STATISTICS AND RELATED TOPICS
M. Csörgö, D.A. Dawson, J.N.K. Rao, A.K.Md.E. Saleh (eds.)

ON EFFICIENT INFERENCE IN SYMMETRIC STABLE LAWS AND PROCESSES

Andrey Feuerverger and Philip McDunnough

Department of Statistics
University of Toronto

1. INTRODUCTION

Distributions which are limits (except for scaling and recentering) of sums of independent identically distributed variates are termed stable. When the variates have second moment the possible limits necessarily are Gaussian and this is the best known case of the central limit theorem. However the wider class of distributions that share in the central limiting feature coincides exactly with the stable laws. These laws, first obtained by P. Levy (1925, 1954) possess a natural interest for statistical applications and for robustness studies in particular. For applications to modelling telephone line noise, see Berger and Mandelbrot (1963), Stuck and Kleiner (1974). Applications to modelling price changes in various financial markets are given by Mandelbrot (1963), Fama (1965), Fielitz and Smith (1972), Samuelson (1975), and Leitch and Paulson (1975). The use of symmetric versus skewed stable distributions, for example, carries implications for investment strategy. Properties of the stable laws are discussed in Gnedenko and Kolmogorov (1954), Feller (1966), Lukacs (1970), and Holt and Crow (1973).

The stable distributions admit unimodal densities having all derivatives but in general these are available only as numerically awkward infinite series. This lack of a closed form for the density has impeded development of statistical methods for this distribution family though a number of ad hoc procedures have been developed. See for example Fama and Roll (1971), Press (1972), Paulson, Halcomb and Leitch (1975), Heathcote (1977), de Haan and Resnick (1980), Brockwell and Brown (1980).

In this paper, we also are concerned with inference for stable distributions, however our interest centers exclusively on procedures which are asymptotically efficient, or at least on procedures whose asymptotic efficiency can be made arbitrarily high. The first indication of efficient inference for the stable laws was given by DuMouchel (1973) who showed that the MLE's were consistent, asymptotically normal and followed the well-known theory for maximum likelihood inference. The maximum likelihood procedure was implemented by DuMouchel (1971, 1975) and subsequently in unpublished independent work by the authors. Because maximum likelihood is so technically cumbersome, alternative asymptotically efficient techniques remain of considerable interest. Since the stable characteristic functions are readily available it is natural to ask if inference can be based directly on these and whether or not efficient procedures exist. This question was studied in Feuerverger and McDunnough (1980, 1979) and has an affirmative answer. The results of these two papers show that under very general conditions statistical procedures based on the empirical characteristic function (e.c.f.) may be used for a wide class of statistical problems and that suitable ecf-based procedures have arbitrarily high asymptotic efficiency. To explore the applicability of these ideas to the problem of inference for the stable laws is one purpose of the present paper.

The outline of our paper is as follows. In §2 we present a new continuous reparametrization of the stable laws and a slight, but useful extension of a result due

to Zolotarev. Certain essential properties of the stable laws are reviewed. In §3-4 we discuss certain numerical aspects of maximum likelihood for these distributions and we present some Monte Carlo results for the symmetric case. In §5 we discuss the ecf and methods of efficient inference in the Fourier domain. The application to stable laws is considered in §6 and some numerical results pertaining to grid selection are obtained. The various procedures discussed extend easily to discrete time linear stable processes and in §7 we provide a brief Monte Carlo study for the stationary AR(1) case. Some unusual results are noted. Our numerical work is confined throughout to the symmetric case; the methods however are entirely general.

2. SOME PROPERTIES OF STABLE DISTRIBUTIONS

The stable distributions are defined through

$$\log \phi(t) = \begin{cases} -|t|^{\alpha}\left\{1 + i\,\beta\,\mathrm{sgn}(t)\tan\left(\frac{\pi\alpha}{2}\right)\right\}, & \alpha \neq 1 \\ -|t|\left\{1 + i\,\frac{2}{\pi}\beta\,\mathrm{sgn}(t)\log|t|\right\}, & \alpha = 1 \end{cases} \tag{2.1}$$

where $\phi(t) = E(\exp(itX))$ is the characteristic function, and $0 < \alpha \leq 2$, $-1 \leq \beta \leq 1$ are shape and skewness parameters. For $\alpha \neq 1$ an alternative representation is sometimes used:

$$\log \phi(t) = -|t|^{\alpha} \exp\{-i\tfrac{\pi}{2}\beta'\,\mathrm{sgn}(t)\cdot K(\alpha)\} \tag{2.2}$$

where $K(\alpha) = 1 - |1-\alpha|$ and $-1 \leq \beta \leq 1$. The two representations disagree on scaling as well as skewness. The relation between β and β' is given in Lukacs (1970, pp. 136-8).

For applications involving the full nonsymmetric class the discontinuity at $\alpha = 1$ is troublesome. We shall show that this discontinuity may be removed. To do so we first write the $\alpha \neq 1$ term of (2.1) in the form

$$-|t|^{\alpha} - i\,\beta\,t|t|^{\alpha-1}\tan\tfrac{\pi\alpha}{2} \quad . \tag{2.3}$$

If we now shift the mean by an amount $\beta\tan\frac{\pi\alpha}{2}$ we obtain

$$-|t|^{\alpha} - i\,\beta\,t(|t|^{\alpha-1} - 1)\tan\tfrac{\pi\alpha}{2} \tag{2.4}$$

or

$$-|t|^{\alpha} - i\,\beta^{*}h(t,\alpha) \tag{2.5}$$

where

$$h(t,\alpha) = \frac{t(|t|^{\alpha-1} - 1)}{\alpha-1} \tag{2.6}$$

and

$$\beta^{*} = \beta(\alpha-1)\tan\tfrac{\pi\alpha}{2} \quad . \tag{2.7}$$

Some analysis now establishes that the function $h(t,\alpha)$ is continuous on R^2 and can be defined by continuity as $t\,\ell n|t|$ when $\alpha = 1$.

