Local asymptotic normality for Student-Lévy processes under high-frequency sampling

Till Massing

October 1, 2018

Abstract

There is considerable interest in parameter estimation in Lévy models. The maximum likelihood estimator is widely used because under certain conditions it enjoys asymptotic efficiency properties. The toolkit for Lévy processes is the local asymptotic normality which guarantees these conditions. Although the likelihood function is not known explicitly, we prove local asymptotic normality for the location and scale parameters of the Student-Lévy process assuming high-frequency data. In addition, we propose a numerical method to make maximum likelihood estimates feasible based on the Monte Carlo expectation-maximization algorithm. A simulation study verifies the theoretical results.

Keywords: Lévy process, Student $t$ distribution, high frequency sampling, local asymptotic normality, Monte Carlo expectation-maximization

2010 Mathematics Subject Classification: 60G51, 62E20, 62F12

1 Introduction

There is considerable interest in Lévy processes and parameter estimation in Lévy models, see, e.g., Masuda (2015) and the references therein. However, estimation is difficult because the transition density often is not available in closed form. This paper deals with parameter estimation for the Student-Lévy process ${X_t}_t \geq 0$ such that $X_1$ is Student $t$ distributed given a sample path. Throughout we are interested in estimating the unknown $\theta = (\mu, \sigma)$, where $\mu$ denotes the location and $\sigma$ the scale parameter, while we assume the degree of freedom $\nu > 1$ to be known. The reason for this assumption is discussed in Section 5. The additional estimation of $\nu$ is left for future research. Furthermore, we discuss estimation of $\theta = (\mu, \sigma)$ for the Skew Student-Lévy process with skewness parameter $\beta$.

As the crude method of moment estimator has poor asymptotic efficiency properties, we focus on maximum likelihood (ML) estimation. The maximum likelihood estimator (MLE) requires the density function (or likelihood function) to be known. In the case of the Student-Lévy process, however, we only know the transition density for the 1-increments. For $t \neq 1$, $X_t$ has no closed-form transition density. Thus, maximum likelihood estimation is difficult both theoretically and practically.

The purpose of this paper is, first, to develop asymptotic theory for the MLE in the Student-Lévy model even though the likelihood function is not given explicitly. Second, we propose a time-efficient numerical method in order to make ML estimation feasible.

Let us introduce some notation. Let $(\Omega, \mathcal{F}, P)$ be a probability space. The Student $t$ distribution $t(\nu, \mu, \sigma^2)$ has density function

$$f_{\nu, \mu, \sigma^2}(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\nu\pi\sigma^2}} \left(1 + \frac{1}{\nu} \left(\frac{x - \mu}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}},$$  

(1)

with $\nu$ degrees of freedom, location parameter $\mu \in \mathbb{R}$ and scale parameter $\sigma > 0$. Aas & Haff (2006) introduced the Generalized Hyperbolic (GH) Skew Student $t$ distribution $t(\nu, \mu, \sigma^2, \beta)$ with density

$$f_{\nu, \mu, \sigma^2, \beta}(x) = \frac{2^{\frac{\nu}{2} - 1} \Gamma\left(\frac{\nu+2}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi}} \exp\left(\beta(x - \mu)\right) \left(\frac{\beta^2}{\nu\sigma^2 + (x - \mu)^2}\right)^{-\frac{\nu+2}{2}},$$  

(2)

$^{*}$Faculty of Economics, University of Duisburg-Essen, Universitätsstr. 12, 45117 Essen, Germany.
E-Mail: till.massing@uni-due.de
with \( \nu \) degrees of freedom, location parameter \( \mu \in \mathbb{R} \), scale parameter \( \sigma > 0 \) and skewness parameter \( \beta \in \mathbb{R}[0,1] \). The GH Skew Student \( t \) distribution is obtained as a limiting case of the Generalized Hyperbolic (GH) distribution, see An & Hall (2006) and their references within.

An \( \mathbb{R} \)-valued process \( \{X_t : t \geq 0\} \) is called a Lévy process if \( X_0 = 0 \) a.s., it has independent and stationary increments, it is stochastically continuous and the path function is càdlàg a.s. A Lévy process \( \{X_t \}_{t \geq 0} \) is called Student-Lévy process if \( \mathcal{L}(X_t) = t(\nu,\mu,\sigma^2) \). Let \( \nu \) be the known degree of freedom and \( \theta = (\mu,\sigma) \in \Theta \), where \( \Theta \) is a bounded convex domain such that its closure \( \mathcal{C} \subset \mathbb{R} \times (0,\infty) \). Let \( (P_\theta; \theta \in \Theta) \) be the family of distributions of \( \{X_t\} \) dependent on the unknown parameter \( \theta \). The Radon-Nikodým derivative \( \frac{dP_\theta}{dP_0} \) denotes the likelihood function (which in case of the Student-Lévy process is well-defined). By \( p_\theta(x|\theta) \) we denote the Lebesgue density of \( X_t \) which is always positive and by \( \ell_n(\theta) \) the log-likelihood function. Since the Student-Lévy process is a pure jump process we denote by \( \Delta X_t := X_t - \lim_{s \uparrow t} X_s \) the jump size at time \( t \).

The Student-Lévy \( \{X_t\} \) process can be constructed by subordination. One-dimensional, (a.s.) non-decreasing Lévy processes are called subordinators. Let \( \{Y_t\} \) be an inverse gamma subordinator, meaning that \( \mathcal{L}(Y_t) = RIG(\nu/2,\nu/2) \), where \( RIG(\alpha,\beta) \) denotes the inverse gamma distribution with density function \( \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left( -\frac{\beta}{x} \right) \), with shape parameter \( \alpha > 0 \) and rate parameter \( \beta > 0 \) and \( \Gamma(\alpha) \) the gamma function of \( \alpha \). Then \( X_t = \sigma Y_t + \mu t \), where \( \{Y_t\} \) is a Brownian motion. The Skew Student-Lévy process \( \{X_t\} \) is defined by \( \mathcal{L}(X_t) = t(\nu,\mu,\sigma^2,\beta) \). The Skew Student-Lévy process \( \{X_t\} \) can be constructed by \( X_t = \sigma Y_t + \beta \sigma^2 Y_t + \mu t \) with an inverse gamma subordinator \( \{Y_t\} \) such that \( Y_t \sim RIG(\nu/2,\nu/2) \).

A useful concept for studying asymptotics is local asymptotic normality (LAN) of a family of probability measures. It says that the logarithm of the likelihood ratio behaves asymptotically as a normal random variable. More precisely, we have

**Definition 1.** A sequence of parametric statistical models \( (P^n_\theta; \theta \in \Theta, n \in \mathbb{N}) \) is said to be locally asymptotic normal (LAN) with rate \( A_n \) and Fisher information matrix \( \mathcal{I}(\theta) \), if for each \( u \in \mathbb{R}^p \) and \( \theta_n := \theta + A_n u \in \Theta \)

\[
\log \frac{dP^n_{\theta_n}}{dP^n_\theta} = \ell_n(\theta_n) - \ell_n(\theta) = u^T A_n \nabla \ell_n(\theta) - \frac{1}{2} u^T \mathcal{I}(\theta) u + o_P(1)
\]

holds true under \( P_\theta \), where \( A_n \nabla \ell_n(\theta) \overset{L^2}{\rightarrow} N_p(0, \mathcal{I}(\theta)) \).

