Parametric estimation for subordinators and induced OU-processes

GEURT JONGBLOED

Vrije Universiteit Amsterdam

FRANK H. VAN DER MEULEN

Institute for Business and Industrial Statistics of the University of Amsterdam

13th December 2005

Running heading: Estimation for subordinators

Abstract

Consider a stationary sequence of random variables with infinitely divisible marginal law, characterized by its Lévy density. We analyze the behavior of a so-called cumulant M-estimator, in case this Lévy density is characterized by a Euclidean (finite-dimensional) parameter. Under mild conditions, we prove consistency and asymptotic normality of the estimator. The estimator is considered in the situation where the data are increments of a subordinator as well as the situation where the data consist of a discretely sampled Ornstein Uhlenbeck process induced by the subordinator. We illustrate our results for the Gamma-process and the Inverse-Gaussian-OU-process. For these processes we also explain how the estimator can be computed numerically.

Key words and Phrases: cumulant, empirical characteristic function, Lévy process, self-decomposable distribution, stationary process.

1 Introduction

Let $Z = (Z_t, t \geq 0)$ be a pure-jump increasing Lévy process, also known as a *subordinator*. The distribution of Z is characterized by its marginal distribution at time $t = 1$ which, by the Lévy-Khintchine representation (Sato (1999), theorem 8.1), is infinitely divisible and has Fourier transform

$$\psi(t) = \mathbb{E}e^{itZ_1} = \exp\left(\int_0^\infty (e^{itx} - 1)\nu(dx)\right), \quad \forall t \in \mathbb{R}.$$

Here the measure ν is a σ-finite measure on $(0, \infty)$ satisfying the integrability condition $\int (x \wedge 1)\nu(dx) < \infty$. The measure ν is called the *Lévy measure* of Z. We assume ν admits a density with respect to Lebesgue measure, which we denote by a. Thus $\nu(dx) = a\,dx$.

Barndorff-Nielsen & Shephard (2001) consider the *Ornstein-Uhlenbeck (OU)-process* driven by Z as a model for stochastic volatility in mathematical finance. For a given number $\lambda > 0$, this process is defined as the stationary solution to the stochastic differential equation

$$dX(t) = -\lambda X(t)dt + dZ(\lambda t).$$

The marginal distribution of X is positive and infinitely divisible with Lévy density that can be expressed in terms of the Lévy measure of Z:

$$\nu^X(dx) = x^{-1}\nu(x, \infty)\,dx =: x^{-1}k(x)\,dx$$

Here k is a decreasing function on $(0, \infty)$. In Jongbloed et al. (2005), a nonparametric estimator for this Lévy measure is introduced and shown to be consistent under appropriate conditions. Restricting the model to a parametric one, asymptotic distribution theory can be established as well. In the present paper we introduce a parametric analogue of this estimator based on a discretely observed OU process. However, since the monotonicity of the function k is not essential for the proofs in this paper (contrary to Jongbloed et al. (2005)), we adopt a slightly more general

approach, also covering parametric estimation of a based directly on the discretely observed process Z.

Suppose we observe a (sufficiently mixing) stationary time series $X_1, X_2, \ldots$ with infinitely divisible marginal distribution. Assume the Lévy density of X_1 is parametrized by a Euclidean parameter $\theta \in \Theta \subseteq \mathbb{R}^k$. Hence, $a = a_\theta$ and $\psi = \psi_\theta$. Let $\theta_0 \in \Theta$ denote the true value of the parameter, which we wish to estimate.

Firstly, we explain our estimation paradigm briefly, deferring the precise details to the next section. As a starting point, we take a sequence of preliminary estimators $(\tilde{\psi}_n, n \geq 1)$ such that for each t, $\tilde{\psi}_n(t) \to \psi_{\theta_0}(t)$, either almost surely or in probability. The most natural example of such an estimator satisfying this consistency condition in many interesting situations, is the *empirical characteristic function*, though other estimators are possible. Now define an estimator $\hat{\theta}_n$ for θ_0 as the minimizer of the random criterion function

$$\theta \mapsto \int (\log \tilde{\psi}_n(t) - \log \psi_\theta(t))^2 w(t) dt \tag{1}$$

over Θ. Here w is a compactly supported weight-function and $\log \psi$ refers to the distinguished logarithm of a characteristic function ψ, the cumulant function. For this reason, we call this estimator a *cumulant M-estimator* (CME). In this paper, we give conditions on the existence of the CME $\hat{\theta}_n$, and prove consistency and asymptotic normality under mild conditions. The asymptotic analysis is restricted to the case where we use the empirical characteristic function as preliminary estimator.

We consider two specific examples for the sequence (X_n). The first is a sequence of increments of Z. Since Z has stationary and independent increments, the differences $X_i = Z_{i\Delta} - Z_{(i-1)\Delta}$ are independent and identically distributed (i.i.d.). Moreover, X_1 is infinitely divisible with Lévy measure $\Delta \nu_{\theta_0}$. Without loss of generality we assume $\Delta = 1$. The second is a stationary sequence obtained by sampling the OU-process driven by Z at equidistant time points. The latter problem is in fact the problem addressed nonparametrically in Jongbloed et al. (2005).

Especially in the i.i.d. setting of increments of Z, a maximum likelihood ap-

proach might seem appropriate. However, a problem is that for many infinitely divisible distributions a closed-form expression for the density of X_i is not available or intractable, whereas the Lévy density a_{θ_0} is often explicit. An example of such infinitely divisible distributions is given by the class of tempered stable distributions (cf. Barndorff-Nielsen & Shephard (2002)), which are used in finance.

Some related work on empirical characteristic function procedures includes Feuerverger & McDunnough (1981), Feuerverger & McDunnough (1981a) and Knight & Satchell (1997). In the first two papers an estimator is obtained by minimizing a quadratic distance between the empirical characteristic function and the supposed parametric characteristic function. Knight & Satchell (1997) take the same approach, with characteristic functions replaced by cumulant functions. In this sense, their estimators are closer to the ones defined here, although they consider a criterion function that differs from (1). Our estimator has the advantage that it works for any consistent sequence of preliminary estimators (in a sense to be made precise later on).

Bayesian estimation for OU-processes driven by a subordinator is discussed in Roberts et al. (2004). A nonparametric estimator for the Lévy measure of a discretely observed Lévy process is given in Figueroa-López & Houdré (2004). Their estimation method is different from ours in the sense that their estimator is explicitly based on the jumps of the process, requiring a continuous monitoring of the process. A parametric estimator is obtained by calibrating the parametric model to the nonparametric estimator.

Here is an outline of the contents of this paper. In the next section, we start with a precise definition of our estimator. Sufficient conditions for consistency are given in section 3. In section 4 we analyze the asymptotic behavior of the CME, in case $\tilde{\psi}_n$ is the empirical characteristic function of the observations. The results are illustrated in section 5, where we suppose we discretely observe a Gamma Lévy process. In section 6 we move to the case of dependent observations from a Lévy

4

driven OU-process. To illustrate the results, we work out the computations involved for the Inverse-Gaussian OU-process. The appendix contains some technical results.

2   Definition of the estimator

Throughout, we denote the true value of the parameter by θ_0 and suppose $\theta_0 \in \Theta \subseteq \mathbb{R}^k$. We write d for Euclidean distance on $\mathbb{R}^k$. We aim to estimate θ_0, based on $X_1, \ldots, X_n$, where X_1 has law π_{θ_0} satisfying

$$\psi_{\theta_0}(t) = \int e^{itx} \pi_{\theta_0}(dx) = \exp\left(\int_0^\infty (e^{itx} - 1)a_{\theta_0}(x)dx\right), \quad \forall t \in \mathbb{R}.$$

To explain our estimation method, we first discuss some characterizations of the distribution of X_1.

By the uniqueness theorem for characteristic functions, π_{θ_0} is determined by ψ_{θ_0}. The following result shows that it suffices to know ψ_{θ_0} on a finite interval containing the origin. A proof can be found in Loéve (1977), chapter 4.

Lemma 1 *Let X be a positive random variable with characteristic function ψ. If ψ_M is the restriction of ψ to an interval $(-M, M)$, then ψ_M determines ψ.*

By a result from complex analysis (see for example Chung (2001), section 7.6), we can attach to each characteristic function without zeros a unique continuous function g such that $e^{g(t)} = \psi(t)$. We call this function the *cumulant function* and denote the cumulant of a particular characteristic function ψ_1 by g_1 or $\log \psi_1$. Since an infinitely divisible characteristic function has no zeros (Sato (1999), lemma 7.5), its cumulant is well-defined. Combining this with the above lemma, we conclude that the distribution of X_1 is determined by its cumulant function on an interval $(-M, M)$. Therefore, all information on θ_0 is present in g_{θ_0}, restricted to an interval $(-M, M)$.

To obtain a consistent estimator for θ_0, we suppose there exists a preliminary estimator $\tilde{\psi}_n$ for ψ_0 such that

$$\forall t \in \mathbb{R} \quad \tilde{\psi}_n(t) \to \psi_{\theta_0}(t), \text{ as } n \to \infty, \tag{2}$$

where the convergence is either almost surely or in probability. Unless stated otherwise, we will assume (2) holds with convergence almost surely. Often, this requirement is fulfilled by the empirical characteristic function (e.c.f.), which is defined by $\tilde{\psi}_n(t) = \int e^{itx} d\mathbb{F}_n(x)$, where $\mathbb{F}_n$ denotes the empirical distribution function of the data. If (2) holds, then, for n sufficiently large, we expect that $\tilde{g}_n = \log \tilde{\psi}_n$ and g_{θ_0} are close in a sense yet to be made precise. Upon comparing these functions on a compact interval around the origin, we obtain an M-estimator. We now work out the details.

Let w be an integrable weight-function with compact support, denoted by S. Assume w is symmetric around the origin and w is strictly positive on a neighborhood of the origin. One may think of $w = \mathbf{1}_{[-m,m]}$ for some $m > 0$. Define the space of square integrable functions w.r.t. $w(t)dt$ by

$$\mathcal{L}^2(w) := \left\{ f : \mathbb{R} \to \mathbb{C} \mid f \text{ is measurable and } \int |f(t)|^2 w(t)dt < \infty \right\}.$$

We define a semi inner-product $\langle \cdot, \cdot \rangle_w$ on $\mathcal{L}^2(w)$ by

$$
\begin{aligned}
\langle f, g \rangle_w &= \Re \int f(t)\overline{g(t)}w(t)dt \\
&= \int \Re f(t)\Re g(t)w(t)dt + \int \Im f(t)\Im g(t)w(t)dt\,,
\end{aligned}
$$

where the bar over g denotes complex conjugate and $\Re$ ($\Im$) the operation of taking the real (imaginary) part of an element of $\mathbb{C}$. For $g \in L^2(w)$ we define a seminorm by $\|g\|_w = \sqrt{\langle g, g \rangle_w}$.

