theta(t) = \int e^{itx}\, dF_\theta(x)$ and its empirical version $\phi_n(t) = \int e^{itx}\, dF_n(x) = (1/n) \sum e^{itx_j}$, respectively.

Focusing on the expectation $\mathscr{E}_\theta g_t$ and the empirical expectation $\hat{\mathscr{E}} g_t$ (equivalently, the transforms G_θ, G_n), consider now the possible procedures of inference concerning the unknown parameter θ. Since $\mathscr{E}_\theta \hat{\mathscr{E}} g_t = \mathscr{E}_\theta g_t$, we may write

$$\hat{\mathscr{E}} g_t = \mathscr{E}_\theta g_t + e_n(t), \qquad t \in T, \tag{1.3}$$

where $\mathscr{E}_\theta e_n(t) = 0$. By the strong law of large numbers, we have $e_n(t) \rightarrow 0$ almost surely for every $t \in T$, and hence any countable collection in T. This suggests the consistency of procedures based on fitting $\mathscr{E}_\theta g_t$ to $\hat{\mathscr{E}} g_t$ by various means. Under further conditions the process $\{e_n(t),\ t \in T\}$ will, asymptotically and in varying degrees, have the properties of a Gaussian process. This normality, combined with a concern for asymptotic efficiency, suggests a variety of procedures for study.

Let $t_1, t_2, \ldots, t_k$ be a fixed finite subset of T. Let $\mathscr{E}_\theta \mathbf{g}$ and $\hat{\mathscr{E}} \mathbf{g}$ be $k \times 1$ vectors having entries $\mathscr{E}_\theta g_{t_j}$, $\hat{\mathscr{E}} g_{t_j}$ respectively, and assumed for the moment to be real. We may indicate now the procedures with which we are concerned. [Some related procedures are investigated in an unpublished thesis by Brant (1982), who also discusses the approximation of the full likelihood via transforms; see also Jarrett (1973).] For convenience, the context here is estimation. We have the following classes of methods:

(A) *Moment (linear) methods.* Estimate θ by solving

$$\mathbf{d}' \mathscr{E}_\theta \mathbf{g} = \mathbf{d}' \hat{\mathscr{E}} \mathbf{g} \tag{1.4}$$

where $\mathbf{d}$ is a $k \times 1$ vector of constants.

(B) *Regression (quadratic) methods.* Estimate θ by minimizing

$$(\hat{\mathscr{E}} \mathbf{g} - \mathscr{E}_\theta \mathbf{g})' \mathbf{Q} (\hat{\mathscr{E}} \mathbf{g} - \mathscr{E}_\theta \mathbf{g}) \tag{1.5}$$

where $\mathbf{Q}$ is a $k \times k$ nonnegative definite matrix of constants.

(C) *k - L methods.* Estimate θ by maximizing the asymptotic normal form of the likelihood of $\hat{\mathscr{E}} \mathbf{g}$. This may be taken as

$$-\frac{k}{2} \log \det \frac{1}{n} \mathbf{\Sigma} - \frac{n}{2} (\hat{\mathscr{E}} \mathbf{g} - \mathscr{E}_\theta \mathbf{g})' \mathbf{\Sigma}_\theta^{-1} (\hat{\mathscr{E}} \mathbf{g} - \mathscr{E}_\theta \mathbf{g}), \tag{1.6}$$

where $\mathbf{\Sigma} = n\, Var_\theta(\hat{\mathscr{E}} \mathbf{g}) = [Cov_\theta(g_{t_i}, g_{t_j})]$, or as just the second term of this expression. [It can be shown that this yields an asymptotically equivalent estimate, in that the two estimates differ only by a factor which is $o(1/\sqrt{n})$.]

The above three procedures are to a certain extent new, and are discussed more fully below. We content ourselves here with remarking that the term "k − L method" is based on the fact that the procedure is restricted to a finite grid of k points in T, and is likelihood-based. Finally, we note that the procedures as described above appear to be of "discrete" type, but have in fact "continuous" analogues. In particular (1.4) may be replaced by

$$\int_T \hat{\mathscr{E}} g_t \, dH(t) = \int_T \mathscr{E}_\theta g_t \, dH(t), \tag{1.7}$$

while (1.5) may be replaced by

$$\iint_{T\times T} (\hat{\mathscr{E}} g_s - \mathscr{E}_\theta g_s)(\hat{\mathscr{E}} g_t - \mathscr{E}_\theta g_t) A(ds, dt), \tag{1.8}$$

where $A(ds, dt)$ is nonegative definite.

Presumably an analogue for (1.6) is possible also, along the lines of Parzen (1961) or Grenander (1981, Ch. 3), and is of great interest statistically, but will not be pursued here. However, it would seem clear that the discrete and continuous versions are intimately connected. In particular, it would be of interest to compare the continuous-form likelihood based on the asymptotic normal law of the transform process with the actual nonasymptotic likelihood of the x_j's.

While our general approach, as well as some of the procedures presented, may be new, many particular and important cases have appeared in earlier work involving, for example, the transforms based on the kernel functions 1 – 3 provided above. The most extensive of this literature is related to the empirical characteristic function (ecf), owing perhaps to the very special properties enjoyed by the Fourier transform.

The first reference to the ecf of which we are aware appears in Parzen (1962), and early applications are provided by Heathcote (1972, 1977), Press (1972, 1975), and Paulson, Halcomb, and Leitch (1975). A systematic study of the ecf with a view towards applications is undertaken in Feuerverger and Mureika (1977). Further probability investigations are provided by Kent (1975), Csörgő (1981), and Marcus (1981). The efficiency of a suitable class of ecf procedures was investigated and proved first in two papers by the authors (1981a and 1981b). These papers, hereafter referred to as FM-1 and FM-2 respectively, provide the principal motivation for the present investigation, and some familiarity with them will prove helpful here. Finally we mention also Feigin and Heathcote (1976), Thornton and Paulson (1977), Tarter (1979), Feuerverger and McDunnough (1980), Koutrouvelis (1980a, 1980b), Koutrouvelis and Kellermeier (1981), Kellermeier (1980), Murota and Takeuchi (1981), and Hall and Welsh (1983).

The empirical moment-generating function (emgf), which is similar to the ecf in certain limited ways, is applied in the studies by Quandt and Ramsey (1978), Leslie and Khalique (1980), and Read (1981). The efficiency of these procedures, not previously known, is resolved below.