That the discontinuity at $\alpha = 1$ can be removed by reparametrization is known; see Chambers, Mallows and Stuck (1977) and DuMouchel (1971). The approach given here seems more direct, however, and results in a parameter domain whose shape (roughly elliptical but having corners at $\alpha = 0,2$) is more consistent with the known behaviour of the densities near $\alpha = 0,2$ for varying skewness. The parametrizations β, β' and β^{*} are in 1-1 correspondence through (2.7) and

$$\beta^* = (1-\alpha)\tan\frac{\pi K(\alpha)\beta'}{2} \quad . \tag{2.8}$$

By verifying that the variation of the difference between characteristic functions within a shrinking neighbourhood in $(\alpha,\beta^*,\mu,\sigma)$ approaches zero we may prove:

Theorem 2.1. Let $p_{\alpha\beta^*}(x)$ be the density corresponding to (2.5). Then the family $\left\{\frac{1}{\sigma}p_{\alpha\beta^*}\left[\frac{x-\mu}{\sigma}\right]\right\}$ varies continuously in the sup-norm over the domain of its $(\alpha,\beta^*,\mu,\sigma)$ parametrization.

The Bergstrom-Feller expansions for the stable densities may be written as follows: if $0 < \alpha < 1$, then

$$f_{\alpha\beta'}(x) = \frac{1}{\pi x}\sum_{k=1}^{\infty}\frac{(-1)^{k-1}\Gamma(\alpha k+1)}{k!}\sin\left[\frac{\pi k\alpha}{2}\{\beta'+\operatorname{sgn}x\}\right]\cdot|x|^{-k\alpha} \tag{2.9}$$

and if $1 < \alpha \le 2$, then

$$f_{\alpha\beta'}(x) = \frac{1}{\pi x}\sum_{k=1}^{\infty}\frac{(-1)^{k-1}\Gamma\left(\frac{k}{\alpha}+1\right)}{k!}\sin\left[\frac{\pi k}{2}\left\{\left[\frac{2-\alpha}{\alpha}\right]\beta'+\operatorname{sgn}x\right\}\right]\cdot|x|^{k} \quad . \tag{2.10}$$

For the asymptotic character of these series and for the case $\alpha = 1$ we refer to Lukacs (1970), §5.8 and 5.9. We remark that from a statistical viewpoint expansions for $\log f$ would be of greater interest.

Several useful observations about (2.9) and (2.10) do not seem to have been made previously. First note that (2.10) converges also if $\alpha > 2$ and secondly that both series remain convergent for arbitrary $-\infty < \beta' < \infty$. The series for $0 < \alpha < 1$ and $1 < \alpha < \infty$ (and β' arbitrary) are closely related and we have the following slightly generalized form of a result due to Zolotarev: For $0 < \alpha < \infty$, but $\alpha \neq \frac{1}{2}, 1$ we have

$$f_{\alpha\beta'}(x) = x^{-1}|x|^{-\alpha}f_{\alpha^{-1},\beta''}(|x|^{-\alpha}) \tag{2.11}$$

where

$$\beta'' = \begin{cases}\dfrac{\alpha\beta'+(\alpha-1)\operatorname{sgn}x}{2\alpha-1} & \text{if } 0<\alpha<1\\[2ex] (2-\alpha)\beta'+(\alpha-1)\operatorname{sgn}x & \text{if } 1<\alpha<\infty \quad .\end{cases}$$

The failure of (2.11) at $\alpha = \frac{1}{2}$ is, curiously, of an inessential kind. For if we reparametrize in (2.10) by replacing $\frac{2-\alpha}{\alpha}\beta'$ by β''' then the expansion (2.9) for some $0<\alpha<1$ and some β' will be related to the expansion (2.10) for α^{-1} and $\beta''' = \alpha\beta'+(\alpha-1)\operatorname{sgn}x$. This relation does not fail at $\alpha = \frac{1}{2}$. The significance of (2.11) is that numerical evaluation for any $0<\alpha<1$ (except $\alpha = \frac{1}{2}$) may be replaced by evaluation at α^{-1}; for small α discrete Fourier transformation is very difficult.

3. MAXIMUM LIKELIHOOD BY INVERSION

Our approach to maximum likelihood estimation for the stable laws is an application of the fast Fourier transform (FFT) algorithm (Cooley and Tukey, 1965). Suppose ϕ is an integrable characteristic function; then the evaluation of the integral

$$f(x) = \int_{-\infty}^{\infty}\phi(t)\,e^{-itx}\,dt \tag{3.1}$$

by means of the FFT effectively restricts x to values on an equispaced grid such as $0, \pm\Delta x, \pm 2\Delta x, \ldots$. If the FFT is based on N points the available range for the density will be $\pm N \cdot \Delta x/2$ with one end-point missing; the corresponding spacing for the frequency variable will then be $\Delta t = 2\pi/N \cdot \Delta x$ and the range will be, not $\pm \pi/\Delta x$, but rather $\pm 2\pi/\Delta x$ with endpoints excluded. The range for t is doubled in this way due to the fact that $\phi(-t) = \overline{\phi(t)}$ so that we have the identity

$$\sum_{n=-N+1}^{N-1} \phi(\Delta t \cdot n) e^{-i\lambda n} = 2\,\mathrm{Re}\left\{ \sum_{n=0}^{N-1} \phi_o(\Delta t \cdot n) e^{-i\lambda n} \right\}$$

