HARNESSING EMPIRICAL CHARACTERISTIC FUNCTION CONVERGENCE BEHAVIOUR

© **Lino Sant**

University of Malta
Department of Statistics and OR, Faculty of Science
e-mail: *lino.sant@um.edu.mt*

Abstract. *Parameter estimation for Lévy processes has generated much research effort lately with a strong injection of interest coming from finance. Within this context the problem can be framed as estimation using increments from an infinitely divisible distribution, for which empirical characteristic functions (ecf) are convenient tools. However convergence of ecf's to Gaussian processes has not been exploited as fully as it might have been. In this paper we go back to strong convergence results derived from the Hungarian construction and use Brownian bridge approximations to construct new estimators. In particular we study one integrated square error estimator tailored to show deference to the variance structure of the corresponding Gaussian process. We prove some of its nice statistical properties and present simulation results obtained through its use.*

1. Introduction

The flexibility offered by Lévy processes for use in modeling has been acknowledged in various fields within the natural sciences, notably physics and chemistry, and in the applied science, with special mention in meteorology and geology. In more recent years applications in finance and insurance have given a big boost in the study and use of Levy processes. The possibility of including distributions with heavy tails as well as paths with jumps were two features which made these processes so attractive. Parameter estimation for Lévy processes progressed a lot with a large number of estimation techniques being proposed and developed over a number of papers. In this paper we are specifically interested in methods using the characteristic function. The Levy-Khinchine representation motivates the interest these methods have aroused. In particular the class of infinitely divisible distributions assume an important role seeing that the independent increments of Lévy processes belong this class. However, lately the interest runs deeper than that as researchers are trying to reconstruct Lévy measures through spectral methods applied to characteristic functions as in Belomestny (2010)[1].

Parzen's (1965)[16] idea of using the the empirical characteristic function for estimation was first used for stable distributions by Press (1972)[18]. Notable contributions to the area are those provided by Paulson, A. S., Holcomb and E. W., Leitch, (1975)[17], Heathcote (1977)[10], Koutrouvelis (1980)[14], Kogon and Williams (1998)[12], Feuerverger and McDunnough (1981a, 1981b)[8, 9].

2. Parameter Estimation of the Characteristic Function

2.1. **Uses of the Empirical Characteristic Function.** The search for good estimators of parameters within the Lévy context has been heavily influenced by earlier research on stable distributions. A characteristic function is defined by $\varphi(t) = \int e^{itx} dF(x) = \varphi^R(t) + i\varphi^I(t)$ and is associated uniquely with some distribution F. The class of characteristic functions for stable distributions happens to be parametrized by $\boldsymbol{\theta}$ a 4-dimensional vector as in $\varphi(t, \boldsymbol{\theta})$. In cases where an explicit formula for the distribution function is not known, characteristic functions are most useful. However the advantages of characteristic function methods in statistics, like robustness and smoothness of the functions involved, have been shown to be considerable in Paulson et al (1975)[17], Yu (2004)[19]. Their use has been quite extensive in model-based hypothesis testing and goodness-of-fit statistics.

In general readings from a Lévy process will give us increments which form a sequence of iid random variables $X_1, ..., X_n$ from an infinitely divisible distribution function F . The empirical characteristic function (ecf) is defined by: $c_t^n = \frac{1}{n} \sum\limits_{j=1}^{n} e^{itX_j}$.

Glivenko-Cantelli assures us that we have strong convergence of this sum of random variables to the characteristic function uniformly in t. Following the development of empirical process theory, a stochastic process Y_t^n can be constructed out of the iid sample: $Y_t^n = \sqrt{n} \left(\frac{1}{n} \sum\limits_{j=1}^{n} e^{itX_j} - \varphi(t) \right)$ which is called the normalized empirical characteristic function. The behaviour of this process was studied extensively from mid-1970's starting with Kent (1975)[11] onwards. The major result was that it converges weakly to a complex Gaussian process under certain conditions. These conditions were refined and related to a number of properties of the limit complex process which we denote by $Z_t = U_t + iV_t$, with U and V being both real processes. Z_t has mean 0 and covariance function given by: $K(s,t) = \varphi(t-s) - \varphi(t)\varphi(-s)$.

One important property, which leading researchers were insisting on, was continuity of sample paths for Z_t , or rather the existence of a version of the limit process which does have continuous paths. This condition guarantees that convergence occurs with reference to the measure generated by the paths of the stochastic process viewed as random elements in the space of continuous functions on some compact subset of $\mathbb{R}$, say $\mathscr{C}([-1,1])$. The insistence that the limit measure has support on this Banach space had deep theoretical implications as discussed in Marcus (1981)[15]. However it is well known that there are Gaussian processes whose sample paths are not continuous in the sense above.