Our other example, where $\mathscr{E}_\theta g_t = F_\theta$ and $\hat{\mathscr{E}} g_t = F_n$, appears simplest in the sense that if use of F_θ is permitted, then many classical inference procedures, such as maximum likelihood, become immediately accessible. However it is useful, and the basic ideas may be illustrated, if we consider this "transform" from the viewpoint of the nonclassical procedures (A) – (C) which concern us here. Thus suppose, for example, that F_n and F_θ are to be used in a moment method of continuous type. Let us write the estimation equation (1.7) in the form

$$\int h(x) F_\theta(x) \, dx = \int h(x) F_n(x) \, dx \tag{1.9}$$

and seek that function $h(x)$ leading to estimators of smallest asymptotic variance. A variational approach to this problem is possible as detailed in FM-1, but not required in the present case. In fact the optimal $h(x)$ depends on the unknown θ_0 and is given by

$$h_{\theta_0}(x) = \frac{\partial^2 \log f_\theta(x)}{\partial \theta_0 \, \partial x}. \tag{1.10}$$

To see this, transform (1.9) using integration by parts into the form

$$\int u(x)\, dF_\theta(x) = \int u(x)\, dF_n(x), \tag{1.11}$$

where $h(x)$ and $u(x)$ are related through

$$h(x) = \frac{du(x)}{dx}. \tag{1.12}$$

The "optimal" $u(x)$ is now easily guessed, since if

$$u(x) = u_\theta(x) = \frac{\partial \log f_\theta(x)}{\partial \theta}, \tag{1.13}$$

then (1.11) may be recognized as just the likelihood equation. Of course, we require $u(\cdot)$ to be a function of x alone, but the optimal choice clearly will be $u_{\theta_0}(x)$. It is easily verified that (1.10) provides an efficient solution for the procedure (1.9).

Further, concerning the example where $\mathscr{E}_\theta g_t = F_\theta$ and $\hat{\mathscr{E}} g_t = F_n$, fix $-\infty < t_1 < t_2 < \cdots < t_k < \infty$ and consider now the discrete moment procedure

$$\mathbf{d}'\mathbf{F}_\theta = \mathbf{d}'\mathbf{F}_n \tag{1.14}$$

where $\mathbf{d}$, $\mathbf{F}_\theta$, $\mathbf{F}_n$ are $k \times 1$ vectors. What is the optimal choice for the vector $\mathbf{d}$ of constants? Using a first-order expansion for F_θ at θ_0, the solution of (1.14) may be given as

$$\hat{\theta} \simeq \theta_0 + \frac{\mathbf{d}'(\mathbf{F}_n - \mathbf{F}_{\theta_0})}{\mathbf{d}' \dfrac{d\mathbf{F}_\theta}{d\theta_0}} \tag{1.15}$$

with asymptotic variance [cf. (2.3)]

$$n\, Var(\hat{\theta}) = \frac{\mathbf{d}' \boldsymbol{\Sigma} \mathbf{d}}{\left[\mathbf{d}' \dfrac{d\mathbf{F}_\theta}{d\theta_0}\right]^2}, \tag{1.16}$$

where $\boldsymbol{\Sigma}$ has (i, j)th entry

$$F_{\theta_0}(\max(t_i, t_j)) - F_{\theta_0}(t_i) F_{\theta_0}(t_j). \tag{1.17}$$

As (1.16) is the ratio of a quadratic form and a squared linear form, it is minimized by

$$\mathbf{d} = \boldsymbol{\Sigma}^{-1} \frac{d\mathbf{F}_\theta}{d\theta_0} \tag{1.18}$$

and attains the minimum value of

$$n\, Var(\hat{\theta}) = \left[\left(\frac{d\mathbf{F}_\theta}{d\theta_0}\right)' \boldsymbol{\Sigma}^{-1} \left(\frac{d\mathbf{F}_\theta}{d\theta_0}\right)\right]^{-1}. \tag{1.19}$$

Using (1.17), the value of (1.19) could now be evaluated, since the component $[F_{\theta_0}(\max(t_i, t_j))]$ has a known inverse, while the component $[F_{\theta_0}(t_i) F_{\theta_0}(t_j)]$ is of unit rank and thus may be adjusted for using Bartlett's identity. We omit these steps here, as the value of (1.19) may be established (as in the following section) by a simpler argument. We note here however that this value [given by the expression (2.5)] can be made arbitrarily close to the Cramér-Rao lower bound by selecting the grid $\{t_j\}$ to be sufficiently fine and extended.

The procedure (1.14) is the discrete version of (1.9), and several questions of real interest now arise. For example, as $\{t_j\}$ becomes finer and more extended, do the values of the entries of $\mathbf{d}$ given by (1.18) trace out a function proportional to (1.10) except for some adjustment for any unevenness of the $\{t_j\}$ spacing? We conjecture that for general kernel functions $g_t(x)$ and under fairly broad conditions a result of this type must be true.

This paper consists of five sections and is primarily concerned with the efficiency of the transform procedures introduced here. Presumably, it should be possible to treat the efficiency of various inferential methods (estimation, testing, etc.) by means of a single unified approach, but how this may be done is not quite clear. Instead, Section 2 is concerned with efficiency in the estimation context, while Section 3 is concerned with efficiency in the testing context. Sections 2 and 3 are restricted to discrete-case considerations. Some treatment of the continuous case is undertaken in Section 4. Finally, a number of remarks and brief examples indicating the generality and usefulness of transform methods is provided in Section 5. The present paper is self-contained, but we mention again that some familiarity with FM-1 and FM-2 may be helpful.

2. EFFICIENCY IN ESTIMATION

In this section we consider discrete estimation procedures based on the transform $G_\theta(t) = \mathscr{E}_\theta g_t(x)$ and its empirical version $G_n(t) = \hat{\mathscr{E}} g_t(x)$. Here the context is parametric with $\theta \in \Theta$, where Θ is a real open interval.