Throughout the rest of the paper, we assume n is large enough such that $\tilde{g}_n = \log \tilde{\psi}_n$ exists on S. Given $\tilde{g}_n$, we define an estimator for θ_0 as the minimizer of

$$\Gamma_n(\theta) := \|g_\theta - \tilde{g}_n\|_w^2 = \int |g_\theta(t) - \tilde{g}_n(t)|^2 w(t)dt$$

over $\Theta \subseteq \mathbb{R}^k$. Since we minimize a weighted difference between cumulants, we call this estimator a *cumulant-M-estimator* (CME). Put $M_n(\theta) := \sqrt{\Gamma_n(\theta)}$. In the following, we write $\|\cdot\|_\infty$ for the supremum norm on S, i.e. for $f : \mathbb{R} \to \mathbb{C}$ bounded, we define $\|f\|_\infty = \sup_{t \in S} |f(t)|$.

Proposition 2 *Assume the parameter-set $\Theta \subseteq \mathbb{R}^k$ is compact. If for every sequence $(\theta_m) \in \Theta$*

$$\theta_m \to \theta \quad \Rightarrow \quad \int_0^\infty (x \wedge 1) \cdot |a_{\theta_m}(x) - a_\theta(x)| dx \to 0, \quad (m \to \infty) \tag{3}$$

then there is a $\hat{\theta}_n \in \Theta$ such that $M_n(\hat{\theta}_n) = \min_{\theta \in \Theta} M_n(\theta)$.

Proof It suffices to prove that the mapping $\theta \mapsto M_n(\theta)$ is continuous. Let $\theta, \xi \in \Theta$. For $t \in S$,

$$
\begin{aligned}
|g_\theta(t) - g_\xi(t)| &= \left| \int_0^\infty (e^{itx} - 1)(a_\theta(x) - a_\xi(x)) dx \right| \\
&\leq |t| \int_0^1 x \, |a_\theta(x) - a_\xi(x)| dx + 2 \int_1^\infty |a_\theta(x) - a_\xi(x)| dx \\
&\leq \max\{|t|, 2\} \int_0^\infty (x \wedge 1) |a_\theta(x) - a_\xi(x)| dx,
\end{aligned}
$$

where we use the inequality $|e^{ix} - 1| \leq \min\{|x|, 2\}$. By the triangle inequality, we find that

$$
\begin{aligned}
|M_n(\theta) - M_n(\xi)| &= \big| \|g_\theta - \tilde{g}_n\|_w - \|g_\xi - \tilde{g}_n\|_w \big| \leq \|g_\theta - g_\xi\|_w \\
&\leq \left(\|g_\theta - g_\xi\|_\infty^2 \int w(t) dt \right)^{1/2} \\
&\leq C \int_0^\infty (x \wedge 1) \cdot |a_\theta(x) - a_\xi(x)| dx,
\end{aligned}
$$

for some constant $C > 0$. Now, continuity of M_n on Θ is a consequence of assumption (3). $\quad\blacksquare$

By a dominated convergence argument, it is often easy to check for (3).

Corollary 3 *Assume the parameter-set $\Theta \subseteq \mathbb{R}^k$ is compact. Suppose there exists a function $A : (0, \infty) \to [0, \infty)$ satisfying $\int_0^\infty (x \wedge 1) A(x) dx < \infty$ (e.g. $A(x) = 1/(x\sqrt{x})$) such that $\sup_{\theta \in \Theta} a_\theta(x) \leq A(x)$, for all $x > 0$. If the mapping $\theta \mapsto a_\theta(x)$ is continuous for all $x > 0$, then $\hat{\theta}_n = \operatorname{argmin}_{\theta \in \Theta} \Gamma_n(\theta)$ exists.*

Remark 4 Even if the Lévy measure ν does not admit a density with respect to Lebesgue measure we may still prove existence of a CME. For example, let Z be a

Poisson process with jumps of size $\theta \in \Theta \subseteq \mathbb{R}$. Then ν_θ is a point-mass at θ and hence

$$|g_\theta(t) - g_\xi(t)| = |e^{i\theta t} - e^{i\xi t}| \le |t| \cdot |\theta - \xi|,$$

which implies that we can adapt the proof of proposition 2 to this case.

3 Consistency

Under appropriate conditions, we expect the random criterion functions M_n to converge to a deterministic function $M : \Theta \to [0, \infty]$. Since roughly, $\tilde{g}_n \sim g_{\theta_0}$ for n large, we expect, under θ_0, that

$$M_n(\theta) \xrightarrow{\text{a.s.}} M(\theta) := \|g_\theta - g_{\theta_0}\|_w.$$

The latter deterministic map is easily seen to be minimized for $\theta = \theta_0$.

Although $\hat{\theta}_n$ is possibly not uniquely defined, the next theorem shows that any choice of $\hat{\theta}_n$ (as minimizer of M_n) is a consistent estimator for θ_0 under mild conditions.

Theorem 5 *Suppose $\Theta \subseteq \mathbb{R}^k$ is compact and (3) holds. Assume the sequence of preliminary estimators is such that (2) holds (for convergence almost surely). If $\pi_\theta \ne \pi_{\theta_0}$ whenever $\theta \ne \theta_0$, then, $\hat{\theta}_n \to \theta_0$, almost surely, as n tends to infinity.*

Proof By lemma 5.2 in Jongbloed et al. (2005), we can strengthen the convergence in (2) to $\sup_{t \in S} |\tilde{\psi}_n(t) - \psi_{\theta_0}(t)| \xrightarrow{\text{a.s.}} 0$ under θ_0. Theorem 7.6.3 in Chung (2001) implies that the uniform convergence of $\tilde{\psi}_n$ to ψ_{θ_0} on S carries over to uniform convergence of $\tilde{g}_n$ to g_{θ_0} on S. That is, $\|\tilde{g}_n - g_{\theta_0}\|_\infty \xrightarrow{\text{a.s.}} 0$. By the triangle inequality we have

$$
\begin{aligned}
\sup_{\theta \in \Theta} |M_n(\theta) - M(\theta)| &= \sup_{\theta \in \Theta} |\|g_\theta - \tilde{g}_n\|_w - \|g_\theta - g_{\theta_0}\|_w| \\
&\le \|\tilde{g}_n - g_{\theta_0}\|_w \le \|\tilde{g}_n - g_{\theta_0}\|_\infty \left(\int w(t)dt \right)^{1/2} \xrightarrow{\text{a.s.}} 0.
\end{aligned}
$$

Furthermore, since $\hat{\theta}_n$ minimizes M_n over Θ,

$$M_n(\hat{\theta}_n) \le M_n(\theta_0) = \|g_{\theta_0} - \tilde{g}_n\|_w \xrightarrow{\text{a.s.}} 0.$$

8

Combining both assertions, we conclude that $M(\hat{\theta}_n) \xrightarrow{\text{a.s.}} 0$ $(n \to \infty)$. Once we have proved

$$\forall \varepsilon > 0 \quad \inf_{\theta \,:\, d(\theta,\theta_0) \geq \varepsilon} M(\theta) > M(\theta_0) = 0, \tag{4}$$

this implies that for each $\omega \in \Omega$ and for every $\varepsilon > 0$, $d(\hat{\theta}_n, \theta_0) < \varepsilon$ for $n \geq N$. That is, $\hat{\theta}_n \xrightarrow{\text{a.s.}} \theta_0$, as $n \to \infty$.

We now prove (4). Suppose $\theta \neq \theta_0$, then by identifiability $\pi_\theta \neq \pi_{\theta_0}$. Using lemma 1, this implies that there exists a neighborhood $U \subseteq S$ around zero, such that $\psi_\theta \not\equiv \psi_{\theta_0}$ on U. Therefore $g_\theta \not\equiv g_{\theta_0}$ on U and hence, by the continuity of the cumulant function, $M(\theta) = \|g_\theta - g_{\theta_0}\|_w > 0$, since we have $U \subseteq S$. Thus $M(\theta) > M(\theta_0) = 0$. Therefore M has a *unique* minimizer on Θ.

By similar inequalities as in the proof of lemma 2, for $\theta, \xi \in \Theta$,

$$|M(\theta) - M(\xi)| \leq \|g_\theta - g_\xi\|_w \leq C \int_0^\infty (x \wedge 1)|a_\theta - a_\xi| dx,$$

for some positive constant C. Hence, by (3), the mapping $\theta \mapsto M(\theta)$ is continuous on Θ. Let $\varepsilon > 0$. The set $V_{\theta_0} := \{\theta \in \Theta \;:\; d(\theta, \theta_0) \geq \varepsilon\}$ is a closed subset of the compact set Θ and hence compact. By continuity, $M(V_{\theta_0})$, the image of V_{θ_0} under M, is compact. Since θ_0 is the *unique* minimizer of M on Θ and $M(\theta_0) = 0$, $0 \notin M(V_{\theta_0})$. Therefore, there exists an $\eta > 0$ such that $M(\theta) \geq \eta$ on V_{θ_0}. This proves (4).

In case we assume (2) holds with convergence in probability instead of convergence almost surely and retain the other assumptions of theorem 5, the proof of the previous theorem can easily be adapted to show that in that case $\sup_{\theta \in \Theta} |M_n(\theta) - M(\theta)| \xrightarrow{\text{p}} 0$. Since (4) remains true, we can apply theorem 5.7 in Van der Vaart (1998), which then asserts that any sequence of estimators $\bar{\theta}_n$ with $M_n(\bar{\theta}_n) \leq M_n(\theta_0) + o_P(1)$ converges in probability to θ_0. (Here, and in the following we use stochastic order symbols: for a sequence of random vectors X_n and a given sequence of random variables R_n we write $X_n = o_P(R_n)$ if $X_n = Y_n R_n$ for a sequence

of random vectors Y_n tending to zero in probability.) Now

$$|M_n(\hat{\theta}_n) - M_n(\theta_0)| \le M_n(\hat{\theta}_n) + |M(\theta_0) - M_n(\theta_0)| = o_P(1) + o_P(1),$$

where the two $o_P(1)$ terms follow from the inequalities in the first two displays of the proof of theorem 5. Therefore, $\hat{\theta}_n \xrightarrow{\mathrm{P}} \theta_0$.

Corollary 6 *Suppose $\Theta \subseteq \mathbb{R}^k$ is compact and (3) holds. Assume the sequence of preliminary estimators satisfies (2) for convergence in probability. If $\pi_\theta \ne \pi_{\theta_0}$ whenever $\theta \ne \theta_0$, then, $\hat{\theta}_n \xrightarrow{\mathrm{P}} \theta_0$, as n tends to infinity.*

4   Asymptotic behavior of the estimator

To derive asymptotic distribution results for the CME, we go through the following steps:

(I) Derive weak convergence of the process $(\sqrt{n}(\tilde{\psi}_n(t) - \psi_{\theta_0}(t)), t \in S)$.

(II) Show Hadamard differentiability of the mapping that attaches to a characteristic function its cumulant. Subsequently, apply the functional Delta method to obtain weak convergence of the process $(\sqrt{n}(\tilde{g}_n(t) - g_{\theta_0}(t)), t \in S)$

(III) Use results from the theory of M- and Z-estimators to derive the asymptotic distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$.

Details will be given along the way.

To simplify results a bit, we assume in this section that our initial estimator is the empirical characteristic function, which we denote by $\tilde{\psi}_n$ (see however remark 16). Throughout this section, we assume all observations are independent, with common distribution π_{θ_0}.