2.2. **Strong Approximations.** In practice one might well be happy working with an empirical characteristic function whose limiting Gaussian process might have paths in the space of right-continuous functions $\mathscr{D}(\mathbb{R})$. Path continuity might not be needed in some applications. There are a lot of interesting properties still around. This can be appreciated by the fact that by construction, Y_t^n has $\varphi(t-s)-\varphi(t)\varphi(-s)$ as its covariance function. One particularly fruitful way of studying the asymptotic behaviour of Y_t^n is provided by recourse to the Hungarian construction of the Brownian bridge and Kiefer process sequence approximations as first set up in Komlos, Major, Tusnady (1975) [13]. This technique was perfected, generalized and applied to many situations to obtain more manageable results by Csörgö (1981)[4].

The starting point is the empirical process $\sqrt{n}(F_n(t) - F(t))$ which can be approximated strongly by a sequence of Brownian bridges B_t^n (to which we limit ourselves) at the following rates:

$$\mathbb{P}\Big[\omega : \sup_{0\le t\le T} |\sqrt{n}(F_n(t) - F(t)) - B^n_{F(t)}| = O\Big(\frac{\log n}{\sqrt{n}}\Big)\Big] = 1 \tag{1}$$

where we assume the sufficient condition given in Csörgö (1981)[5], namely:

Condition 4. *For some $\alpha > 0$, $x^\alpha F(-x) + x^\alpha(1 - F(x)) = O(1)$ when $x \to \infty$*

holds. These Brownian bridges live on the same probability space and thus can be used to approximate the empirical process on a set of probability 0.

Under this same condition, following Csörgö, we have a similar result for empirical characteristic functions. For an underlying probability space which is large enough to allow suitable constructions of the various processes involved, there exists a sequence of Brownian bridges B_t^n defined on the same probability space for which we define the corresponding Fourier transform, written as a stochastic integral: $Z_t^n = \int_{-\infty}^{\infty} e^{itx} dB^n_{F(x)}$ such that:

$$\mathbb{P}\big[\omega : \sup_{T_1\le t\le T_2} |Y_t^n - Z_t^n| = O\big(\frac{(\log n)^{(\alpha+1)/\alpha+2}}{n^{\alpha/(2\alpha+4)}}\big)\big] = 1 \text{ where } -\infty < T_1 < T_2 < \infty. \tag{2}$$

2.3. **The Gaussian Limit Process.** It is not hard to see that the Csörgö perspective gives us another expression for the limit process Z_t, which is of course the same process introduced earlier on:

$$Z_t = \int_{-\infty}^{\infty} e^{itx} dB_{F(x)} = \int_0^1 e^{itF^{-1}(y)} dB_y = U_t + iV_t \tag{3}$$

Having an explicit form of the limit process, we can do a lot of computations with it for estimation purposes. We can experiment through simulation to get a good picture of the more probable paths of the process. The plots in Figure 1 shown give an idea of how the paths more likely to be generated by the normalized ecf look like for a process whose increments gave gamma distributed random variables.

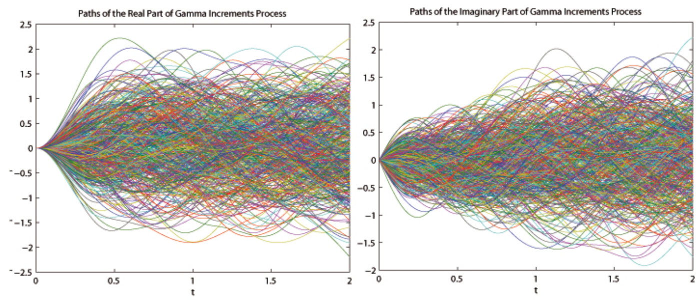

Fig. 1. Paths of a Gaussian Limit Process

We can treat the characteristic empirical function, $\mathbb{P}$ almost surely and hence distributionally , as $O\big(\frac{(\log n)^{(\alpha+1)/\alpha+2}}{n^{\alpha/(2\alpha+4)}}\big)$ close to the stochastic integral with respect to a Brownian bridge. If the distribution function F or its inverse is not known, computation-wise we are still not defeated. We could approximate F^{-1} by the empirical quantile process obtained from F_n^{-1} whose approximation by Brownian bridges runs parallel to the one above and has been extensively studied by another Csörgö, Miklos (1983)[6]

If we only know the functional form of the characteristic function, as in the case of stable distributions, then we could apply the inverse Fourier transform on the characteristic function.

2.4. **Estimation using the Characteristic Function.** There are quite a few estimation techniques that have been developed to obtain estimates of parameters of the characteristic function proper using the ecf. We mention briefly two important ones and concentrate more on the technique which is closest in spirit to the ones we are proposing here.

The natural idea for using the ecf in estimation is to define some distance d between the empirical characteristic function c_t^n and any characteristic function $\varphi(\boldsymbol{\theta})$, call it $d(c_t^n, \varphi(\boldsymbol{\theta}))$ or some suitably defined functional of the difference between the two functions, and measure this distance cumulatively over some subset O of the set over which t varies. For instance if O is finite, $O = \{t_1, t_2, ...t_K\}$ we could use $G(\theta) = \sum_{k=1}^{K} d(c_{t_k}^n, \varphi(t_k, \boldsymbol{\theta}))$ as the discrepancy measure between the ecf and a particular characteristic function over O. Then we compute the values of the parameter vector θ_0 which minimizes this discrepancy and declare the corresponding vector to be our estimate: $\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{argmin} G(\boldsymbol{\theta})$.