where ϕ_o is identical to ϕ except that $\phi_o(0) = \frac{1}{2}$

Now, given a complex sequence $X(0), X(1), \ldots, X(N-1)$, the FFT algorithm produces the sequence $\sum_{n=0}^{N-1} X(n)\, e^{-i\lambda n}$ for $\lambda = \frac{2\pi s}{N}$, $s = 0, 1, \ldots, (N-1)$. Therefore if the FFT is applied to the sequence

$$\tfrac{1}{2}, \phi(\Delta t), \phi(2\Delta t), \ldots, \phi((N-1)\cdot\Delta t),$$

and if the real parts of the resulting sequence are multiplied by $2/N\cdot\Delta x$ we obtain - except for the effects of truncation and discretization of the integral - values of the density

$$f(0), f(\Delta x), f(2\Delta x), \ldots, f(-2\Delta x), f(-\Delta x).$$

Note the circular format with values for f on the negative axis occuring at the end of the sequence.

The effect of truncation (e.g. Brillinger, 1975, Ch. 3) is that we obtain a convoluted form of the transform required. One possibility would be to use a tapering function; the one due to Bohman (1960) seems especially appropriate. DuMouchel (1971, p. 35) gives a better resolution and shows how the truncation effect may be eliminated using a "wrapped summation" method. In our work we used a 10% cosine taper (Tukey, 1967) and found this gave satisfactory results provided α was not less than about 0.5.

The effect of discretization (e.g. Brillinger, 1975, §5.11) is that we obtain an aliased version $\sum_{j=-\infty}^{\infty} f(x + jN\Delta x)$ of the density. Since $N\Delta x$ typically is not small, de-aliasing could be achieved using the asymptotic expansion for $|x| \to \infty$; in fact this same expansion is required also for the tails where the inversion is inaccurate numerically. DuMouchel (1971) replaced the Fourier integral by $N/2$ intervals and quadratically interpolated $\phi(t)$ in each interval (Filon's method). Our approach to de-aliasing was based on the fortuitously discovered observation - which is supported by simple numerical arguments - that within a suitable range the aliasing error essentially is constant and thus may be determined immediately by subtracting the known exact expression $f_{\alpha,\beta'}(0) = \pi^{-1}\Gamma(1+\alpha^{-1})$ for the density at $x = 0$ from the corresponding FFT determined value. Further details about this as well as other aspects of this paper may be found in an unpublished technical report by the authors.

4. MAXIMUM LIKELIHOOD SIMULATION STUDY

Using the methods in §3 a maximum likelihood procedure was developed in FORTRAN on the University of Toronto IBM 360/170. Versions for both the symmetric and non-symmetric laws were produced but only the symmetric case was subjected to

simulation. A noteworthy feature of the programs, particularly for comparison with the ecf procedures of §5-6, is their essential technical complexity and the considerable programming effort required. On this point, see also DuMouchel (1971, 1974).

The stable variates required for the simulation study were generated using the algorithm RSTAB (Chambers, Mallows and Stuck, 1976) corrected for an error noticed in the published function D2: the fourth DATA line should read "&,.18001 33704 07390 023 D3" (cf. approximation 1801 in Hart et. al., 1968). The FFT subroutine used was the November 1967, Bell Laboratories, Murray Hill version of AR1DFT written by W.M. Gentleman and G. Sande. The FFT was used to obtain the standardized density at the current estimate $\hat{\alpha}$ as well as at $\hat{\alpha} \pm \Delta\alpha$ where $\Delta\alpha = 0.025$. The data was then subjected to standardization at the current estimates $\hat{\mu}$ and $\hat{\sigma}$ and as well at $\hat{\mu} \pm \Delta\mu$, $\hat{\sigma} \pm \Delta\sigma$ where $\Delta\mu = .05$ and $\Delta\sigma = .05$. In this manner, the likelihood was calculated on the subset of the $3 \times 3 \times 3$ grid of parameter points needed in order to perform a Newton-Raphson procedure. We used $\Delta x = 0.1$ and $N = 1024$ and found single precision adequate. For $|x| \le 7.5$ values for the density were obtained using quadratic interpolation on the FFT values which then were corrected for aliasing as described in §3. For $|x| > 7.5$ values of the density were determined using three terms of the asymptotic series. These values and rules were determined after numerical experimentation.

The simulation study spanned the values $\alpha = 0.7\ (.1)\ 1.9$ and sample sizes $N = 50, 100, 200$ of standard variates, with each sample being used once only. (Note N now is no longer the FFT length.) The initial estimates were taken to be the actual true values and five full iterations of the Newton-Raphson procedure were carried out for sample sizes $N = 50, 100$, and four iterations for $N = 200$. For $N = 100$ and 200, fifty trials ($n = 50$) were conducted and for $N = 50$ we conducted $n = 100$ trials.

Table 4.1 summarizes the results of this Monte Carlo. For the three parameters in each cell we give the sample average and the sample standard deviation of the estimates resulting from the n trials. We also give the value n for the number of trials; whenever n differs from the value declared at the top of the table, this indicates failures of the MLE procedure to terminate normally. This could occur if the latest update exceeded the boundaries of the parameter space or if an unacceptable level of numerical instability was detected. The results of table 4.1 are in good agreement with the asymptotic calculations given in DuMouchel (1975). A detailed discussion appears in the technical report mentioned above.