Step (I)

Define for $t \in \mathbb{R}$

$$Y_n(t) = \sqrt{n}(\tilde{\psi}_n(t) - \psi_{\theta_0}(t)).$$

10

(Presently, we suppress the dependence of Y_n on θ_0 for ease of notation. Later on, we will explicitly write $Y_n^{\theta_0}$.)

By the multivariate central limit theorem it follows that for $t_1 < \cdots < t_k$,

$$(Y_n(t_1), \ldots, Y_n(t_k)) \rightsquigarrow N_k(0, \Sigma(\theta_0)), \quad n \to \infty,$$

where, if X is a random variable with distribution π_{θ_0},

$$\Sigma^{\theta_0}|_{j,k} = \mathbb{E}(e^{it_j X} - \psi_{\theta_0}(t_j))\overline{(e^{it_k X} - \psi_{\theta_0}(t_k))} = \psi_{\theta_0}(t_j - t_k) - \psi_{\theta_0}(t_j)\psi_{\theta_0}(-t_k). \quad (5)$$

We aim to prove that there exists a centered Gaussian process Y with covariance as in (5) such that Y_n converges weakly to Y in the space $\ell^\infty(S)$, where

$$\ell^\infty(S) = \{z : S \to \mathbb{C} \ : \ \|z\|_\infty = \sup_{t \in S} |z(t)| < \infty\},$$

is the space of bounded complex valued functions on S, equipped with the supremum norm. By weak convergence we mean

$$E^* h(Y_n) \to E h(Y), \quad n \to \infty$$

for every bounded and continuous function $h : \ell^\infty(S) \to \mathbb{R}$. Here E^* denotes outer expectation. (We use outer expectations, since elements of $\ell^\infty(S)$ may not be Borel-measurable. See for instance Van der Vaart (1998), chapter 18.) If the sequence Y_n converges weakly to Y, we write $Y_n \rightsquigarrow Y$.

The result below follows from Giné and Zinn (1986). The ε-covering number of a set A for a semi-metric ρ, denoted by $N(\varepsilon, A, \rho)$, is defined as the minimal number of ρ-balls of radius ε needed to cover the set A.

Theorem 7 (Giné and Zinn (1986), chapter 4, theorem 6.1) *Let X be a random variable with distribution π_{θ_0}. Define for $s, t \in \mathbb{R}$*

$$\sigma(s,t) := \left(\mathbb{E}|e^{itX} - e^{isX}|^2 \right)^{1/2} = 2\left(\mathbb{E}\left[\sin^2 \tfrac{1}{2}(t-s)X \right] \right)^{1/2}.$$

If

$$\int_0^{\varepsilon_0} (\log N(\varepsilon, S, \sigma))^{1/2} d\varepsilon < \infty, \quad (6)$$

11

for some small $\varepsilon_0 > 0$, then Y_n converges weakly in the space $\ell^\infty(S)$ to a complex-valued centered Gaussian process Y with covariance structure

$$cov(Y_t, Y_s) = \psi_{\theta_0}(t - s) - \psi_{\theta_0}(t)\psi_{\theta_0}(-s). \tag{7}$$

Let B denote a standard Brownian Bridge and define a process B^0 by

$$B^0(x) = B(\pi_{\theta_0}([0, x])), \quad x \geq 0. \tag{8}$$

The limit process can be identified as the process Y', defined by

$$Y'(t) := \int e^{itx} dB^0(x), \quad t \in S, \tag{9}$$

where B denotes a standard Brownian Bridge. To see this, note that Y' is a centered Gaussian process with the same covariance function as Y. Since weak limits are unique in a distributional sense, it follows that Y_n converges weakly in $\ell^\infty(S)$ to Y'.

Since $F := (C(S), \|\cdot\|_\infty)$ is a closed subset of $(\ell^\infty(S), \|\cdot\|_\infty)$, we have by the Portmanteau theorem

$$P(Y' \in F) \geq \limsup_n P(Y_n \in F) = 1.$$

Therefore Y' is almost surely continuous and by lemma 18.13 in Van der Vaart (1998) the weak convergence of Y_n of Y in space $\ell^\infty(S)$ carries over to the space $C(S)$. Hence, $Y_n \rightsquigarrow Y'$ in $(C(S), \|\cdot\|_\infty)$. We summarize this result in a corollary.

Corollary 8 *If the entropy condition of theorem 7 holds true, then Y_n converges weakly in $(C(S), \|\cdot\|_\infty)$ to the centered Gaussian process Y', as defined in (9).*

A simple moment condition suffices for (6).

Lemma 9 *If for some $\alpha > 0$*

$$\int |x|^\alpha \pi_{\theta_0}(dx) < \infty, \tag{10}$$

then the entropy condition (6) is satisfied and the conclusion of theorem 7 holds.

A proof is given in the appendix.

Step (II)

Next, we consider the weak convergence of the empirical cumulant process, defined by

$$U_n(t) := \sqrt{n}(\tilde{g}_n(t) - g_{\theta_0}(t)) = \sqrt{n}(\log \tilde{\psi}_n(t) - \log \psi_{\theta_0}(t)).$$

We use the functional Delta method (see e.g. Van der Vaart (1998), chapter 20). Let Ψ denote the set of infinitely divisible characteristic functions and let T be the mapping that assigns to $\psi \in \Psi$ its corresponding cumulant function. We write $G = T(\Psi)$ to denote the set of cumulants. Recall $\|\cdot\|_\infty$ refers to the supremum norm on S.

Proposition 10 *The mapping $T : (\Psi, \|\cdot\|_\infty) \to (G, \|\cdot\|_\infty)$ is Hadamard differentiable tangentially to $C^0 := \{f : \mathbb{R} \to \mathbb{C} \mid f \text{ is continuous and } f(0) = 0\}$. Its Hadamard derivative at ψ is given by $T'_\psi(\varphi) = \varphi/\psi$ for $\varphi \in C^0$.*

Proof Fix $\psi \in \Psi$ and $\varphi \in C^0$. Let φ_ε be such that $\psi + \varepsilon\varphi_\varepsilon \in \Psi$ for all small $\varepsilon > 0$ and such that $\|\varphi_\varepsilon - \varphi\|_\infty \to 0$, as $\varepsilon \downarrow 0$. By the triangle inequality

$$\left\|\frac{T(\psi + \varepsilon\varphi_\varepsilon) - T(\psi)}{\varepsilon} - \frac{\varphi}{\psi}\right\|_\infty \leq \left\|\frac{T(\psi + \varepsilon\varphi_\varepsilon) - T(\psi)}{\varepsilon} - \frac{\varphi_\varepsilon}{\psi}\right\|_\infty + \left\|\frac{\varphi - \varphi_\varepsilon}{\psi}\right\|_\infty. \quad (11)$$

By the defining property of the distinguished logarithm, we have for any $\psi_1, \psi_2 \in \Psi$ that $T(\psi_1) - T(\psi_2) = T(\psi_1/\psi_2)$. Therefore, the first term on the right-hand-side of the preceding display equals

$$\frac{1}{\varepsilon}\left\|T\left(1 + \varepsilon\frac{\varphi_\varepsilon}{\psi}\right) - \varepsilon\frac{\varphi_\varepsilon}{\psi}\right\|_\infty = \frac{1}{\varepsilon}\sup_{t \in S}\left|T\left(1 + \varepsilon\frac{\varphi_\varepsilon}{\psi}\right)(t) - \varepsilon\frac{\varphi_\varepsilon(t)}{\psi(t)}\right|. \quad (12)$$

Define $u_\varepsilon(t) := \varphi_\varepsilon(t)/\psi(t)$. Since $\{|u_\varepsilon(t)|, t \in S\}$ is compact, we can choose ε small enough such that $\varepsilon u_\varepsilon(t) \in B_{1/2}(0) = \{z \in \mathbb{C} : |z| \leq 1/2\}$ for all $t \in S$. Within $B_{1/2}(0)$ the following inequality holds: $|\log(1 + w) - w| \leq \frac{1}{2}|w|^2/(1 - |w|)$ (see Remmert (1991), section 5.4). Using this inequality we see that (12) is bounded by

$$\frac{1}{\varepsilon}\sup_{t \in S}\left|\frac{1}{2}\frac{\varepsilon^2(u_\varepsilon(t))^2}{1 - \varepsilon u_\varepsilon(t)}\right| \leq \frac{1}{2\varepsilon}\frac{\varepsilon^2 M_\varepsilon^2}{1 - \varepsilon M_\varepsilon} \leq \frac{1}{2}M_\varepsilon^2\varepsilon,$$

13

where $M_\varepsilon = \|u_\varepsilon\|_\infty$. Now $M_\varepsilon \leq \left\|\frac{\varphi - \varphi_\varepsilon}{\psi}\right\|_\infty + \left\|\frac{\varphi}{\psi}\right\|_\infty$. We conclude that once we have proved that $\left\|\frac{\varphi - \varphi_\varepsilon}{\psi}\right\|_\infty \to 0$ as $\varepsilon \downarrow 0$ and $\left\|\frac{\varphi}{\psi}\right\|_\infty < \infty$, then both terms on the right-hand-side of (11) can be made arbitrarily small by letting ε tend to zero.

Since an infinitely divisible characteristic function has no zeros, every $\psi \in \Psi$ is bounded away from zero on compacta. Therefore $\|1/\psi\|_\infty = \sup_{t \in S} |1/\psi(t)| \leq C$ for some positive constant C and

$$\|(\varphi - \varphi_\varepsilon)/\psi\|_\infty \leq \|1/\psi\|_\infty \|\varphi_\varepsilon - \varphi\|_\infty \leq C\|\varphi_\varepsilon - \varphi\|_\infty,$$

which tends to zero as ε tends to zero. The same argument shows that $\|\varphi/\psi\|_\infty \leq C\|\varphi\|_\infty < \infty$. Hence, the left-hand-side of (11) can be made arbitrarily small, which means that T is Hadamard differentiable at ψ with T'_ψ as stated.

By the functional Delta method (Van der Vaart (1998), theorem 20.8) we now obtain

Corollary 11 *If $Y_n \rightsquigarrow Y$ in $(C(S), \|\cdot\|_\infty)$, then $U_n \rightsquigarrow U$ in $(C(S), \|\cdot\|_\infty)$, where*

$$U = T'_\psi(Y) = Y/\psi.$$

The process U is centered Gaussian with covariance function

$$cov(U(t), U(s)) = [\psi_{\theta_0}(t - s) - \psi_{\theta_0}(t)\psi_{\theta_0}(-s)]/[\psi_{\theta_0}(t)\psi_{\theta_0}(-s)], \quad t, s \in S.$$

In the following we will write $U_n^{\theta_0}$ and U^{θ_0}, to avoid confusion with other thetas appearing.

Step (III)

To derive the asymptotic behavior of $\hat{\theta}_n$, we define conditions which enable us to define $\hat{\theta}_n$ as a Z-estimator. By this we mean that we define conditions under which $\hat{\theta}_n$ is a point for which all partial derivatives of $\theta \mapsto \Gamma_n(\theta)$ are (nearly) zero. The following assumption suits our purposes well. We denote by θ_i the i-th coordinate of a vector $\theta \in \Theta \subseteq \mathbb{R}^k$.

