Figure 4

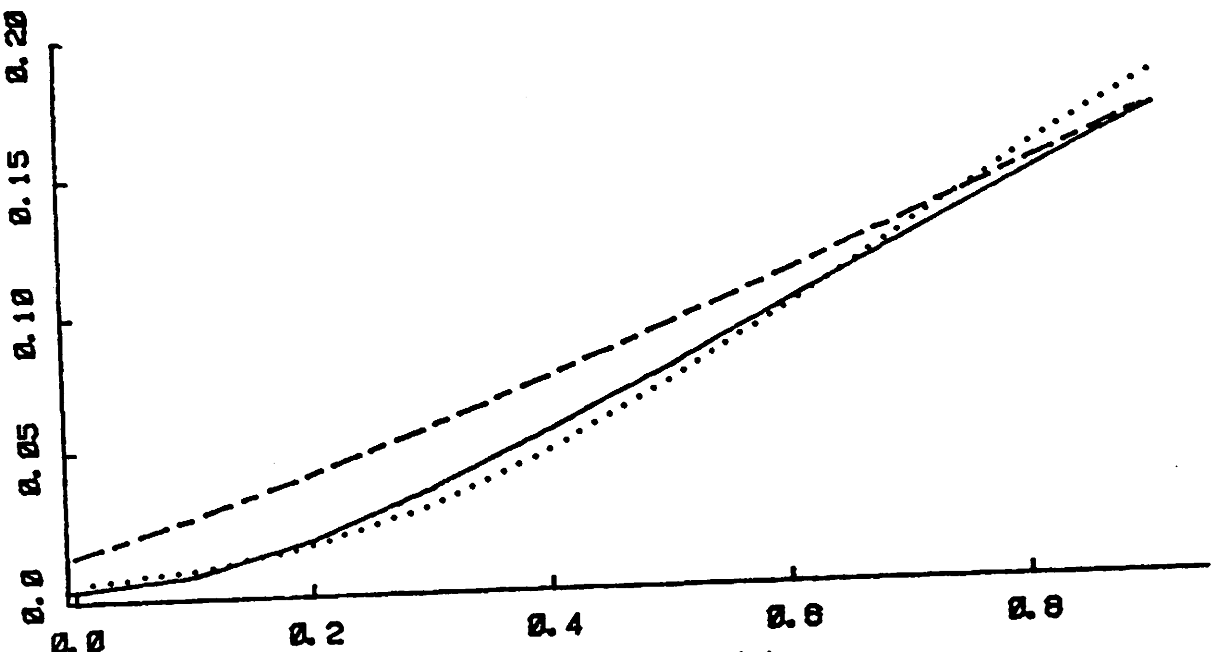

Figure 4(a): Lower tails from Figure 3(a)

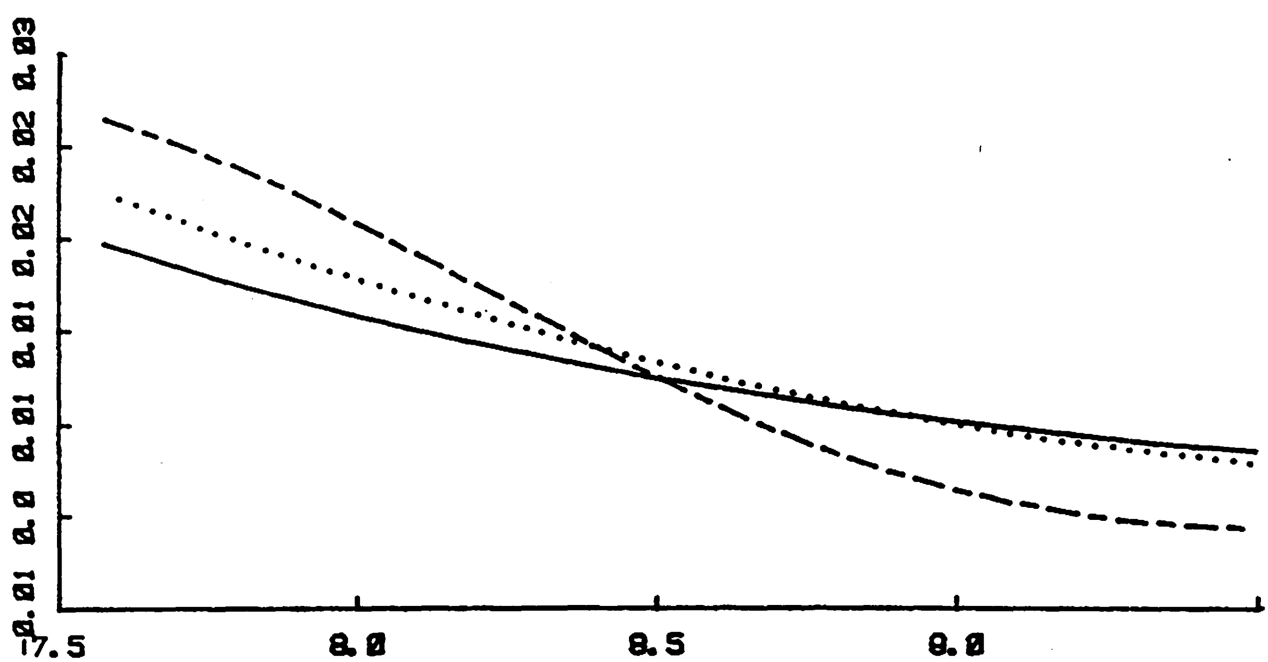

Figure 4(b): Upper tails from Figure 3(b)