5. FOURIER METHODS FOR INFERENCE

Suppose $X_1, X_2, \ldots, X_n$ are iid variates with density in $\{f_\theta(x)\}$ where θ is a real univariate parameter. The equation of maximum likelihood may be written in the form

$$\int_{-\infty}^{\infty} \frac{\partial \log f_\theta(x)}{\partial \theta}\, d\left(F_n(x) - F_\theta(x)\right) = 0 \qquad (5.1)$$

where F_θ is the cdf of f_θ and $F_n(x)$ is the empirical cdf. Define now the following transformed quantities: the characteristic function

$$c_\theta(t) = \int e^{itx}\, dF_\theta(x)\ ; \qquad (5.2)$$

the empirical characteristic function (ecf)

TABLE 4.1
Simulation Results for Maximum Likelihood
Estimation for Symmetric Stable Laws

		N = 50 n = 100 (# iter = 5)		N = 100 n = 50 (# iter = 5)		N = 200 n = 50 (# iter = 4)	
$\alpha = 1.9$	μ	-	-	.030	.119	-.035	.122
	α	-	-	-	-	1.870	.063
	σ	-	-	-	-	.987	.062
	n	26		23		30	
$\alpha = 1.8$	μ	-	-	-.046	.137	-.023	.103
	α	-	-	1.780	.098	1.758	.094
	σ	-	-	.987	.087	.987	.064
	n	53		36		41	
$\alpha = 1.7$	μ	.018	.247	.048	.133	.021	.097
	α	-	-	1.629	.124	1.689	.097
	σ	-	-	.985	.094	.993	.077
	n	68		43		45	
$\alpha = 1.6$	μ	.008	.174	.002	.128	.017	.101
	α	1.582	.173	1.593	.160	1.608	.119
	σ	.997	.159	1.005	.088	.995	.066
	n	90		46		50	
$\alpha = 1.5$	μ	.019	.233	.034	.113	-.037	.116
	α	1.492	.162	1.474	.154	1.503	.107
	σ	.989	.152	.983	.136	.989	.069
	n	90		50		50	
$\alpha = 1.4$	μ	.036	.221	-.002	.177	.012	.110
	α	1.434	.183	1.391	.149	1.399	.092
	σ	1.028	.162	.959	.101	1.006	.056
	n	93		50		50	
$\alpha = 1.3$	μ	-.024	.223	-.005	.152	.011	.112
	α	1.302	.194	1.320	.137	1.324	.106
	σ	.970	.163	.983	.123	1.105	.094
	n	93		50		50	
$\alpha = 1.2$	μ	.020	.205	-.054	.168	.001	.123
	α	1.229	.187	1.191	.139	1.209	.092
	σ	1.012	.179	1.010	.132	1.003	.094
	n	93		50		50	
$\alpha = 1.1$	μ	-.022	.189	.031	.163	.021	.093
	α	1.146	.139	1.088	.109	1.107	.096
	σ	1.039	.169	.995	.136	1.005	.097
	n	88		49		50	
$\alpha = 1.0$	μ	-.013	.198	-.005	.135	-.007	.101
	α	1.061	.158	1.003	.123	.997	.077
	σ	.989	.188	1.022	.161	1.000	.108
	n	89		50		50	
$\alpha = .9$	μ	-.018	.187	-.032	.149	-.014	.084
	α	.958	.140	.911	.102	.927	.081
	σ	.996	.203	1.015	.141	1.033	.104
	n	82		50		50	
$\alpha = .8$	μ	-	-	-.038	.146	.007	.096
	α	-	-	.808	.084	.804	.052
	σ	-	-	.992	.158	1.028	.136
	n	72		46		50	
$\alpha = .7$	μ	-	-	.031	.122	.015	.083
	α	-	-	.740	.059	.717	.043
	σ	-	-	1.026	.168	1.025	.111
	n	39		32		42	

$$c_n(t) = \int e^{itx}\, d F_n(x) = \frac{1}{n} \sum_{j=1}^{n} e^{itx_j} \tag{5.3}$$

and the inverse transform of the score

$$w_\theta(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\partial \log f_\theta(x)}{\partial \theta} e^{-itx}\, dx \ . \tag{5.4}$$

Our starting point is to note that under very general conditions, we may apply the Parseval theorem to (5.1) to obtain

$$\int_{-\infty}^{\infty} w_\theta(t) \left(c_n(t) - c_\theta(t)\right) dt = 0 \ . \tag{5.5}$$

This is the Fourier domain version of the likelihood equation. Note that $w_\theta(t)$ must usually be regarded as a generalized function. (The multiparameter extension is straightforward.) This result at once suggests that the empirical characteristic function may have valuable applications; procedures based on the ecf and their asymptotic efficiency were explored in Feuerverger and McDunnough (1980, 1979).

Consider the process $c_n(t)$. We are indebted to R.A. Mureika for pointing out the following result which is due to an anonymous referee and which generalizes a result of Feuerverger and Mureika (1977):

Theorem 5.1: If F is any distribution function and $\log T_n = o(n)$ then

$$P\{\lim_{n\to\infty} \sup_{|t|\le T_n} |c_n(t) - c(t)| = 0\} = 1 \ .$$

Proof. For fixed $\varepsilon > 0$ let A be such that $F(-A) + 1 - F(A) \le \varepsilon$, and replace $c_n(t) - c(t)$ by

$$\gamma(t, X_1, \ldots, X_n) \equiv \frac{1}{n} \sum_{|X_j| \le A} e^{itx_j} - \int_{[-A,A]} e^{itx}\, d F(x) \ .$$

Clearly

$$|\gamma(t, X_1, \ldots, X_n) - \gamma(t', X_1, \ldots, X_n)| \le |t_1 - t_2|\, A \ .$$