Assumption 12 *(i).* Condition (6) is satisfied, so that the process Y_n converges weakly and, by corollary 11, the process U_n as well.

(ii). For each $i = 1, \dots, k$ the partial derivative $\theta \mapsto g_\theta^i := \frac{\partial}{\partial \theta_i} g_\theta$ exists. Moreover, the mapping $(\theta, t) \mapsto g_\theta^i(t)$ is jointly continuous on $\Theta \times S$.

Combining the second part of this assumption with dominated convergence we get

$$\frac{\partial}{\partial \theta_i} \Gamma_n(\theta) = 2\langle g_\theta^i, g_\theta - \tilde{g}_n \rangle_w =: \Psi_n^i(\theta), \quad 1 \le i \le k.$$

Hence, $\hat{\theta}_n$ is a (near) zero of the random criterion function

$$\Psi_n(\theta) := (\Psi_n^1(\theta), \dots, \Psi_n^k(\theta))',$$

Since $\tilde{g}_n$ converges almost surely uniformly on S to g_{θ_0} , we expect that $\hat{\theta}_n$ converges to a zero of

$$\Psi_{\theta_0}(\theta) := (2\langle g_\theta^1, g_\theta - g_{\theta_0} \rangle_w, \dots, 2\langle g_\theta^k, g_\theta - g_{\theta_0} \rangle_w)'.$$

Note that θ_0 is such a zero. The actual proof follows the line of thought as in theorem 5.21 of Van der Vaart (1998). For the statement of the next theorem, we need a few definitions. Define $\mathbb{H}_{n,\theta_0} : \Theta \subseteq \mathbb{R}^k \to \mathbb{R}^k$ by

$$
\begin{aligned}
\mathbb{H}_{n,\theta_0}(\theta) : \quad &= \quad \sqrt{n}(\Psi_n(\theta) - \Psi_{\theta_0}(\theta)) \\
&= \quad (-2\langle g_\theta^1, \sqrt{n}(\tilde{g}_n - g_{\theta_0}) \rangle_w, \dots, -2\langle g_\theta^k, \sqrt{n}(\tilde{g}_n - g_{\theta_0}) \rangle_w) \\
&= \quad -2(\langle g_\theta^1, U_n^{\theta_0} \rangle_w, \dots, \langle g_\theta^k, U_n^{\theta_0} \rangle_w).
\end{aligned}
$$

Denote the i-th coordinate of $\mathbb{H}_{n,\theta_0}$ by $\mathbb{H}_{n,\theta_0}^i$. We will often write $\mathbb{H}_n(\theta)$ instead of $\mathbb{H}_{n,\theta_0}(\theta)$. In case we do so, we should keep in mind that U_n is always assumed to depend on θ_0 (not on the general θ, which is the argument of $\mathbb{H}_n$). For $\theta \in \Theta$, define $z_\theta : \mathbb{R}^2 \to \mathbb{C}$ by

$$z_\theta(t, x) := \frac{e^{itx}}{\psi_\theta(t)}. \tag{13}$$

Theorem 13 *Assume the mapping $\theta \mapsto \Psi_{\theta_0}(\theta)$ is differentiable at θ_0 with non-singular derivative matrix $\dot{\Psi}_{\theta_0}(\theta_0)$. Assume*

$$\mathbb{H}_n(\hat{\theta}_n) - \mathbb{H}_n(\theta_0) \xrightarrow{\mathrm{P}} 0. \tag{14}$$

15

If $\hat{\theta}_n \xrightarrow{\mathrm{P}} \theta_0$ and $\Psi_n(\hat{\theta}_n) = o_P(n^{-1/2})$ ($\hat{\theta}_n$ is a near zero of Ψ_n), then

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = -\dot{\Psi}_{\theta_0}^{-1} \mathbb{H}_n(\theta_0) + o_P(1), \tag{15}$$

where the sequence $\{\mathbb{H}_n(\theta_0)\}_n$ is tight. Moreover, the sequence $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is asymptotically normal with mean zero and covariance matrix $\dot{\Psi}_{\theta_0}(\theta_0)^{-1} \Sigma^{\theta_0} (\dot{\Psi}_{\theta_0}(\theta_0)^{-1})'$, where

$$
\begin{aligned}
\Sigma^{\theta_0}|_{i,j} &= 4 \int \langle g_{\theta_0}^i, z_{\theta_0}(\cdot, x) \rangle_w \langle g_{\theta_0}^j, z_{\theta_0}(\cdot, x) \rangle_w d\pi_{\theta_0}(x) \\
&\quad - 4 \left(\int \langle g_{\theta_0}^i, z_{\theta_0}(\cdot, x) \rangle_w d\pi_{\theta_0}(x) \right) \left(\int \langle g_{\theta_0}^j, z_{\theta_0}(\cdot, x) \rangle_w d\pi_{\theta_0}(x) \right). \tag{16}
\end{aligned}
$$

Proof First note that

$$
\begin{aligned}
\mathbb{H}_n(\hat{\theta}_n) &= \sqrt{n} \Psi_n(\hat{\theta}_n) + \sqrt{n}(\Psi_{\theta_0}(\theta_0) - \Psi_{\theta_0}(\hat{\theta}_n)) - \sqrt{n} \Psi_{\theta_0}(\theta_0) \\
&= \sqrt{n}((\Psi_{\theta_0}(\theta_0) - \Psi_{\theta_0}(\hat{\theta}_n)) + o_P(1). \tag{17}
\end{aligned}
$$

Since $\theta \mapsto \Psi_{\theta_0}(\theta)$ is differentiable near θ_0, we have for $h \in \mathbb{R}^k$, that $R(h) := \Psi_{\theta_0}(\theta_0 + h) - \Psi_{\theta_0}(\theta_0) - \dot{\Psi}_{\theta_0}(h)$ satisfies $R(h) = o(\|h\|)$, as $h \to 0$. Since $\hat{\theta}_n - \theta_0 = o_P(1)$ this implies (Van der Vaart (1998), lemma 2.12)

$$\sqrt{n}(\Psi_{\theta_0}(\hat{\theta}_n) - \Psi_{\theta_0}(\theta_0)) - \dot{\Psi}_{\theta_0}(\theta_0) \sqrt{n}(\hat{\theta}_n - \theta_0) = \sqrt{n} o_P(\|\hat{\theta}_n - \theta_0\|).$$

Inserting (17) into this expression and using (14) gives

$$\sqrt{n} \dot{\Psi}_{\theta_0}(\theta_0)(\hat{\theta}_n - \theta_0) + \sqrt{n} o_P(\|\hat{\theta}_n - \theta_0\|) = -\mathbb{H}_n(\hat{\theta}_n) + o_P(1) = -\mathbb{H}_n(\theta_0) + o_P(1). \tag{18}$$

If we define for each $\theta \in \Theta$ the mapping $\Lambda_\theta : (C(S), \| \cdot \|_\infty) \to \mathbb{R}^k$ by

$$\Lambda_\theta(f) = (-2 \langle g_\theta^1, f \rangle_w, \ldots, -2 \langle g_\theta^d, f \rangle_w)', \quad f \in C(S),$$

then $\mathbb{H}_n(\theta) = \Lambda_\theta(U_n^{\theta_0})$. By lemma 22 from the appendix, the mapping Λ_θ is continuous. Since $U_n^{\theta_0} \rightsquigarrow U^{\theta_0}$ in $C(S)$ we have by the continuous mapping theorem $\Lambda(U_n^{\theta_0}) \rightsquigarrow \Lambda(U^{\theta_0})$ in $\mathbb{R}^k$, which means that

$$\mathbb{H}_n(\theta) \rightsquigarrow \Lambda_\theta(U^{\theta_0}) = -2(\langle g_\theta^1, U^{\theta_0} \rangle_w, \ldots, \langle g_\theta^k, U^{\theta_0} \rangle_w)' =: \mathbb{H}_{\theta_0}(\theta).$$

16

As a consequence, the sequence $\{\mathbb{H}_n(\theta_0)\}_n$ is tight.

By lemma 24 and assumption 12 we have

$$\int |\langle g_{\theta_0}^i, z_{\theta_0}(\cdot, x)\rangle_w|^2 d\pi_{\theta_0}(x) < \infty, \quad \forall\, 1 \leq i \leq k,$$

which enables us to apply lemma 23. This yields that the distribution of $\mathbb{H}(\theta_0)$ is multivariate normal with mean zero and covariance Σ^{θ_0} as specified in the statement of the theorem.

By invertibility of the matrix $\dot{\Psi}_{\theta_0}$ we obtain from (18)

$$\sqrt{n}\|\hat{\theta}_n - \theta_0\| \leq \|\dot{\Psi}_{\theta_0}^{-1}\|\sqrt{n}\|\hat{\theta}_n - \theta_0\| = O_P(1) + o_P(\sqrt{n}\|\hat{\theta}_n - \theta_0\|).$$

This implies that $\hat{\theta}_n$ is $\sqrt{n}$-consistent. Inserting this in (18), we obtain that $\sqrt{n}\dot{\Psi}_{\theta_0}(\hat{\theta}_n - \theta_0) = -\mathbb{H}_n(\theta_0) + o_P(1)$. Now multiplication of both sides with $\dot{\Psi}_{\theta_0}^{-1}$ gives (15). (The remainder term still converges to zero in probability, since matrix multiplication is a continuous operation.) The final statement of the theorem follows from the weak convergence of $\mathbb{H}_n(\theta_0)$ to $\mathbb{H}(\theta_0)$.

Remark 14 Under regularity conditions,

$$\dot{\Psi}_{\theta_0}(\theta)|_{i,j} = \frac{\partial}{\partial \theta_j} \Psi_{\theta_0}^i(\theta) = 2\langle g_\theta^{ji}, g_\theta - g_{\theta_0}\rangle_w + 2\langle g_\theta^i, g_\theta^j\rangle_w,$$

where $g_\theta^{ji} = \frac{\partial}{\partial \theta_j \partial \theta_i} g_\theta$. In that case $\dot{\Psi}_{\theta_0}(\theta_0)|_{i,j} = 2\langle g_{\theta_0}^i, g_{\theta_0}^j\rangle_w$.

It remains to give conditions under which (14) holds. Since convergence in probability of a vector is equivalent to convergence in probability of each of its components, it suffices to prove $\mathbb{H}_n^i(\hat{\theta}_n) - \mathbb{H}_n^i(\theta_0) \xrightarrow{\text{P}} 0$ for each $i \in \{1, \ldots, k\}$. As seen in the proof of theorem 13

$$\mathbb{H}_n(\theta) \rightsquigarrow \mathbb{H}_{\theta_0}(\theta), \quad \theta \in \Theta, \tag{19}$$

where

$$\mathbb{H}_n^i(\theta) = -2\langle g_\theta^i, U_n^{\theta_0}\rangle_w \qquad \text{and} \qquad \mathbb{H}_{\theta_0}^i(\theta) = -2\langle g_\theta^i, U^{\theta_0}\rangle_w.$$

Just as for $\mathbb{H}_n$ we will drop the θ_0 on $\mathbb{H}_{\theta_0}$, for sake of readability, thus we write $\mathbb{H}$ instead of $\mathbb{H}_{\theta_0}$. The next lemma shows that we can strengthen the convergence in (19) to (14) under mild conditions.