Let $t_1, t_2, \ldots, t_k$ be fixed points in T, and let $\hat{\mathscr{E}}\mathbf{g}$, $\mathscr{E}_\theta \mathbf{g}$ be $k \times 1$ vectors having entries $\hat{\mathscr{E}} g_{t_j}$ and $\mathscr{E}_\theta g_{t_j}$, respectively ($g$ being assumed real here). Then, except for a constant, the asymptotic form of the log likelihood for $\hat{\mathscr{E}}\mathbf{g}$ is given by

$$\frac{1}{2}\log\left|\frac{1}{n}\boldsymbol{\Sigma}_\theta\right| - \frac{n}{2}(\hat{\mathscr{E}}\mathbf{g} - \mathscr{E}_\theta\mathbf{g})'\boldsymbol{\Sigma}_\theta^{-1}(\hat{\mathscr{E}}\mathbf{g} - \mathscr{E}_\theta\mathbf{g}), \tag{2.1}$$

where $\boldsymbol{\Sigma}_\theta = n\, Var(\hat{E}\mathbf{g}) = Var(\mathbf{g})$ and has entries $Cov(g_{t_i}(x), g_{t_j}(x))$. The form (2.1) suggests estimating the true θ_0 by that $\hat{\theta}_{kL}$ which minimizes the criterion function

$$(\hat{\mathscr{E}}\mathbf{g} - \mathscr{E}_\theta\mathbf{g})'\boldsymbol{\Sigma}_\theta^{-1}(\hat{\mathscr{E}}\mathbf{g} - \mathscr{E}_\theta\mathbf{g}). \tag{2.2}$$

We refer to this as a k-L *procedure,* since it involves, essentially, an (asymptotic) likelihood based on k fixed points. We now utilize Theorem 6.1 of FM-2, with the correction that the covariance matrix in part (b) should read

$$\frac{1}{n}\left(\frac{\partial F}{\partial \theta}\right)^{-1}\frac{\partial F}{\partial T}\boldsymbol{\Sigma}\frac{\partial F'}{\partial T}\left(\frac{\partial F'}{\partial \theta}\right)^{-1}.$$

It follows that under mild conditions, $\hat{\theta}_{kL}$ is asymptotically $\mathrm{N}(\theta_0, \sigma_{kL}^2/n)$, where

$$\sigma_{kL}^{-2} = \frac{\partial \mathscr{E}_\theta\mathbf{g}'}{\partial\theta}\boldsymbol{\Sigma}_\theta^{-1}\frac{\partial\mathscr{E}_\theta\mathbf{g}}{\partial\theta} \tag{2.3}$$

evaluated at $\theta = \theta_0$. (Explicit reference to this will often be omitted.)

We consider here the problem of when σ_{kL}^2 can be made arbitrarily close to the Cramér-Rao bound $I^{-1}(\theta_0)$, where

$$I(\theta) = \mathscr{E}\left(\frac{\partial \log f_\theta(x)}{\partial \theta}\right)^2. \tag{2.4}$$

This has already been demonstrated in FM-1, FM-2 for the ecf which we may represent here in the form

$$g_t(x) = \begin{cases} \sin tx, & t > 0, \\ \cos tx, & t \leq 0 \end{cases}.$$

A further example is furnished by

$$g_t(x) = \begin{cases} 1, & x \leq t, \\ 0, & x > t, \end{cases}$$

associated with the distribution function. Here (2.3) may be evaluated directly, but it is easier to appeal to the multinomial nature of $\hat{\mathscr{E}}\mathbf{g}$. Thus [e.g. see Rao (1965), Section 5e] we find that

$$\sigma_{kL}^{-2} = \sum_{j=1}^{k+1} P_j\left(\frac{\partial \log P_j}{\partial \theta}\right)^2, \tag{2.5}$$

where $P_j = G(t_j) - G(t_{j-1})$ and $G(t_0) = 0$, $G(t_{k+1}) = 1$. This is just a discrete approximation to $I(\theta)$, and so the k-L procedure again permits arbitrarily high efficiency.

In order to treat the general case, it is convenient, as in FM-1, to introduce a related method which we refer to as a *generalized moment procedure*. Take $\mathbf{d}$ to be a fixed $k \times 1$ vector, and consider the equation

$$\mathbf{d}'(\hat{\mathscr{E}}\mathbf{g} - \mathscr{E}_\theta \mathbf{g}) = 0. \tag{2.6}$$

As in FM-1, FM-2, this yields an estimator $\hat{\theta}_M$ which is asymptotically $\mathrm{N}(\theta_0, \sigma_M^2/n)$, where

$$\sigma_M^2 = \frac{\mathbf{d}'\boldsymbol{\Sigma}_0\mathbf{d}}{\left(\mathbf{d}'\dfrac{\partial \mathscr{E}_\theta \mathbf{g}}{\partial \theta_0}\right)^2}. \tag{2.7}$$

This is minimized by taking

$$\mathbf{d} = \mathbf{d}_0 = \boldsymbol{\Sigma}_\theta^{-1}\frac{\partial \mathscr{E}_\theta \mathbf{g}}{\partial \theta_0}, \tag{2.8}$$

and in this case, (2.7) is seen to reduce to (2.3). Of course, the optimal weights $\mathbf{d}_0$ depend on the unknown θ_0, but may be estimated. The implied two stage procedure will have the same asymptotic distribution as that which uses the optimal $\mathbf{d}_0$, provided only that a consistent estimate of $\mathbf{d}_0$ is used.

Continuing now, and with a view to more general two-stage procedures, consider a class $\mathscr{H}$ of functions $h(\cdot)$ and the corresponding estimation equations

$$\hat{\mathscr{E}}h(x) = \mathscr{E}_\theta h(x), \tag{2.9}$$

where $\hat{\mathscr{E}}h(x) = n^{-1}\Sigma_{j=1}^n h(x_j)$. For an $h(x) \in \mathscr{H}$ we take $\hat{\theta}_h$ to be a consistent estimate satisfying (2.9) for n sufficiently large. (Simple conditions which guarantee the existence of such a $\hat{\theta}_h$ are given in Theorem 6.1 of FM-2.) It is straightforward to show that $\hat{\theta}_h$ is asymptotically $\mathrm{N}(\theta_0, \sigma_h^2/n)$, where

$$\sigma_h^2 = \frac{Var_{\theta_0}[h(x)]}{\left(\dfrac{\partial}{\partial \theta_0}\mathscr{E}_\theta h(x)\right)^2}. \tag{2.10}$$

Recall now the so-called regression identity which, in obvious notation, may be written as

$$\mathscr{E}(y_1 - \mu_1)^2 = \mathscr{E}[(y_1 - \mu_1) - \beta(y_2 - \mu_2)]^2 + \beta^2\sigma_2^2, \tag{2.11}$$

where $\beta = \sigma_{12}/\sigma_2^2$. If we set $y_1 = \partial \log f_\theta(x)/\partial\theta_0$ and $y_2 = h(x)$, then we obtain the well-known identity

$$\begin{aligned} I(\theta_0) &= E_{\theta_0}\left[ah(x) + b - \frac{\partial \log f_\theta(x)}{\partial\theta_0}\right]^2 \\ &\quad + \frac{Cov_{\theta_0}^2\left(h(x), \dfrac{\partial \log f_\theta(x)}{\partial\theta_0}\right)}{Var_{\theta_0}[h(x)]} \\ &= \sigma_h^{-2} + \mathscr{E}_{\theta_0}\left[ah(x) + b - \frac{\partial \log f_\theta(x)}{\partial\theta_0}\right]^2, \end{aligned} \tag{2.12}$$

where a, b are constants depending on h and θ_0 and have the property of minimizing the last term in (2.12). Now consider a sequence of functions $h_l \in \mathscr{H}$. Note that (2.9), (2.10) are unaltered if h_l is replaced by $ah_l + b$ where a, b are constants. Then by (2.12) we have $\sigma_{h_l}^2 \to I^{-1}(\theta_0)$ if and only if there are constants a_l, b_l such that

$$a_l h_l(x) + b_l \to \frac{\partial \log f_\theta(x)}{\partial\theta_0},$$

where convergence is in $L_2(f_{\theta_0})$.