Thus, it suffices to show that

$$\lim_{n\to\infty} \sup_{|k| \le \frac{AT_n}{\varepsilon}} \left|\gamma\left(\frac{k\varepsilon}{A}, X_1, \ldots, X_n\right)\right| = 1 \quad \text{w.p.1.}$$

But

$$P\left\{ \sup_{|k| \le \frac{AT_n}{\varepsilon}} \left|\gamma\left(\frac{k\varepsilon}{A}, X_1, \ldots, X_n\right)\right| \ge \varepsilon \right\}$$

$$\le \frac{AT_n}{\varepsilon} \sup_t P\{|\gamma(t, X_1, \ldots, X_n)| \ge \varepsilon\} \ .$$

The result follows using standard exponential bounds: the latter probability decreases exponentially, because $n \cdot \gamma$ is the sum of n iid bounded random variables with mean zero. □

Define $Y_n(t) = \sqrt{n}(c_n(t) - c(t))$. As $c_n(t)$ is a sum of iid bounded processes we

have at once $E\,Y_n(t) = 0$ and $\operatorname{cov}(Y_n(s),Y_n(t)) = E\,Y_n(s)\overline{Y_n(t)} = c(s-t) - c(s)\overline{c(t)}$ and the covariance structure of the real and imaginary parts:

$$\begin{aligned}
\operatorname{Cov}(\operatorname{Re} Y_n(s),\operatorname{Re} Y_n(t)) &= \tfrac{1}{2}[\operatorname{Re} c(s-t)+\operatorname{Re} c(s+t)] - \operatorname{Re} c(s)\operatorname{Re} c(t)\\
\operatorname{Cov}(\operatorname{Re} Y_n(s),\operatorname{Im} Y_n(t)) &= \tfrac{1}{2}[\operatorname{Im} c(s-t)+\operatorname{Im} c(s+t)] - \operatorname{Re} c(s)\operatorname{Im} c(t) \qquad (5.6)\\
\operatorname{Cov}(\operatorname{Im} Y_n(s),\operatorname{Im} Y_n(t)) &= \tfrac{1}{2}[\operatorname{Re} c(s-t)-\operatorname{Re} c(s+t)] - \operatorname{Im} c(s)\operatorname{Im} c(t) \quad .
\end{aligned}$$

Let $Y(t)$ be a zero mean complex Gaussian process having covariance structure identical to Y_n. By the central limit theorem Y_n converges in distribution to Y at finite numbers of points. Feuerverger and Mureika (1977) prove the weak convergence of $Y_n(t)$ to $Y(t)$ in any finite interval provided that $E|X|^{1+\delta} < \infty$. Csorgo (1980) shows that the moment condition is not easily removed and gives a general treatment of convergence questions. Necessary and sufficient conditions for the weak convergence are given by Marcus (1980). Feuerverger and McDunnough (e.g. lemma 2.1 of 1980) show that weak convergence of the ecf process is not critical for many statistical purposes since one can exploit the essentially simple stochastic structure of $c_n(t)$ to obtain needed results. A quadratic version of the quoted lemma 2.1 may be proved upon evaluating the limiting cumulants:

Lemma 5.2 Let $A(t_1,t_2)$ be a function having bounded variation on R^2. Then

$$n \cdot \iint (c_n(t_1) - c(t_1))\,(c_n(t_2) - c(t_2))\; A(dt_1,dt_2) \xrightarrow{\mathcal{D}} \iint Y(t_1)\,Y(t_2)\,A(dt_1,dt_2) \quad .$$

The result holds also if matching factors in the integrands are conjugated.

Turning to inference, a comprehensive discussion of asymptotically efficient or arbitrarily highly efficient procedures based on the ecf is given in Feuerverger and McDunnough (1980, 1979). Here we emphasize only two of these - the harmonic-regression procedure, and the k-L procedure, both of "discrete type". The k-L procedure is so-called because it is of likelihood type, and based on a fixed number k of ecf points. Let $0 < t_1 < \ldots < t_k$ be this fixed grid. Define $\underset{\sim}{z}_0 = (\operatorname{Re} c(t_1),\ldots,\operatorname{Re} c(t_k)\,,\operatorname{Im} c(t_1),\ldots,\operatorname{Im} c(t_k))'$ and let $\underset{\sim}{z}_n$ be its empirical counterpart. Letting $n^{-1}\Sigma$ be the covariance matrix of $\underset{\sim}{z}_n$, the entries of Σ will be given by (5.6). The k-L procedure estimates a vector parameter $\underset{\sim}{\theta}$ by maximizing the asymptotic normal form of the log-likelihood of $\underset{\sim}{z}_n$. This may be taken either as

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

or as just the second term of this expression. Under very general conditions, the asymptotic efficiency of the k-L procedure can be made arbitrarily close to the Cramer-Rao bound by selecting the grid $\{t_j\}$ to be sufficiently fine and extended. The harmonic-regression procedure is equivalent, asymptotically, to the k-L procedure and involves finding $\underset{\sim}{\theta}$ by fitting $\underset{\sim}{z}_\theta$ to $\underset{\sim}{z}_n$ using nonlinear least squares and any consistent estimate of the asymptotically optimal weights. This is carried out using a first order expansion for $\underset{\sim}{z}_\theta$; to preserve the asymptotic properties, a single iteration starting from any consistent estimates suffices. The ease with which this procedure may be implemented contrasts sharply with the methods of §3.