Figure 3

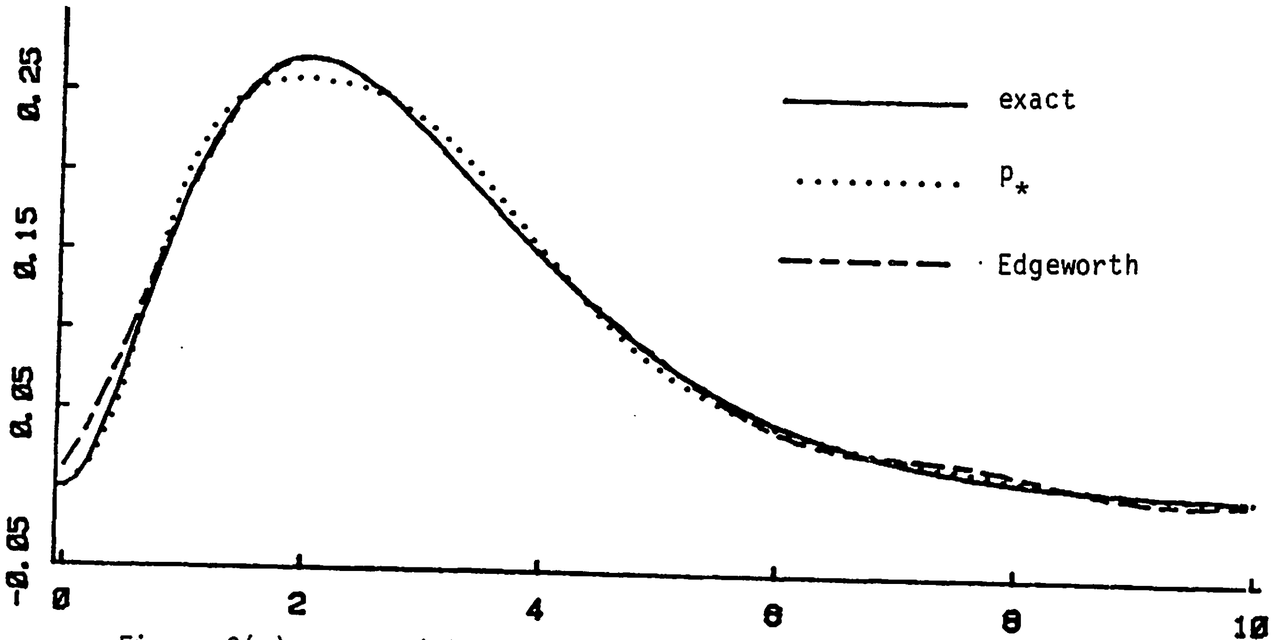

Figure 3(a): Gamma(3) density and approximations

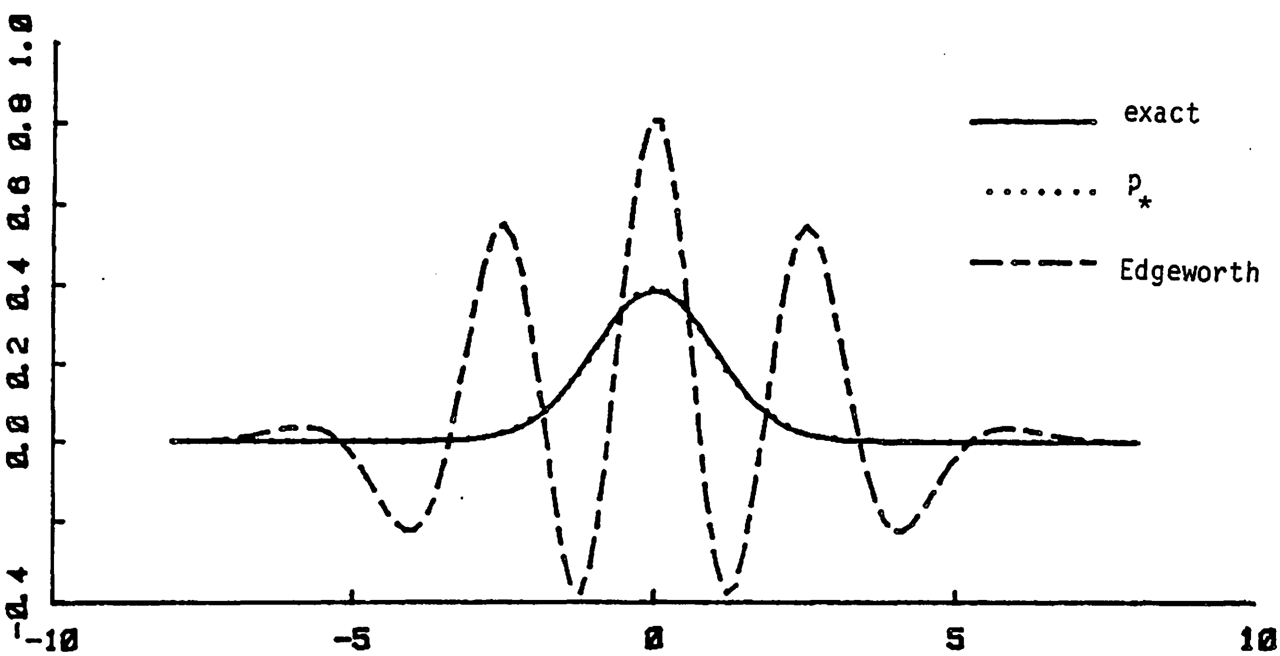

Figure 3(b): Contaminated Gaussian density and approximations

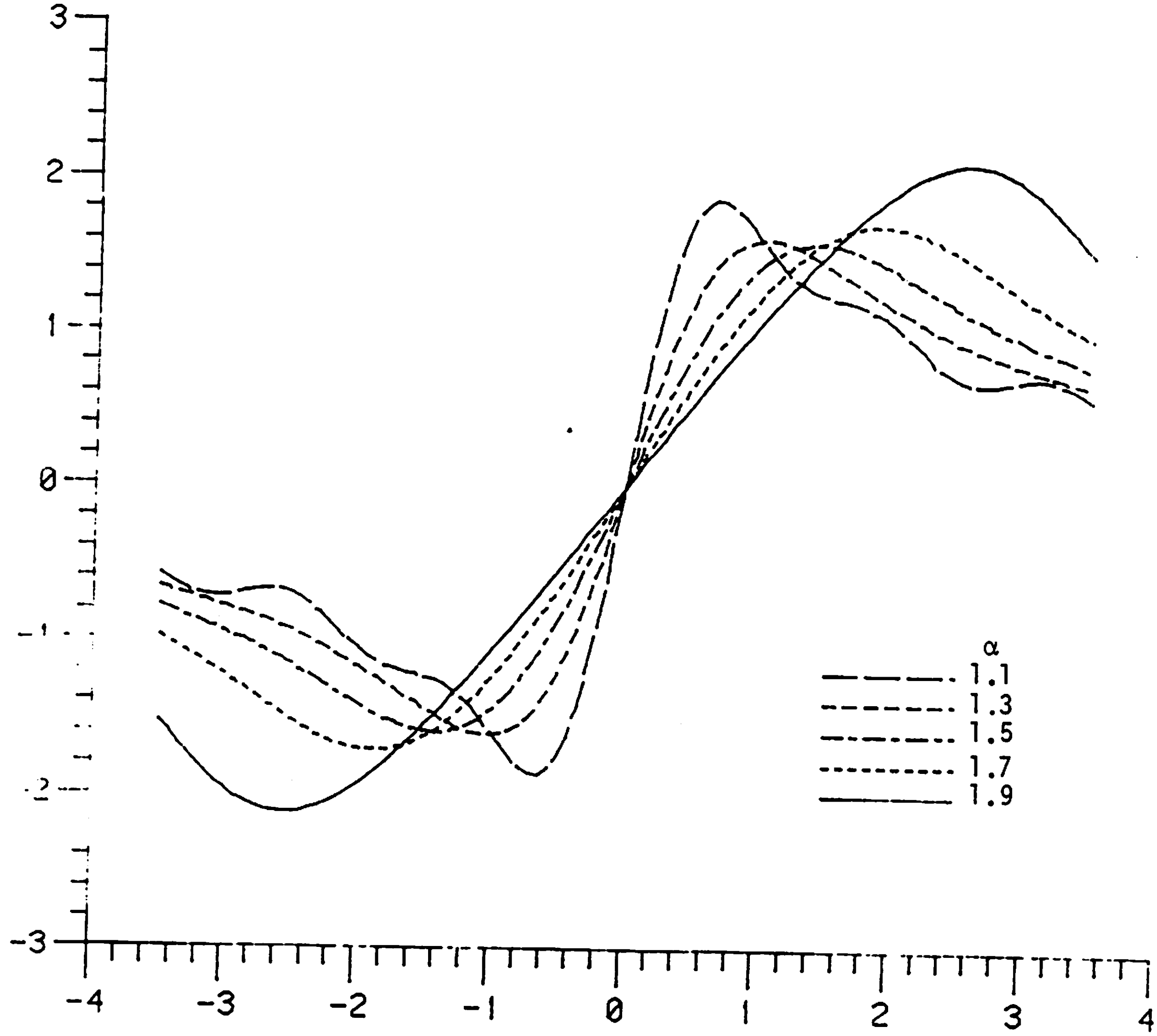

Figure 2: Stable Models (α=1.1(.2)1.9) $J_*(\tau=.3, k=15)$

Figure 1 (cont.)