This technique was developed quite a bit in Feurverger and Mureika (1977) [7] and was even analyzed through the GMM perspective by various authors Yu (2004) [19]. Within the GMM paradigm discrepancies are expressed vectorially as $\boldsymbol{d}\mathrm{B}\boldsymbol{d'}$for some suitably selected weighting matrix **B**. Carrasco et al (2007)[2] take this idea further by proposing to take as **B** a suitably selected operator working on a Hilbert space and also to go from summation over time instants to integration over time into what they call a continuum of moment conditions CGMM. The intimate dependence of these methods on characteristic functions can best be appreciated in Carrasco and Kotchoni (2010) [3] but it is the integrating moment conditions over t which is the more interesting to us. However, we shall not take this point further.

Of direct interest to us is the use of the related integrated square error function for parameter estimation technique successfully underpinned by a theory Heathcote (1977)[10]. A distance function between the ecf and a characteristic function is defined as the weighted integral of the square of the modulus of the difference. Its minimum gives the estimator: $\hat{\boldsymbol{\theta}} = \underset{\theta}{argmin} \int_{-\infty}^{\infty} |c_t^n(\boldsymbol{\theta}) - \varphi(\boldsymbol{\theta})|^2 dG(t)$. The development of the theory parallels that of the maximum likelihood method. But it has well-known problems of poor efficiency in comparison with this same method Yu (2004) [19] . Usually the weighting function is blamed on the choice of the weighting function. And this is where we strike. The choices of weighting functions were made to be dependent only on t with absolutely no consideration of the characteristic functions itself.

2.5. **A New Type of Estimators.** We propose a class of estimators which are designed to exploit the strong Brownian bridge approximations. Such approximations can be useful for proving statistical properties of the estimators as well as for providing ways to compute associated asymptotic distributions through simulation. We shall define functionals of the type $J^n(\boldsymbol{\theta}) = \int_0^T \zeta(\boldsymbol{\theta}, \varphi, Y_t^n)dt$, which when suitably normalized, will converge strongly to a functional J of the Gaussian limit process. Furthermore, passage to the asymptotic limit can be made to proceed through estimators of the type $\widehat{\boldsymbol{\theta}}$, given by $J^n(\widehat{\boldsymbol{\theta}}) = \underset{\boldsymbol{\theta}}{inf} J^n(\boldsymbol{\theta})$ and will lead to $J(\boldsymbol{\theta_0})$ the value at the true parameter vector $\boldsymbol{\theta_0}$. In contrast with the squared integrated error type estimators our estimator involves φ more intricately in the integrand ζ.

As examples we give:

$$J_1^n(\boldsymbol{\theta}) = \int_0^T \frac{(U_t^n(\boldsymbol{\theta}))^2}{\frac{1}{2}(1+\varphi^R(2t,\boldsymbol{\theta}) - \varphi^R(t,\boldsymbol{\theta})^2}dt \text{ and } J_2^n(\boldsymbol{\theta}) = \int_0^T \frac{(V_t^n(\boldsymbol{\theta}))^2}{\frac{1}{2}(1-\varphi^R(2t,\boldsymbol{\theta})) - \varphi^I(t,\boldsymbol{\theta})^2}dt.$$

And for our estimators we define: $\widehat{\boldsymbol{\theta}}_k = \underset{\theta}{\operatorname{argmin}}\ J_k(\boldsymbol{\theta})$

The form of ζ has been purposely constructed so as to reflect the variance structure of the limit processes (real and imaginary). The integrand converges to a χ^2 distributed random variable at the true value of the parameter. It can be suitably selected according to the type of distribution under investigation or to capture the features considered important. Besides appropriate statistical properties which may be needed to ensure the required asymptotic convergence, η could be chosen so that simulation techniques can be applied on the corresponding stochastic integral of Brownian bridges to obtain numerical values for the required distributions. These estimators are more general than the integrated square error estimators in including directly the characteristic function in the "weighting"function for the integral. From now onwards we shall work with the minimizing function:

$$J^n(\boldsymbol{\theta}) = \int_0^T \frac{|Y_t^n(\boldsymbol{\theta})|^2}{1 - |\varphi(t, \boldsymbol{\theta})|^2} dt \tag{4}$$

which has been designed to penalize mismatches between the variance of the normalized ecf and the variance given by the $\boldsymbol{\theta}$ choice. In some sense we are forcing on our choice of estimates a variance structure on the normalized empirical characteristic function which is close to that of the limit process. As an extension of this idea we propose another estimator, which enforces the covariance structure more rigorously as follows, while it seeks for the minimum of the functional: $J^n(\boldsymbol{\theta}) = \int_0^T \int_0^T \frac{|Y_t^n(\boldsymbol{\theta})\overline{Y_s^n(\boldsymbol{\theta})}|^2}{\varphi(t-s,\boldsymbol{\theta}) - \varphi(t,\boldsymbol{\theta})\varphi(-s,\boldsymbol{\theta})} dsdt$ Working with this estimator may be a bit cumbersome, but from some simulation work we conducted, the results obtained were very encouraging. We shall revert to proving results for the simpler estimator 4. We prove a number of results about its statistical properties most of which should apply to similarly defined estimators along the lines indicated above.