This last result, together with the asymptotic distribution equivalence, noted earlier, of the k-L and generalized moment precedures, leads directly to the following:

THEOREM 2.1. *The k-L procedure applied to $\{G_n(t) = \hat{\mathscr{E}}g_t(x)\}$ admits arbitrarily high asymptotic efficiency if and only if the closure in $L_2(f_{\theta_0})$ of the space of functions $\sum_{j=1}^k d_j g_{t_j}(x)$, where $k, t_1, t_2, \ldots, t_k$ are arbitrary, includes the true score $\partial \log f_\theta(x)/\partial\theta_0$.*

We note that as θ_0 is unknown, we generally would require the stated criterion to hold regardless of the value of θ_0 in Θ. Due to questions concerning selection of the $\{t_j\}$ in practice, we generally would require that the theorem hold also if the $\{t_j\}$ are restricted to an arbitrary countable and dense grid $T' \subset T$.

3. EFFICIENCY IN TESTING

We consider first the problem of testing, in the context of goodness of fit, the hypothesis $H: \theta = \theta_0$ on the basis of a sample $x_1, x_2, \ldots, x_n$ from a true distribution specified by $\theta = \theta_1$. As in the k-L procedure for estimation, we select k points $t_1, t_2, \ldots, t_k$ in T corresponding to the functions $g_{t_1}(x), \ldots, g_{t_k}(x)$. The asymptotic normality of $\hat{\mathscr{E}}(\mathbf{g})$, where $\mathbf{g}$ has entries $g_{t_j}(x)$ and $\hat{\mathscr{E}}$ denotes empirical expectation, suggests the use of a quadratic distance

$$D_{\mathbf{Q}}^2 = n(\hat{\mathscr{E}}\mathbf{g} - E_0\mathbf{g})'\mathbf{Q}(\hat{\mathscr{E}}\mathbf{g} - E_0\mathbf{g}), \tag{3.1}$$

where $E_0\mathbf{g}$ is evaluated using $\theta = \theta_0$ and $\mathbf{Q}$ is a nonnegative definite matrix of constants. Large values of $D_{\mathbf{Q}}^2$ will constitute evidence against H. The observed level of significance (OLS) of this test is the statistic

$$\text{OLS} = P_{\theta_0}(D_{\mathbf{Q}}^2 \geq \text{observed value}), \tag{3.2}$$

so that small values of OLS constitute evidence against H.

One measure of the goodness of this test is its approximate Bahadur slope [a discussion of the exact Bahadur slope may be found in Brant (1982)], defined for $\theta_1 \neq \theta_0$ as

$$c(\theta_1) \stackrel{\text{a.s.}}{=} \lim_{n\to\infty}\left(-\frac{2}{n}\log \widehat{\text{OLS}}\right), \tag{3.3}$$

where $\widehat{\text{OLS}}$ is the approximation to OLS obtained through use of the asymptotic distribution of $\hat{\mathscr{E}}\mathbf{g}$ in the calculation of (3.2). This measure is discussed in Bahadur (1960), where it is shown that for real θ

$$c(\theta_1) \leq 2J(\theta_0, \theta_1), \tag{3.4}$$

where $J(\theta_0, \theta_1)$ is the Kullback-Liebler information number. Moreover, equality is obtained for the likelihood ratio test. Now $\hat{\mathscr{E}}\mathbf{g}$ is asymptotically $\mathbf{N}(\mathscr{E}_1\mathbf{g}, \boldsymbol{\Sigma}_1/n)$ where $\mathscr{E}_1\mathbf{g}$ is evaluated under $\theta = \theta_1$, and $\boldsymbol{\Sigma}_1 = n\, Var_{\theta_1}(\hat{\mathscr{E}}\mathbf{g})$. Since $D_{\mathbf{Q}}^2 = nd_{\mathbf{Q}} + o(n)$, where $d_{\mathbf{Q}}^2 = (\mathscr{E}_1\mathbf{g} - \mathscr{E}_0\mathbf{g})'\mathbf{Q}(\mathscr{E}_1\mathbf{g} - \mathscr{E}_0\mathbf{g})$ we have

$$c(\theta_1) = \lim_{n\to\infty}\left(-\frac{2}{n}\log P(\mathbf{z}'\mathbf{A}\mathbf{z} \geq nd_{\mathbf{Q}})\right) \tag{3.5}$$

where $\mathbf{z}$ consists of k independent N(0, 1) variables and $\mathbf{A} = \boldsymbol{\Sigma}_0^{\frac{1}{2}}\mathbf{Q}\boldsymbol{\Sigma}_0^{\frac{1}{2}}$, where $\boldsymbol{\Sigma}_0 = n\, Var_{\theta_0}(\hat{\mathscr{E}}\mathbf{g})$. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k$ be the eigenvalues of $\mathbf{A}$. Then by Lemma (2.4) of Gregory (1980) we have

$$\begin{aligned} c(\theta_1) &= \lim_{n\to\infty}\left[-\frac{2}{n}\log P\left(\sum_{j=1}^{k}\lambda_j z_j^2 \geq nd_{\mathbf{Q}}\right)\right] \\ &= d_{\mathbf{Q}}/\lambda_1. \end{aligned} \tag{3.6}$$

It follows that the approximate Bahadur efficiency of the $D_{\mathbf{Q}}^2$ test is

$$e_B(\theta_1) = \frac{d_{\mathbf{Q}}}{2\lambda_1 J(\theta_0, \theta_1)} \tag{3.7}$$