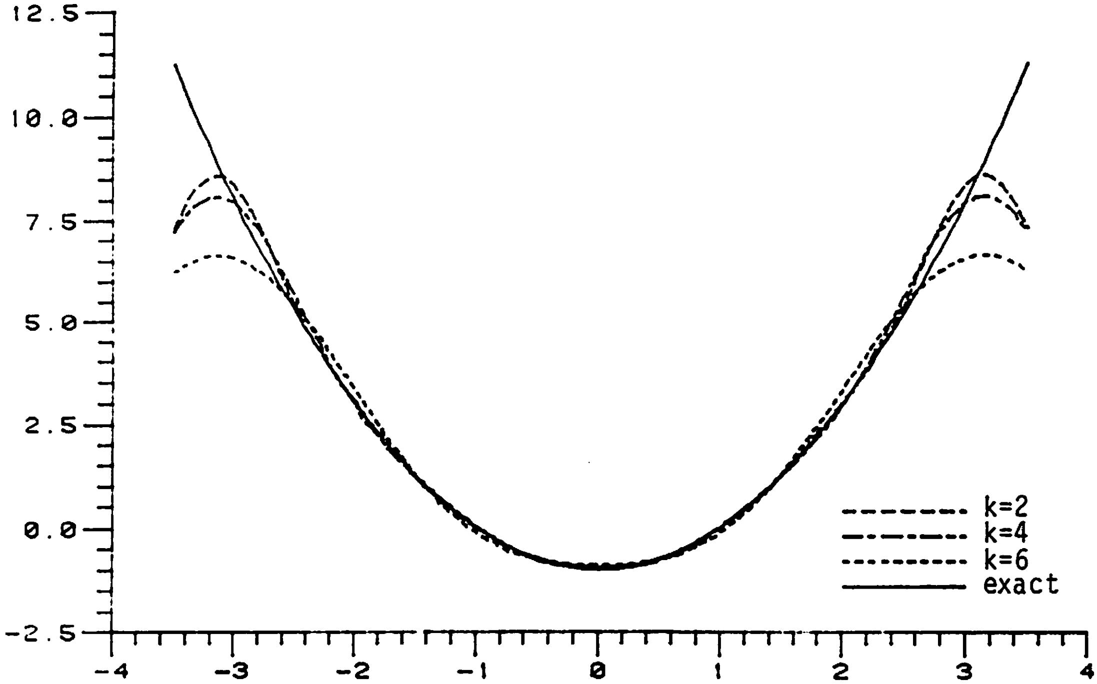

Figure 1(c): Gaussian Model: . K and $K_*(\tau=1;\ k=2,4,6)$

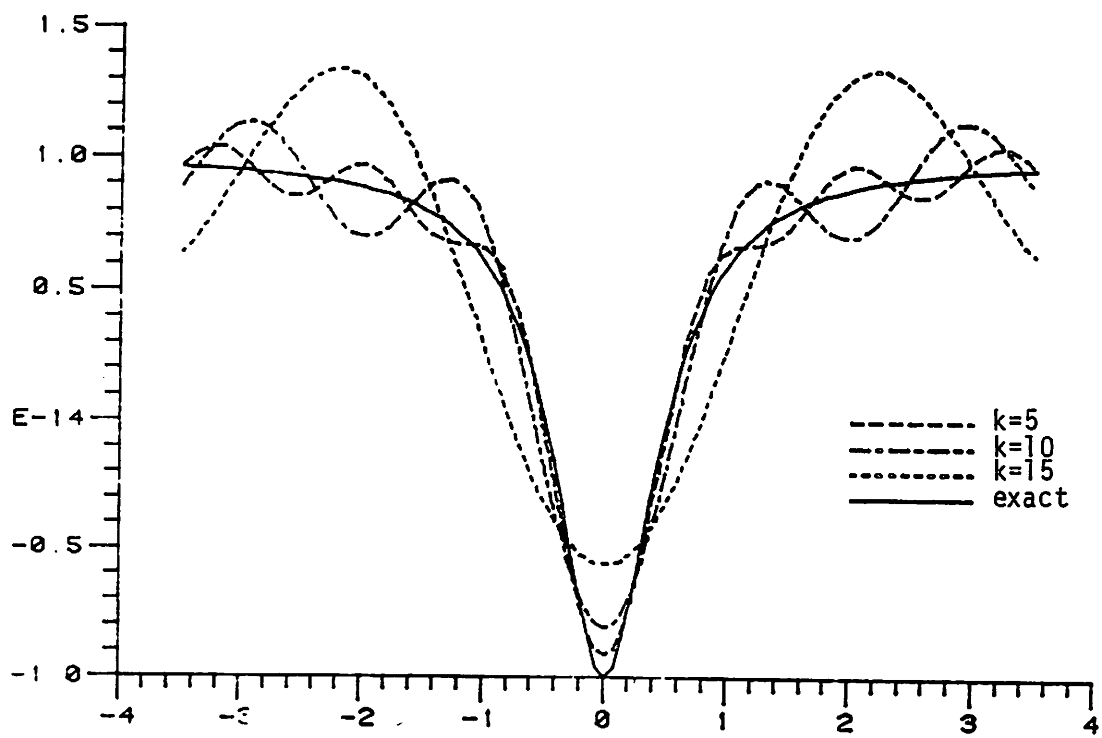

Figure 1(d): Cauchy Model: K and $K_*(\tau=1/3;\ k=5,10,15)$

Figure 1

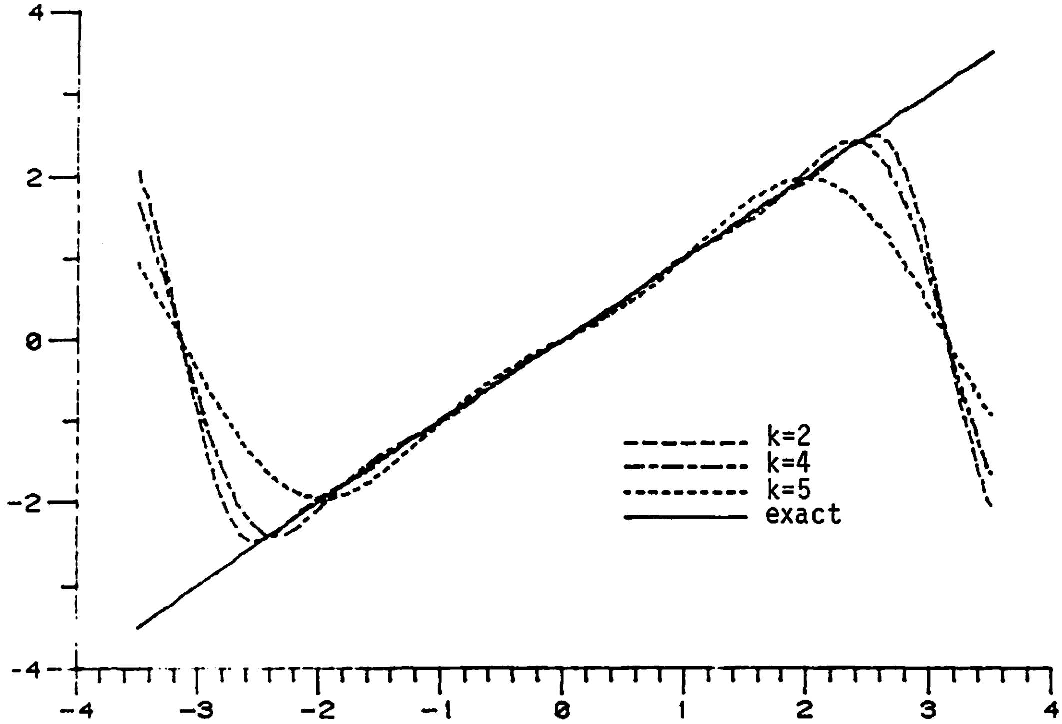

Figure 1(a): Gaussian Model: J and $J_*(\tau=1;\ k=2,4,6)$

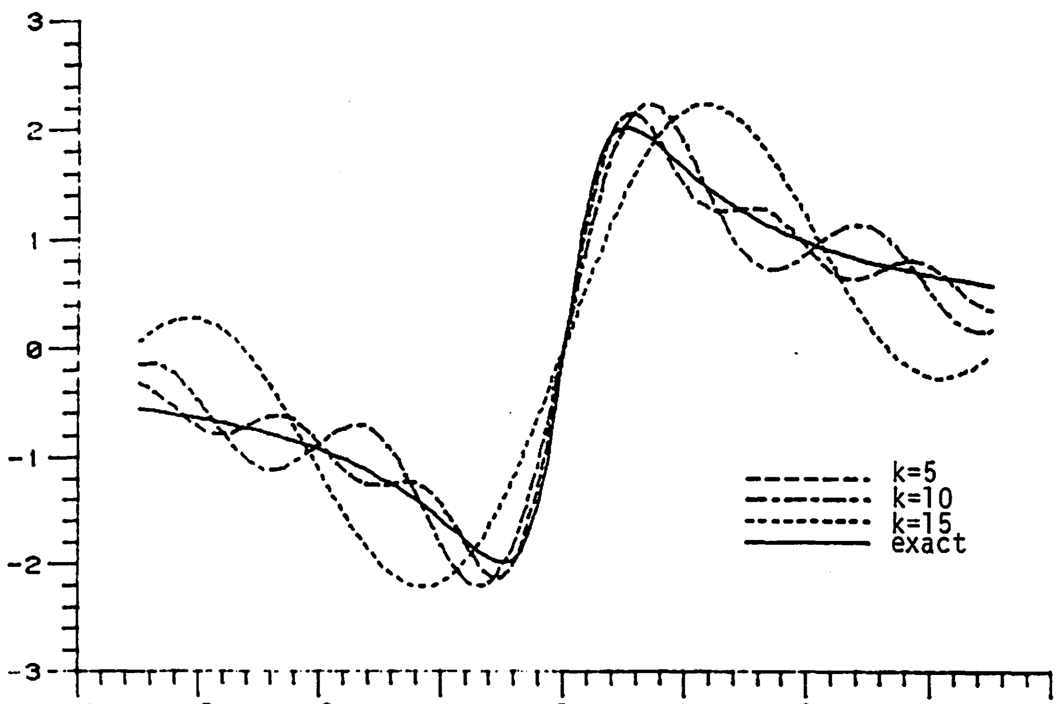

Figure 1(b): Cauchy Model: J and $J_*(\tau=1/3;\ k=5,10,15)$

Table 3

Tail Probabilities and Approximations

Lower Tails

x	True	F_*	Relative Error	G.C. Type A	Relative Error	Edgeworth	Relative Error
.05	.000	.000	658.6%	.230	> 10^6%	-.010	-51796.3%
.10	.000	.000	162.5%	.178	114862.6%	-.009	-6091.1%
.15	.001	.001	59.7%	.122	24078.3%	-.008	-1653.0%
.20	.001	.001	21.6%	.061	5254.9%	-.006	-621.3%
.25	.002	.002	4.1%	-.002	-192.5%	-.004	-274.8%
.30	.004	.003	-4.6%	-.068	-2000.1%	-.001	-132.4%
.40	.008	.007	-10.8%	-.207	-2717.5%	.005	-32.8%
.50	.014	.013	-10.7%	-.350	-2535.6%	.014	-5.5%
.60	.023	.021	-8.3%	-.491	-2225.1%	.024	2.5%
.70	.034	.032	-5.4%	-.623	-1925.7%	.036	4.5%

Upper Tails

x	True	F_*	Relative Error	G.C. Type A	Relative Error	Edgeworth	Relative Error
7.50	.020	.022	8.4%	.282	1290.9%	.021	2.9%
7.75	.017	.018	6.2%	.295	1663.5%	.016	-6.4%
8.00	.014	.014	3.6%	.252	1729.1%	.011	-18.4%
8.25	.011	.011	.7%	.179	1478.7%	.008	-30.9%
8.50	.009	.009	-2.1%	.099	970.5%	.005	-41.3%
8.75	.008	.007	-4.2%	.031	312.6%	.004	-46.7%
9.00	.006	.006	-5.2%	-.017	-364.9%	.003	-45.0%
9.25	.005	.005	-4.7%	-.043	-940.3%	.003	-35.1%
9.50	.004	.004	-2.3%	-.051	-1324.8%	.003	-17.6%
9.75	.003	.003	2.3%	-.047	-1476.6%	.004	5.6%

Table 2

Information Calculations for the Binomial Poisson Convolution Model

Parameters			Exact (I) and Approximate (I^*) Calculations					
N	θ	λ	I^*_θ	I_θ	I^*_λ	I_λ	$I^*_{\theta\lambda}$	$I_{\theta\lambda}$
20	.50	10	26.64667	27.06326	.06581	.06691	1.30205	1.34342
20	.50	50	7.31844	7.32785	.01816	.01818	.36290	.36336
20	.90	10	37.28053	37.51228	.08471	.08504	1.65376	1.66239
20	.90	50	7.94860	7.95635	.01929	.01931	.38534	.38568
100	.50	10	286.04950	286.04951	.02878	.02878	2.84876	2.84876
100	.50	50	133.72869	133.72869	.01334	.01334	1.33136	1.33136
100	.90	10	538.53335	538.53335	.05362	.05362	5.15320	5.15320
100	.90	50	172.88912	172.88912	.01696	.01696	1.68880	1.68880

Table 1

Asymptotic Efficiencies (I^*/I) for
Stable Family Location Estimates

Stable exponent (α)	Order of trigonometric approximation (k)		
	5	10	15
1.1	.86225	.97722	.99158
1.3	.96261	.99465	.99514
1.5	.99426	.99717	.99718
1.7	.99889	.99929	.99929
1.9	.99863	.99888	.99889

JACKSON, D. (1930). Theory of Approximation. American Mathematical Society Colloquium Publication 11.

JARRETT, R. G. (1973). Efficiency and estimation in asymptotically normal distribution. Doctoral Thesis, Univ. of London.

SCLOVE, S. L. and VAN RYZIN, J. (1969). Estimating the parameters of a convolution, J. Roy. Statist. Soc., Ser. B., pp. 181-191.

SPROTT, D. A. (1983). Estimating the parameters of a convolution by maximum likelihood. J. Amer. Statist. Assoc., 78, pp. 457-463.

REFERENCES

BROCKWELL, P. J. and BROWN, B. M. (1982). High efficiency estimation for the positive stable laws. J. Amer. Statist. Assoc., 76, pp. 626-631.

BRYANT, J. L. and PAULSON, A. S. (1979). Some comments on characteristic function based estimators. Sankhya A., 41, pp. 109-116.

CHAMBERS, R. L. and HEATHCOTE, C. R. (1981). On the estimation of slope and the identification of outliers in linear regression. Biometrika, 68, pp. 21-33.

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

FEUERVERGER, A. and McDUNNOUGH, P. (1981a). On the efficiency of empirical characteristic function procedures. J. Roy. Statistic Soc., Ser. B, 43, pp. 20-27.

FEUERVERGER, A. and McDUNNOUGH, P. (1981b). On some Fourier methods of inference. J. Amer. Statist. Assoc., 76, pp. 379-387.

FEUERVERGER, A. and McDUNNOUGH, P. (1981c). On efficient inference in symmetric stable laws and processes. Statistics and Related Topics: Proceedings of the International Symposium on Statistics 1980, New York: North-Holland.

FEUERVERGER, A. and McDUNNOUGH, P. (1982). On statistical transform methods and their efficiency. preprint.

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

HAMPEL. F. (1973). Some small sample asymptotics. Proceedings of the Prague Symposium on Asymptotic Statistics, Vol. II, pp. 109-126.

4. exactly recovers $p(x)$ when J is a polynomial of order less than or equal to k, for example when P is exponential or Gaussian. (The Gram Charlier and Edgeworth expansions recover Gaussian distributions only after an appropriate standardization of the moments).

The approximation is applicable whenever the moments of P exist and uniquely characterize P, and are known to sufficient precision to compute accurately the coefficients determining $\tilde{\ell}_*$.