2.6. **Basic results.** First a few definitions and elementary results:

Let $Y_t^n = \sqrt{n}(U_t^n + iV_t^n)$ so that $\sqrt{n}U_t^n = \mathfrak{Re}(Y_t^n)$ and $\sqrt{n}V_t^n = \mathfrak{Im}(Y_t^n)$.

The following equations hold:

$$\mathbb{E}[Y_t^n(\boldsymbol{\theta_0})] = 0 \text{ and } \mathbb{E}[Y_t^n(\boldsymbol{\theta_0})\overline{Y_s^n}(\boldsymbol{\theta_0})] = \varphi(t-s, \boldsymbol{\theta_0}) - \varphi(t, \boldsymbol{\theta_0})\varphi(-s, \boldsymbol{\theta_0})) \tag{5a}$$

$$\mathbb{E}[|Y_t^n|^2(\boldsymbol{\theta_0})] = 1 - |\varphi(t, \boldsymbol{\theta_0})|^2 \tag{5b}$$

$$\lim_{n\to\infty} c_t^n = \varphi(t, \boldsymbol{\theta_0}) \ \mathbb{P} \text{ a.s. uniformly in } t \tag{5c}$$

$$\lim_{n\to\infty} U_t^n(\boldsymbol{\theta_0}) = 0 = \lim_{n\to\infty} V_t^n(\boldsymbol{\theta_0}) \ \mathbb{P} \text{ a.s. uniformly in } t \tag{5d}$$

$$\mathbb{P}a.s. - \lim_{n\to\infty} Y_t^n(\boldsymbol{\theta_0}) = Z_t \tag{5e}$$

$$n\mathbb{V}ar[U_t^n(\boldsymbol{\theta_0})] = \frac{1}{2}(1 + \varphi^R(2t, \boldsymbol{\theta_0})) - \varphi^R(t, \boldsymbol{\theta_0})^2 \tag{5f}$$

$$n\mathbb{V}ar[V_t^n(\boldsymbol{\theta_0})] = \frac{1}{2}(1-\varphi^R(2t,\boldsymbol{\theta_0})) - \varphi^I(t,\boldsymbol{\theta_0})^2 \tag{5g}$$

$$n\mathbb{E}[U_t^n(\boldsymbol{\theta_0})V_t^n(\boldsymbol{\theta_0})] = \frac{1}{2}(\varphi^i(2t,\boldsymbol{\theta_0})-1) - \varphi^R(t,\boldsymbol{\theta_0})\varphi^I(t,\boldsymbol{\theta_0}) \tag{5h}$$

$$\mathbb{V}ar[Y_t^n(\boldsymbol{\theta_0})] = 1 - |\varphi(t,\boldsymbol{\theta_0})|^2 \tag{5i}$$

$$\frac{\partial U_t^n}{\partial \boldsymbol{\theta}} = \frac{\partial \varphi^R(t)}{\partial \boldsymbol{\theta}}, \frac{\partial V_t^n}{\partial \boldsymbol{\theta}} = \frac{\partial \varphi^I(t)}{\partial \boldsymbol{\theta}} \tag{5j}$$

2.7. **Consistency of BB Estimator.** We denote our estimator by $\widehat{\boldsymbol{\theta}} = \underset{\theta}{\text{argmin}}\ J^n(\boldsymbol{\theta})$, with J as in 4 from now onwards, and we shall refer to it as the BB estimator. To simplify our proofs parametrization will involve only one variable. The generalization to vector $\boldsymbol{\theta}$ will be straightforward. We write: $\frac{1}{n}J^n(\theta) = \int_0^T \eta(t,\theta)dt = \int_0^T \frac{|c_t^n - \varphi(t,\theta)|^2}{1-|\varphi(t,\theta)|^2}dt = \int_0^T \frac{(U_t^n)^2+(V_t^n)^2}{1-|\varphi(t,\theta)|^2}dt$ Note that $\mathbb{E}[\eta(\theta_0)] = \frac{1}{n}$ and $\underset{n\to\infty}{lim}\,\eta(t,\boldsymbol{\varphi_0}) = 0$ $\mathbb{P}$ a.s. uniformly in $t \in [0,T]$.
We shall need some other conditions which ensure that the integrals we use exist:

Condition 5. $\displaystyle\int_0^T \frac{|\frac{\partial \varphi^R}{\partial \boldsymbol{\theta}}|^2 + |\frac{\partial \varphi^I}{\partial \boldsymbol{\theta}}|^2}{1-|\varphi(t)|^2}dt < \infty.$

Condition 6. $\displaystyle\int_0^T \frac{1}{(1-|\varphi(t)|^2)^2}dt < \infty.$

Condition 7. *The usual regularity conditions, allowing the interchange of the integral and the differential operators, hold for integrands used.*

Theorem 1. *Under conditions 4, 5 , 6 and 7 the BB estimator in 4 is a strongly consistent estimator of* $\boldsymbol{\theta}$.