The harmonic regression procedure was implemented for the parametrization

$\underset{\sim}{\theta} = (\mu,\sigma,\alpha)$ for the symmetric stable laws: $c_\theta(t) = e^{i\mu t} \cdot e^{-|\sigma t|^\alpha}$. We used centered variates $\hat{X}_j = (X_j - \hat{\mu})/\hat{\sigma}$ where $\hat{\mu}$, $\hat{\sigma}$ were current estimates: for symmetric families this gives a convenient block-diagonal structure for Σ . In particular the $2k\times 3$ regression may then be separated into a $k\times 2$ and a $k\times 1$ regression - the former involving only α and σ , and the latter only μ . The procedure was tested extensively and proved to be well behaved, however, because the question of optimal gridpoints is not resolved we did not carry out a Monte Carlo study for this procedure.

6. FOURIER GRIDPOINTS FOR THE SYMMETRIC STABLE LAWS

For fixed k the optimal $\{t_j\}$ depends on the unknown $\theta = (\mu,\sigma,\alpha)$. Since iteration takes place at standardized variates, we may presume without loss of generality that $\mu = 0$, $\sigma = 1$; the dependence on α is therefore the one of greatest interest. In general, for fixed k we propose to use those points which minimize the asymptotic variances (at the current estimates). For arbitrary spacing, the asymptotic covariance of the estimators is $\underset{\sim}{G}' \Sigma^{-1} \underset{\sim}{G}$ where $\underset{\sim}{G} = \left(\frac{\partial Z_{\underset{\sim}{\theta}}}{\partial \mu}, \frac{\partial Z_{\underset{\sim}{\theta}}}{\partial \sigma}, \frac{\partial Z_{\underset{\sim}{\theta}}}{\partial \alpha} \right) \Big|_{\theta = (0,1,\alpha)}$. Actually the optimal $\{t_j\}$ depends on which parameter variance, or which joint-criterion we wish to minimize (the determinant of the covariance matrix being one possibility). Now the updating algorithm is such that the μ adjustment depends only on the imaginary ecf, while the σ , α adjustments depend only on the real part. (Note that the increased computational burden in using distinct grids for the real and imaginary components is very slight. The components for the imaginary parts covariances in (5.6) will now differ from those for the real parts.) The important tradeoff therefore takes place on the real axis between σ and α .

The question of optimal gridpoints requires numerical treatment. For the case of uniform spacing we refer to Feuerverger and McDunnough (1979), especially table 1. According to these results, α is the parameter least amenable to uniform spacing, particularly for small α , and is hence the parameter of greatest interest here. Adjustments were made to the programme which calculates the asymptotic variances. By an iterative procedure of arbitrary starts, sequential optimization using steepest ascent on lattices, and further checking, grid-points were obtained which appear to be optimal (asymptotically) for the estimation of α . These were obtained for $k = 2, 3, 4$ and 5 with $\alpha = 1.0\ (.1)\ 1.9$ and are given in table 6.1.

Table 6.2 compares the asymptotic values of $n \cdot \mathrm{var}(\hat{\alpha})$ for α-optimal spacing of $k = 2, 3, 4$ and 5 points. The Cramer-Rao bound value ranges shown are determined from DuMouchel (1975). It may be noted that the change from optimal *uniform* spacing to optimal spacing involves a sharp improvement in the asymptotic efficiences. As before, however, for fixed k , efficiencies are seen to decrease with α ; in particular, for $k = 5$ the efficiency is seen to drop below 90% for $\alpha < 1.1$. We may remark that while the *optimal* spacings become costlier and more difficult to determine as k increases, the results of table 6.1 provide a useful guide in determining good spacings for larger k . In practice there are no special difficulties in using values of $k = 10$ or even 20 . For further results and discussion (and an indication of the intrinsic complexity of the k-dimensional surfaces optimized here) we refer the reader to our technical report.

7. A SIMULATION STUDY FOR AR(1) STABLE PROCESSES

The closure under convolutions property of the stable distributions provides a

α	k = 2		k = 3			k = 4				k = 5				
	t_1	t_2	t_1	t_2	t_3	t_1	t_2	t_3	t_4	t_1	t_2	t_3	t_4	t_5
1.0	.09	1.83	.03	.14	1.73	.03	.16	1.48	2.43	.016	.07	.20	1.4	2.4
1.1	.11	1.70	.04	.17	1.62	.04	.18	1.39	2.22	.022	.09	.23	1.3	2.1
1.2	.13	1.59	.05	.19	1.53	.05	.21	1.30	2.03	.026	.10	.25	1.3	2.0
1.3	.15	1.50	.06	.22	1.44	.07	.24	1.23	1.87	.032	.12	.28	1.2	1.8
1.4	.18	1.41	.07	.24	1.36	.08	.26	1.17	1.72	.042	.14	.31	1.1	1.7
1.5	.20	1.33	.09	.27	1.29	.048	.15	.31	1.28	.050	.16	.33	1.1	1.6
1.6	.22	1.26	.10	.29	1.22	.058	.17	.33	1.20	.062	.18	.36	1.0	1.4
1.7	.24	1.17	.12	.32	1.14	.070	.19	.36	1.12	.074	.20	.37	1.0	1.3
1.8	.25	1.09	.13	.33	1.05	.082	.21	.38	1.04	.082	.21	.39	0.9	1.1
1.9	.27	0.95	.15	.35	0.92	.100	.23	.40	.90	.096	.22	.38	1.0	1.6

TABLE 6.1 α-optimal spacings for k = 2 to k = 5

TABLE 6.2

Asymptotic values of $N \cdot VAR(\hat{\alpha})$
for α-optimal spacings

α	k = 2	k = 3	k = 4	k = 5
1.0	1.8327	1.5890	1.4637	1.393
1.1	2.0993	1.8426	1.7207	1.648
1.2	2.3432	2.0826	1.9710	1.898
1.3	2.5478	2.2924	2.1974	2.127
1.4	2.6918	2.4524	2.3774	2.314
1.5	2.7508	2.5363	2.4831	2.427
1.6	2.6942	2.5120	2.4691	2.434
1.7	2.4839	2.3402	2.3079	2.290
1.8	2.0698	1.9682	1.9464	1.940
1.9	1.3721	1.3152	1.3038	1.284

class of highly tractable linear stationary processes. With an MLE algorithm already available, it is not very difficult to carry out maximum likelihood estimation for stable autoregressive processes. We give here the results of a brief simulation study for the AR(1) case.