Now suppose θ is defined on a real interval. Then of particular relevance in large samples is $e_B = \lim_{\theta_1\to\theta_0} e_B(\theta_1)$. Indeed, Wieand (1976) has shown that e_B corresponds, under general conditions, to the limiting ($\alpha \to 0$) Pitman effienciency. For the $D_{\mathbf{Q}}^2$ test, Wieand's results hold provided there is a neighborhood N_{θ_0} of θ_0 and a constant M such that for all $n > M/d_{\mathbf{Q}}^2$

$$P_{\theta_1}(|D_{\mathbf{Q}}^2/n - d_{\mathbf{Q}}| \geq \varepsilon d_{\mathbf{Q}}) \leq \delta \qquad \text{for all} \quad \theta_1 \in N_{\theta_0} \tag{3.8}$$

for every preassigned $\varepsilon > 0$ and $0 < \delta < 1$. Now, using Markov's inequality, we have

$$P_{\theta_1}(|D_{\mathbf{Q}}^2/n - d_{\mathbf{Q}}| \geq \varepsilon d_{\mathbf{Q}}) \leq \frac{\mathscr{E}_{\theta_1}|D_{\mathbf{Q}}^2/n - d_{\mathbf{Q}}|}{\varepsilon d_{\mathbf{Q}}}. \tag{3.9}$$

Setting $\mathbf{Q} = \mathbf{P}'\boldsymbol{\Lambda}\mathbf{P}$ where $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_k)$ and corresponding eigenvectors forming the columns of $\mathbf{P}$, we have

$$D_{\mathbf{Q}}^2/n = \bar{\mathbf{h}}'\boldsymbol{\Lambda}\bar{\mathbf{h}} = \Sigma\lambda_j\bar{h}_j^2 \tag{3.10}$$

and

$$d_{\mathbf{Q}} = \mathbf{a}'\boldsymbol{\Lambda}\mathbf{a} = \Sigma\lambda_j a_j^2, \tag{3.11}$$

where $\bar{h}_j$ and a_j are the entries of $\bar{\mathbf{h}} = \mathbf{P}(\hat{\mathscr{E}}\mathbf{g} - \mathscr{E}_0\mathbf{g})$ and $\mathbf{a} = \mathbf{P}(\mathscr{E}_1\mathbf{g} - \mathscr{E}_0\mathbf{g})$. Then (3.9) is

$$\leq \frac{1}{\varepsilon d_{\mathbf{Q}}} \sum_{j=1}^{k} \lambda_j E_1 \left| \bar{h}_j^2 - a_j^2 \right|$$

$$\leq \frac{1}{\varepsilon d_{\mathbf{Q}}} \sum \lambda_j \, Var_{\theta_1}^{\frac{1}{2}} \bar{h}_j \cdot \mathscr{E}_1^{\frac{1}{2}} \left| \bar{h}_j + a_j \right|$$

$$\leq \frac{1}{\sqrt{n}\, \varepsilon d_{\mathbf{Q}}} \sum \lambda_j \, Var_{\theta_1}^{\frac{1}{2}} h_j(x) \left(\mathscr{E}_1 \left| \frac{h_j(x)}{n} \right| + |a_j| \right)^{\frac{1}{2}}, \tag{3.12}$$

where $h_j(x)$ denotes one of the n terms constituting the average $\bar{h}_j(x)$. Now since the $h_j(x)$'s are simple affine transformations of the $g_{t_j}(x)$'s (the transformations not involving θ_1), and the a_j's of the $\mathscr{E}_1 g_{t_j}(x)$'s, it follows that (3.8) will hold if

$$\mathscr{E}_1[g_{t_j}(x)]^2 \leq M' \qquad \text{for all} \quad \theta_1 \in N_{\theta_1} \text{ and } j. \tag{3.13}$$

This condition is very mild, and whenever it is satisfied, the efficiencies quoted will have the dual interpretation of Bahadur and Pitman efficiencies.

Provided θ is identifiable from $E_\theta(\mathbf{g})$, the $D_{\mathbf{Q}}^2$ test is consistent. For $\theta_1 \neq \theta_0$ implies $D_{\mathbf{Q}}^2 \xrightarrow{\text{a.s.}} \infty$. Furthermore, an "optimal" test amongst all $D_{\mathbf{Q}}^2$ tests may be arrived at by choosing that $\mathbf{Q}$ which maximizes the (Pitman or Bahadur) efficiency e_B. For real θ we have, as $\theta_1 \to \theta_0$,

$$2J(\theta_0, \theta_1) \simeq I(\theta_0)(\theta_1 - \theta_0)^2 \tag{3.14}$$

and

$$d_{\mathbf{Q}} \simeq (\theta_1 - \theta_0)^2 \frac{\partial \mathscr{E}\mathbf{g}'}{\partial \theta_0} \mathbf{Q} \frac{\partial \mathscr{E}\mathbf{g}}{\partial \theta_0}, \tag{3.15}$$

so that

$$e_B = \frac{\dfrac{\partial \mathscr{E}\mathbf{g}'}{\partial \theta_0} \mathbf{Q} \dfrac{\partial \mathscr{E}\mathbf{g}}{\partial \theta_0}}{I(\theta_0) \cdot \lambda_1}. \tag{3.16}$$

Now, setting $\mathbf{y} = \Sigma_0^{\frac{1}{2}} \partial \mathscr{E}\mathbf{g} / \partial \theta_0$, we have

$$e_B = \frac{\mathbf{y}' \mathbf{A} \mathbf{y}}{I(\theta_0) \cdot \lambda_1} \tag{3.17}$$

and this is maximized by taking $A = I$. It follows that the optimal $\mathbf{Q}$ is simply Σ_0^{-1} and the optimal $D_{\mathbf{Q}}^2$ test is based on

$$D_{\mathbf{Q}}^2 = n\hat{\mathscr{E}}\mathbf{g} - \mathscr{E}_0\mathbf{g})' \Sigma_0^{-1} (\hat{\mathscr{E}}\mathbf{g} - \mathscr{E}_0 \mathbf{g}), \tag{3.18}$$

with corresponding efficiency

$$\frac{1}{I(\theta_0)} \left(\frac{\partial \mathscr{E}\mathbf{g}'}{\partial \theta_0} \Sigma_0^{-1} \frac{\partial \mathscr{E}\mathbf{g}}{\partial \theta_0} \right) \tag{3.19}$$

The results of Section 2 now imply that (3.19) may be made arbitrarily close to one if and only if the score function $\partial \log f_\theta(x) / \partial \theta_0$ may be approximated in the mean-square sense by affine transformations of the g's. Finally, turning to the general problem of testing a simple hypothesis of a vector paramter $\boldsymbol{\theta}$, it is clear from the results above that the likelihood-ratio test applied to the asymptotic likelihood of $\hat{\mathscr{E}} g_{t_1}(x), \ldots, \hat{\mathscr{E}} g_{t_k}(x)$ will have arbitrarily high asymptotic efficiency compared to the likelihood-ratio test based on $x_1, \ldots, x_n$ if and only if $\partial \log f_\theta / \partial \boldsymbol{\theta}_0$ can be approximated by means of linear combinations of the g_t's.