Proof. Firstly we observe that, assuming continuity of φ with respect to θ, $\displaystyle\underset{n\to\infty}{lim}\frac{1}{n}\mathbb{E}[J^n(\theta)] = \int_0^T \frac{|\varphi(t,\theta_0)-\varphi(t,\theta)|^2}{1-|\varphi(t,\theta)|^2}dt.$ $\mathbb{P}a.s.$ giving us $\displaystyle\underset{n\to\infty}{lim}\frac{1}{n}J(\theta_0) = 0, \mathbb{P}a.s.$, which minimum value is achieved only at θ_0 by the properties of characteristic functions. So that this minimum has to become isolated as n increases. The nature of the functions, whose minima we are chasing, and the above allow us to conclude that the values of θ giving us the minimum are random variables which have to converge to the value of θ for which the ultimate limit 0 is achieved. In other words the estimator 4 converges strongly to θ_0. □

2.8. **Asymptotic Distribution of the BB estimator.** We next set about proving the main result of this paper. We set the arguments in the sequel and present the theorem at the end of the section.

To make the notation a little less cumbersome we shall take our vector of parameters $\boldsymbol{\theta}$ as one-dimensional. Generalizing all our results to the multi-dimensional case is elementary.

Applying Taylor's theorem:
$\frac{\partial\eta}{\partial\theta}(\widehat{\theta}) = \frac{\partial\eta}{\partial\theta}(\theta_0) + (\widehat{\theta}-\theta_0)\frac{\partial^2\eta}{\partial^2\theta}(\theta_0 + \lambda(\widehat{\theta}-\theta_0))$ for some $|\lambda| < 1$
Also by the definition of the estimator: $\int_0^T \frac{\partial\eta}{\partial\theta}(\widehat{\theta})dt = 0$

For the derivations which follow we are evaluating all functions at $\theta = \theta_0$.

$$\int_0^T \frac{\partial\eta}{\partial\theta}(\theta_0)dt = 2\int_0^T \frac{U_t^n\frac{\partial\varphi^R}{\partial\theta} + V_t^n\frac{\partial\varphi^I}{\partial\theta}}{1-|\varphi(t)|^2}dt + 2\int_0^T \eta\frac{\frac{\partial\varphi^R}{\partial\theta}\varphi^R + \frac{\partial\varphi^I}{\partial\theta}\varphi^I}{1-|\varphi(t)|^2}dt$$

Both integrands in the RHS tend $\mathbb{P}$ almost surely to 0. The first term's asymptotic behaviour is given by:
$\sqrt{n}\int_0^T \frac{U_t^n\frac{\partial\varphi^R}{\partial\theta} + V_t^n\frac{\partial\varphi^I}{\partial\theta}}{1-|\varphi(t)|^2} \to \int_0^T \frac{U_t\frac{\partial\varphi^R}{\partial\theta} + V_t\frac{\partial\varphi^I}{\partial\theta}}{1-|\varphi(t)|^2}$ and it dominates the second term by an order of $n^{1/2}$. Also $\mathbb{E}\left[\frac{\partial\eta}{\partial\theta}\right] = \frac{\frac{\partial\varphi^R}{\partial\theta}\varphi^R + \frac{\partial\varphi^I}{\partial\theta}\varphi^I}{n(1-|\varphi(t)|^2)}$