The following examples give some indication of the quality of approximation afforded by p_*, in comparison with its competitors. In Figure 3 and 4, p_*, based on the least-squares 11^{th} degree polynomial approximation to $J(x)$ (so that ℓ_* is 12^{th} degree) for a Gamma(3) distribution, is compared to the 12^{th} degree Edgeworth expansion. The corresponding Gram-Charlier expansion fared so poorly that it could not be plotted on the same scale, and hence was excluded. As the plot indicates, p_* is superior to the Edgeworth expansion in the tails, but is worse in the middle. Table 3, which gives tail probabilities (determined by numerical integration in the case of p_*) bears this out. In addition, the techniques were applied to the approximation of the density of $.95 \cdot N(0,1) + .05 \cdot N(0,9)$ (see Figure 3). The Edgeworth expansion fails to converge in this case, while the behaviour of the Gram-Charlier expansion is even wilder in its fluctuations.

[Figures 3 & 4 here]

[Table 3 here]

role the Gaussian distribution in density approximation. Hampel (1973) notes of the Gaussian density f that "f'(x)/f(x) is a linear function, and the simplest function which in turn determines the normal distribution".

The possibility of approximating J(x) by a polynomial arises since

$$\eta_k = \text{cov}\{J(x), x^k\}$$

$$= k\mu_{k-1} ,$$

under mild regularity conditions.

Letting $\underset{\sim}{\eta} = (\eta_1, \eta_2, \ldots, \eta_k)^T$, $\underset{\sim}{t} = (1,2,3,\ldots,k)$ and $\underset{\sim}{\mu} = E\{x^{\underset{\sim}{t}}\}$ and $\Sigma = \text{var}(x^{\underset{\sim}{t}})$, with entries $\Sigma_{ij} = \mu_{i+j} - \mu_i\mu_j$, we can write $J_*(x) = \underset{\sim}{\eta}^T\Sigma^{-1}(x^{\underset{\sim}{t}} - \underset{\sim}{\mu})$. Integrating termwise yields the improper form

$$\tilde{\ell}_*(x) = \underset{\sim}{\eta}^T\Sigma^{-1}(D^{-1}x^{(\underset{\sim}{t}+\underset{\sim}{1})} - \underset{\sim}{\mu})$$

where $D = \text{diag}(\underset{\sim}{t}+\underset{\sim}{1})$, where $\underset{\sim}{1}$ is a k-vector of 1's. The convergence of $\ell_*(x)$ holds if J(x) lies in $L_2(dP)$, since $J_* \to J$ in L_2 implies the L_1 result, and since

$$\left|\int_0^x [J(u)-J_*(u)]\cdot du\right| = |\ell(x) - \ell(0) - \ell_*(x)|$$

$$\leq \frac{1}{\lambda(x)} \int |J(x)-J_*(x)|p(x)dx$$

where $\lambda(x) = \inf\{p(u);\ 0 \leq u \leq x\}$. This rather crude upper bound ($\lambda(x)$ may be rather small) is still sufficient, based on the L_1 convergence noted, to confirm the pointwise convergence of $\tilde{\ell}_*(x)$ to $\ell(x)-\ell(0)$, as long as $\lambda(x) \neq 0$. By taking $p_*(x) = c_*\exp(\ell_*(x))$, where c_* is the normalizing constant determined by numerical integration, we achieve an approximation to p(x) that:

1. is easy to compute,
2. is always positive,
3. minimizes the error of approximation on an appropriate scale, and

which may be approximated by

$$S_*(x;\theta)^2 - \ddot{\phi}(\theta)^*\Sigma(\theta)^{-1}[\exp i\underset{\sim}{t}x - \underset{\sim}{\phi}(\theta)].$$

Note that $S_*(x;\theta)^2$ is a trigonometric polynomial, and that we can write

$$S_*(x;\theta)^2 = \dot{\underset{\sim}{\phi}}(\theta)^*\Sigma(\theta)^{-1}\exp i\underset{\sim}{t}x\cdot(\exp i\underset{\sim}{t}x)^*\Sigma(\theta)^{-1}\dot{\underset{\sim}{\phi}}(\theta),$$

so that an alternative approximation to the observed information is

$$I_{*n} = n\{\dot{\underset{\sim}{\phi}}(\hat{\theta})^*\Sigma(\hat{\theta})^{-1}\hat{\Phi}\Sigma(\hat{\theta})^{-1}\dot{\underset{\sim}{\phi}}(\hat{\theta}) - \ddot{\underset{\sim}{\phi}}(\hat{\theta})^*\Sigma(\hat{\theta})^{-1}[\hat{\underset{\sim}{\phi}} - \underset{\sim}{\phi}(\hat{\theta})]\},$$

where $\hat{\Phi}$ has entries,

$$\hat{\Phi}_{ij} = \hat{\phi}([i-j]\cdot\tau).$$

7. Probability Calculations Based on Transforms

As noted in Section 6, $p_*(x) = \exp=\{\ell_*(x)\}$ yields a new approximation to p(x) (in what follows we shall neglect the parameter θ, which is irrelevant). There ℓ_* was based on the c.f., $\phi(t)$, whereas here we will consider a version of ℓ_* based on the moments, $\mu_k = E\{x^k\}$, k=1,2,3,..., and will compare the resultant $p_*(x)$ with the Gram Charlier Type A series, and the closely related Edgeworth expansion, both of which provide moment based approximations to p(x). Briefly, the latter expansions arise from best $L_2(\omega dx)$ polynomial approximations to $h(x) = p(x)/\omega(x)$ where $\omega(x)$ is the standard Gaussian density. Minimizing errors of approximation on the implied scale of measurement does not always yield satisfactory results. The approximations may fail to converge, or may be deficient in the tails, assuming negative values for example. The logarithmic scale, by comparison, is natural in any context where the relative, rather than absolute, error in approximation is relevant. The appearance of $J(x) = -p'(x)/p(x)$ in motivating ℓ_* and p_* is not unnatural either, relative to the central

$$\lim_{k\to\infty}\left[\frac{|\Sigma_k(\theta)|}{|\Sigma_{k-1}(\theta)|}\right]^{\frac{1}{2}} = \frac{2\pi}{\tau}\exp(b_0(\theta)) \quad .$$

This establishes the convergence of $b_*(\theta)$ given in (6.4), to $b_0(\theta)$, completing the proof of the theorem.

Examples considered by the author show that, as in the case of $S_*(x;\theta)$, k need not be inordinately big for $\ell_*(x;\theta)$ to be virtually indistinguishable from $\ell(x;\theta)$. Thus $\ell_*(x;\theta)$ is of general utility. For example, if the empirical characteristic function $\hat{\phi}(t)$ is substituted for exp(itx) in $\ell_*(x;\theta)$, the result is a convenient transform based means of describing the log likelihood surface. As well, $\ell_*(x;\theta)$ may serve as a basis for approximating $p(x;\theta)$, a possibility which is further explored in section 7, using the version of ℓ_k based on the moments.

6.4 Related Forms

Other forms which are commonly encountered in connection with likelihood calculations are the Fisher Information, $\mathcal{I}(\theta)$, and the observed information, I_n, defined in terms of

$$I(x;\theta) = -\frac{\partial^2 p(x;\theta)}{\partial\theta^2} \quad .$$

Fundamental to all our results is the fact that $\mathcal{I}_*(\theta) = \mathring{\phi}(\theta)^*\Sigma(\theta)^{-1}\mathring{\phi}(\theta)$ provides an approximation to $\mathcal{I}(\theta)$. As well, it may be noted that the behaviour of

$$I_{*n} = -\left.\frac{\partial S_{*n}}{\partial\theta}\right|_{\theta=\hat{\theta}} \quad ,$$

where $S_{*n} = \sum_{i=1}^{n} S_*(x_i;\theta)$, is analogous to that of I_n, and may be regarded as an approximation to it.

Alternately, it may be noted that

$$I(x;\theta) = -\frac{\partial^2 \log p(x;\theta)}{\partial\theta^2} = S(x;\theta)^2 - \frac{\ddot{p}(x;\theta)}{p(x;\theta)} \quad ,$$

$\ell_k(x;\theta) = \sum_{j=-k}^{k} b_j(\theta) \exp(ij\tau x)$, converges uniformly to $\ell_\tau(x;\theta)$ on I_τ, as well as in the $L_2(dx_\tau)$ sense. Consider then,

$$\tilde{\ell}_*(x;\theta) = -i \sum_{\substack{j=-k \\ j\neq 0}}^{k} \frac{a_{jk}(\theta)}{j\tau} \exp(ij\tau x),$$

given previously by (6.3). Letting $\tilde{\ell}_k(x;\theta) = \ell_k(x;\theta) - b_0(\theta)$,

$$\frac{\tau}{2\pi}\int_{I_\tau} [\tilde{\ell}_*(x;\theta) - \tilde{\ell}_k(x;\theta)]^2\, dx$$

$$= \sum_{\substack{j=-k \\ j\neq 0}}^{k} \frac{|a_j(\theta)-a_{jk}(\theta)|^2}{(k\tau)^2}$$

$$\leq \frac{1}{\tau^2} \sum_{j=-k}^{k} |a_j(\theta) - a_{jk}(\theta)|^2 \quad ,$$

which tends to 0, as $k\to\infty$, i.e., $\tilde{\ell}_*(x;\theta) - \tilde{\ell}_k(x;\theta)$ tends to 0 in $L_2(dx_\tau)$. Since $\tilde{\ell}_k(x;\theta)$ tends in $L_2(dx_\tau)$ to $\tilde{\ell}_\tau(x;\theta) = \ell_\tau(x;\theta) - b_0(\theta)$, by the above, the same holds for $\tilde{\ell}_*(x;\theta)$. As well, for all x in I_τ ,

$$|\tilde{\ell}_*(x;\theta) - \ell_k(x;\theta)|$$

$$= \left| \sum_{\substack{j=-k \\ j\neq 0}}^{k} [a_j(\theta) - a_{jk}(\theta)]\, \frac{\exp(ij\tau x)}{j\tau} \right|$$

$$\leq \frac{1}{\tau} \left[\sum_{j=-k}^{k} |a_j(\theta) - a_{jk}(\theta)|^2 \cdot 2 \sum_{j=1}^{k} j^{-2} \right] ,$$

by Cauchy-Schwarz. Since $\sum_{j=1}^{\infty} j^{-2}$ converges, and the first sum tends to 0, the above confirms the pointwise convergence of $\tilde{\ell}_*(x;\theta)$ to $\tilde{\ell}(x;\theta) - \ell(x;\theta) - b_0(\theta)$, as $k\to\infty$.