The LAN concept implies many useful properties, including the asymptotic normality and asymptotic efficiency of likelihood-based estimation. It was introduced by Le Cam (1960) and since then has been applied in various statistical models. Le Cam & Lo Yang (1990) provided a concise introduction to the topic. Because Lévy processes have a diverse structure, a universal LAN theory is lacking: the very different forms of the likelihood function make analysis difficult, for instance, if the likelihood function \( p_t(x|\theta) \) does not exist in closed form (as is the case for the Student-Lévy process). However, there are some specific cases for which the LAN does exist. Examples for special Lévy models include Masuda (2009b) for the gamma subordinator and the inverse Gaussian subordinator, Kawai & Masuda (2011) for the Meixner Lévy process, and Kawai & Masuda (2013) for the normal inverse Gaussian Lévy process, Kawai (2015) for the variance gamma Lévy process, Aït-Sahalia & Jacod (2008) as well as Masuda (2009a) derived LAN results for non-Gaussian stable Lévy processes. More recently Ivanenko et al. (2015) investigated locally stable Lévy processes, i.e., \( \mathcal{L}(h^{-1}X_h) \) weakly tends to an \( \alpha \)-stable distribution as \( h \to 0 \), which contain the Student-Lévy process as a special case. For more comments on locally stable processes, see below. Masuda (2015) provided an excellent detailed overview and summarized many of the results to be found in the literature.

For the purposes of estimation, it is important to clarify the structure of the available data and the meaning of large sample theory. There are three different senses in which we may sample a path of \( \{X_t\} \):

- **Sampling the path** \( \{X_t\}_{t \in [0,T]} \) in continuous-time. This means that we observe the whole path for any time \( t \in [0,T] \). Here asymptotic theory assumes \( T \to \infty \). In this setting, some parameters may be estimated without error.
- **Sampling \( \{X_t\} \) at discrete and low-frequency time points** \( \{t^n_k\}_{k=0,\ldots,n} \subseteq (0,\infty) \) such that
  
  \[
  0 = t^n_0 < t^n_1 < \cdots < t^n_n := T_n
  \]
  

  2
for each \( n \in \mathbb{N} \) and the sampling intervals \( \Delta_k^n t := t^*_k - t^*_{k-1} \) satisfy
\[
\liminf_{n \to \infty} \min_{1 \leq k \leq n} \Delta_k^n t > 0,
\]
which requires that \( T_n \to \infty \).

- Sampling \( \{X_t\} \) at discrete time points \( \{t^*_k\}_{k=0,\ldots,n} \) but with high-frequency, i.e.,
\[
h_n := \max_{1 \leq k \leq n} \Delta_k^n t \to 0,
\]
as \( n \to \infty \). Here \( T_n \) does not need to tend to infinity and, moreover, may even be fixed as \( T \equiv T_n \).

We mainly consider the case where the step sizes are of equal length \( h_n \equiv \Delta_k^n t \) for each \( 1 \leq k \leq n \).

The main difference between high-frequency and low-frequency sampling is that in the former case the differences between the observation times \( h_n \) become arbitrarily small. For the latter, this is not the case. Here, the endpoint \( T_n \) must tend to infinity. A simple example of sampling at low-frequency is given by the scheme \( t^*_k = k \). This means we sample the 1-increments of the path, which in the case of the Student-Lévy process are Student \( t \) distributed. As \( T_n \to \infty \) we obtain classic asymptotic theory for the estimation of Student \( t \) random variables. For any other low-frequency sampling scheme, the theory becomes more involved; see Remark 3. Most references above mainly under low-frequency sampling (and some under high-frequency sampling) in certain special cases. Woerner (2001) derived some LAN results under low-frequency sampling (and some under high-frequency sampling) in certain special cases.

This paper also mainly focuses on high-frequency sampling. The contribution to the literature is to derive the LAN property for \((\mu, \sigma)\) for high-frequency sampling in the Student-Lévy model. Moreover, we discuss why there is no such LAN result for the Skew Student-Lévy process when jointly estimating \((\mu, \sigma, \beta)\), but only for \((\mu, \sigma)\). Results for the other schemes are relegated to the end of Section 2.

The second contribution of this paper is more practical. As is the case even for the plain Student \( t \) distribution ML estimation becomes numerically feasible using the Expectation-Maximization (EM) algorithm introduced by Dempster et al. (1977). Since Rubin (1983) (see also Little & Rubin (2014)) applied the EM algorithm to the Student \( t \) distribution, many extensions have been developed. For example, Liu & Rubin (1995) described ML estimation of the unknown \( \nu \) by Expectation-Conditional Maximization Either (ECME; see Liu & Rubin (1994)). Nadarajah & Kotz (2008) summarized some of the most important methods for the Student \( t \) distribution. McLachlan & Krishnan (2007) is a standard reference for the EM algorithm.

Returning to the Student-Lévy process, we here propose a Monte Carlo EM (MCEM) algorithm. The MCEM algorithm was initially developed by Wei & Tanner (1990) and replaces one or both of the E- and the M-steps with a Monte Carlo variant. Details are discussed in Section 3 below. We aim to estimate \((\mu, \sigma)\) given a sample with density \( p_t(x|\theta) \) where \( t \neq 1 \) and possibly is smaller.

As already mentioned, we consider \( \nu \) to be known in this paper. Possible extensions are discussed in Section 5.

The remainder of this paper is organized as follows: Section 2 states and proves the LAN result for the Student-Lévy process. Numerical methods such as the MCEM algorithm are discussed in Section 3. In Section 4 we test these methods in Monte Carlo experiments. Section 5 concludes.

## 2 Main results

Under the high-frequency sampling scheme with observation times \( \{t^*_k\} \) and observed points \( \{X_{t^*_k}\} \) we define the \( k \)th increments of \( \{X_t\} \) as
\[
\Delta_k^n X := X_{t^*_k} - X_{t^*_{k-1}}, \quad k = 1, \ldots, n.
\]
\( \Delta_k^n X \) are i.i.d. with density function \( p_{h_n}(x|\theta) \). We define the log-likelihood function by
\[
\ell_n(\theta) := \sum_{k=1}^n \log p_{h_n}(\Delta_k^n X|\theta).
\]
We write
\[
g_{nk}(\theta) := \nabla \log p_{h_n}(\Delta_k^n X|\theta) = \left( \frac{\partial}{\partial \mu} p_{h_n}(\Delta_k^n X|\theta), \frac{\partial}{\partial \sigma} p_{h_n}(\Delta_k^n X|\theta) \right)^T.
\]
We now state the main result for the non-skew Student-Lévy process. The Skew Student-Lévy process will be considered in Corollary 1.

**Theorem 1.** Let \( \{X_t\} \) be a Student-Lévy process such that \( \mathcal{L}(X_1) = t(\nu, \mu, \sigma^2) \) (with known \( \nu > 1 \)). Consider a sample \( (X_{kh_n})_{1 \leq k \leq n} \) with a sequence \( \{h_n\}_{n \in \mathbb{N}} \) of positive step sizes. If \( h_n \to 0 \) as \( n \to \infty \), the LAN property holds true for each \( \theta = (\mu, \sigma) \in \Theta \) with rate

\[
A_n := \text{diag} \left( \frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}} \right)
\]

and Fisher information

\[
\mathcal{I}(\theta) := \begin{pmatrix}
\frac{1}{\nu \sigma^2} & 0 \\
0 & \frac{1}{2 \nu \sigma^2}
\end{pmatrix}.
\]

In particular, \( \mathcal{I}(\theta) \) is positive definite for each \( \theta \in \Theta \) and, moreover, the maximum likelihood estimator \( \hat{\theta} \) exists and is asymptotically normal:

\[
A_n^{-1}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \mathcal{I}(\theta)^{-1}) \quad \text{as} \quad n \to \infty.
\]

The positive definiteness of \( \mathcal{I}(\theta) \) implies that \( \mathcal{I}(\theta)^{-1} \) exists. \( \hat{\theta} \) is asymptotically efficient because it attains the Cramér-Rao bound asymptotically.