4. THE CONTINUOUS CASE

The procedures in the previous two sections are based on a finite set $t_1, t_2, \ldots, t_k$ of T and hence are of "discrete" type. However, the procedures have natural "continuous" analogues in the case where $g_t(x)$ is a measurable function of t. We treat only the testing context in any detail here; brief remarks on the estimation context appear at the end of the section. In particular, we shall discuss tests of the hypothesis $H:\theta = \theta_0$ based on large deviations of

$$\hat{D}^2 = n\int\int(\hat{\mathscr{E}}g_s - \mathscr{E}_0 g_s)(\hat{\mathscr{E}}g_t - \mathscr{E}_0 g_t)B(s,t)\,ds\,dt \tag{4.1}$$

for some positive semidefinite integrable kernel $B(s,t)$. The asymptotic distribution of $\hat{D}^2$ under H is, under general conditions, given by the distribution of

$$D^2 = \int\int z(s)z(t)B(s,t)\,ds\,dt, \tag{4.2}$$

where $\{z(t)\}$ is a zero-mean Gaussian process with

$$Ez(s)z(t) = E(g_s g_t) - Eg_s Eg_t \equiv K(s,t). \tag{4.3}$$

Now suppose $B(s,t)$ is continuous. Then, by Mercer's theorem,

$$B(s,t) = \Sigma\,\mu_i\phi_i(s)\phi_i(t), \tag{4.4}$$

where μ_i, ϕ_i are the eigenvalues and corresponding orthonormal eigenvectors of the kernel $B(s,t)$:

$$\mu_i\phi_i(t) = \int\phi_i(s)K(s,t)\,ds. \tag{4.5}$$

We therefore have

$$D^2 = \Sigma\,\mu_i z_i^2, \tag{4.6}$$

where the

$$z_i = \int z(s)\phi_i(s)\,ds \tag{4.7}$$

are independent N(0, 1) variables. Furthermore

$$\hat{D}^2 = \Sigma\,\mu_i\hat{z}_i^2, \tag{4.8}$$

where

$$\hat{z}_i = \int z_n(s)\phi_i(s)\,ds \tag{4.9}$$

and

$$z_n(s) = \sqrt{n}(\hat{\mathscr{E}}g_s - \mathscr{E}_0 g_s). \tag{4.10}$$

The $\hat{z}_i$ have zero mean and unit variance, and are uncorrelated. The results (4.6) and (4.8) may be noted as being analogous to the decomposition of the Cramér–von Mises statistic given by Durbin and Knott (1972). Consequently, each of the individual components, of (4.6) represents a certain aspect of the departure from H so that the corresponding estimates in (4.8) may be of value in much the same way as components in an analysis of variance.

Turning now to questions of efficiency, since $\hat{D}^2$ may be approximated by statistics D_Q^2 of discrete type, then in view of the results in the previous section, tests based on $\hat{D}^2$ should have high efficiency compared to the likelihood-ratio test. Perhaps more surprising [but see also (2.9) and (2.10) of FM-2] is that we may in general restrict ourselves to kernels $B(s,t)$ of unit rank. Thus, taking $B(s,t) = b(s)b(t)$, we have

$$\hat{D}^2 = \left(\int b(s)(\hat{\mathscr{E}}g_s - \mathscr{E}_0 g_s)\,ds\right)^2 \tag{4.11}$$

and

$$D^2 = \left(\int b(s)z(s)\,ds\right)^2 \sim \sigma_b^2\chi_1^2, \tag{4.12}$$

where

$$\sigma_b^2 = \int\int b(s)b(t)K(s,t)\,ds\,dt. \tag{4.13}$$

Let $d_b = [\int(\mathscr{E}_1 g_t - \mathscr{E}_0 g_t)b(t)\,dt]^2$. Then, arguing essentially as before, we find

$$\lim_{n\to\infty}\left(-\frac{2}{n}\log P(\hat{D}^2 > nd_b)\right) = \frac{d_b}{\sigma_b^2}, \tag{4.14}$$

so that the limiting $(\theta_1 \to \theta_0)$ approximate Bahadur efficiency is given by

$$e_b = \frac{\left[\int b(t)\dfrac{\partial \mathscr{E}g_t}{\partial\theta_0}\,dt\right]^2}{I(\theta_0)\int\int b(s)b(t)K(s,t)\,ds\,dt}. \tag{4.15}$$

We may now apply a variational argument as in FM-1 to find that e_b is maximized when $b(t) = b_0(t)$, where $b_0(\cdot)$ satisfies

$$\int b_0(s)K(s,t)\,ds = \frac{\partial \mathscr{E}g_t}{\partial\theta_0}, \tag{4.16}$$

which gives a maximal efficiency

$$\begin{aligned} e_{\max} &= I^{-1}(\theta_0)\int b_0(t)\frac{\partial \mathscr{E}g_t}{\partial\theta_0}\,dt \\ &= I^{-1}(\theta_0)\int \frac{\partial f_\theta(x)}{\partial\theta_0}h(x)\,dx, \end{aligned} \tag{4.17}$$

where

$$h(x) = \int b_0(t)g_t(x)\,dt. \tag{4.18}$$

We may now see from (4.17) that a sufficient condition for $e_{\max} = 1$ is that (to within a constant)

$$h(x) = \frac{\partial \log f_\theta(x)}{\partial\theta_0}, \tag{4.19}$$

so that, in addition to (4.16), $b_0(t)$ satisfies

$$\frac{\partial \log f_\theta(x)}{\partial\theta_0} = \int b_0(t)g_t(x)\,dt. \tag{4.20}$$

Actually (4.20) implies (4.16), as may be seen directly:

$$\int b_0(s)K(s,t)\,ds$$

$$= \int b_0(s)\mathscr{E}_0(g_s g_t)ds - \int b_0(s)\mathscr{E}_0 g_s \mathscr{E}_0 g_t\, ds$$

$$= \mathscr{E}_0 g_t \int b_0(s)\mathscr{E}_0 g_s\, ds - \mathscr{E}_0 g_t \mathscr{E}_0 \int b_0(s) g_s\, ds$$

$$= \mathscr{E}_0\left(g_t \frac{\partial \log f_\theta(x)}{\partial \theta_0}\right) + \mathscr{E}_0 g_t \cdot \mathscr{E}_0\left(\frac{\partial \log f_\theta}{\partial \theta_0}\right)$$

$$= \int g_t(x)\frac{\partial f_\theta(x)}{\partial \theta_0}dx + 0$$

$$= \frac{\partial}{\partial \theta_0}\mathscr{E} g_t.$$

Hence the condition (4.20) is sufficient for $e_{\max} = 1$. It is worth noting that (4.20) will also be necessary whenever the family $\{f_\theta(x)\}$ is complete in the sense of having no unbiased estimator of zero. To see this, note that if two different $h(x)$ functions give the same value for (4.17), then, by completeness, the difference of these functions must be a constant. To sum up, the continuous procedure (4.11) will be efficient provided that $b(t) = b_0(t)$ satisfying (4.20).