The term $b_0(\theta)$ may be approximated based on the result in Grenander and Szego (1958, p.45), which yields that where

$$a_j(\theta) = \frac{\tau}{2\pi}\int_{I_\tau} J_\tau(x;\theta)\ \exp(-ij\tau x)\,dx.$$

Note that $a_0(\theta) = 0$, and that the partial sums $J_k(x;\theta) = \sum_{j=-k}^{k} a_j(\theta)\exp(ij\tau x)$

will tend in $L_2(dx_\tau)$ (the restriction of $L_2(dx)$ to I_τ) to $J_\tau(x;\theta)$, i.e.

$$\int_{I_\tau} [J_k(x;\theta) - J_\tau(x;\theta)]^2\ dx \to 0,$$

as $k\to\infty$. Now $J_*(x;\theta)$ is itself of the form $J_*(x;\theta) = \sum_{j=-k}^{k} a_{jk}(\theta)\exp(ij\tau x)$, the coefficients being given implicitly by (6.2). Since $J_*(x;\theta)$ tends to $J(x;\theta)$ as well, let us consider the relationship of $a_{jk}(\theta)$ to $a_j(\theta)$ as $k\to\infty$. If $p(x;\theta) > 0$ on R^1, and is continuous in x, then $p_\tau(x;\theta)$ will be bounded away from 0 in $I_\tau = [-\pi/\tau, \pi/\tau]$, say $p_\tau(x;\theta) \geq \lambda(\theta) > 0$. (If $p(x;\theta)$ has support on $(0,\infty)$ only, defining $p_\tau(x;\theta)$ on $[\tau, \tau+\frac{2\pi}{\tau}]$ permits a similar argument). This implies that

$$\int [J_*(x;\theta) - J_\tau(x;\theta)]^2\ dP_\theta$$

$$\geq \lambda(\theta)\int_{I_\tau} [J_*(x;\theta) - J_\tau(x;\theta)]^2\ dx,$$

so that J_* tends to J_τ in $L_2(dx_\tau)$, as well. This fact can be expressed in terms of co-efficients as

$$\sum_{j=-k}^{k} |a_j(\theta) - a_{jk}(\theta)|^2 \to 0$$

as $k\to\infty$.

Now consider $\ell_\tau(x;\theta) = \log p_\tau(x;\theta)$, whose Fourier co-efficients are given by $b_j(\theta) = a_j(\theta)/ij\tau$, for $j \neq 0$, and $b_0(\theta) = \frac{\tau}{2\pi}\int_{I_\tau} \ell_\tau(x;\theta)\ dx$. Whenever $J_\tau(x;\theta)$ exists and is bounded, the Fourier expansion of $\ell_\tau(x;\theta)$, since $\text{Im}(J_*) = 0$, whenever $\underset{\sim}{t}$ is as above. Note that in the complex domain it is customary to define $\text{cov}\{u,v\} = E\{u\bar{v}\} - E\{u\}E\overline{\{v\}}$, so that $\Gamma(\theta)$, above, has entries $\Gamma_{ij} = \phi([i-j]\tau;\theta))$. $\ell_*(x;\theta)$ is obtained by integrating the trigonometric terms of $J_*(x;\theta)$, yielding the intermediate form,

$$\tilde{\ell}_*(x;\theta) = \underset{\sim}{\phi}(\theta)^* D_t \Sigma_k(\theta)^{-1} D_t^{-1} \exp(i\underset{\sim}{t}x) , \tag{6.3}$$

which approximates $\ell(x;\theta)$ up to a term in θ, which may be approximated by

$$b_*(\theta) = \frac{1}{2}\{\log|\Sigma_k(\theta)| - \log|\Sigma_{k-1}(\theta)| - \log(\tfrac{2\pi}{\tau})\} . \tag{6.4}$$

The result below pertains to

$$\ell_*(x;\theta) = \tilde{\ell}_*(x;\theta) + b_*(\theta) .$$

<u>Theorem 6.8</u>:

If $p(x;\theta) > 0$ on R^1 and is continuous and differentiable in x for all θ, then

$$\lim_{\tau\to 0}\lim_{k\to\infty} |\ell_*(x;\theta) - \ell(x;\theta)| = 0$$

for all x and θ.

<u>Proof</u>:

We begin by noting that $J_*(x;\theta)$ is a slight generalization of $J_*(z)$ of section 3.1, and converges in $L_2(dP_\theta)$ to $J(x;\theta)$ as long as $J(x;\theta)$ lies in this space. By the arguments of the previous section, $J_*(x;\theta)$ approximates $J_\tau(x;\theta)$, the form corresponding to $p_\tau(x;\theta)$. Let us consider the convergence of J_* more closely. $J_\tau(x;\theta)$ is periodic, with natural domain $I_\tau = [-\pi/\tau,\pi/\tau]$ and as such, possesses an ordinary Fourier expansion given by

$$J_\tau(x;\theta) = \sum_{j=-\infty}^{\infty} a_j(\theta) \exp(ij\tau x)$$

Proof:

$$|x^2 h(x)| = |\int_{R^1} x^2 \exp(-ixt)\eta(t)\,dt|$$

$$\leq \int_{R^1} |\eta''(t)\,dt| \quad ,$$

by the Fourier inversion formula, and integration by parts.
Thus, a strong version of the order condition, i.e., with α=2, can be verified easily, in terms of $\phi(t;\theta)$, $\mathring{\phi}(t;\theta)$, et cetera. The condition that $\phi(t;\theta)$ be twice differentiable implies the existence of the second moment of x, for all θ. Thus this result may not be applied in such interesting cases as x being stably distributed.

6.2 The Log Likelihood Function

In the continuous case it is possible to approximate $\ell(x;\theta) = \log p(x;\theta)$ via a trigonometric polynomial, $\ell_*(x;\theta)$, facilitating examination of the log likelihood surface. The approximate form arises from considering $J(x;\theta) = \frac{\partial \log p(x;\theta)}{\partial x}$, and $J_*(x;\theta) = E^*\{J(x;\theta)\,|\,1, \exp i\underset{\sim}{t}x\}$, which may be obtained explicitly since

$$E_\theta\{J(x;\theta) \exp itx\} = \int p'(x;\theta) \exp(itx)\,dx$$

$$= it\phi(t;\theta) \ ,$$

integrating by parts. $J_*(x;\theta)$ can be written as

$$J_*(x;\theta) = [i\, D_t\, \underset{\sim}{\phi}(\theta)]^* \Sigma_k(\theta)^{-1}[\exp i\underset{\sim}{t}x - \underset{\sim}{\phi}(\theta)] \ , \qquad (6.2)$$

where $\underset{\sim}{t} = (-k\tau,\ldots,-\tau,\tau,\ldots,k\tau)$, $D_t = \text{diag}(\underset{\sim}{t})$, $\underset{\sim}{\phi}(\theta) = \phi(\underset{\sim}{t};\theta)$ and $\Sigma_k(\theta) = \Gamma(\theta) - \underset{\sim}{\phi}(\theta)\underset{\sim}{\phi}(\theta)^*$. The switch to complex form is merely notational, By construction, $S_*(x,\theta)$ is the best $L_2(dP_\theta^\tau)$ approximation to $S_\tau(x;\theta)$ as a k^{th} degree trigonometric polynomial so that the above is

$$\leq \frac{1}{m_\tau}\int_{I_\tau}[q_k(x;\theta) - S_\tau(x;\theta)]^2 p_\tau(x;\theta)\,dx$$
$$\leq \frac{1}{m_\tau}\left[\frac{KA_\tau}{k\tau}\right]^2$$

Applying theorem yields

$$\sup_{x\in I_\tau}|S_\tau(x;\theta) - S_*(x;\theta)| \leq \frac{KA_\tau}{\tau}[4(\tau/m_\tau)^{\frac{1}{2}}\, k^{-\frac{1}{2}} + 5k^{-1}] \text{ ,}$$

uniformly in $\theta\epsilon\Theta_c$. Since τ is fixed in the above, it is clearly possible to choose k so that $|S_\tau(x;\theta) - S_*(x;\theta)| \leq \frac{\varepsilon}{2}$, and the proof is complete.

It must be pointed out that the conditions on $p(x;\theta)$ and its derivatives, though not stringent, are nontheless troublesome in that their verification depends on knowledge of their asymptotic behaviour. (The nonvanishing of $p(x;\theta)$ and its boundedness on compact sets will generally be easy to verify, and indeed will usually be implicit in the formulation of the model.) Ideally, these conditions should be restated in terms of $\phi(t;\theta)$, since this is what is known. Some progress in this direction can be made by noting that the Fourier transforms of $p(x;\theta)$, $\dot{p}(x;\theta)$, $p'(x;\theta)$, and $\dot{p}'(x;\theta)$ are, under regularity conditions, $\phi(t;\theta)$, $\dot{\phi}(t;\theta)$, $t\cdot\phi(t;\theta)$, and $t\cdot\dot{\phi}(t;\theta)$. The following lemma is thus relevant.

<u>Lemma 6.7</u>:

If h(x) is an integrable function with Fourier transform, $\eta(t) = \int_R h(x)\exp(i\,tx)dx$ and $\eta(t)$ is twice differentiable everywhere, then

$$h(x) \leq \left[\int_{R^1} |\eta''(t)|\,dt\right] |x|^{-2} \quad .$$

Then

$$\sup_{x \in I_\tau} |h(x) - s(x)| \leq 4[k\tau d]^{\frac{1}{2}} + 5\delta.$$

We may now proceed directly to establish theorem 6.1. We first consider $p_\tau(x;\theta)$, the wrapped version of $p(x;\theta)$, defined as in lemma 6.2, as well as its derivative, $\dot{p}_\tau(x;\theta)$, which exists by corollary 6.4. Without loss of generality, we assume $R_C = [-C,C] \times \Theta_C$. Since $p_\tau(x;\theta)$ and $\dot{p}_\tau(x;\theta)$ tend uniformly on R_C to $p(x;\theta)$ and $\dot{p}(x;\theta)$, by appeal to lemma 6.2 and its corollaries, and since $p(x;\theta)$ is bounded below, away from 0, on R_C, this implies that $S_\tau(x;\theta) = \dfrac{\partial \log p_\tau(x;\theta)}{\partial\theta} = \dfrac{\dot{p}_\tau(x;\theta)}{p_\tau(x;\theta)}$ tends uniformly to $S(x;\theta)$ on the same set. Choose $\tau < \min[\pi/B, \pi/C]$ so that

$$|S_\tau(x;\theta) - S(x;\theta)| < \varepsilon/2 \ ,$$

for $(x,\theta) \in R_C$.