**Remark 1.** The time horizon \( T_n \) does not need to tend to infinity. It may possibly be fixed \( T = T_n \). This is in contrast to, e.g., scaled Brownian motions with drift where the maximum likelihood estimator for \( \mu \) is not even consistent if \( T_n \) does not tend to infinity. This can easily be visualized by a short simulation. The LAN for scaled Brownian motions with drift has rate \( (\sqrt{T_n}, \sqrt{n}) \) for \( (\mu, \sigma) \) (see, e.g., Kawai 2013).

**Remark 2.** Although the Student-Lévy process is clearly not stable, it has the same asymptotic Fisher information as a stable Cauchy-Lévy process. (In fact, the Cauchy-Lévy process is a special case of the Student-Lévy process with \( \nu = 1 \). As all increments are Cauchy distributed, ML estimation reduces to the standard i.i.d. Cauchy case (Haas 1966) which we do not consider here.) Masuda (2009a) derived the LAN property for symmetric stable Lévy processes. A Lévy process whose \( 1 \)-increments are Cauchy-Lévy distributed fulfills the LAN with Fisher information (5) and rate (4).

Masuda (2009a) showed that the following conditions of Lemma 1 are sufficient for the LAN to hold true. In the proof of Theorem 1 we will verify that conditions (i) – (iii) are satisfied under the assumptions of Theorem 1.

**Lemma 1.** Assume the following conditions hold true as \( n \to \infty \):

(i) \( n \mathbb{E}_\theta [A_n g_{n_1}(\theta)] A_n] \rightarrow \mathcal{I}(\theta) \),

(ii) \( n \mathbb{E}_\theta [A_n g_{n_1}(\theta)]^2 \to 0 \),

(iii) \( n \left( \sup_{\theta \in \Theta} \mathbb{E}_\theta \left[ |A_n \nabla g_{n_1}(\theta)^T A_n|^2 + |A_n g_{n_1}(\theta)|^2 \right] \right) \to 0 \).

Then the LAN holds true.

Assumption (iii) implies the Lindeberg condition

\[
\sum_{k=1}^n \mathbb{E}_\theta \left[ |A_n g_{nk}(\theta)|^2 ; |A_n g_{nk}(\theta)| > \varepsilon \right] \to 0
\]

for every \( \varepsilon > 0 \). This allows us to apply the central limit theorem.

The following lemmas are needed in order to prove Theorem 1. The first two show the locally stable behavior of inverse gamma subordinator's increments and Student-Lévy increments. The third gives a bound for the density function of the inverse gamma subordinator's increments.
Lemma 2. Let \( \{Y_t\} \) be an inverse gamma subordinator such that \( \mathcal{L}(Y_1) = R\Gamma(\nu/2, \nu/2) \). As \( n \to \infty \)
\[
\frac{Y_{hn}}{h_n^2} \xrightarrow{d} \text{Lévy}(0, \nu),
\]
i.e., the Lévy distribution with density function
\[
\sqrt{\frac{\nu}{2\pi}} e^{-\frac{\nu}{2x^2}} x \in \mathbb{R}.
\]

Proof. Again, as in the proof of Lemma 3, we use convergence of the characteristic function. The characteristic function of \( Y_t \) is given by
\[
\left(\frac{2(-i\frac{n}{2}u)}{\Gamma(\frac{\nu}{2})}K_{\nu/2}\left(\sqrt{-2iu}\right)\right)^t
\]
Hence \( \frac{Y_{hn}}{h_n^2} \) has the characteristic function
\[
\left(\frac{2(-i\frac{n}{2}u)}{h_n^2}\right)^{h_n} K_{\nu/2}\left(\sqrt{-2ih_n^2u}\right)^{h_n}.
\]
Similarly, as in the proof of Lemma 3
\[
K_{\nu/2}\left(\sqrt{-2ih_n^2u}\right)^{h_n} \sim h_n^{\nu} e^{-\sqrt{-2ih_n^2u}} h_n \to e^{-2i\nu u}
\]
for \( n \to \infty \). Of course, \( e^{-2i\nu u} \) is the characteristic function of the Lévy\((0, \nu)\) distribution.

Lemma 3. Let \( \{X_t\} \) be a Student-Lévy process such that \( \mathcal{L}(X_1) = t(\nu, \mu, \sigma^2) \). As \( n \to \infty \)
\[
\frac{X_{hn} - h_n\mu}{h_n\sigma} \xrightarrow{d} \text{Cauchy}(0, \sqrt{\nu}),
\]
i.e., the Cauchy distribution with density function
\[
\frac{1}{\pi \sqrt{\nu} \left(1 + \frac{x^2}{\nu}\right)}, \quad x \in \mathbb{R}.
\]

Proof. The result follows from Lemma 2 since
\[
\frac{X_{hn} - h_n\mu}{h_n\sigma} \xrightarrow{d} B_{h_n^{-2}Y_{hn}}.
\]

Lemma 4. Let \( R\Gamma(y|\alpha, \beta) \) denote the density function for the \( R\Gamma(\alpha, \beta) \) distribution and let \( R\Gamma^{*t}(y|\alpha, \beta) \) denote its \( t \)-fold convolution. Then for \( \nu > 1 \) and any \( t > 0 \) there exists a \( K > 0 \) such that for all \( 0 < y < K \),
\[
R\Gamma^{*t}\left( y \left| \frac{\nu}{2}, \frac{\nu}{2}\right. \right) > R\Gamma\left( y \left| \frac{\nu}{2}, \frac{i^2\nu}{2}\right. \right), \quad t > 1,
\]
\[
R\Gamma^{*t}\left( y \left| \frac{\nu}{2}, \frac{\nu}{2}\right. \right) < R\Gamma\left( y \left| \frac{\nu}{2}, \frac{i^2\nu}{2}\right. \right), \quad t < 1.
\]

Proof. Girón & del Castillo (2001) showed that
\[
R\Gamma^{*2}\left( y \left| \frac{\nu}{2}, \frac{\nu}{2}\right. \right) = \sum_{i=0}^{\nu-1} w_i R\Gamma\left( y \left| \frac{\nu}{2} + i, \frac{2i\nu}{2}\right. \right)
\]
for odd $\nu > 1$ with $w_i \geq 0$ and $\sum_i w_i = 1$. The cumbersome formulas for $w_i$ can be found in \cite{Girón & del Castillo 2001}. For small $y$ it holds true that

$$\sum_{i=0}^{\nu-1} w_i R\Gamma \left( y \left| \nu, 2^2 \nu \right. \right) > R\Gamma \left( y \left| \nu, 2^2 \nu \right. \right),$$

since $R\Gamma \left( y \left| \nu, 2^2 \nu \right. \right) > R\Gamma \left( y \left| \nu, 2^2 \nu \right. \right)$ for all $i \geq 1$. The difference $R\Gamma^{*2} \left( y \left| \nu, \frac{\nu}{2} \right. \right) - R\Gamma \left( y \left| \nu, 2^2 \nu \right. \right)$ is increasing in $\nu$. Hence, by continuity in $\nu$,

$$R\Gamma^{*2} \left( y \left| \nu, \frac{\nu}{2} \right. \right) > R\Gamma \left( y \left| \nu, 2^2 \nu \right. \right),$$

for all $\nu > 1$ and small $y$. By induction,

$$R\Gamma^{*m} \left( y \left| \nu, \frac{\nu}{2} \right. \right) > R\Gamma \left( y \left| \nu, 2^2 \nu \right. \right),$$

for all integers $m \geq 2$, for $y$ small enough.