Finally, we remark that while phrased in the context of testing, the results of this section relate directly to estimation as well. For consider a "continuous" moment-type estimation procedure given by the equation

$$\int b(t)(\hat{\mathscr{E}} g_t - \mathscr{E}_\theta g_t)\,dt = 0. \tag{4.21}$$

Then, a differential argument gives the asymptotic variance of the resulting estimator as the inverse of the expression (4.15) but without the $I(\theta_0)$ factor. The arguments subsequent to (4.15) then apply nearly verbatim, and we find that the estimator corresponding to (4.21) attains, asymptotically, the Cramér-Rao bound, provided $b(t) = b_0(t)$ satisfying (4.20). The optimal $b_0(t)$ depends on the unknown θ_0 but, of course, may be estimated, and the resulting two-stage procedure is asymptotically efficient. It is worth noting that (4.21) may be written in the form

$$\int\int b(t)g_t(x)\,dt\,d[F_n(x) - F_\theta(x)] = 0 \tag{4.22}$$

and becomes, with $b(t) = b_0(t)$ satisfying (4.20),

$$\int \frac{\partial \log f_\theta(x)}{\partial \theta_0} d[F_n(x) - F_\theta(x)] = 0, \tag{4.23}$$

which may be compared with the likelihood equation.

5. SOME REMARKS AND APPLICATIONS

1.

The broader context for the considerations of this paper involves the Hilbert spaces $L_2(f_\theta)$, $\theta \in \Theta$. Finiteness of Fisher information means that for every θ the score function

$\partial \log f_\theta(x)/\partial\theta$ is an element of $L_2(f_\theta)$. The equality (2.12) may be viewed as providing the loss of information associated with replacing maximum likelihood by the procedure (2.9). With a, b chosen to minimize $\mathscr{E}_{\theta_0}[ah(x) + b - \partial \log f_\theta(x)/\partial\theta_0]^2$, this quantity represents the squared distance between the score function and its projection onto the manifold associated with $h(x)$ and measures the information loss. Hence, it may be seen that the efficiency of k-L-type procedures depends upon whether the score functions are approximable, in the manner of Theorem 2.1, in the spaces $L_2(f_\theta)$. This statement involves a collection of Hilbert spaces (indexed by θ). However, it appears that there is no need to deal with the collection aspect per se, but only with the individual spaces. Often we will have a countable set of functions $g_t(x)$, which will be complete in all the spaces. In these cases the efficiency requirement will be met. However, note that completeness is not required — only that within each space the manifold spanned by the functions should include the score.

2.

If F is a distribution function having bounded support, say on $(0, 2\pi)$ for convenience, then, as is well known, the functions e^{inx} for integers n are complete in $L_2(F)$. Thus if the family $\{f_\theta\}$ is supported on $(0, 2\pi)$, it follows that the integer coordinates of the empirical characteristic function will suffice for purposes of asymptotic efficiency in this case.

3.

Using straightforward arguments it is possible to establish that any countable collection of the functions e^{itx} where the values of t are dense in $(-\infty, \infty)$ is complete in any space of the type $L_2(F)$. Using Theorem 2.1, we are therefore led to a proof of the efficiency of ecf procedures alternative to the proofs given in FM-1, FM-2.

4.

According to Theorem 4.12 of Kufner and Kadlec (1971), we have that the functions x^n, $n = 0, 1, 2, \ldots$, are complete in $L_2(f_\theta)$ provided that

$$f_\theta(x) < ce^{-\beta|x|} \quad \text{for} \quad |x| > R. \tag{5.1}$$

Using Theorem 2.1, we therefore have

THEOREM 5.1. *If for every $\theta \in \Theta$ there are c, β, and R such that* (5.1) *holds, then the k-L procedure applied to the empirical moments, $\Sigma_{i=1}^n x_i^j$, where $j = 1, 2,, \ldots, k$, is arbitrarily highly efficient.*

Note that for this result to be useful, it is necessary that values of $\mathscr{E}_\theta x^n$ be known.

The Gaussian case affords an interesting example. If $x \sim N(\mu, \sigma^2)$, then it is well known that

$$\mathscr{E}x^{2n} = \sigma^{2n} \sum_{l=0}^{n} \frac{(2n)!}{2^{n-l}(2l)!(n-l)!}\left(\frac{\mu}{\sigma}\right)^{2l} \tag{5.2}$$

and

$$\mathscr{E}x^{2n+1} = \sigma^{2n+1} \sum_{l=0}^{n} \frac{(2n+1)!}{2^{n-l}(2l+1)!(n-l)!}\left(\frac{\mu}{\sigma}\right)^{2l+1} \tag{5.3}$$

for positive integers n. Hence the covariance structure, and the form of the k-L procedure, for the first k empirical moments can be determined, though with some difficulty. A simpler approach may be based on the sufficiency of Σx_i and Σx_i^2. For $k > 2$, the density function for Σx_i^j, $j = 1, 2, \ldots, k$, has (in obvious notation) the form

$$h_\theta(\Sigma x_i, \ldots, \Sigma x_i^k) = h_\theta^{(1)}(\Sigma x_i, \Sigma x_i^2) h^{(2)}(\Sigma x_i^3, \ldots, \Sigma x_i^k \mid \Sigma x_i, \Sigma x_i^2), \tag{5.4}$$

where the last factor is independent of θ. It follows that the asymptotic normal form of the distribution for the empirical moments will also factor in this form and that the k-L procedure for $k > 2$ is in fact identical to the procedure for $k = 2$. This latter is now easily shown to reduce to the usual MLE procedure.

5.