Suppose $X(t)$ is a discrete stationary process satisfying

$$(X(t) - \mu) = a(X(t-1) - \mu) + e(t)$$

where the $e(t)$ are iid variates of the form σS_α where S_α is the standardized symmetric stable law. Then

$$(X(t) - \mu) = \sum_{j=0}^{\infty} a^j e(t-j) \sim \sigma(1 - |a|^\alpha)^{-\frac{1}{\alpha}} S_\alpha \quad .$$

The joint density for a length T of the stationary process

$$f(x_1, x_2, \ldots, x_T) = f(x_1) \prod_{j=2}^{T} f(x_j | x_{j-1})$$

$$= \frac{1}{\sigma^*} f_\alpha \left(\frac{x_1 - \mu}{\sigma^*} \right) \frac{1}{\sigma^{T-1}} \prod_{j=2}^{T} f_\alpha \left(\frac{(x_j - \mu) - a(x_{j-1} - \mu)}{\sigma} \right)$$

where $\sigma^* = \sigma(1 - |a|^\alpha)^{-1/\alpha}$ and where f_α is the density of the standard stable variate. We may therefore easily carry out the full (unconditional) MLE procedure for such AR processes by adapting the methods of §3-4.

Tables 7.1 and 7.2 summarize the results of a limited simulation study for the cases $\alpha = 1.9, 1.8, 1.6, 1.0$ and $a = 0, .25, .5$ for series lengths of $T = 200$ and $T = 500$. The data series used were constructed with $\mu = 0$, $\sigma = 1$ and initial estimates were taken at the true values and followed by five Newton-Raphson iterations. Convergence was, in general, extremely rapid. We carried out $n = 25$ trials for each cell; values $n < 25$ indicate that some trials did not terminate normally (see §4). As before, the tables show the mean and standard deviation of the n trials for each parameter.

Of particular interest, though not apparent in these tables, is an instability in the $Var(\hat{a})$ values produced by the MLE procedure, with many of these values being exceedingly small. This phenomenon occurs because the

accuracy of estimates for AR coefficients clearly is conditional on whether or not there are some extreme outliers present to help us. This is a numerical confirmation of the limiting infinite Fisher information per observation for this parameter: see Hannan and Kanter (1977) and Kanter and Steiger (1974). A related observation is made by Cox (1966).

TABLE 7.1
Simulation Results for Symmetric Stable Time Series of Length T = 200
with 25 Trials per Cell and 5 Iterations per Trial

		a = 0		a = .25		a = .5	
$\alpha = 1.9$	μ	.013	.109	-.054	.166	-.017	.153
	α	1.877	.051	1.853	.055	1.859	.064
	σ	.984	.049	1.003	.045	.981	.046
	a	.024	.042	.248	.066	.499	.037
	n	16		18		18	
$\alpha = 1.8$	μ	.032	.090	-.005	.105	.011	.082
	α	1.792	.132	1.815	.082	1.826	.115
	σ	.998	.057	.989	.067	.990	.058
	a	.000	.052	.237	.072	.486	.041
	n	25		25		25	
$\alpha = 1.6$	μ	-.021	.129	.007	.131	.039	.275
	α	1.633	.109	1.609	.107	1.644	.112
	σ	.996	.048	1.014	.061	1.005	.061
	a	.014	.049	.256	.043	.499	.044
	n	25		25		25	
$\alpha = 1.0$	μ	-.032	.109	-.048	.133	.023	.190
	α	1.007	.080	.986	.084	1.008	.065
	σ	.985	.082	.989	.087	1.005	.102
	a	-.003	.009	.251	.008	.501	.007
	n	23		22		25	

TABLE 7.2
Simulation Results for Symmetric Stable Time Series of Length T = 500
with 25 Trials per Cell and 5 Iterations per Trial

		a = 0		a = .25		a = .5	
$\alpha = 1.9$	μ	.024	.074	.001	.094	-.011	.140
	α	1.895	.039	1.879	.048	1.887	.056
	σ	1.012	.044	.994	.031	.998	.049
	a	.013	.053	.247	.044	.504	.032
	n	19		18		20	
$\alpha = 1.8$	μ	.023	.173	.023	.080	.036	.127
	α	1.783	.063	1.807	.074	1.824	.069
	σ	.999	.033	1.003	.049	.984	.037
	a	.005	.029	.253	.032	.507	.035
	n	25		25		25	
$\alpha = 1.6$	μ	.016	.052	.007	.092	.002	.166
	α	1.612	.071	1.598	.061	1.604	.062
	σ	.991	.039	1.001	.056	1.005	.038
	a	.000	.023	.254	.031	.494	.019
	n	25		25		24	
$\alpha = 1.0$	μ	-.006	.072	.017	.092	-.030	.146
	α	1.013	.052	1.010	.046	1.006	.032
	σ	1.013	.063	1.014	.045	.980	.057
	a	.000	.005	.249	.006	.501	.003
	n	25		25		25	

Acknowledgments
The research of both authors is supported by NSERC. The manuscript was typed by Margot Burrows.

8. REFERENCES

[1] Berger, J.M. and Mandelbrot, B., A new model for error clustering in telephone circuits, I.B.M. Journal of Research and Development 7 (1963) 224-236.