Obviously, $(R\Gamma^{*m})^{*m} \left( y \left| \nu, \frac{\nu}{2} \right. \right) \equiv R\Gamma \left( y \left| \nu, \frac{\nu}{2} \right. \right)$. By (11),

$$R\Gamma^{*m} \left( y \left| \nu, \frac{\nu}{2}, \frac{\nu}{m^2} \right. \right) > R\Gamma \left( y \left| \nu, \frac{\nu}{2}, \frac{\nu}{m^2} \right. \right) = (R\Gamma^{*m})^{*m} \left( y \left| \nu, \frac{\nu}{2} \right. \right),$$

which implies

$$R\Gamma^{*m} \left( y \left| \nu, \frac{\nu}{2} \right. \right) < R\Gamma \left( y \left| \nu, \frac{\nu}{2}, \frac{\nu}{m^2} \right. \right),$$

for all integer $m \geq 2$, for $y$ small enough. (11) and (12) together with the infinite divisibility of the inverse gamma distribution imply the claim for all $t > 0$. \hfill $\Box$

The next lemma clarifies the asymptotic behavior of the density of Student-Lévy increments.

**Lemma 5** \cite{Berg & Vignat 2008}. Let $p_t(x|\nu, \theta)$ be the transition density of $X_t$, where $\{X_t\}$ is the Student-Lévy process. Then for any $t > 0$, $\nu > 0$ and $\theta \in \mathbb{R} \times (0, \infty)$,

$$p_t(x|\nu, \theta) \sim \frac{C_{\nu, \theta}}{\sigma^{x-h_\theta}} \left( x-\frac{h_\theta}{\sigma} \right)^{-\nu-1},$$

as $|x| \to \infty$, where $C_{\nu}$ is a constant only depending on $\nu$. \hfill (13)

\cite{Berg & Vignat 2008} proved this statement actually for $p_t(x|\nu, 0, 1) \sim \frac{C_{\nu, \theta}}{|x|^{\nu+1}}$. We generalize this using $p_t(x|\nu, \theta) = \frac{1}{\theta} p_t \left( \frac{x-h_\theta}{\sigma} | \nu, 0, 1 \right)$. Using these lemmas we now prove Theorem 1. There are two main ideas in the proof. First, we use the fact that the Student-Lévy process is a subordinated Brownian motion. Second, we apply Monte Carlo integration techniques to treat complicated integrals.

**Proof of Theorem 1.** We prove the theorem by checking the assumptions of Lemma 1. Before that, we prove boundedness in order to be able to apply the bounded convergence theorem. Note that $p_{h_n}(x|\theta) > 0$ for any $x \in \mathbb{R}$, $h_n > 0$, $\theta \in \Theta$, $\nu > 0$. We start with the first entry of $\nabla \log p_{h_n}(X_{h_n}|\theta)$

$$E_\theta \left[ \left( \frac{\partial}{\partial \mu} \log p_{h_n}(X_{h_n}|\theta) \right)^2 \right] = \int_{-\infty}^{\infty} \left( \frac{\partial}{\partial \mu} p_{h_n}(x|\theta) \right)^2 p_{h_n}(x|\theta) dx.$$ 

Observe that

$$p_{h_n}(x|\theta) = \int_0^\infty p_{h_n}(x, y|\theta) dy,$$
where \( p_n(x,y|\theta) \) is the joint density of \( \sigma B_{Y_n} + h_n \mu \), where \( B \) is a Brownian motion. Thus,

\[
p_n(x|\theta) = \int_0^\infty p_n(x,y|\theta)dy = \int_0^\infty N(x|h_n \mu, \sigma^2 y)p_n(y)dy,
\]

where \( N(x|h_n \mu, \sigma^2 y) \) denotes the density of the \( N(h_n \mu, \sigma^2 y) \)-distribution and \( p_n(y) := R \Gamma^{h_n}(y - \frac{\nu}{2}, \frac{\nu}{2}) \) the density of the (unobserved) subordinator \( Y_n \). Next,

\[
\frac{\partial}{\partial \mu} \int_0^\infty N(x|h_n \mu, \sigma^2 y)p_n(y)dy = \int_0^\infty \frac{\partial}{\partial \mu} N(x|h_n \mu, \sigma^2 y)p_n(y)dy,
\]

since \( \left| \frac{\partial}{\partial \mu} N(x|h_n \mu, \sigma^2 y) \right| \leq C_{h_n}/\sigma \) for a constant \( C > 0 \) independent of \( y \) and \( \theta \). But \( \int_0^\infty \frac{1}{\sigma} p_n(y)dy < \infty \) is implied by the uniform integrability, which we show below. \((15)\) can be proven analogously for \( \frac{\partial}{\partial \sigma} \) using that \( \Theta \) is a compact set.

We now show that \( \left| \frac{\partial}{\partial \mu} p_n(x|\theta) \right| \) is uniformly bounded in \( x \) and \( h_n \). This implies

\[
\lim_{n \to \infty} E_\theta \left[ \left( \frac{\partial}{\partial \mu} \log p_n(X_{n|h_n}|\theta) \right)^2 \right] = E_\theta \left[ \lim_{n \to \infty} \left( \frac{\partial}{\partial \mu} \log p_n(X_{n|h_n}|\theta) \right)^2 \right].
\]

Note that \( p_n(x|\theta) \sim \frac{G_{h_n,x}(x)}{\sigma \left( \frac{x-h_n \mu}{\nu} \right)^{\nu}} \), as \( x \to +\infty \) (see Lemma 5). This carries over to \( \frac{\partial}{\partial \sigma} p_n(x|\theta) \sim \frac{\partial}{\partial \sigma} \left( \frac{G_{h_n,x}(x)}{\left( \frac{x-h_n \mu}{\nu} \right)^{\nu}} \right) \) by applying \((15)\) in the proof of Theorem 2 of Berg & Vignat (2008). Hence it holds true that

\[
\frac{\partial}{\partial \mu} p_n(x|\theta) \sim \frac{\partial}{\partial \mu} \frac{G_{h_n,x}(x)}{\left( \frac{x-h_n \mu}{\nu} \right)^{\nu}} \to 0,
\]

as \( x \to +\infty \), and analogously for \( x \to -\infty \) which implies that \( \left| \frac{\partial}{\partial \mu} p_n(x|\theta) \right| \) is bounded in \( x \).

Furthermore,

\[
\frac{\partial}{\partial \mu} \frac{p_n(x|\theta)}{p_n(x|\theta)} \sim \frac{\sigma}{\left( \frac{x-h_n \mu}{\nu} \right)^{\nu}} =: C_n(x) \to 0
\]

as \( n \to \infty \), for any \( x \), see (i) below. Moreover, \( \sup_x |C_n(x)| = \frac{1}{\sqrt{2}\sigma} \) implies uniform convergence. By the continuity of \( \frac{\partial}{\partial \mu} p_n(x|\theta) \) in \((x,h_n)\) there exists a constant \( C > 0 \) such that

\[
\left| \frac{\partial}{\partial \mu} p_n(x|\theta) \right| < C,
\]

for all \( x \in \mathbb{R} \), and all \( h_n \in (0, \infty) \). The proof of the boundedness of \( \frac{\partial}{\partial \sigma} \log p_n(x|\theta) \) works in a very similar fashion.