Lemma 15 *Assume Θ is compact. If $\hat{\theta}_n$ is a sequence in Θ such that $\hat{\theta}_n \xrightarrow{\text{P}} \theta_0$, then $\mathbb{H}_n(\hat{\theta}_n) - \mathbb{H}_n(\theta_0) \xrightarrow{\text{P}} 0$.*

Proof To prove $\mathbb{H}_n^i(\hat{\theta}_n) - \mathbb{H}_n^i(\theta_0) \xrightarrow{\text{P}} 0$ $(1 \leq i \leq k)$, we check the conditions of lemma 25 below.

Let (θ_m) be a sequence in Θ converging to θ. By assumption 12*(ii)*, and compactness of S and Θ, dominated convergence gives that $\|g_{\theta_m}^i - g_\theta^i\|_w \to 0$ $(m \to \infty)$. By the Cauchy-Schwartz inequality

$$|\mathbb{H}^i(\theta_m) - \mathbb{H}^i(\theta)| \leq 2\|g_{\theta_m}^i - g_\theta^i\|_w \|U^{\theta_0}\|.$$

The right-hand-side of this expression tends to zero, as $m \to \infty$, since almost all sample paths $t \mapsto U^{\theta_0}(t)$ are continuous on S. This shows continuity of $\theta \mapsto \mathbb{H}^i(\theta)$, a.s.

Let $\|\cdot\|_\Theta$ denote the supremum norm on Θ. Define $\Lambda : C(S, \|\cdot\|_\infty) \to C(\Theta, \|\cdot\|_\Theta)$ by $[\Lambda(f)](\theta) = \langle g_\theta^i, f \rangle_w$. If $f_n \to f$ in $C(S)$, then

$$\|\Lambda(f_n) - \Lambda(f)\|_\Theta = \sup_{\theta \in \Theta} |\langle g_\theta^i, f_n - f \rangle_w| \leq \sup_{\theta \in \Theta} \|g_\theta^i\|_w \|f_n - f\|_w \to 0,$$

which proves continuity of Λ. Since $U_n^{\theta_0} \rightsquigarrow U^{\theta_0}$ in $C(S)$, the continuous mapping theorem implies $\Lambda(U_n^{\theta_0}) \rightsquigarrow \Lambda(U^{\theta_0})$ in $C(\Theta)$. That is, $\mathbb{H}_n^i \rightsquigarrow \mathbb{H}^i$ in $C(\Theta)$. Now apply lemma 25.

Remark 16 The line of proof given in this section holds for more general preliminary estimators than the empirical characteristic function. We now point out at which places in the proof we use that the preliminary estimator is the empirical characteristic function. Firstly, step **(I)** was simplified thanks to theorem 7. Secondly, if we were to use another preliminary estimator than the empirical characteristic function, we would still obtain that $\sqrt{n}(\hat{\theta}_n - \theta_0) = O_P(1)$ (by theorem 13), although it would be much harder to find the precise limiting distribution of this sequence. In particular, the derivation of (16) depends heavily on the expression for Y in terms of a Brownian bridge.

Remark 17 Ideally, we would like to choose the weight-function w such that the asymptotic variance of $\sqrt{n}\hat{\theta}_n$ (as given in theorem 13) is minimal. However, even if the parameter is one-dimensional, this problem is difficult by the complicated form of the asymptotic variance. Moreover, this optimal weight-function will depend on the unknown parameter that needs to be estimated.

We now point out some related estimation procedures that are based on characteristic functions and cumulants, where similar problems occur. The K-L-procedure of Feuerverger & McDunnough (1981) runs as follows:

(i) Choose a set of points $T = \{t_1, \ldots, t_q\}$. Set

$$
\begin{aligned}
z_\theta &= (\Re\psi_\theta(t_1), \ldots, \Re\psi_\theta(t_q), \Im\psi_\theta(t_1), \ldots, \Im\psi_\theta(t_q))' \\
z_n &= (\Re\psi_n(t_1), \ldots, \Re\psi_n(t_q), \Im\psi_n(t_1), \ldots, \Im\psi_n(t_q))',
\end{aligned}
$$

for $\tilde{\psi}_n$ the empirical characteristic function.

(ii) Under conditions, $\sqrt{n}(z_n - z_{\theta_0})$ converges weakly to a Normal distribution with mean zero and covariance-matrix Σ^{θ_0}. Define $\hat{\theta}_n$ as the maximum-likelihood estimator for this Normal distribution, with Σ^{θ_0} replaced by a consistent estimator, say Σ^n. Thus, $\hat{\theta}_n$ is the minimizer of $\theta \mapsto (z_n - z_\theta)'(\Sigma^n)^{-1}(z_n - z_\theta)$.

For the one-dimensional case, Feuerverger & McDunnough (1981) point out that the variance of $\sqrt{n}\hat{\theta}_n$ can be made arbitrarily close to the Cramer-Rao-bound, by choosing the grid points in T sufficiently fine and extended. For the multivariate case, which is considered in Feuerverger & McDunnough (1981a) this is not so clear. Most arguments are given in a heuristic way. If we take a regularly spaced grid, i.e. $t_j = j\tau$ for some $\tau > 0$, it is proposed to take τ as the minimizer of the determinant of Σ^{θ_0}. We doubt whether minimizing the determinant is a sensible criterium.

Knight & Satchell (1997) follow exactly the same approach as in Feuerverger & McDunnough (1981a), but with the characteristic function replaced by the cumulant function. The same remarks hold for the choice of the weights.

For our method, numerical results indicate that a good choice for the weight-function can improve the results obtained so far. In Jongbloed et al. (2005), section 7.1, a data-adaptive choice for w is proposed, which could also be used for the numerical examples in this paper. However, then our asymptotic results are not valid as stated.

To conclude, since there is no universal rule for choosing the grid-points in the K-L-procedure, as well as the weight-function for the CME, a fair numerical comparison of both models is hard.

5 Example: discrete observations from a Gamma-process

In this section we give an example in which we consider existence, consistency and asymptotic behavior of the CME. Furthermore, we explain how the estimator can be approximated numerically.

Suppose we discretely observe a Gamma process. Statistically, this is equivalent to observing a sample $X_1, \ldots, X_n$ with common law $\pi_\theta \sim Gamma(c, \alpha)$, where $\theta = (c, \alpha) \in (0, \infty)^2$. The cumulant of π_θ is given by

$$g_\theta(t) = -c \log(1 - \alpha^{-1} it).$$

The Lévy density is given by $a_\theta(x) = c x^{-1} e^{-\alpha x} \mathbf{1}_{\{x \geq 0\}}$, which is continuous in θ for all $x > 0$. Let $\Theta \subseteq (0, \infty)^2$ be compact and suppose the true value of the parameter θ_0 is in Θ. If we take a sequence of preliminary estimators which satisfies (2), then, by corollary 3, $\hat{\theta}_n = \mathrm{argmin}_{\theta \in \Theta} \Gamma_n(\theta)$ exists. From theorem 5 we obtain consistency: $\hat{\theta}_n \xrightarrow{\text{a.s.}} \theta_0$.

Next, we turn attention to the asymptotics of $\hat{\theta}_n$. For the rest of this example, we use the empirical characteristic function as a preliminary estimator. Clearly, condition (10) is satisfied. Whence the empirical cumulant process U_n converges

weakly. Both partial derivatives of g_θ exist and are given by

$$g_\theta^1(t) = \frac{\partial}{\partial c}g_\theta(t) = -\log(1 - \alpha^{-1}it)$$

$$g_\theta^2(t) = \frac{\partial}{\partial \alpha}g_\theta(t) = -c\frac{-t^2 + i\alpha t}{\alpha(\alpha^2 + t^2)}.$$

Joint continuity of the partial derivatives in (t, c, α) is easily seen. We conclude that assumption (12) is fulfilled. Now, we want to apply theorem 13 to obtain asymptotic normality of $\hat{\theta}_n$.

By existence and continuity of all partial derivatives of $\theta \mapsto g_\theta$ it follows that the mapping $\theta \mapsto \Psi_{\theta_0}(\theta)$ is differentiable near θ_0, for $\theta_0 \in \Theta$. Since also all second order partial derivatives exist and are continuous, we have $\dot{\Psi}_{\theta_0}(\theta_0)|_{i,j} = 2\langle g_{\theta_0}^i, g_{\theta_0}^j\rangle_w$ (see remark 14). Then, we will generally have that the matrix $\dot{\Psi}_{\theta_0}$ is non-singular. By lemma 15, condition (14) is easily satisfied. Thus,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \rightsquigarrow N((0,0)', \dot{\Psi}_{\theta_0}(\theta_0)^{-1}\Sigma^{\theta_0}(\dot{\Psi}_{\theta_0}(\theta_0)^{-1})'), \tag{20}$$

with Σ^{θ_0} as specified in (16).

The estimator can easily be computed numerically. First note the linearity property $g_\theta(\cdot) = g_{(c,\alpha)}(\cdot) = cg_{(1,\alpha)}(\cdot)$. This property makes the numerical optimization problem relatively easy. The objective function can be written as

$$\Gamma_n(\theta) \equiv \Gamma_n(g_\theta) = \|g_\theta - \tilde{g}_n\|_w^2 = \|cg_{(1,\alpha)} - \tilde{g}_n\|_w^2.$$

First we minimize $\Gamma_n(\theta)$ for fixed α with respect to c. This is easy, since $c \mapsto \Gamma_n((c,\alpha))$ is quadratic. Taking the partial derivative with respect to c gives

$$\frac{\partial}{\partial c}\Gamma_n((c,\alpha)) = 2c\|g_{(1,\alpha)}\|_w^2 - 2\langle g_{(1,\alpha)}, \tilde{g}_n\rangle_w$$

and by equating this expression to zero we obtain

$$c_n(\alpha) = \frac{\langle g_{(1,\alpha)}, \tilde{g}_n\rangle_w}{\|g_{(1,\alpha)}\|_w^2}.$$

as a minimizer for fixed α. Now we can minimize $\alpha \mapsto \Gamma_n((c_n(\alpha), \alpha))$ numerically by using a Fibonacci search algorithm. Denote the minimizer by $\hat{\alpha}_n$, then $\hat{c}_n = c_n(\hat{\alpha}_n)$.

21

The asymptotic covariance can be obtained from theorem 13 by numerical integration. We took $w(\cdot) = \mathbf{1}_{[-2,2]}(\cdot)$ and $\theta_0 = (3,2)$. Applying Simpson's rule gives that

$$\dot{\Psi}_{\theta_0}(\theta_0)^{-1}\Sigma^{\theta_0}(\dot{\Psi}_{\theta_0}(\theta_0)^{-1})' \approx \begin{pmatrix} 25.5691 & 17.3076 \\ 17.3076 & 13.6966 \end{pmatrix}. \tag{21}$$

This value can be compared with the asymptotic variance we would find if we would apply maximum likelihood. This value is given by the inverse of the Fisher information matrix, which equals

$$I_{\theta_0}^{-1} = \frac{\alpha_0^2}{c_0\Upsilon(1,c_0) - 1} \begin{pmatrix} c_0/\alpha_0^2 & 1/\alpha_0 \\ 1/\alpha_0 & \Upsilon(1,c_0) \end{pmatrix} \approx \begin{pmatrix} 16.2336 & 10.8224 \\ 10.8224 & 8.5483 \end{pmatrix}. \tag{22}$$

Here Υ denotes the derivative of the logarithm of the Gamma-function. This shows that the CME is less efficient than the MLE. However, there is a gain in robustness, by the use of the empirical characteristic function, see Feuerverger & McDunnough (1981a).