[2] Bohman, H., Approximate Fourier analysis of distribution functions, Ark. Mat. 4 (1960) 99-157.

[3] Brillinger, D.R., Time Series: Data analysis and theory (Holt, Rinehart and Winston, New York, 1975).

[4] Brockwell, P.J. and Brown, B.M., Efficient estimation for the positive stable laws. Submitted for publication.

[5] Chambers, J.M., Mallows, C.L. and Stuck, B.W., A method for simulating stable random variables, J. Amer. Statist. Assoc. 71 (1976) 340-344.

[6] Cooley, J.W. and Tukey, J.W., An algorithm for the machine calculation of complex Fourier series, Math. Comp. 19 (1965) 297-301.

[7] Cox, D.R., The null distribution of the first serial correlation coefficient, Biometrika 53 (1966) 623-626.

[8] Csorgo, S., Limit behaviour of the empirical characteristic function, Annals Probab., to appear.

[9] de Haan, L. and Resnick, S.I., A simple asymptotic estimate for the index of a stable distribution, J. Royal Statist. Soc. (B) 42 (1980) 83-87.

[10] DuMouchel, W.H., Stable distributions in statistical inference: 2 - Information from stably distributed samples, J. Amer. Statist. Assoc. 70 (1975) 386-393.

[11] DuMouchel, W.H., Stable distributions in statistical inference: 3 - Estimation of the parameter of a stable distribution by the method of maximum likelihood. Unpublished manuscript. Presented to NBER-NSF Conference on Bayesian Econometrics, Ann Arbor, Michigan, April 1974.

[12] DuMouchel, W.H., On the asymptotic normality of maximum-likelihood estimates when sampling from a stable distribution, Annals of Statist. 1 (1973) 948-957.

[13] DuMouchel, W.H., Stable distributions in statistical inference, Ph.D. dissertation, Yale University.

[14] Fama, E.F. and Roll, R., Parameter estimates for symmetric stable distributions, J. Amer. Statist. Assoc. 66 (1971) 331-338.

[15] Fama, E.F. and Roll, R., Some properties of symmetric stable distributions, J. Amer. Statist. Assoc. 63 (1968) 817-836.

[16] Fama, E.F., The behaviour of stock market prices, J. Bus. 38 (1965) 34-105.

[17] Feller, W., An Introduction to Probability Theory and its Application, Vol. 2 (Wiley, New York, 1966).

[18] Feuerverger, A. and McDunnough, P., On some Fourier methods for inference, submitted to J. Amer. Statist. Assoc., 1979.

[19] Feuerverger, A. and McDunnough, P., On the efficiency of empirical characteristic function procedures, J. Royal Statist. Soc., Part 3 (1980) to appear.

[20] Feuerverger, A. and Mureika, R.A., The empirical characteristic function and its applications, Annals. Statist. 5 (1977) 88-97.

[21] Fielitz, B.D. and Smith, E.W., Asymmetric stable distributions of stock price changes, J. Amer. Statist. Assoc. 67 (1972) 813-814.

[22] Gnedenko, B.V. and Kolmogorov, A.N., Limit Distributions for Sums of Independent Random Variables, (trans. by K.L. Chung)(Addison-Wesley, Massachusetts, 1954).

[23] Hannan, E.J. and Kanter, M., Autoregressive processes with infinite variance, J. Appl. Prob. 14 (1977) 411-415.

[24] Hart, J.F. et al, Computer Approximations (Wiley, New York, 1968).

[25] Heathcote, C.R., The integrated squared error estimation of parameters, Biometrika 64 (1977) 255-264.

[26] Holt, H.R. and Crow, E.L., Tables and graphs of the stable probability density functions, J. Res. of the Nat. Bur. of Standards 77B (1973) 143-198.

[27] Kanter, M. and Steiger, W.L., Regression and autogression with infinite variance, Adv. Appl. Prob. 6, (1974) 768-783.

[28] Leitch, R.A. and Paulson, A.S., Estimation of stable law parameters: stock, price behaviour application, J. Amer. Statist. Assoc. 70 690-697.

[29] Lévy, P., Théorie de l'addition des Variables Aléatoires, 2nd ed., (Gauthier-Villars, Paris, 1954).
[30] Lévy, P., Calcul des Probabilités, (Gauthier-Villars, Paris, 1925).
[31] Lukacs, E., Characteristic Functions, (Hafner, Connecticut, 1970).
[32] Mandelbrot, B., The variation of certain speculative prices, J. Bus. of the Univ. of Chicago 26 (1963), 394-419.
[33] Mandelbrot, B., The Pareto-Levy law and the distribution of income, Int. Econ. Rev. 1 (1960), 79-106.
[34] Marcus, M.B., Weak convergence of the empirical characteristic function, Annals Probab., to appear.
[35] Paulson, A.S., Halcomb, E.W. and Leitch, R.A., The estimation of the parameters of the stable laws, Biometrika 62 (1975), 163-170.
[36] Press, S.J., Estimation in univariate and multivariate stable distributions, J. Amer. Statist. Assoc. 67 (1972), 842-846.
[37] Samuelson, P., The mathematics of speculative price, S.I.A.M. Review 15, No. 1 (1973), 1-41.
[38] Stuck, B.W. and Kleiner, B., A statistical analysis of telephone noise, Bell System Technical Journal 53, No. 7 (1974), 1263-1312.
[39] Tukey, J.W., An introduction to the calculations of numerical spectrum analysis, in Advanced Seminar on Spectral Analysis of Time Series, ed. B. Harris, (Wiley, New York, 1967, 25-46).