Now consider $S_\tau(x;\theta)$ and its derivative,

$$S_\tau'(x;\theta) = \frac{p_\tau'(x;\theta)\ \dot{p}_\tau(x;\theta) - p_\tau(x;\theta)\ \dot{p}_\tau'(x;\theta)}{p_\tau(x;\theta)^2}$$

Again, by lemma 6.2 and its corollaries, the functions in the numerator exist and are continuous on $I_\tau \times \Theta_C$, and hence bounded on that set. By assumption $p_\tau(x;\theta) \geq p(x;\theta) \geq m_\tau > 0$ on $I_\tau \times \Theta_C$, so that $|S_\tau'(x;\theta)| < A_\tau$ on $I_\tau \times \Theta_C$ for some $A_\tau < \infty$. By appeal to theorem 6.5 then, for each k=1,2,3,..., there exist trigonometric polynomials, $q_k(x;\theta)$, satisfying

$$|q_k(x;\theta) - S_\tau(x;\theta)| < \frac{KA_\tau}{k\tau}$$

Defining $S_*(x;\theta)$ as in the statement of the theorem, we let

$$\begin{aligned} d_k(\theta) &= \int_{I_\tau} [S_*(x;\theta) - S_\tau(x;\theta)]^2\, dx \\ &\leq \frac{1}{m_\tau} \int_{I_\tau} [S_*(x;\theta) - S_\tau(x;\theta)]^2\, p_\tau(x;\theta)\, dx \quad . \end{aligned}$$

Proofs:

The results follow by uniform convergence of the series defining each of the "wrapped" forms.

The following two theorems due to Jackson (1930) relate mean square and uniform convergence for trigonometric polynomials, and will allow us to construct $S_*(x;\theta)$, using the τ indicated above, so that

$$|S_*(x;\theta) - S_\tau(x;\theta)| < \varepsilon/2.$$

Theorem 6.5 follows directly from Jackson's theorem 1, p.2, while Theorem 6.6 is a slight generalization of theorem IIa, p.84.

Theorem 6.5:

Let $h(x)$ be a differentiable function of period $2\pi/\tau$, such that $|h'(x)|<A$ for all x. Then there exists a trigonometric polynomial of degree k, $q(x) = a_0 + \sum_{j=1}^{k} (a_j \cos j\tau x + b_j \sin j\tau x)$

$$\sup_{x\in R^1} |q(x) - h(x)| \leq \frac{KA}{k\tau},$$

where K is an absolute constant.

Theorem 6.6:

Let $h(x)$ be continuous, and let $r(x)$ and $s(x)$ be k^{th} degree trigonometric polynomials in τx such that

$$\sup_{x\in I_\tau} |h(x) - r(x)| < \delta$$

and

$$\int_{I_\tau} |h(x) - s(x)|^2 dx \leq d.$$

Proof:

For $x \in I_\tau$, consider the modulus of the partial sum,

$$\left|\sum_{|j|\geq k} h(x+2\pi j/\tau)\right| \leq A \sum_{|j|\geq k} |x+2\pi j/\tau|^{-\alpha}$$

$$\leq 2A \sum_{j=k}^{\infty} \left|\frac{\pi}{\tau}[2j-1]\right|^{-\alpha}$$

$$\leq A\ \tau^{\alpha}[2\pi^{-\alpha} \sum_{j=k}^{\infty} j^{-\alpha}]$$

for $k>0$. This confirms uniform convergence of the sum defining $h_\tau(x)$, and hence, its existence and continuity. Setting $k=1$ in the above yields, for $x\in I_\tau$,

$$|h_\tau(x) - h(x)| \leq A\ \tau^{\alpha}[2\pi^{-\alpha} \sum_{j=1}^{\infty} j^{-\alpha}]\ ,$$

from which conclusion 2 follows directly. (When h is defined and continuous on R_*, similar results apply by considering $h_\tau(x)$ on $[1/\tau,(2\pi+1)/\tau]$ as $\tau\to 0$.)

Corollary 6.3:

If h' exists and satisfies the conditions of the lemma as well, then $h'_\tau(x) = \frac{\partial h_\tau(x)}{\partial x}$ exists, and equals the termwise taken derivative, i.e.,

$$h'_\tau(x) = \sum_{j=-\infty}^{\infty} h'(x+2\pi j/\tau),$$

justifying the otherwise ambiguous notation.

Corollary 6.4:

If $h(x;\theta)$ depends continuously on $\theta\in\Theta$, and satisfies condition (6.1) uniformly in θ, $h_\tau(x;\theta)$ exists and is a continuous function on $I_\tau \times \Theta$. If $\dot{h}(x;\theta) = \frac{\partial h(x;\theta)}{\partial\theta}$ exists and satisfies a similar condition, $\dot{h}_\tau(x;\theta)$ exists and is well defined.

Let $R_c \subseteq R^1 x \Theta$ be compact. Then, for all $\varepsilon > 0$ there exist k and τ such that for $\underset{\sim}{t} = <\tau, 2\tau, \ldots, k\tau>$,

$$S_*(x;\theta) = E^*\{S(x;\theta) | 1, \sin \underset{\sim}{t}x, \cos \underset{\sim}{t}x\}$$

satisfies $|S_*(x;\theta) - S(x;\theta)| < \varepsilon$ for all $(x,\theta) \in R_c$.

Proof:

The approach to choosing τ and k described in Section 4 is reflected in the outline of the proof below. We consider first the "wrapped" version of $S(x;\theta)$, $S_\tau(x;\theta)$, choosing τ small enough that $|S_\tau(x;\theta) - S(x;\theta)| < \varepsilon/2, (x,\theta) \in R_c$. To do this we require the following lemma, 6.2, and its corollaries, 6.3 and 6.4

Lemma 6.2:

Let $h(x)$ be a continuous, real valued function on R^1, and suppose that there exist $A, B < \infty$ and $\alpha > 1$ such that

$$|h(x)| \leq A\, |x|^{-\alpha} \text{ for } |x| > B. \tag{6.1}$$

Define the $2\pi/\tau$ periodic function

$$h_\tau(x) = \sum_{j=-\infty}^{\infty} h(x + \frac{2\pi j}{\tau}).$$

Then, for $\tau < \pi/B$,

1. $h_\tau(x)$ is a well defined, continuous function on $I_\tau = [-\pi/\tau, \pi/\tau]$.
2. $h_\tau(x)$ converges almost uniformly to $h(x)$ as $\tau \to 0$, in the sense that $\lim_{\tau \to 0} [\sup_{x \in I_\tau} |h_\tau(x) - h(x)|] \to 0$.

6. Fourier Likelihood Methods

We have seen that when distributions of a parametric family are specified by the characteristic function, $\phi(t;\theta) = E_\theta\{\cos tx\} + iE_\theta\{\sin tx\}$, the quantities, $S(x;\theta)$ and $\mathcal{I}(\theta)$, may be approximated. In this section we see that $\ell(x;\theta) = \log p(x;\theta)$ may be similarly approximated, and that furthermore, the sense of approximation of the above forms may be strengthened, given regularity conditions. Thus likelihood methods may conveniently be implemented under these assumptions. We consider first the approximation of $S(x;\theta)$.

6.1 The Score Function

The arguments regarding the convergence of the form $S_*(x;\theta)$, and related forms, have been stated in terms of mean square convergence, which in general does not imply pointwise convergence, let alone the uniform result. For these forms to provide a reliable basis for likelihood calculations, it is clearly desirable to show some kind of uniform convergence on $R^1 \times \Theta$. We demonstrate such behaviour in the case of $S_*(x;\theta)$(under regularity conditions); similar arguments apply to the other forms considered.

Theorem 6.1.

Let P_θ, $\theta\in\Theta$, be a set of distributions on R^1 with associated densities, $p(x;\theta)$, satisfying:

1. $p(x;\theta)$ is bounded above and away from 0, on any compact subset of $R^1 \times \Theta$.
2. $p(x;\theta)$, $\dot{p}(x;\theta)$, and $\dot{p}'(x;\theta) = \frac{\partial^2 p(x;\theta)}{\partial x \partial\theta}$ exist, are continuous in x and θ, and for compact subsets of Θ satisfy conditions of the form

$$h(x;\theta) \leq A|x|^{-\alpha} \quad \text{for } |x| > B$$

uniformly in θ, with $\alpha > 1$ and $A,B < \infty$.

5.2 Binomial-Poisson convolution

If $x = y+z$ where y is binomial, with parameters N and θ, and z is Poisson, with mean λ, independent of y, then the probability distribution function of x is $p(x;\theta,\lambda) = e^{-\lambda}\sum_y \binom{N}{y}\theta^y(1-\theta)^{N-y}\frac{\lambda^{x-y}}{(x-y)!}$, whereas its c.f. has the simpler form

$$\phi(x;\theta,\lambda) = [(1-\theta)+\theta\exp(it)]^N\exp\{\lambda[\exp(it)-1]\} .$$

The problem of estimation for this family has been considered by Sclove and Van Ryzin (1969), who note that maximum likelihood is "computationally intractable". Sprott (1983) points out that this is not entirely true, owing to the relationships,

$$\frac{\partial p(x;\theta,\lambda)}{\partial\theta} = \frac{[(x-N\theta)p(x;\theta,\lambda) - \lambda\, p(x-1;\theta,\lambda)]}{\theta(1-\theta)}$$

and

$$\frac{\partial p(x;\theta,\lambda)}{\partial\theta} = -p(x;\theta,\lambda) + p(x-1;\theta,\lambda) .$$

Using these facts,

$$S_\theta(x;\theta,\lambda) = \frac{\partial \log p(x;\theta,\lambda)}{\partial\theta} \quad \text{and} \quad S_\lambda(x;\theta,\lambda) = \frac{\partial \log p(x;\theta,\lambda)}{\partial\lambda}$$

can be expressed in terms of x and $u(x;\theta,\lambda) = \frac{p(x-1;\theta,\lambda)}{p(x;\theta,\lambda)}$.

Sprott exploits these results to facilitate likelihood calculations. The calculation of $I_\theta(\theta,\lambda) = \text{var}\{S_\theta\}$, $I_\lambda(\theta,\lambda) = \text{var}\{S_\lambda\}$ and $I_{\theta\lambda}(\theta,\lambda) = \text{cov}\{S_\theta,S_\lambda\}$ can be simplified similarly by using $\text{var}(x) = N\theta(1-\theta)+\lambda$, $\text{cov}(x,u) = 1$, and $\text{var}(u) = \{\sum_{x=1}^{\infty} \frac{p(x-1;\theta,\lambda)^2}{p(x;\theta,\lambda)}\} - 1$. Alternately, the straightforward multi-parameter extensions of I_* can be constructed to yield approximations. Some numerical results are given in Table 2, for various choices of N, θ, and λ, with $\tau=2\pi/(N+\lambda+5\lambda^{\frac{1}{2}})$, based on a conservative bound on the domain of interest for x.