For a sample of size $n = 1000$ we computed both the CME and MLE 20 times. The resulting scatterplot is given in figure 1. By (20), for n large, the CME $\hat{\theta}_n$ has approximately a $N_2(\theta_0, \frac{1}{n}A_{\theta_0})$ distribution, where $A_{\theta_0} := \dot{\Psi}_{\theta_0}^{-1}\Sigma^{\theta_0}(\dot{\Psi}_{\theta_0}^{-1})'$. If we define the ellipse

$$\mathcal{E}_{n,\alpha} := \left\{ \theta \in \mathbb{R}^2 \ : \ (\theta - \theta_0)'(A_{\theta_0})^{-1}(\theta - \theta_0) \leq \chi_{2,\alpha}/n \right\},$$

then, for n large, $P(\hat{\theta}_n \in \mathcal{E}_{n,\alpha}) \approx 1 - \alpha$. Here $\chi_{2,\alpha}$ denotes the upper α-quantile of the χ^2-distribution with two degrees of freedom. In figure 1 we added a contour plot of $\mathcal{E}_{20,0.05}$ for both the CME and the MLE.

Figure 2 shows the estimated canonical function and density in case we have respectively $n = 100$ (dashed curve) and $n = 1000$ (solid curve) observations. The true canonical function and true density are also plotted (dotted line).

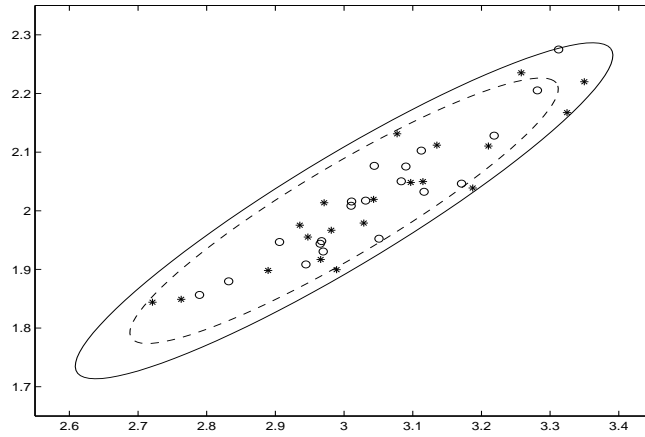For $n = 100$ we computed the CME 100 times. Denote the estimates by $q_{100} =$

Figure 1: Gamma$(3, 2)$ distribution, Scatterplot of parameter estimates for 20 i.i.d. data sets of sample-size $n = 1000$. Starred: CME, circled: MLE. Solid ellipse: 95%-contour plot for the CME. Dashed ellipse: 95%-contour plot for the MLE.
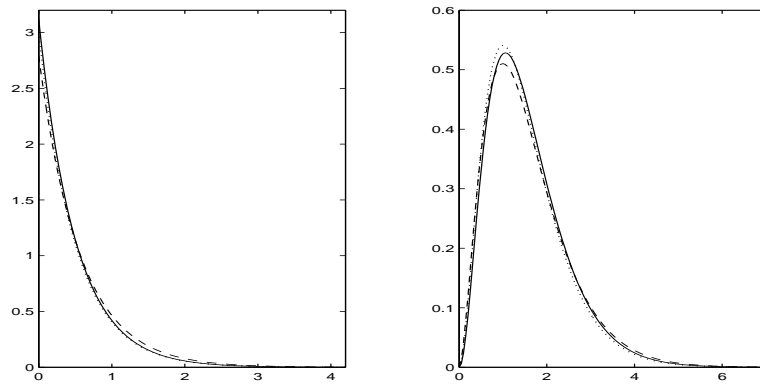

Figure 2: Gamma$(3, 2)$ distribution, Left figure: $x \mapsto xa(x)$. Right figure: density function. Dashed: $n = 100$, solid: $n = 1000$, dotted: true.

$(\hat{\theta}_{100}^{(1)}, \ldots, \hat{\theta}_{100}^{(100)})$, then

$$\text{mean}(q_{100}) = \begin{pmatrix} 3.1228 \\ 2.0836 \end{pmatrix} \quad \text{and} \quad 100 * \text{cov}(q_{100}) = \begin{pmatrix} 23.4063 & 16.1694 \\ 16.1694 & 12.6988 \end{pmatrix}.$$

Similarly, for $n = 1000$ we got

$$\text{mean}(q_{1000}) = \begin{pmatrix} 3.0038 \\ 2.0048 \end{pmatrix} \quad \text{and} \quad 1000 * \text{cov}(q_{1000}) = \begin{pmatrix} 25.6449 & 16.2889 \\ 16.2889 & 12.3042 \end{pmatrix}.$$

6 Parametric estimation for OU-processes driven by a subordinator

Suppose Z is a subordinator without drift. For a given $\lambda > 0$ define the process X as the stationary solution to the stochastic differential equation

$$dX(t) = -\lambda X(t)dt + dZ(\lambda t), \quad t \geq 0. \tag{23}$$

We call X an *Ornstein-Uhlenbeck (OU) process*, driven by Z. The theorem below shows that a stationary OU-process exists if the Lévy measure of the driving subordinator Z satisfies an integrability condition. In that case the marginal distribution of X is infinitely divisible and its Lévy measure can be expressed explicitly by the Lévy measure of the underlying process Z.

Theorem 18 *Suppose Z is an increasing Lévy process with Lévy measure ρ (which is by definition the Lévy measure of $Z(1)$). For $t \geq 0$, let $P_t(x, \cdot)$ denote the transition kernel of the (Markov) process X, when started at x. If ρ satisfies the integrability condition*

$$\int_2^\infty \log x \rho(dx) < \infty,$$

then $P_t(x, \cdot)$ converges weakly to a limit distribution π as $t \to \infty$ for each $x \in (0, \infty)$ and each $\lambda > 0$. Moreover, π is infinitely divisible with Lévy measure $\nu(dx) = x^{-1}k(x)dx$, for k the decreasing function defined by $k(x) = \rho(x, \infty)\mathbf{1}_{(0,\infty)}(x)$. Furthermore, π is the unique invariant probability distribution of X.

24

For a proof, see Sato (1999), theorem 17.5 and corollary 17.9. The theorem in fact implies that π is *self-decomposable* (Sato (1999), section 15), and within this context, k is referred to as the *canonical function*. Note that due to the special scaling in (23), π does not depend on λ.

We assume that ρ, the Lévy measure of Z, is parametrized by $\theta \in \Theta \subseteq \mathbb{R}^k$ and that we wish to estimate the true value of the parameter (θ_0) based on discrete-time observations $X_0, X_\Delta, \ldots, X_{(n-1)\Delta}$ $(\Delta > 0)$ from X, as defined by (23). By theorem 18 this problem is the same as the problem considered so far, except for the fact that the observations are not independent. Nevertheless, the results on existence and consistency of the CME of sections 2 and 3 remain valid. We remark that the empirical characteristic function satisfies (2), since a stationary OU-process is β-mixing (Jongbloed et al. (2005), theorem 3.1). In section 6.1 we consider asymptotic normality for the CME. In section 6.2 we apply the results to an Inverse-Gaussian OU-process.

6.1 Adaptations to the proof of asymptotic normality of the CME

Part **I** of section 4 needs some adaptations, which we shall now work out. We start with a result due to Rio (1998). Denote the L_2-norm with respect to a measure Q by $\|\cdot\|_{2,Q}$. For the definition of an image admissible class we refer to the appendix, definition 26.

Theorem 19 (Rio (2000), theorem 1) *Suppose $(X_n, n \in \mathbb{Z})$ is a stationary time-series with β-mixing coefficients β_n satisfying $\sum_{n>0} \beta_n < \infty$. Let $\mathcal{F}$ be a class of image admissible functions. Suppose*

(i) there exists an envelope function F for the class $\mathcal{F}$ for which

$$\int_0^1 \beta^{-1}(u) Q_F^2(1-u) du < \infty. \tag{24}$$

Here Q_F denotes the quantile function of $|F(X_0)|$ and $\beta^{-1}(u) = \inf\{k \in \mathbb{N} : \beta_k \leq u\}$.

(ii)

$$\int_0^1 \sqrt{H_2^F(\varepsilon, \mathcal{F}) \log(1/\varepsilon)} d\varepsilon < \infty, \tag{25}$$

where

$$H_2^F(\varepsilon, \mathcal{F}) := \log(\sup_Q N(\varepsilon \|F\|_{2,Q}, \mathcal{F}, \|\cdot\|_{2,Q}) \vee 2),$$

and the supremum is taken over all finite discrete measures Q on $\mathbb{R}$ (recall the definition of N given before theorem 7).

If we let P denote the common law of the observations and P_n the empirical measure of the first n observations, then $\sqrt{n}(P_n - P)$ converges weakly in $\ell^\infty(\mathcal{F})$ to a Gaussian process with covariance function Γ.

We remark that a nice expression for the covariance function Γ in the above theorem does not exist.

We verify the conditions of this theorem for $X_n \equiv X_{n\Delta}$ the discretely observed OU-process and $\mathcal{F}$ the class of functions defined by

$$\mathcal{F} := \{f_t \mid f_t(x) = e^{itx}, x \in \mathbb{R}, t \in S\}, \tag{26}$$

Since

$$\ell^\infty(\mathcal{F}) := \{z : \mathcal{F} \to \mathbb{R} \ : \ \|z\|_{\mathcal{F}} := \sup_{f \in \mathcal{F}} |z(f)| = \sup_{t \in S} |z(f_t)| < \infty\},$$

it is natural to identify f_t with $t \in S$ and the space $\ell^\infty(\mathcal{F})$ with $\ell^\infty(S)$.

Corollary 20 *Let $X_0, X_\Delta, \ldots, X_{(n-1)\Delta}$ be observations from the stationary OU-process. If (10) holds, then the stochastic process $(Y_n(t), t \in S)$ converges weakly in the space $\ell^\infty(S)$ to a centered Gaussian process Y.*

Proof First note that $\mathcal{F}$ has envelope function $F \equiv 1$. By lemma 27 from the appendix, $\mathcal{F}$ is image admissible. If (10) holds, then the OU-process is geometrically ergodic (Masuda (2004), theorem 4.3). That is, there exists a constant c such that $\beta_n \le e^{-c\Delta n}$. Therefore,

$$\beta^{-1}(u) \le \inf\{k \in \mathbb{N} \ : \ e^{-c\Delta k} \le u\} = \lceil -\tfrac{1}{c\Delta} \log u \rceil \le -\tfrac{1}{c\Delta} \log u + 1.$$

Furthermore, for all $x \in (0,1)$, $Q_F(1-x) = \inf\{u \; : \; P(1 \leq u) \geq 1 - x\} = 1$. These combined results show that condition (24) is satisfied if (10) holds.