Next, we prove that \( h_n N(x|h_n \mu, \sigma^2 Y_n) \), \( h_n \frac{\partial}{\partial \mu} N(x|h_n \mu, \sigma^2 Y_n) \) and \( h_n \frac{\partial}{\partial \sigma} N(x|h_n \mu, \sigma^2 Y_n) \) are uniformly integrable for \( n \in \mathbb{N} \). This implies that

\[
\lim_{n \to \infty} \int_0^\infty h_n N(X_{n|h_n}|\mu, \sigma^2 y)p_n(y)dy = \int_0^\infty \lim_{n \to \infty} h_n N(X_{n|h_n}|\mu, \sigma^2 y)p_n(y)dy,
\]

or, if replacing integrals with its Monte Carlo estimators,

\[
\lim_{n \to \infty} \lim_{B \to \infty} \frac{1}{B} \sum_{b=1}^B h_n N(X_{b|h_n}|\mu, \sigma^2 Y_{b|h_n}) = \lim_{B \to \infty} \frac{1}{B} \sum_{b=1}^B \lim_{n \to \infty} h_n N(X_{b|h_n}|\mu, \sigma^2 Y_{b|h_n}),
\]

and analogously for the integrals containing \( \frac{\partial}{\partial \mu} \) and \( \frac{\partial}{\partial \sigma} \). Uniform integrability can be proven using uniform integrability test functions \((16)\). If for a \( \varphi: [0, \infty) \to [0, \infty) \) with \( \lim_{z \to \infty} \varphi(z) = \infty \) it
holds true that $\sup_{\omega \in \mathcal{E}} E[\varphi((h_n X_{\omega})(h_n \mu, \sigma^2 Y_{\omega}))] < \infty$, then we have uniform integrability. We choose $\varphi(z) = z^2$. Since $p_{h_n}(y)$ is not available in closed form, we use two approximations. The first approximation $p_{h_n}(y) := R(\nu/2, h_n^2 \nu/2)$ is motivated by the fact that for $\nu = 1$, $p_{h_n}(y) = p_{\nu_n}^*(y)$ for all $y > 0$, $h_n > 0$. Additionally, the approximation $p_{h_n}(y)$ is chosen because $p_{h_n}(y) < p_{\nu_n}^*(y)$ for small $y > 0$ and $h_n < 1$ (Lemma 5). The second approximation makes use of Lemma 2 in Massing (2018), which states that the Lévy measure of the inverse gamma subordinator is bounded from above by the explosive Lévy measure of the $1/2$-stable Lévy subordinator with density function $p_{\theta_n}(y) := R(\nu/1, h_n^2 \nu/2)$ for every $h_n$ if $\nu > 1$. This implies that $p_{h_n}(y) < p_{\nu_n}^*(y)$ for large values of $y$. Hence for all $n$ there exist $K_n^{(1)}, K_n^{(2)} > 0$ such that

$$
\int_0^\infty h_n^2 N(X_{h_n} | h_n \mu, \sigma^2 y)^2 p_{h_n}(y)dy \\
= \int_0^{K_n^{(1)}} h_n^2 N(X_{h_n} | h_n \mu, \sigma^2 y)^2 p_{h_n}(y)dy + \int_{K_n^{(1)}}^{K_n^{(2)}} h_n^2 N(X_{h_n} | h_n \mu, \sigma^2 y)^2 p_{h_n}(y)dy \\
+ \int_{K_n^{(2)}}^\infty h_n^2 N(X_{h_n} | h_n \mu, \sigma^2 y)^2 p_{h_n}(y)dy \\
\leq C \left( \int_0^{K_n^{(1)}} h_n^2 \bar{p}_{h_n}(y)dy + \int_{K_n^{(1)}}^{K_n^{(2)}} h_n^2 \bar{p}_{h_n}(y)dy + \int_{K_n^{(2)}}^\infty h_n^2 \bar{p}_{h_n}(y)dy \right) \\
\leq C \left( \int_0^{\infty} h_n^2 \bar{p}_{h_n}(y)dy + (K_n^{(2)} - K_n^{(1)}) \frac{h_n^2}{K_n^{(1)}} p_{h_n}(K_n^{(1)}) + \int_{0}^{\infty} \frac{h_n^2}{y} \bar{p}_{h_n}(y)dy \right) \\
= C \left( 1 + (K_n^{(2)} - K_n^{(1)}) \frac{h_n^2}{K_n^{(1)}} p_{h_n}(K_n^{(1)}) + \frac{1}{\nu} \right),
$$

where $C > 0$ is a finite constant independent of $y$ and $n$ which may vary from line to line. $(K_n^{(2)} - K_n^{(1)}) \frac{h_n^2}{K_n^{(1)}} p_{h_n}(K_n^{(1)})$ converges to zero for $n \to \infty$ since $K_n^{(1)}, K_n^{(2)} \to 0$ and $\frac{h_n^2}{K_n^{(1)}} = \mathcal{O}(1)$. This implies uniform integrability for $h_n N(X_{h_n} | h_n \mu, \sigma^2 Y_{h_n})$.

We also show uniform integrability for $\frac{\partial}{\partial \mu} h_n N(X_{h_n} | h_n \mu, \sigma^2 y)$ by

$$
\int_0^\infty h_n^2 \left( \frac{\partial}{\partial \mu} N(X_{h_n} | h_n \mu, \sigma^2 y) \right)^2 p_{h_n}(y)dy \\
\leq C \left( \int_0^{\infty} h_n^2 (h_n - h_n \mu)^2 \bar{p}_{h_n}(y)dy + (K_n^{(2)} - K_n^{(1)}) \frac{h_n^2 (h_n - h_n \mu)^2}{(K_n^{(1)})^3} p_{h_n}(K_n^{(1)}) \\
+ \int_{K_n^{(1)}}^{K_n^{(2)}} h_n^2 (h_n - h_n \mu)^2 \bar{p}_{h_n}(y)dy \right) \\
\leq C \left( \frac{(X_{h_n} - h_n \mu)^2}{h^2} + (K_n^{(2)} - K_n^{(1)}) \frac{h_n^2 (h_n - h_n \mu)^2}{(K_n^{(1)})^3} p_{h_n}(K_n^{(1)}) + \frac{(X_{h_n} - h_n \mu)^2}{h^2} \right),
$$

and by Lemma 5 $(X_{h_n} - h_n \mu)^2 \overset{d}{=} \tilde{X}^2$, $\tilde{X} \sim \text{Cauchy}(0, \sqrt{\sigma})$ and $\tilde{X}^2$ is a.s. finite. Uniform integrability for $\frac{\partial}{\partial \mu} h_n N(X_{h_n} | h_n \mu, \sigma^2 y)$ can be proven analogously.

We are now able to check the assumptions (i)–(iii) of Lemma 1.

(i) We here derive the limiting form (5) of the Fisher information matrix. The expression (14) is crucial and is used multiple times subsequently. We begin with the first entry $\mathcal{I}^{(1)}(\theta)$.

$$
\lim_{n \to \infty} n E_\theta \left[ \left( A^{(1)}_{n} g_{h_n}(\theta) \right)^2 \right] = \lim_{n \to \infty} E_\theta \left[ \left( \frac{\partial}{\partial \mu} \log p_{h_n}(\Delta_{n} X | \theta) \right)^2 \right] = \lim_{n \to \infty} E_\theta \left[ \left( \frac{\partial}{\partial \mu} p_{h_n}(X_{h_n} | \theta) \right)^2 \right]. \tag{18}
$$
We again make use of the fact that the Student-Lévy density can be expressed as (14). Hence, we can write (18) as

\[
\lim_{n \to \infty} E\theta \left[ \frac{\left( \frac{\partial}{\partial \mu} \int_0^\infty N(X_{h_n} | h_n \mu, \sigma^2 y) p_{h_n}(y) dy \right)^2}{\left( \int_0^\infty N(X_{h_n} | h_n \mu, \sigma^2 y) p_{h_n}(y) dy \right)^2} \right] = \lim_{n \to \infty} E\theta \left[ \frac{\left( \int_0^\infty \frac{\partial}{\partial \mu} N(X_{h_n} | h_n \mu, \sigma^2 y) p_{h_n}(y) dy \right)^2}{\left( \int_0^\infty N(X_{h_n} | h_n \mu, \sigma^2 y) p_{h_n}(y) dy \right)^2} \right].
\]