Furthermore

$$\begin{aligned}\int_0^T \frac{\partial^2\eta}{\partial\theta^2}(\theta_0)dt &= 2\int_0^T \frac{(\frac{\partial\varphi^R}{\partial\theta})^2 + (\frac{\partial\varphi^I}{\partial\theta})^2 + U_t^n\frac{\partial^2\varphi^R}{\partial\theta^2} + V_t^n\frac{\partial^2\varphi^I}{\partial\theta^2}}{1-|\varphi(t)|^2}dt \\ &+ 4\int_0^T \Big(\frac{U_t^n\frac{\partial\varphi^R}{\partial\theta} + V_t^n\frac{\partial\varphi^I}{\partial\theta}}{1-|\varphi(t)|^2}\Big)\Big(\frac{\frac{\partial\varphi^R}{\partial\theta}\varphi^R + \frac{\partial\varphi^I}{\partial\theta}\varphi^I}{1-|\varphi(t)|^2}\Big)dt \\ &+ 2\int_0^T \frac{\partial\eta}{\partial\theta}\frac{\frac{\partial\varphi^R}{\partial\theta}\varphi^R + \frac{\partial\varphi^I}{\partial\theta}\varphi^I}{1-|\varphi(t)|^2}dt \\ &+ 2\int_0^T \eta\frac{\frac{\partial^2\varphi^R}{\partial\theta^2}\varphi^R + \frac{\partial^2\varphi^I}{\partial\theta^2}\varphi^I + (\frac{\partial\varphi^R}{\partial\theta})^2 + (\frac{\partial\varphi^I}{\partial\theta})^2}{1-|\varphi(t)|^2}dt \\ &+ 4\int_0^T \eta\frac{\frac{\partial\varphi^R}{\partial\theta}\varphi^R + \frac{\partial\varphi^I}{\partial\theta}\varphi^I}{(1-|\varphi(t)|^2)}\Big(\frac{\frac{\partial\varphi^R}{\partial\theta}\varphi^R + \frac{\partial\varphi^I}{\partial\theta}\varphi^I}{1-|\varphi(t)|^2}\Big)dt\end{aligned}$$

so that $\lim\limits_{n\to\infty} \mathbb{E}[\int\limits_0^T \frac{\partial^2\eta}{\partial\theta^2}(\theta_0)dt] = \int\limits_0^T \frac{|\varphi(t)|^2}{1-|\varphi(t)|^2}dt$
where all the terms on the right hand side are evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta_0}$.

So going back to the result derived from Taylor's theorem and using the results above and denoting the first term by K_n, we have:

$0 = K_n + (\widehat{\theta} - \theta_0)\frac{\partial^2\eta}{\partial^2\theta}(\theta_0 + \lambda(\widehat{\theta} - \theta_0))$ so that $\sqrt{n}(\widehat{\theta} - \theta_0) = \frac{-\sqrt{n}K_n}{\frac{\partial^2\eta}{\partial^2\theta}(\theta_0+\lambda(\widehat{\theta}-\theta_0))}$ Under the regularity assumptions, the denominator tends $\mathbb{P}a.s.$ to $\int_0^T \frac{|\varphi(t)|^2}{1-|\varphi(t)|^2}dt$ while the numerator is dominated by $W = \int_0^T \frac{U_t\frac{\partial\varphi^R}{\partial\theta}+V_t\frac{\partial\varphi^I}{\partial\theta}}{1-|\varphi(t)|^2}dt$ which is the sum of two centred normal random variables with covariance $C = \int_0^T \frac{\frac{1}{2}(\varphi^i(2t)-1)-\varphi^R(t)\varphi^I(t)}{(1-|\varphi(t)|^2)^2}\frac{\partial\varphi^R}{\partial\theta}\frac{\partial\varphi^I}{\partial\theta}dt$ But coming from U_t^n and V_t^n we can use the strong approximations using Brownian bridges we mentioned before.

Vectorizing our parameters, we have random vector $\boldsymbol{W}$ and matrix $\mathbb{C}$:

$$\boldsymbol{W} = \int\limits_0^T \frac{U_t\frac{\partial\varphi^R}{\partial\boldsymbol{\theta}} + V_t\frac{\partial\varphi^I}{\partial\boldsymbol{\theta}}}{1-|\varphi(t)|^2}dt \text{ and } \mathbb{C} = \int\limits_0^T \frac{\frac{1}{2}(\varphi^i(2t)-1) - \varphi^R(t)\varphi^I(t)}{(1-|\varphi(t)|^2)^2}\frac{\partial\varphi^R}{\partial\boldsymbol{\theta}}(\frac{\partial\varphi^I}{\partial\boldsymbol{\theta}})'dt$$

We comment again on the ability to work out numerically to excellent accuracy all the quantities we may require from random vector $\boldsymbol{W}$. The integrand can be simulated through the use of simulated paths from Brownian bridge. The parts needed from the characteristic function can be obtained as the corresponding quantities in $\varphi(t, \widehat{\boldsymbol{\theta}})$. Generating lots of proxy values for this random vector will allow us to approximate its variance, for instance, or obtain values for its distribution function. This same approximation $\varphi(t, \widehat{\boldsymbol{\theta}})$ can give us values for the entries of $\mathbb{C}$.

We state in generality the relevant theorem :

Theorem 2. *Given iid sequence $X_1, ..., X_n$, from a distribution with characteristic function $\varphi(t, \boldsymbol{\theta_0})$, and $T > 0$, under assumptions 1, 2, 3, 4 and :*

Condition 8. $\frac{\partial^2\varphi^R}{\partial\boldsymbol{\theta}^2}$ *and* $\frac{\partial^2\varphi^I}{\partial\boldsymbol{\theta}^2}$ *are dominated by a Lebesgue integrable functions over* $[0, T]$

the estimator: $\hat{\boldsymbol{\theta}} = \underset{\theta}{argmin} \int\limits_0^T \frac{|Y_t^n(\boldsymbol{\theta})|^2}{1-|\varphi(t,\boldsymbol{\theta})|^2}dt$ *is an asymptotically unbiased , consistent estimator of* $\boldsymbol{\theta_0}$ *for which the random vector* $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta_0})$ *converges* $\mathbb{P}$ *almost surely to a centred normally distributed random vector which has the same distribution as random vector* $\boldsymbol{W}$ *with covariance matrix* $\mathbb{C}$.