The class $\mathcal{F}$ is Lipschitz in the parameter, in the sense that $|f_t(x) - f_s(x)| \leq d(s,t)R(x)$, with $d(s,t) = |s-t|$ and $R(x) = |x|$. We have

$$N(\varepsilon \|R\|_{2,Q}, \mathcal{F}, \|\cdot\|_{2,Q}) \leq N(\varepsilon, S, d) \sim \frac{1}{\varepsilon}. \tag{27}$$

The inequality follows as in theorem 2.7.11 Van der Vaart & Wellner (1996). The proof is easy: let $t_1, \ldots, t_p$ be an ε-net for d for S, then for each $t \in S$ there is a t_i such that $d(t_i, t) \leq \varepsilon$. Now

$$\|f_t - f_{t_i}\|_{2,Q} \leq d(t, t_i)\|R\|_{2,Q} \leq \varepsilon \|R\|_{2,Q}.$$

Taking the supremum over all finite discrete measures Q in (27) shows that $\sup_Q N(\varepsilon \|R\|_{2,Q}, \mathcal{F}, \|\cdot\|_{2,Q})$ is bounded by a function of order $1/\varepsilon$. Thus (25) is satisfied and the above theorem applies to our case.

Remark 21 To gain some insight to the conditions involved in theorem 19, we compute for a fixed $t \in \mathbb{R}$, $v_t^n := \text{var} Y_n(t)$. Define random variables (W_j) by $W_j = e^{itX_j\Delta} - \mathbb{E}_\theta e^{itX_j\Delta}$. Then

$$v_t^n = \text{var}(\sqrt{n}\,\overline{W}_n) = \sum_{h=-n}^{n}\left(\frac{n-|h|}{n}\right)\gamma_W(h),$$

where $\gamma_W(h) = \text{cov}(W_0, W_h)$ denotes the auto-covariance function of the stationary process (W_j). Therefore, we can bound v_t^n by

$$
\begin{aligned}
v_t^n &\leq \gamma_W(0) + 2\sum_{h=1}^{n}\left(\frac{n-|h|}{n}\right)|\gamma_W(h)| \\
&\leq \gamma_W(0) + 4\int_0^1 \sum_{h=1}^{n}\mathbf{1}_{\{u<\beta_W(h)\}}Q_{|W_0|}^2(1-u)du \\
&\leq \gamma_W(0) + 4\int_0^1 \beta_X^{-1}(u)Q_{|W_0|}^2(1-u)du,
\end{aligned}
$$

where the second inequality follows from theorem 1.1 in Rio (2000). The last inequality follows since the β-mixing numbers of W are smaller than the β-mixing numbers of X.

Thus condition (24) ensures that the variance of the limiting process is finite. Furthermore, the entropy condition measures the size of the class $\mathcal{F}$, which should not be too large.

Next, we go through steps **(I)** up to **(III)** in case we have dependent observations by pointing out where the differences with the case of independent observations occur. For **(I)**, we see that the limit process Y is still Gaussian, but with covariance function Γ, for which we do not have a nice closed form expression. Accordingly, the limit process U in part **(II)** remains a Gaussian process, but with a different (more complicated) covariance structure as for the case with independent observations.

In the proof of theorem (13) we still obtain weak convergence of $\mathbb{H}_n$, but with a different limit $\mathbb{H}$. In this case it is much harder to find the weak limit of $\sqrt{n}(\hat{\theta}_n - \theta_0)$. Despite this, under similar conditions as in the independent observations case, we do get weak convergence of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ to a normal limiting distribution.

6.2 Example: discrete observations from an Inverse Gaussian-OU-process

Suppose we have observations $X_0, \ldots, X_{(n-1)\Delta}$ from a stationary OU-process with Inverse Gaussian marginal law, which we denote by π_θ. Hence, for $\theta = (\delta, \gamma) \in (0, \infty) \times [0, \infty)$, $\pi_\theta \sim IG(\delta, \gamma)$. The density of π_θ is given by

$$f_{(\delta,\gamma)}(x) = \frac{1}{\sqrt{2\pi}} \delta e^{\delta\gamma} x^{-3/2} \exp(-(\delta^2 x^{-1} + \gamma^2 x)/2) \mathbf{1}_{\{x>0\}},$$

and if $\gamma > 0$ the cumulant function is given by

$$g_{(\delta,\gamma)}(t) = \delta\gamma - \delta|\gamma^2 - 2it|^{1/2} \exp(-\tfrac{i}{2}|\arctan(2t/\gamma^2)|).$$

Let $\Theta \subset (0, \infty) \times [0, \infty)$ be compact and suppose the true value of the parameter θ_0 is in Θ. The Lévy density is given by $a_\theta(x) = x^{-1} k_\theta(x)$, where the canonical function is given by $k_\theta(x) = \frac{1}{\sqrt{2\pi}} \delta x^{-1/2} \exp(-\gamma^2 x/2)$. It is easy to see that $\theta \mapsto a_\theta(x)$ is continuous for each $x > 0$. Assume the sequence of preliminary estimators
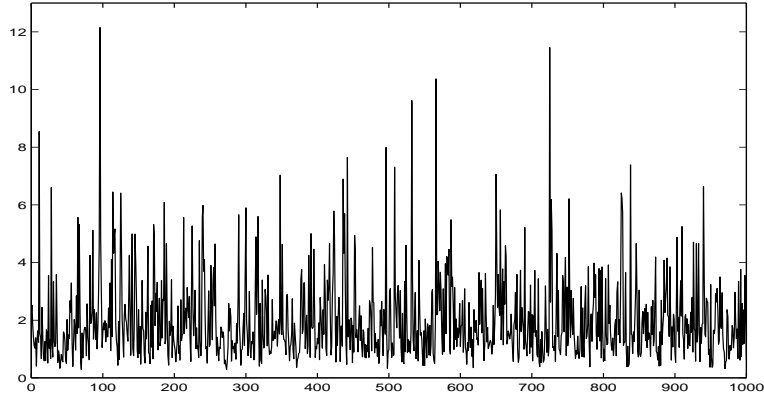
Figure 3: Simulation of a stationary OU-process with Inverse Gaussian$(2, 1)$ marginal law.

satisfies (2). By corollary 3, $\hat{\theta}_n = \mathrm{argmin}_{\theta \in \Theta} \Gamma_n(\theta)$ exists. From theorem 5 we obtain consistency: $\hat{\theta}_n \xrightarrow{\text{a.s.}} \theta_0$. If we take the empirical characteristic function as a preliminary estimator, we can verify the conditions for $\sqrt{n}(\hat{\theta}_n - \theta_0)$ to converge weakly to a 2-dimensional Normal distribution. Since this is basically the same problem as in the Gamma-example of section 5, we omit discussion on this.

The numerical issues are also similar to the Gamma-case. This time we have the linearity property $g_\theta(\cdot) = g_{(\delta,\gamma)}(\cdot) = \delta g_{(1,\alpha)}(\cdot)$. Therefore, the objective function can be written as $\Gamma_n(\theta) = \|\delta g_{(1,\gamma)} - \tilde{g}_n\|_w^2$. Minimizing $\Gamma_n(\theta)$ for fixed γ with respect to δ gives

$$\delta_n(\gamma) = \frac{\langle g_{(1,\gamma)}, \tilde{g}_n \rangle_w}{\|g_{(1,\gamma)}\|_w^2}.$$

Now we can minimize $\gamma \mapsto \Gamma_n(\delta_n(\gamma), \gamma)$ numericallly by using a Fibonacci search algorithm. Denote the minimizer by $\hat{\gamma}_n$, then $\hat{\theta}_n = (\delta_n(\hat{\gamma}_n), \hat{\gamma}_n)$.

Figure 3 shows a simulated path of a stationary Inverse-Gaussian OU-process with $(\delta, \gamma) = (2, 1)$ and intensity parameter $\lambda = 2$ (by this we mean that Z is constructed such that $\pi \sim IG(2, 1)$). We took observations at $t = 0, \ldots, 999$ (i.e. $\Delta = 1$) and computed the CME based on the first 100 and all 1000 observations. We took $w(\cdot) = \mathbf{1}_{[-2,2]}(\cdot)$. Figure 4 shows estimated canonical function and density in both cases. (dashed curve: $n = 100$, solid curve: $n = 1000$). The true canonical
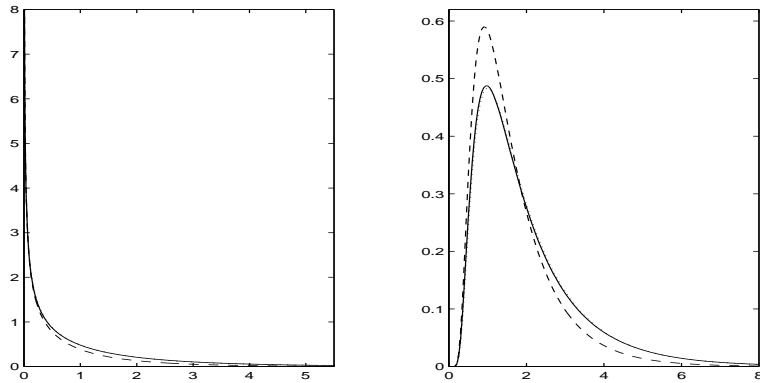
Figure 4: Inverse Gaussian$(2, 1)$ distribution, Data from the OU-process. Left figure: canonical function. Right figure: density function. Dashed: $n = 100$, solid: $n = 1000$, dotted: true.

function and true density are also plotted (dotted line). The estimated parameters corresponding to this figure are

$$\begin{pmatrix} \hat{\delta}_{100} \\ \hat{\gamma}_{100} \end{pmatrix} = \begin{pmatrix} 1.9864 \\ 1.2039 \end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix} \hat{\delta}_{1000} \\ \hat{\gamma}_{1000} \end{pmatrix} = \begin{pmatrix} 1.9600 \\ 0.9877 \end{pmatrix}.$$

So the main reason for the bad fit of the density-function in case $n = 100$ is due to a relatively bad estimation of γ. This has less influence on the fit of the canonical function, since in the expression for k, γ appears only in the exponent.

References

Barndorff-Nielsen, O.E. & Shephard, N. (2001) Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics, *J. Roy. Statist. Ser. B* **63**, 167-241.

Barndorff-Nielsen, O.E. & Shephard, N. (2002) *Normal modified stable processes*, Teor. Ĭmovīr. Mat. Stat. **65** 1-19, 2001. Translation in Theory Probab. Math. Statist. **65**, 1-20.

Chung, K.L. (2001) *A course in probability theory*, Academic Press, 3rd Edition.

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

Feuerverger, A. & McDunnough, P. (1981a) On some Fourier Methods for Inference, *J. Amer. Statist. Assoc.* **76**, 379-387.

Figueroa-López, E. & Houdré, C. (2004) Nonparametric estimation for Lévy processes with a view towards mathematical finance, submitted to *Ann. Statist.*

Giné, E. and Zinn, J. (1986) *Lectures on the central limit theorem for empirical processes* Lecture notes in mathematics **1221**, 50-113.