(19)

The bounded convergence theorem implies that (19) equals

\[
E\theta \left[ \lim_{n \to \infty} h_n \left( \int_0^\infty \frac{\partial}{\partial \mu} N(X_{h_n} | h_n \mu, \sigma^2 y) p_{h_n}(y) dy \right)^2 \right] = E\theta \left[ \lim_{n \to \infty} h_n \left( \int_0^\infty \frac{\partial}{\partial \mu} N(X_{h_n} | h_n \mu, \sigma^2 y) p_{h_n}(y) dy \right)^2 \right].
\]

(20)

Since the density \( p_{h_n}(y) \) of \( Y_{h_n} \) is unknown, we cannot compute the inner integrals of (20) directly. Therefore, we use the approach of Monte Carlo integration. Let \( \{Y_{h_{n,1}}, \ldots, Y_{h_{n,B}}\} \) be independent inverse gamma subordinators, each independent of \( (X_{h_{k,n}})_{1 \leq k \leq n} \), such that \( Y_{h_{n,b}} \) has density function \( p_{h_n}(y) \). (See Massing [2018] for more information on simulation of the inverse gamma subordinator.) Then, a.s.,

\[
\lim_{n \to \infty} \int_0^\infty \frac{\partial}{\partial \mu} h_n N(X_{h_n} | h_n \mu, \sigma^2 y) p_{h_n}(y) dy = \lim_{n \to \infty} \lim_{B \to \infty} \frac{1}{B} \sum_{b=1}^B \frac{\partial}{\partial \mu} h_n N(X_{h_n} | h_n \mu, \sigma^2 Y_{h_{n,b}})
\]

and

\[
\lim_{n \to \infty} \int_0^\infty h_n N(X_{h_n} | h_n \mu, \sigma^2 y) p_{h_n}(y) dy = \lim_{n \to \infty} \lim_{B \to \infty} \frac{1}{B} \sum_{b=1}^B h_n N(X_{h_n} | h_n \mu, \sigma^2 Y_{h_{n,b}}).
\]

By uniform integrability of \( h_n N(X_{h_n} | h_n \mu, \sigma^2 y) \) and \( \frac{\partial}{\partial \mu} h_n N(X_{h_n} | h_n \mu, \sigma^2 y) \) in \( y \) and \( n \), (20) is equal to

\[
E\theta \left[ \left( \lim_{B \to \infty} \frac{1}{B} \sum_{b=1}^B \lim_{n \to \infty} \frac{\partial}{\partial \mu} h_n N(X_{h_n} | h_n \mu, \sigma^2 Y_{h_{n,b}}) \right)^2 \right] = E\theta \left[ \left( \lim_{B \to \infty} \frac{1}{B} \sum_{b=1}^B \lim_{n \to \infty} h_n N(X_{h_n} | h_n \mu, \sigma^2 Y_{h_{n,b}}) \right)^2 \right].
\]

(21)

Now,

\[
N(X_{h_n} | h_n \mu, \sigma^2 Y_{h_{n,b}}) = \frac{1}{\sqrt{2\pi\sigma^2 Y_{h_{n,b}}}} \exp \left( -\frac{(X_{h_n} - h_n \mu)^2}{2\sigma^2 Y_{h_{n,b}}} \right) = \frac{1}{h_n} \frac{1}{\sqrt{2\pi\sigma^2 Y_{h_{n,b}}/h_n^2}} \exp \left( -\frac{(X_{h_n} - h_n \mu)^2}{2\sigma^2 Y_{h_{n,b}}/h_n^2} \right).
\]
By Lemmas 3 and 4 we know that \( \frac{X_b - h_n \mu}{h_n \sigma} \xrightarrow{\text{d}} X \sim \text{Cauchy}(0, \sqrt{\nu}) \) and \( \frac{Y_n b}{h_n} \xrightarrow{\text{d}} Y_b \sim \text{Lévy}(0, \nu) \) for any \( b \). Then,

\[
1 \sqrt{2 \pi \sigma^2 \frac{Y_n b}{h_n}} \exp \left( -\frac{(X_n - h_n \mu)^2}{2 \sigma^2 h_n^2 \frac{Y_n b}{h_n}} \right) \xrightarrow{\text{d}} 1 \sqrt{2 \pi \sigma^2 Y_b} \exp \left( -\frac{\hat{X}^2}{2 Y_b} \right),
\]

as \( n \to \infty \). Analogously,

\[
\frac{\partial}{\partial \mu} N(X_n | h_n \mu, \sigma^2 Y_n b) = \frac{1}{\sqrt{2 \pi \sigma^2 Y_n b}} \exp \left( -\frac{(X_n - h_n \mu)^2}{2 \sigma^2 Y_n b} \right) \frac{1}{\sigma} \frac{1}{\sigma h_n Y_n b h_n} \xrightarrow{\text{d}} \frac{1}{\sqrt{2 \pi \sigma^2 Y_b}} \exp \left( -\frac{\hat{X}^2}{2 Y_b} \right) \frac{\hat{X}}{\sigma Y_b}.
\]

Using this expression in (21) (note that the 1/h_n factors cancel out), (21) equals

\[
E_\theta \left[ \left( \lim_{B \to \infty} \frac{1}{B} \sum_{b=1}^{B} \lim_{n \to \infty} \frac{\partial}{\partial \mu} h_n N(X_n | h_n \mu, \sigma^2 Y_n b) \right)^2 \right]
= E_\theta \left[ \left( \lim_{B \to \infty} \frac{1}{B} \sum_{b=1}^{B} \lim_{n \to \infty} h_n N(X_n | h_n \mu, \sigma^2 Y_n b) \right)^2 \right]
= E_\theta \left[ \left( \int_0^{\infty} \frac{1}{\sqrt{2 \pi \sigma^2 y}} \exp \left( -\frac{\hat{X}^2}{2 y} \right) \frac{\hat{X}}{\sigma^2} p_Y(y) dy \right)^2 \right]
\]

by reversing the Monte Carlo integration argument. \( p_Y(y) \) denotes the density function of \( Y \sim \text{Lévy}(0, \nu) \). The inner integrals of (22) can be computed explicitly, that is

\[
E_\theta \left[ \left( \frac{2 \sqrt{\pi} \sqrt{\frac{\sigma^2}{\pi \sigma^2 + \nu}}} \pi \sigma^2 \left( x^2 + \nu \right) \right)^2 \right] = \int_{-\infty}^{\infty} \frac{4 x^2}{\sigma^2 \left( x^2 + \nu \right)} p_\hat{X}(x) dx = \frac{1}{2 \sigma^2},
\]

where \( p_\hat{X}(x) \) is the density function of \( \frac{X}{\sqrt{\nu}} \) distribution.