A comment at this stage should be made about the efficiency of the estimator under consideration. The rate of convergence given in the theorem above is clearly of the order of maximum likelihood, which asymptotically goes towards the optimal Cramér-Rao bound. We are technically in the same situation here.

3. Simulation Studies

Having obtained reassuring resultsabout our estimator, we next present results involving simulations using estimator 4. As a general guide, we tried to compare results from BB with those from maximum likelihood. MLE is the best there is in the business on a number of issues for a wide spectrum of distributions. So the comparison should be a stiff test for the viability of BB. Of primary importance, at this stage of preliminary testing, was the size of bias and of sampling variance. We should also mention the frequency of the data points, which naturally depend on the application, should also somehow come into the picture. Financial time series and climate statistics usually have data with very high frequency. But there are many other applications with more meagre datasets. Here we do just a preliminary exercise to check whether it is worthwhile to work further with BB. The choices of the parameters were not guided by some deep considerations and consequently they should be digested with caution.

We took samples with size varying in the medium range, 100 in steps of 100 to 500. Simulations with 5000 strong sample were also conducted to have a feel for how fast the convergence studied above moves in practice. Having started our discussion from a Lévy context , it makes only sense that we look at infinitely divisible distributions where MLE works well : normal and gamma. Tables 1 and 2 show clearly that as far as bias is concerned it is minimal for both estimators, in many cases the BB estimate being better. The situation with variance as expected is slightly in favour of MLE but not by much and furthermore as the sample dize increases the discrepancy in favour of MLE diminishes.

Table 1. Normally Distributed RV's

True parameters are $\mu = -1.32$, $\sigma^2 = 3.2$ and $T = 2$									
Sample Size	MLE means of $\widehat{\mu}$	MLE means of $\widehat{\sigma^2}$	BB means of $\widehat{\mu}$	BB means of $\widehat{\sigma^2}$	MLE variance of $\widehat{\mu}$	MLE variance of $\widehat{\sigma^2}$	BB variance of $\widehat{\mu}$	BB variance of $\widehat{\sigma^2}$
100	-1.2851	3.1813	-1.2966	3.1985	0.1150	0.0467	0.1215	0.0724
200	-1.3122	3.1770	-1.3150	3.1658	0.0563	0.0315	0.0624	0.0544
300	-1.2789	3.1813	-1.2834	3.1949	0.0345	0.0169	0.0393	0.0248
400	-1.3112	3.1722	-1.3142	3.1884	0.0217	0.0122	0.0271	0.0168
500	-1.3093	3.1791	-1.3038	3.1811	0.0211	0.0129	0.0266	0.0181
5000	-1.2982	3.1970	-1.3016	3.2011	0.0018	0.0013	0.0020	0.0018

Table 2. Gamma Distributed RV's

True parameters are $\alpha = 5.3$, $\sigma^2 = 4.2$ and $T = 2$								
Sample	MLE means of		BB means of		MLE variance of		BB variance of	
Size	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\alpha}$	$\widehat{\beta}$
100	5.3403	4.2770	5.3256	4.3063	0.6331	0.4170	0.8207	0.5412
200	5.3560	4.2033	5.3866	4.2137	0.3008	0.1794	0.5121	0.3207
300	5.3633	4.1724	5.3125	4.2146	0.2010	0.1326	0.2598	0.1836
400	5.3461	4.1600	5.3329	4.1793	0.1080	0.0691	0.1820	0.1133
500	5.3210	4.2144	5.3232	4.2213	0.1006	0.0634	0.1583	0.1001
5000	5.3057	4.2012	5.3157	4.1944	0.0129	0.0085	0.0195	0.0133

We also repeated the exercise with a stable distribution. The picture is very similar to the one we have just described for the other two distributions, though in this case the passage to the limit is more rough! Again the choice of parameters was casual as these results are preliminary in nature. The comparison here cannot be made with the MLE of course! So we used a method described in Koutrouvelis (1980)[14] to provide us with estimates from the same data for comparative purposes. Results can be seen in Table 3.