[Table 2 here]

strategy entertained will be the only one feasible.

With regard to other kernel sets, such as G_E, yielding the m.g.f., less is known; numerical issues similar to those affecting G_M indicate that G_E is not a practical choice. Overall, the author's experience has been that G_H leads to the most numerically reliable forms, favouring the choice of trigonometric moments over the more familiar quantities as a basis for approximate calculations.

5. Examples

5.1 Stable Laws

To illustrate the potential utility of the forms considered so far we consider their application to aspects of inference regarding the family of symmetric stable laws, specified by the c.f.'s, $\phi(t;\alpha) = \exp\{-\frac{|t|^\alpha}{2}\}$, $1 \leq \alpha \leq 2$. Only two members of this family have closed form density representations, the Cauchy (α=1) and the Gaussian (α=2). These cases provide convenient reference points, since the exact forms can be computed explicitly and compared with the approximations. This is done for J_* and K_* with k chosen to illustrate the rate of convergence. In accordance with the approach in section 4, the domain of approximation in the Cauchy case (note nonstandard scale) is taken to be $-3\pi \leq x \leq 3\pi$, leading to the use of $\tau = 1/3$; in the Gaussian case we consider $-\pi < x < \pi$, with τ=1.

[Figure 1 here]

In addition, the forms, J_*, are illustrated for intermediate members of the stable family in Figure 3, based on the conservative choice of k=15 and τ=.3. The relevant efficiencies, I^*/I, were calculated, using the essentially exact results of Dumouchel (1975), and are given in Table 1.

[Figure 2 here]

[Table 1 here]

to replacing P by a "wrapped" version, P^{τ}, the distribution induced by identifying points on R^1 on $2\pi/\tau$ spaced grids. The corresponding density $p_\tau(x)$ can be written as $p_\tau(x) = \sum_{j=-\infty}^{\infty} p(x+2\pi j/\tau)$, and $\phi(j\tau)$ can be seen to yield the j^{th} Fourier co-efficient of $p_\tau(x)$, $a_j = [2\pi/\tau]^{-1} \int_{-\pi/\tau}^{\pi/\tau} \exp(ij\tau x)p_\tau(x)dx$. If τ is held fixed and $k\to\infty$, the forms, h_* considered thus far, will converge to wrapped versions h_τ. For instance $S_*(x;\theta)$ will tend to $S_\tau(x;\theta) = \frac{\partial \log p_\tau(x;\theta)}{\partial\theta}$, which can be viewed as the projection of $S(x;\theta)$ on $\{\exp(ij\tau x),\ j=-\infty,\dots,\infty\}$. Thus, in general, if τ can be chosen so that P^τ represents a reasonable approximation to P, so that h^τ approximates h, it is practical to fix τ, and then choose k to obtain a reasonable degree of convergence. In the completely general parametric case, the range of interest of $S(x;\theta)$ may depend on θ, as in the case of a scale parameter, it is usually desirable to re-scale the random variable x in accordance with θ, as is accomplished through the use of K_* in the scale case. The examples indicate that moderate values of k often suffice; it is also useful to note, however, that due to the boundedness of the trigonometric functions, ill-conditioning problems associated with large k values are less prevalent than in the case of the moments.

It must be noted that the general strategy outline for choosing $\underset{\sim}{g}(x)$ can in no way be considered "optimal". In individual instances, it is not, in fact, the completeness of G that is of the essence, but how the function of interest, h(x), may be best approximated by elements of G. For example, in the Gaussian location problem, arbitrarily good approximations to the score, $S(x;\theta) = x - \theta$, are obtained by considering the pair, $g(x) = (\cos\tau x, \sin\tau x)^T$, as $\tau\to 0$, since

$$x - \theta \simeq \frac{\sin(\tau[x-\theta])}{\tau} = \tau^{-1}[\cos\theta \sin x - \sin\theta \cos x] ,$$

for τ small. In the case where the techniques considered are of interest, however, the form of h(x) will be largely unaccessible, so that the general

4. Choosing the Kernel Vector, $\underset{\sim}{g}$.

The convergence of the approximations considered thus far depend on the completeness properties of the kernel set G in the space, $L_2(dP)$, and on the particular choice of $\underset{\sim}{g}$ from G. We therefore consider G_M, G_H and G_E individually.

In the case of $G_M = \{x^t, t=1,2,\ldots\}$, uniqueness of the moments is sufficient to ensure completeness. The choice $\underset{\sim}{g}(x) = \langle x, x^2, \ldots, x^k\rangle^T$ is natural, though to achieve an adequate degree of convergence a fairly large k may be required. The instability of the higher order moments whose use is thus implied is balanced automatically by the choice of coefficients in $S_*(x,\theta)$, which correspond to optimal weights. The numerical problems associated with solving $\Sigma\, \underset{\sim}{d} = \underset{\sim}{\lambda}$ for d (explicit inversion of Σ being unnecessary and numerically undesirable), are a greater issue, since Σ may be ill conditioned for k large. In examples considered by the author, $k \leq 10$ is usually sufficient, though an intelligent choice of scale is still necessary to avoid numerical inaccuracy.

The completeness of $G_H = \{\cos tx, \sin tx, t \in R\}$ holds under completely arbitrary assumptions. Any element of $L_2(dP)$ may be approximated by a continuous function with compact support, which in turn can be approximated by a continuous, periodic function, which may be uniformly approximated by trigonometric polynomials. A convenient choice for $\underset{\sim}{g}(x)$ is to take $\underset{\sim}{g}(x) = (\cos \tau x, \ldots, \cos k\tau x, \sin \tau x, \ldots, \sin k\tau x)^T$, in which case approximations reduce to trigonometric polynomials. Since $2\pi/\tau$-periodic functions can be so approximated, if the domain of interest of the "target" function lies in an interval $M \pm L$, the choice $\tau = 2\pi/L$ is indicated. The approximations considered will then depend on the characteristic function, $\phi(t)=E\{\cos tx\} + iE\{\sin tx\}$, at the points, $t=j\tau$, $j=1,\ldots,k$. This approach is equivalent explicit (D, $\underset{\sim}{\gamma}$, and Σ being fixed), and since it also provides an approximation to $J(x;\theta) = \frac{\partial p(x;\theta)}{\partial x}$, a function that is useful in the more general context. Of course $I_* = \underset{\sim}{\gamma}^T D^T \Sigma^{-1} D \underset{\sim}{\gamma}$ yields an approximation to the (constant) Fisher information.

When observations are of the form $x=\theta z$, z specified as before, the relevant score function has the form $S(x;\theta) = \theta^{-1}K(x/\theta)$, where $K(z) = z \cdot J(z) - 1$. Again $K_*(z) = E^*\{K(z)|1, \underset{\sim}{g}(z)\}$ is generally available. Note that

$$K(z) = -\frac{zp'(z)+p(z)}{p(z)} \quad ,$$

so that

$$\text{cov}\{g(z),K(z)\} = -\int [zp'(z)+p(z)]g(z)dz$$

$$= -zp(z)g(z)\Big|_{-\infty}^{\infty} + \int zg'(z)p(z)dz \ .$$

Again the first term disappears in general, and the reader may check that the latter integral may be explicitly given in terms of γ in all the interesting cases. For example if $g(z) = \exp tz$,

$$\int zg'(z)p(z)dz = t\int z \exp tz \, p(z)dz$$

$$= t\int \frac{\partial[\exp tz \, p(z)]}{\partial t} dz$$

$$= t \cdot \gamma'(t) \ ,$$

assuming regularity. Thus $\lambda = \text{cov}\{\underset{\sim}{g}(z), K(z)\}$ is generally obtainable, so that $\theta^{-1}K_*(x/\theta) = \theta^{-1}\underset{\sim}{\lambda}^T\Sigma^{-1}[\underset{\sim}{g}(x/\theta) - \underset{\sim}{\gamma}]$ provides a suitably invariant approximation to $S(x;\theta)$, while the quantity $I_*(\theta) = \theta^{-2}\underset{\sim}{\lambda}^T\Sigma^{-1}\underset{\sim}{\lambda}$ provides a corresponding approximation to the information. In general, the algebraic form of K_* will closely follow that of S_*; K_* represents the projection of $S(x;\theta)$ on $\underset{\sim}{g}(x/\theta)$, while S_*, the projection on $\underset{\sim}{g}(x)$.

of $I_*(\theta)$ and $I(\theta)$, can be made arbitrarily close to the Cramer-Rao lower bound.

It must be noted that the sense of approximation considered here is context dependent, since the form of $S_*(x;\theta)$ depends on the underlying P_θ. However, depending on G, stronger forms of convergence, e.g. almost uniform, are demonstrable, as seen in section 6.