We continue with the computation of \( \mathcal{S}^{(22)}(\theta) \). Analogously to the computation above,

\[
\lim_{n \to \infty} n E_\theta \left[ \left( A^{(22)}_n \sigma_n^2 \left( \theta \right) \right)^2 \right]
= \lim_{n \to \infty} E_\theta \left[ \left( \frac{\partial}{\partial \sigma} p_{h_n}(X_n | \theta) \right)^2 \right]
= E_\theta \left[ \left( \lim_{B \to \infty} \frac{1}{B} \sum_{b=1}^{B} \lim_{n \to \infty} \frac{\partial}{\partial \sigma} h_n N(X_n | h_n \mu, \sigma^2 Y_n b) \right)^2 \right].
\]

(23)
Then,

\[
\frac{\partial}{\partial \sigma} \left[ N(X_{n_n} | h_n \mu, \sigma^2 Y_{h_n b}) \right]
= \frac{1}{\sqrt{2 \pi \sigma^2 Y_{h_n b}}} \exp \left( -\frac{(X_{h_n} - h_n \mu)^2}{2 \sigma^2 Y_{h_n b}} \right) \left( \frac{1}{\sigma^2} \right)
\]

\[
- \frac{1}{\sqrt{2 \pi \sigma^2 Y_{h_n b}}} \exp \left( -\frac{(X_{h_n} - h_n \mu)^2}{2 \sigma^2 Y_{h_n b}} \right) \left( \frac{1}{\sigma^2} \right)
\]

\[
= \frac{1}{h_n \sigma^2} \frac{1}{\sqrt{2 \pi \sigma^2 Y_{h_n b}}} \exp \left( -\frac{(X_{h_n} - h_n \mu)^2}{2 \sigma^2 h_n^2 Y_{h_n b}} \right) \left( \frac{1}{\sigma^2 h_n^2 Y_{h_n b}} \right)
\]

\[
= \frac{1}{\sigma^2} \frac{1}{\sqrt{2 \pi}} \exp \left( -\frac{\bar{X}^2}{2 \bar{Y}_b} \right) \left( \frac{\bar{X}^2}{\bar{Y}_b} - 1 \right)
\]

as \( n \to \infty \). Therefore, \( \left[ 22 \right] \) equals

\[
E_\theta \left[ \left( \int_0^\infty \frac{1}{\sqrt{2 \pi y}} \exp \left( -\frac{\bar{X}^2}{2 y} \right) \left( \frac{\bar{X}^2}{y} - 1 \right) p_Y(y) dy \right)^2 \right]
\]

\[
= E_\theta \left[ \left( \frac{\bar{X}}{\sqrt{\pi \sigma}} \right)^2 \left( \frac{\bar{X}}{\sigma (X^2 + \nu)} \right)^2 \right]
\]

\[
= \frac{1}{2 \sigma^2}.
\]

For the off-diagonal elements \( \mathcal{g}^{(12)}(\theta) = \mathcal{g}^{(21)}(\theta) \) observe that

\[
\lim_{n \to \infty} n E_\theta \left[ \left( A_n^{(11)} g_n^{(1)}(\theta) \right) \left( A_n^{(22)} g_n^{(2)}(\theta) \right) \right]
\]

\[
= E_\theta \left[ \left( \frac{2 \sqrt{\bar{X}}}{\pi \sigma \sqrt{X^2 + \nu}} \right) \left( \frac{\bar{X}^2 - \nu \sqrt{\nu}}{\pi \sigma (X^2 + \nu)^2} \right) \right] = 0.
\]

(ii) Use

\[
\lim_{n \to \infty} E_\theta \left[ \frac{\partial}{\partial \theta} p_{h_n}(X_{h_n} | \theta) \right] = E_\theta \left[ \frac{2 \bar{X}}{\sigma (X^2 + \nu)} \right] = 0,
\]

and

\[
\lim_{n \to \infty} E_\theta \left[ \frac{\partial}{\partial x} p_{h_n}(X_{h_n} | \theta) \right] = E_\theta \left[ \frac{\bar{X}^2 - \nu \sqrt{\nu}}{\sigma (X^2 + \nu)} \right] = 0,
\]

which implies (ii).

(iii) We continue with verifying the Lindeberg condition. First,

\[
n E_\theta \left[ \left( A_n^{(11)} g_n^{(1)}(\theta) \right)^4 \right] \sim \frac{1}{n} E_\theta \left[ \frac{16 \bar{X}^4}{\sigma^4 (X + \nu)^4} \right] = \frac{1}{n} \frac{3}{8 \nu^2 \sigma^4},
\]
\[
\frac{1}{n} E_{\theta} \left[ \left( A_n^{(22)} g_{n1}^{(2)}(\theta) \right)^4 \right] \sim \frac{1}{n} E_{\theta} \left[ \frac{(\bar{X} - \nu)^4}{\sigma^4(X^2 + \nu)^4} \right] = \frac{1}{8} \frac{3}{n},
\]

which both converge to zero for \( n \to \infty \). The compactness of \( \Theta \) implies that

\[
\lim_{n \to \infty} \sup_{\theta \in \Theta} n E_{\theta} \left[ |A_n g_{n1}(\theta)|^4 \right] = 0.
\]

Second, we have a look at the entries of the Hessian matrix \( \nabla (g_{n1}(\theta)^T) \). Note that \( h_n \frac{\partial^2}{\partial \mu^2} N(X, h_n \mu, \sigma^2 Y_{h_n b}) \), 
\( h_n \frac{\partial^2}{\partial \sigma^2} N(X, h_n \mu, \sigma^2 Y_{h_n b}) \) and \( h_n \frac{\partial^2}{\partial \mu \partial \sigma} N(X, h_n \mu, \sigma^2 Y_{h_n b}) \) are uniformly integrable, analogously to the terms above, thus

\[
h_n \frac{\partial^2}{\partial \mu^2} N(X, h_n \mu, \sigma^2 Y_{h_n b}) 
= \frac{1}{\sqrt{2\pi\sigma^2 Y_b}} \exp \left( -\frac{\bar{X}^2}{2\sigma^2 Y_b} \right) \left( \frac{\bar{X}^2}{\sigma^2 Y_b} - \frac{1}{\sigma^2 Y_b} \right),
\]

and

\[
h_n \frac{\partial^2}{\partial \sigma^2} N(X, h_n \mu, \sigma^2 Y_{h_n b}) 
= \frac{1}{\sqrt{2\pi\sigma^2 Y_b}} \exp \left( -\frac{\bar{X}^2}{2\sigma^2 Y_b} \right) \left( -\frac{2\bar{X}^2}{\sigma^2 Y_b} + \frac{2\bar{X}^4}{\sigma^4 Y_b} - \frac{3\bar{X}^2}{\sigma^2 Y_b} \right),
\]

and cross partial derivative

\[
h_n \frac{\partial^2}{\partial \mu \partial \sigma} N(X, h_n \mu, \sigma^2 Y_{h_n b}) 
= \frac{1}{\sqrt{2\pi\sigma^2 Y_b}} \exp \left( -\frac{\bar{X}^2}{2\sigma^2 Y_b} \right) \left( \frac{\bar{X}^3}{\sigma^2 Y_b} - \frac{3\bar{X}}{\sigma^2 Y_b} \right).
\]