Jongbloed, G. Meulen, F.H. van der, Vaart & A.W. van der (2004) Nonparametric inference for Lévy driven Ornstein-Uhlenbeck processes. *Bernoulli* **11**, 759-791.

Knight, J.L. & Satchell, S.E. (1997) The cumulant generating function method: implementation and asymptotic efficiency, *Econometric Theory* **13**, 170-184.

Loève, M. (1977) *Probability Theory I*. Springer, 4th edition.

Masuda, H. (2004) On multidimensional Ornstein-Uhlenbeck processes driven by a general Lévy process *Bernoulli* **10**, 97-120.

Papaspiliopoulos, O., Dellaportas, P. and Roberts, G. (2004) Bayesian inference for non-Gaussian Ornstein-Uhlenbeck stochastic volatility processes, *J. Roy. Statist. Ser. B* **66**, 369-393.

Protter, P. (2003) *Stochastic Integration and Differential Equations*, Second edition, Springer.

Remmert, R. (1991) *Theory of Complex Functions*, Springer.

Rio, E. (2000) *Théorie asymptotique des processus aléatoires faiblement dependants*, In: Collection Mathématiques & Applications, vol. **31**, Springer.

Rio, E. (1998) Processus empiriques absolument réguliers et entropie universelle, *Probab. Theory Related Fields* **111**, 585-608.

Sato, K.I. (1999) *Lévy processes and infinitely divisible distributions*. Cambridge University Press.

Vaart, A.W. van der (1998) *Asymptotic Statistics*, Cambridge University Press.

Vaart, A.W. van der & Wellner, J.A. (1996) *Weak Convergence and Empirical Processes*, Springer.

F.H. van der Meulen

Institute for Business and Industrial Statistics of the University of Amsterdam

Plantage Muidergracht 24

1018 TV Amsterdam

The Netherlands

meulen@science.uva.nl

7   Appendix

Proof of lemma 9 Let $\varepsilon > 0$ and $k_\alpha = \mathbb{E}|X|^\alpha$. For each $\alpha \in (0, 1]$ and $x \geq 0$ we have $\sin(x) \leq x^\alpha$. Therefore, $\mathbb{E}\sin^2|\varepsilon x| \leq \mathbb{E}|\varepsilon X|^{2\alpha} = \varepsilon^{2\alpha} k_{2\alpha}$. Taking $\varepsilon = |t - s|/2$ gives

$$\sigma(s, t) \leq 2\left|\frac{t - s}{2}\right|^\alpha \sqrt{k_{2\alpha}} = 2^{1-\alpha}|t - s|^\alpha \sqrt{k_{2\alpha}} =: C_\alpha|t - s|^\alpha.$$

Put $d_\alpha(s, t) = |s - t|^\alpha$, then $N(C_\alpha \varepsilon, \sigma, S) \leq N(\varepsilon, d_\alpha, S) \leq K\varepsilon^{-1/\alpha}$, for some positive constant K, depending on S. Since the right hand side of this display satisfies the entropy condition (6), this suffices.

Lemma 22 *For each $\theta \in \Theta$, the mapping $\Lambda_\theta : (C(S), \|\cdot\|_\infty) \to \mathbb{R}^k$ defined by*

$$\Lambda_\theta(f) = (-2\langle g_\theta^1, f\rangle_w, \ldots, -2\langle g_\theta^k, f\rangle_w)', \quad f \in C(S),$$

is continuous.

Proof By the Cauchy-Schwarz inequality, we have

$$\|\Lambda_\theta(f)\|_2^2 = \sum_{i=1}^k |-2\langle g_\theta^i, f\rangle_w|^2 \leq 4\sum_{i=1}^k \|g_\theta^i\|_w^2\|f\|_w^2.$$

Now if $\|f_n - f\|_\infty \to 0$, then (since $\|f_n - f\|_w^2 \le \|f_n - f\|_\infty^2 \int w(t)dt$) also $\|f_n - f\|_w \to 0$ which implies (by the inequality in the preceding display) that $\Lambda_\theta(f_n - f) \to 0$ in $\mathbb{R}^k$. Finally, by linearity of Λ_θ it follows that $\Lambda_\theta(f_n) \to \Lambda_\theta(f)$.

For the next lemma, recall the definition of B^0 as given in (8). Let U^{θ_0} denote the process from section (4), based on independent observations.

Lemma 23 *(i) Let U^{θ_0} be defined as in corollary 11. For $f \in C(\mathbb{R}, \mathbb{C})$,*

$$\langle f, U^{\theta_0} \rangle_w \overset{a.s.}{=} \int \langle f, z_{\theta_0}(\cdot, x) \rangle_w dB^0(x),$$

where z_{θ_0} is as defined in (13).

(ii) Let $f_i \in C(\mathbb{R}, \mathbb{R})$ $(i = 1, \ldots, k)$ be such that $\int f_i(x)^2 d\pi_{\theta_0}(x) < \infty$. The vector $V = (V_1, \ldots, V_k)' \in \mathbb{R}^k$ with $V_i = \int f_i(x)dB^0(x))$ possesses a $N_k(0, \Xi_{\theta_0})$ distribution, where

$$\Xi^{\theta_0}|_{i,j} = cov(V_i, V_j) = \int f_i(x)f_j(x)d\pi_{\theta_0}(x) - \left(\int f_i(x)\pi_{\theta_0}(x) \right) \left(\int f_j(x)\pi_{\theta_0}(x) \right). \tag{28}$$

Proof The proof of *(i)* is straightforward:

$$
\begin{aligned}
\langle f, U^{\theta_0} \rangle_w &= \Re \int f(t)\overline{U^\theta(t)}w(t)dt = \Re \int f(t) \int \overline{z_{\theta_0}(t, x)}dB^0(x)w(t)dt \\
&\overset{a.s.}{=} \int \Re \int f(t)\overline{z_{\theta_0}(t, x)}w(t)dt dB^0(x) = \int \langle f, z_{\theta_0}(\cdot, x) \rangle_w dB^0(x),
\end{aligned}
$$

where we use the stochastic Fubini theorem (Protter (2003), p. 207) at the third equality sign. The second assertion follows by first noting that for $a \in \mathbb{R}^k$,

$$a'V = \int \left(\sum_{i=1}^k a_i f_i(x) \right) dB^0(x).$$

For $f \in C(\mathbb{R}, \mathbb{R})$ satisfying $\int |f(x)|^2 d\pi_{\theta_0}(x) < \infty$, we have that $\int f(x)dB^0(x)$ is normally distributed with mean zero and variance

$$\int f^2(x)d\pi_{\theta_0}(x) - \left(\int f(x)d\pi_{\theta_0}(x) \right)^2,$$

which follows from elementary rules for Ito-integrals. Therefore, $a'V$ is normally distributed with mean zero and variance

$$\int \left(\sum_{i=1}^{k} a_i f_i(x) \right)^2 d\pi_{\theta_0}(x) - \left(\int \left(\sum_{i=1}^{k} a_i f_i(x) \right) d\pi_{\theta_0}(x) \right)^2 = a' \Xi^{\theta_0} a,$$

with Ξ^{θ_0} as in (28).

Lemma 24 *If $h : \mathbb{R} \to \mathbb{C}$ is continuous, then*

$$\int |\langle h, z_{\theta_0}(\cdot, x) \rangle_w|^2 d\pi_{\theta_0}(x) < \infty,$$

Proof The above expression equals

$$\int \left| \Re \int h(t) \overline{z_{\theta_0}(t,x)} w(t) dt \right|^2 d\pi_{\theta_0}(x) \leq \int \int \left| h(t) \overline{z_{\theta_0}(t,x)} w(t) dt \right|^2 d\pi_{\theta_0}(x)$$

$$\leq \int \int |e^{itx}|^2 d\pi_{\theta_0}(x) |h(t)|^2 \left| \frac{1}{\psi_{\theta_0}(t)} \right|^2 w(t) dt \leq \int |h(t)|^2 \left| \frac{1}{\psi_{\theta_0}(t)} \right|^2 w(t) dt.$$

This is finite, since both h and ψ_{θ_0} are continuous, ψ_{θ_0} is bounded away from zero on S, and S is compact.

The next result is similar to lemma 19.24 in Van der Vaart (1998).

Lemma 25 *Suppose that $H_n \rightsquigarrow H$ in the space $\ell^\infty(\Theta)$. Assume almost all sample paths $\theta \mapsto H(\theta)$ are continuous on Θ. If $\hat{\theta}_n$ is a sequence in Θ such that $\hat{\theta}_n \xrightarrow{\text{P}} \theta_0$, then $H_n(\hat{\theta}_n) - H_n(\theta_0) \xrightarrow{\text{P}} 0$.*

Proof Since $\hat{\theta}_n \xrightarrow{\text{P}} \theta_0$ in Θ and $H_n \rightsquigarrow H$ in $\ell^\infty(\Theta)$ we have $(H_n, \hat{\theta}_n) \rightsquigarrow (H, \theta_0)$ in $\ell^\infty(\Theta) \times \Theta$.

Define a mapping $\varphi : \ell^\infty(\Theta) \times \Theta \to \mathbb{R}$ by $\varphi(z, \theta) = z(\theta) - z(\theta_0)$. The function φ is continuous with respect to the product semimetric at every point (z, θ) such that $\eta \mapsto z(\eta)$ is continuous at θ. If $(z_n, \theta_n) \to (z, \theta)$, then $z_n \to z$ uniformly on Θ and hence $z_n(\theta_n) = z(\theta_n)| + o(1) \to z(\theta)$, if z is continuous at θ.

Since H is assumed to be continuous on Θ a.s., we have by the continuous mapping theorem that $\varphi(H_n, \hat{\theta}_n) \rightsquigarrow \varphi(H, \theta_0)$. This means that $H_n(\hat{\theta}_n) - H_n(\theta_0) \rightsquigarrow H(\theta_0) - H(\theta_0) = 0$. The lemma follows, since convergence in distribution and convergence in probability are the same for a degenerate limit.

Definition 26 *Let $\mathcal{F}$ be a class of functions on a measurable space $(X, \mathcal{B}(X))$. $\mathcal{F}$ is called image-admissible if there exist a locally compact space with countable base Y, with its Borel-σ-algebra $\mathcal{B}(Y)$, and a surjective mapping $T : Y \to \mathcal{F}$, for which*

$$T(y)(x) : (X \times Y, \mathcal{B}(X) \otimes \mathcal{B}(Y)) \to (\mathbb{C}, \mathcal{B}(\mathbb{C}))$$

is measurable.

Lemma 27 *If we take $X = \mathbb{R}$ and $Y = S$ and $T(x,y) = e^{iyx}$, then $\mathcal{F}$, as defined in (26) is image-admissible.*

Proof If we define $\tilde{T} : S \to \mathcal{F}$ by the mapping $x \mapsto [\tilde{T}(y)](x)$, then $\tilde{T}$ is surjective. The mapping $T : (\mathbb{R} \times S, \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(S)) \to (\mathbb{C}, \mathcal{B}(\mathbb{C}))$ defined by $T(x,y) = [\tilde{T}(y)](x)$ is continuous in the product topology and hence measurable with respect to the product σ-algebra $\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(S)$.