3.1 Location and Scale. When $x \sim \theta+z$, z possessing a continuous distribution, P, with density with respect to Lebesgue measure, p(z), the relevant score function has the form $S(x;\theta) = J(x-\theta)$, where $J(z) = -\frac{\partial \log p(z)}{\partial z} = -\frac{p'(z)}{p(z)}$. If the distribution of z is specified via $\gamma(t) = E\{g(t,z)\}$, then it is appropriate to consider $J_*(z) = E^*\{J(z)|1,\underset{\sim}{g}(z)\}$. In general, $E\{J(z)\} = 0$, and

$$\begin{aligned} \text{cov}\{J(z),\ g(z)\} &= -\int g(z)p'(z)dz \\ &= -p(z)g(z)\Big|_{-\infty}^{\infty} + \int p(z)g'(z)dz, \end{aligned} \tag{3.2}$$

assuming differentiability of the kernel function g(z). The important kernel classes consist of differentiable functions, and furthermore, are closed (or, rather, span{G} is closed) under this operation. As well, $\underset{\sim}{g}(z)$ will typically be chosen so that span$\{\underset{\sim}{g}(z)\}$ possesses the same closure. Since the first term in (3.2) will disappear in most cases, it follows that in general, one may express $\underset{\sim}{\lambda} = \text{cov}\{\underset{\sim}{g}(z),\ J(z)\}$ as $D\underset{\sim}{\gamma}$, where D is a matrix of constants and $\underset{\sim}{\gamma} = E\{\underset{\sim}{g}(z)\}$. Thus it is possible to form $J_*(z)$, and take as an approximation to $S(x;\theta)$, $J_*(x-\theta) = [D\underset{\sim}{\gamma}]^T \Sigma^{-1}[\underset{\sim}{g}(x-\theta) - \underset{\sim}{\gamma}]$, where $\Sigma = \text{var}\{\underset{\sim}{g}(z)\}$. This form will in fact coincide with that derived from the more general approach in most cases, since it is generally the case that $\underset{\sim}{g}(x-\theta) = M(\theta)\underset{\sim}{g}(x)$, $M(\theta)$ a kxk matrix, i.e. span$\{\underset{\sim}{g}(z)\}$ is invariant under translations. This "pivotal" form is useful in that the translation invariance of subsequently derived procedures is made for any function, $g(x)$. Thus, $\underset{\sim}{\lambda} = \text{cov}_\theta\{S(x;\theta), \underset{\sim}{g}(x)\} = \dot{\underset{\sim}{\gamma}}(\theta)$, where $\underset{\sim}{\gamma}(\theta) = E_\theta\{\underset{\sim}{g}(x)\}$ and $\dot{}$ denotes differentiation (componentwise) with respect to θ. Since $E_\theta\{S(x;\theta)\} = 0$, we have that for any selection of $\underset{\sim}{g}(x)$ from G,

$$S_*(x;\theta) = E^*\{S(x;\theta)|1, \underset{\sim}{g}(x)\} = \dot{\underset{\sim}{\gamma}}(\theta)^T \Sigma(\theta)^{-1}[\underset{\sim}{g}(x) - \underset{\sim}{\gamma}(\theta)] \; .$$

Since the expression is linear in $\underset{\sim}{g}(x)$, the resulting approximation to the maximum likelihood equations based on a sample of size n, can be written as

$$\dot{\underset{\sim}{\gamma}}(\theta)\Sigma(\theta)^{-1}[\hat{\underset{\sim}{\gamma}} - \underset{\sim}{\gamma}(\theta)] = 0, \tag{3.1}$$

where $\hat{\gamma}$ is a vector of observations from $\hat{\gamma}(t) = n^{-1}\sum_{i=1}^{n} g(t,x_i)$, the empirical transform. (The above estimating equation is closely related to that considered by Feuerverger and McDunnough (1981a,b and 1982), whose investigations have motivated the author's development of the forms considered here.) Later examples show that the dimension of $\underset{\sim}{g}$ need not be inordinately big for S_* to provide a useful approximation to S, so that the above may provide convenient vehicle for likelihood calculations.

Observe, as well, that by the orthogonalities implied by the projection operation,

$$\text{var}_\theta\{S(x;\theta)\} = \text{var}_\theta\{S_*(x;\theta)\} + \text{var}_\theta\{S(x;\theta) - S_*(x;\theta)\}.$$

Since the latter term can be made arbitrarily small (given the conditions mentioned previously), $I_*(\theta) = \dot{\underset{\sim}{\gamma}}(\theta)^T\Sigma^{-1}\dot{\underset{\sim}{\gamma}}(\theta) = \text{var}_\theta\{S_*(x;\theta)\}$ can be regarded as a simply derived approximation to the Fisher information, $I(\theta) = \text{var}_\theta\{S(x;\theta)\}$. This method of approximating the Fisher information is essentially that proposed in Jarrett (1973), and applied in Brockwell and Brown (1982). In addition, standard arguments yield that the asymptotic variance of a consistent solution to (3.1) is $[nI_*(\theta)]^{-1}$, which, owing to the relationship may be written as $\underset{\sim}{d}^T \underset{\sim}{g}(x)$, where $\underset{\sim}{d} = (d_1, \ldots, d_k)^T$. h_* may be viewed as the projection (regression) of h on span$\{\underset{\sim}{g}(x)\}$ in the Hilbert space, $L_2(dP)$, induced by the measure P. The coefficients of h_* are easily seen to be given by $\underset{\sim}{d}_* = \Gamma^{-1}\underset{\sim}{\eta}$, where $\Gamma = E\{\underset{\sim}{g}(x)\cdot\underset{\sim}{g}(x)^T\}$ and $\underset{\sim}{\eta} = E\{h(x)\cdot\underset{\sim}{g}(x)\}$. We shall adopt the notation, $h_*(x) = E^*\{h(x)|\underset{\sim}{g}(x)\}$, motivated by the formal similarity to conditional expectation.

It is customary to include a constant term in the approximation, and this is most conveniently accomplished by centering the variables, leading to the form

$$h_*(x) = E^*\{h(x)|1, \underset{\sim}{g}(x)\} = E\{h(x)\} + \underset{\sim}{\lambda}^T \Sigma^{-1}[\underset{\sim}{g}(x) - \underset{\sim}{\gamma}],$$

where $\underset{\sim}{\lambda} = \text{cov}\{\underset{\sim}{g}(x), h(x)\}$, $\Sigma = \text{var}\{\underset{\sim}{g}(x)\}$, and $\underset{\sim}{\gamma} = E\{\underset{\sim}{g}(x)\} = \gamma(\underset{\sim}{t})$. By the foregoing remarks, $\underset{\sim}{\gamma}$, Γ, and Σ are given explicitly by the transform $\gamma(t)$. We shall encounter a number of instances where the remaining coefficients in h_* are easily determined, facilitating the approximation of h(x) and related functions (derivatives, anti-derivatives). Note that the approximation, $h_*(x)$, can be made arbitrarily close to h(x), in the sense considered, in the case that h(x) is contained in the L_2 closure of span {G}. In many cases, this set is the whole of $L_2(dP)$, i.e. G is complete, so that any $h(x) \in L_2(dP)$ may be so approximated.

3. The Score Function and Fisher's Information

As previously mentioned, one obvious approach to the approximation of the score function, $S(x,\theta)$, or the information, $I(\theta)$, is to proceed by numerically approximating $p(x;\theta)$, based on $\gamma(t;\theta)$, the corresponding transform. However, the following approach is more direct.

Observe that, under mild regularity conditions,

$$E_\theta\{S(x;\theta)\cdot g(x)\} = \frac{\partial E_\theta\{g(x)\}}{\partial\theta}$$

here are computationally and statistically efficient. A unified approach to the problem is possible based on certain general properties of transforms, outlined below.

2. Probability Transforms

A probability transform, $\gamma(t)$, corresponding to a real valued random variable, x, is defined through an underlying class of real valued functions, $G = \{g(t,x),\ t \in T\}$, the kernel class, by $\gamma(t) = E\{g(t,x)\}$. The important transforms include the characteristic function (c.f.), the moment generating function (m.g.f.), and the sequence of moments, with kernel classes $G_H = \{\sin tx, \cos tx, t \in R^1\}$, $G_E = \{\exp tx, t \in R^1\}$, and $G_M = \{x^t, t=0,1,2,..\}$, respectively, though others (e.g. the fractional moments) exist. Note that some transforms do not always completely characterize the underlying distribution, P, though we shall generally restrict ourselves to situations when this is the case. The common property of these kernel classes that facilitates the subsequent investigation, is their closure under multiplication. This implies that for two elements of G, $g(t_1,x)$ and $g(t_2,x)$, $E\{g(t_1,x)\cdot g(t_2,x)\} = E\{g(s,x)\} = \gamma(s)$, for some $s = s(t_1,t_2)$. Thus γ yields the joint moment structure of G, in an explicit manner and the following approach to the construction of approximations is often available.

Suppose h(x) is a function of interest, which is not conveniently obtainable in closed form. The best linear approximation (predictor) of h(x) in terms of the set of kernel functions indexed by $\underset{\sim}{t} = (t_1,\ldots,t_k)^T$, is defined to be the form $h_*(x) = \sum_{i=1}^{k} d_i g(t_i,x)$ minimizing the mean square error $\int_{R^1} [h_*(x) - h(x)]^2 dP$. For convenience, we shall define the vector-valued function, $\underset{\sim}{g}(x)$, given by $\underset{\sim}{g}(x) = g(\underset{\sim}{t},x) = (g(t_1,x),\ldots, g(t_k,x))^T$ (adopting the notational convention that a function of a scalar argument, when applied to a vector, yields the corresponding function of a vector argument) so that h_*

Approximate Likelihood and Probability Calculations Based on Transforms

1. Introduction

In applied settings, probability distributions are almost always specified by density or distribution functions. Occasionally, however, distributions arise which are only conveniently represented by transforms, such as the characteristic function, or moment generating fuction. The absence of closed form expressions for the relevant densities complicates the implementation of standard methods of inference in even the simple case considered here, where we assume replicated observations on x, a real valued random variable with distribution P_θ, specified up to some unknown, real valued parameter, θ. Likelihood methods, for example, require the evaluation of the density ($p(x;\theta)$) based forms, $\ell(x;\theta) = \log p(x;\theta)$, $S(x;\theta) = \frac{\partial \log p(x;\theta)}{\partial\theta}$, and $I(\theta) = \mathrm{var}\{S(x;\theta)\}$. In this paper we shall explore certain transform based approximations to these forms which facilitate likelihood estimation, and the associated asymptotic methods, when $p(x;\theta)$ is unavailable or intractable. As well, we shall consider the use of the exponential transformation of the approximation to $\ell(x;\theta)$ as an alternative to more familiar density approximations, such as the Edgeworth and Gram-Charlier expansions.

Approaches taken in previous investigations have relied on either numerical inversion of the transform representations (Dumouchel, 1975; Feuerverger and McDunnough, 1981c) or on non-standard estimation schemes based on the empirical version of the relevant transform (Brockwell and Brown, 1982; Bryant and Paulson, 1979; Chambers and Heathcote, 1981). The first approach leads to considerable computational complexity, while the second produces estimates with sub-optimal properties. By contrast the estimation methods discussed

SUMMARY

Probability distributions, especially in applications, are generally specified via density functions; alternative representations, including the characteristic function, moment generating function, and sequence of moments, are most commonly encountered in theoretical settings. These alternative means of specification do, however, give rise to the construction of certain approximations that can facilitate the implementation of likelihood methods or the calculation of probabilities, even when the density is not available in closed form. The feasibility of a unified treatment of this topic stems from a number of properties shared by probability transforms in general.

Approximate Likelihood and Probability Calculations Based on Transforms

(Transform-based Statistical Calculations)

by

Rollin Brant
University of Minnesota

Technical Report No. 423
August 1983

AMS 1980 subject classification
62F99, 65U05

Key words and phrases: Least squares approximations, transform-based methods, likelihood, Fisher information, density expansions.