The so-called positive stable laws provide a striking application of the methods proposed here. According to Lemma 2.1 of Brockwell and Brown (1981), the functions $x^{-j\beta}$, $j = 0, 1, 2, \ldots$, are complete in the spaces $L_2(F)$ associated with the positive stable laws having index $\alpha \geq \beta/(1 + \beta)$. Further, the expected values [*ibid.*, Equation (4)], and hence the covariance structure of the negative moments, are easily determined. A k-L procedure based on the negative moments is therefore easily implemented and will have arbitrarily high asymptotic efficiency. In fact we may note from the results of Brockwell and Brown that a very few moments generally suffice for high efficiency.

6.

For the empirical mgf we have the following quite general result:

THEOREM 5.2. *Suppose the family* $\{f_\theta\}$ *satisfies the conditions* (5.1). *Then the empirical mgf* $\hat{\mathscr{E}} e^{tx}$ *admits arbitrarily high asymptotic efficiency. Further, the grid* $\{t_j\}$ *may be restricted arbitrarily close to the origin.*

The proof of this result is fairly straightforward. The condition implies that the functions x^n, $n = 0, 1, 2, \ldots$, and $e^{tx}, t \in (-\tau, \tau)$ (for some $\tau > 0$) are elements of the space $L_2(f_{\theta_0})$. By Kufner and Kadlec (1971), Theorem 4.12, the functions x^n are complete in $L_2(f_\theta)$. However, these may be approximated by means of the e^{tx}. In particular we have, using exponential bounds and the dominated-convergence theorem,

$$\lim_{t \to 0} \mathscr{E}_{\theta_0} \left(\frac{e^{tx} - \sum_{j=0}^{n-1} \frac{(tx)^j}{j!}}{t^n} - x^n \right)^2 = 0$$

for all $n \geq 0$. Consequently the arbitrarily high asymptotic efficiency of the k-L procedure applied to $\hat{\mathscr{E}} e^{t_j x}$ follows from Theorem 2.1.

REFERENCES

Bahadur, R.R. (1960). Stochastic comparison of tests. *Ann. Math. Statist.*, 31, 276–295.

Brant, R.F. (1982). Some statistical methods based on transforms. Ph.D. thesis, University of Toronto.

Brockwell, P.J., and Brown, B.M. (1981). High efficiency estimation for the positive stable laws. *J. Amer. Statist. Assoc.*, 76, 626–631.

Bryant, J.L., and Paulson, A.S. (1979). Some comments on characteristic function based estimators. *Sankhyā Ser. A*, 41, 109–116.

Csörgő, S. (1981). Limit behaviour of the empirical characteristic function. *Ann. Probab.*, 9, 130–144.

Durbin, J., and Knott, M. (1972). Components of the Cramer–von Mises statistics. *J. Roy. Statist. Soc. Ser. B*, 34, 290–307.

Feigen, P.D., and Heathcote, C.R. (1976). The empirical characteristic function and the Cramer–von Mises statistic. *Sankhyā Ser. A*, 38, 309–325.

Feuerverger, A., and Mureika, R.A. (1977). The empirical characteristic function and its applications. *Ann. Statist.*, 5, 88–97.

Feuerverger, A., and McDunnough, P. (1981a). On the efficiency of empirical characteristic function procedures. *J. Roy. Statist. Soc. Ser. B*, 43, 20–27.

Feuerverger, A. and McDunnough, P. (1981b). On some Fourier methods for inference. *J. Amer. Statist. Assoc.*, 76, No. 374, pp. 379–387.

Feuerverger, A., and McDunnough, P. (1980). On efficient inference in symmetric stable laws and processes. *Proceedings of the International Symposium on Statistics and Related Topics, Ottawa, May 1980.* (A.K.Md.E. Saleh et al., *eds.*), North Holland, Amsterdam.

Gregory, G.G. (1980). On efficiency and optimality of quadratic tests. *Ann. Statist.*, 8, 116–131.

Grenander, U. (1981). *Abstract Inference.* Wiley, New York.

Hall, P., and Welsh, A.H. (1983). A test for normality based on the empirical characteristic function. *Biometrika*, 70, 485–489.

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

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

Jarrett, R.G. (1973). Efficiency and estimation in asymptotically normal distributions. Doctoral Thesis, University of London.

Kellermeir, J. (1980). The empirical characteristic function and large sample hypothesis testing. *J. Multivariate Anal.*, 10, 78–87.

Kent, J.T. (1975). A weak convergence theorem for the empirical characteristic function. *J. Appl. Probab.*, 12, 515–523.

Koutrouvelis, I.A. (1980a). A goodness-of-fit test of simple hypothesis based on the empirical characteristic function. *Biometrika*, 67, 238–240.

Koutrouvelis, I.A. (1980b). Regression-type estimation of the parameters of stable laws. *J. Amer. Statist. Assoc.*, 75, 918–928.

Koutrouvelis, I., and Kellermeier, J. (1981). A goodness of fit test based on the empirical characteristic function when parameters must be estimated, *J. Roy. Statist. Soc. Ser. B*, 42, 173–176.

Kufner, A., and Kadlec, J. (1971). *Fourier Series* (English translation by G.A. Toombs). Iliffe, London.

Leslie, R.T., and Khalique, A. (1980). Parametric estimation via Laplace transform. Unpublished manuscript.

Marcus, M.B. (1981). Weak convergence of the empirical characteristic function. *Ann. Probab.*, 9, 194–201.

Murota, K., and Takeuchi, K. (1981). The Studentized empirical characteristic function and its application to test for the shape of a distribution. *Biometrika*, 68, 55–66.

Parzen, E. (1961). An approach to time series analysis. *Ann. Math. Statist.*, 32, 951–989.

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 paramters 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.

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

Quandt, R.E., and Ramsey, J.B. (1978). Estimating mixtures of normal distributions and switching regressions. *J. Amer. Statist. Assoc.*, 73, 730–752.

Rao, C.R. (1965). *Linear Statistical Inference and Its Applications.* Wiley, New York.

Read, R.R. (1981). Representation of certain covariance matrices with application to asymptotic efficiency. *J. Amer. Statist. Assoc.*, 76, 148–154.

Tartar, M.E. (1979). Trigonometric maximum likelihood estimation and application to the analysis of incomplete survival information. *J. Amer. Statist. Assoc.*, 74, 132–139.

Thornton, J.C., and Paulson, A.S. (1977). Asymptotic distributions of characteristic function–based estimators for the stable laws. *Sankhyā Ser. A*, 39, 341–354.

Wieand, H.S. (1976). A condition under which the Pitman and Bahadur approaches to efficiency coincide. *Ann. Statist.*, 4, 1003–1011.

Received 7 December 1982
Revised 12 September 1983
Accepted 27 January 1984

Department of Statistics
University of Toronto
Toronto, Ontario I5S 1A1