Table 3. Stable Distributed RV's

True parameters are $\alpha = 1.3$, $\beta = 0.2$, $\gamma = 1.5$, $\delta = 2.2$and $T = 2$								
Sample	Koutrouvelis Method means of				BB means of			
Size	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\gamma}$	$\widehat{\delta}$	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\gamma}$	$\widehat{\delta}$
100	1.3134	0.1868	1.4677	2.1716	1.3017	0.2002	1.4800	2.2098
200	1.2997	0.2039	1.4873	2.2951	1.2746	0.2028	1.4839	5.9524
300	1.2814	0.2039	1.4913	2.3515	1.2784	0.1844	1.4974	2.3635
400	1.2932	0.2158	1.4802	2.2910	1.2920	0.2127	1.4877	2.3354
500	1.2927	0.2121	1.4885	2.2734	1.2846	0.2151	1.4905	2.2883
	Koutrouvelis method variance of				BB variance of			
	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\gamma}$	$\widehat{\delta}$	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\gamma}$	$\widehat{\delta}$
100	0.0237	0.0719	0.0346	0.7752	0.0223	0.0791	0.0311	0.9740
200	0.0139	0.0461	0.0175	0.6410	0.0146	0.0437	0.0167	0.7981
300	0.0077	0.0260	0.0104	0.3785	0.0106	0.0236	0.0112	0.9468
400	0.0059	0.0209	0.0071	0.1878	0.0069	0.0202	0.0077	0.2971
500	0.0044	0.0119	0.0070	0.1416	0.0058	0.0173	0.0071	0.1956

4. Conclusion

Starting from a literature review of clever ecf uses in estimation problems for Lévy processes, one could well have a look at the integrated squared error method with two ideas in mind:

- The Brownian bridge approximation to the empirical characteristic functiom can be put to use more effectively.
- particular features of the type of characteristic function at hand could be incorporated suitably in the function whose mimnimum gives us the estimator

This strategy has worked well with our choice of estimator. The BB estimator has a variance-proxy term built out of the characteristic function embedded within the error function. Results obtained theoretically for this estimator give us an asymptotic behaviour close to that of the maximum likelihood. A few preliminary exercises using simulated data also gave promising results. More work needs to be done with the latter numerical efforts. Moreover, the ideas can be extended and particularized to specific distributions and Lévy process contexts so that more efficient and numerically stable methods can be devised.

References

1. Belomestny, D. 2010. Spectral Estimation of the fractional order of a Lévy process. *Ann. Statist.*, 38, No. 1, pp. 317–351
2. Carrasco, M. and Chernov, Mikhail and Florens, Jean-Pierre and Ghysels, E. 2007. Efficient estimation of general dynamic models with a continuum of moment conditions. *Journal of econometrics*, 140 (2), pp. 529–573
3. Carrasco, M. and Kotchoni, R. 2010. Efficient Estimation using the Characteristic Function. *Working paper.* Université de Montréal.
4. Csörgö, S. 1981a. *Strong Approximations in Probability and Statistics. Academic Press.*
5. Csörgö, S. 1981b. Limit behavior of the empirical distribution function. *Annals of Probability*, Vol. 9, No. 1, pp. 130–144.
6. Csörgö, M. 1983. Quantile Processes with Statistical Applications. *SIAM, Philadelphia.*
7. Feuerverger, A., Mureika, R. A. 1977. The empirical characteristic function and its applications. *Ann. Statist.*, 5, pp. 88–97.
8. Feuerverger, A., McDunnough, P. 1981a. On efficient inference in symmetric stable laws and processes, in *Statistics and Related Topics. New York: North Holland*, pp. 109–122.
9. Feuerverger, A., McDunnough, P. 1981b. On some Fourier methods for inference. *J. Am. Statist. Assoc.*, 76, pp. 379–387.
10. Heathcote, C. R. 1977. The integrated square error estimation of parameters. *Biometrika.*, 64 (2), pp. 255–64.
11. Kent, J. 1975. A Weak Convergence theorem for the empirical characteristic function. *J. Appl. Probability*, 12, pp. 512–523.

12. Kogon, S. M. and Williams, D.B. 1998. Characteristic function based estimation of stable distribution parameters, *A Practical Guide to Heavy Tails*. Birkhauser, pp. 311–325.
13. Komlos, J., Major, P., and Tusnady, G. 1975. An approximation of partial sums of independent random variables and the sample distribution function. *Z. Wahrsch. Verw. Gebeite*, 32, pp. 111–131.
14. Koutrovelis, I.A. 1980. Regression-type estimation of the parameters of stable laws. *J. Am. Statist. Assoc.*, 75, pp. 918–928.
15. Marcus, M.M. 1977. Weak convergence of the characteristic function. *The Annals of Probability*, Vol. 9, No. 2, pp. 194–201.
16. Parzen, E. .1962. On estimation of a probability density function and mode. *Ann. Math. Statist.*, 33, pp. 1065–1076
17. Paulson, A.S., Holcomb, E.W., and Leitch, A.H. 1977. The integrated square error estimation of parameters. *Biometrika.*, 62, pp. 163–70.
18. Press, S.J. 1972. Estimation in univariate and multivariate stable distributions. *J. Am. Statist. Assoc.*, 67, pp. 842–846
19. Yu, J. 2004. Empirical Characteristic function estimation and its applications. *Econometric Reviews*, Vol. 23, No. 2, pp. 93–123.