Hence, by repeating the Monte Carlo integration argument,

\[
n E_{\theta} \left[ A_n^{(11)} \frac{\partial^2}{\partial \mu^2} \log p_{h_n}(X, h_n \mu) A_n^{(11)} \right] 
= \frac{1}{n} E_{\theta} \left[ \frac{\partial}{\partial \mu^2} p_{h_n}(X, h_n \mu) - \left( \frac{\partial}{\partial \mu} p_{h_n}(X, h_n \mu) \right)^2 \right] 
\sim \frac{1}{n} E_{\theta} \left[ \frac{\partial^2}{\partial \mu^2} \log p_{h_n}(X, h_n \mu) A_n^{(11)} \right] 
= \frac{1}{n} E_{\theta} \left[ \frac{\partial^2}{\partial \mu^2} \log p_{h_n}(X, h_n \mu) \right] 
= \frac{7}{n 8\nu^2 \sigma^4},
\]

and

\[
n E_{\theta} \left[ A_n^{(22)} \frac{\partial^2}{\partial \sigma^2} \log p_{h_n}(X, h_n \mu) A_n^{(22)} \right] 
= \frac{1}{n} E_{\theta} \left[ \frac{\partial}{\partial \sigma^2} p_{h_n}(X, h_n \mu) - \left( \frac{\partial}{\partial \sigma} p_{h_n}(X, h_n \mu) \right)^2 \right] 
\sim \frac{1}{n} E_{\theta} \left[ \frac{\partial^2}{\partial \sigma^2} \log p_{h_n}(X, h_n \mu) A_n^{(22)} \right] 
= \frac{1}{n} E_{\theta} \left[ \frac{\partial^2}{\partial \sigma^2} \log p_{h_n}(X, h_n \mu) \right] 
= \frac{4\bar{X}^2}{\sigma^2 (X^2 + \nu)^2}.
\]
The LAN converging to zero as does not hold true, since the Fisher information result because in this case the sampling interval

\[ \frac{n}{\sqrt{\pi \sigma}} \]

The proposition requires that

\[ n \to \infty \]

Consider a sample

\[ n E_\theta \left[ A_n^{(1)} \frac{\partial^2}{\partial \mu \partial \sigma} \log p_{h_n}(X_{h_n}[\theta]) A_n^{(2)} \right] \]

\[ = \frac{1}{n} \frac{5}{8 \sigma^4}. \]

and

\[ n E_\theta \left[ \left( \frac{\partial^2}{\partial \mu \partial \sigma} p_{h_n}(X_{h_n}[\theta]) \right) \frac{\partial}{\partial \sigma} p_{h_n}(X_{h_n}[\theta]) \right] \]

\[ \sim \frac{1}{n} E_\theta \left[ \left( \frac{2\sqrt{\pi}}{\pi \sigma^3 (X^2 + \nu)} \left( \frac{(X^2 - \nu)^2}{\sqrt{\pi}} \right)^2 \right) \right] \]

\[ = \frac{1}{n} \frac{7}{8 \sigma^4}, \]

as \( n \to \infty \). This implies

\[ \lim_{n \to \infty} \sup_{\theta \in \Theta} \left| A_n \nabla (g_{n1}(\theta)^T) A_n \right|^2 = 0, \]

since the matrix norm is the Frobenius norm. This completes the proof.

\[ \square \]

Having obtained the result for the non-Skew Student-Lévy process, a natural question is if it can be extended to the Skew Student-Lévy process.

**Proposition 1.** Let \( \{X_t\} \) be the Skew Student-Lévy process such that \( \mathcal{L}(X_t) = t(\nu, \mu, \sigma^2, \beta) \) (with known \( \nu \)). The LAN \( \text{LAN} \) property for \( \theta = (\mu, \sigma, \beta) \in \Theta \) with \( \Theta \) bounded and convex such that \( \Theta \subset \mathbb{R} \times (0, \infty) \times (\mathbb{R} \setminus \{0\}) \) does not hold true, since the Fisher information \( \mathcal{I}(\theta) \) does not exist for \( A_n = \text{diag}(\sqrt{n}, \sqrt{n}, \sqrt{n}^{-1}) \) converging to zero as \( n \to \infty \), i.e., \( \sqrt{n} h_n \to \infty \) as \( n \to \infty \).

The proposition requires that \( \sqrt{\mu h_n^2} = \sqrt{\nu h_n} \to \infty \) as \( n \to \infty \), implying that the upper boundary of the sampling interval \( T_n \) has to tend to infinity. If this is not given (e.g., if \( T \equiv T_n \)) we also have no LAN result because in this case \( A_n \) is divergent.

**Proof of Proposition 1.** Consider a sample \( \{X_{h_n}\}_{1 \leq k \leq n} \). First, similarly to the proof of Lemma 2 we check local stability of the Skew Student-Lévy process. \( X_1 \sim t(\nu, \mu, \sigma^2, \beta) \) has the characteristic function

\[ K_{\nu/2} (\sqrt{\nu} \sigma \sqrt{u^2 - 2i\beta u}(\sqrt{\nu} \sigma)^{\nu/2}(u^2 - 2i\beta u)^{\nu/4} e^{iuu} \]

\[ \Gamma \left( \frac{\nu}{2} \right) 2^{\nu/2-1} \]

see v. Hammerstein (2010). Then \( X_{h_n(\cdot) - h_n(\cdot)} \) has the characteristic function

\[ K_{\nu/2} \left( \sqrt{\nu} \sigma \sqrt{u^2 - 2i\beta u} \rho_{h_n(\cdot)} \right) \left( \sqrt{\nu} \sigma \left( u^2 - 2i\beta u \rho_{h_n(\cdot)} \right)^{\nu/4} \right) \]

\[ \rho_{h_n(\cdot)} \]

All terms, except the one with the Bessel function, tend to unity as \( n \to \infty \). As above we have

\[ K_{\nu/2} \left( \sqrt{\nu} \sqrt{\frac{u^2}{h_n^2} - 2i\beta \frac{u}{h_n}} \right) h_n \]

\[ \sim \left[ \frac{\pi}{2 \sqrt{2} \sqrt{\frac{u^2}{h_n^2} - 2i\beta \frac{u}{h_n}}} \exp \left( -\sqrt{\nu} \sqrt{\frac{u^2}{h_n^2} - 2i\beta \frac{u}{h_n}} \right) \right] h_n \]

13
\[
\sim \exp \left(-\sqrt{\nu} \sqrt{u^2 - 2\nu \sigma u h_n} \right)
\rightarrow \exp \left(-\sqrt{\nu} |u| \right),
\]
for \(h_n \to 0\), which is the characteristic function of \(\tilde{X} \sim \text{Cauchy}(0, \sqrt{\nu})\).

Second, we focus on \(\frac{\partial}{\partial \beta} p_{\theta_n}(X_{h_n} | \theta)\). Recall that the Skew Student-Lévy process \(\{X_t\}\) can be constructed by \(X_t = \sigma B_{\nu} + \beta \sigma^2 Y_t + \mu t\) with an inverse gamma subordinator \(\{Y_t\}\) such that \(Y_t \sim \Gamma(\nu/2, \nu/2)\) and a Brownian motion \(B\). Then \(X_t | Y_t \sim N(\mu t + \beta \sigma^2 Y_t, \sigma^2 Y_t)\). Thus,
\[
\frac{\partial}{\partial \beta} N(X_{h_n} | h_n \mu + \beta \sigma^2 Y_{h_n}, \sigma^2 Y_{h_n}) = \exp \left(-\frac{(X_{h_n} - h_n \mu - \beta \sigma^2 Y_{h_n})^2}{2 \sigma^2 Y_{h_n}} \right) \frac{X_{h_n} - h_n \mu - \beta \sigma^2 Y_{h_n}}{\sqrt{2 \pi \sigma^2 Y_{h_n}}}
\]
\[
\xrightarrow{\beta} \exp \left(-\frac{\tilde{X}^2}{2Y_{h_n}} \right) \frac{\tilde{X}}{\sqrt{2 \pi Y_{h_n}}}
\]
This is because \(\frac{X_{h_n} - h_n \mu - \beta \sigma^2 Y_{h_n}}{h_n \sigma} \overset{\beta}{\sim} \frac{X_{h_n} - h_n \mu}{h_n \sigma} \overset{\nu}{\sim} \frac{Y_{h_n}}{h_n \nu}\) as \(h_n \nu \to 0\). Using
\[
\int_0^\infty \exp \left(-\frac{\tilde{X}^2}{2y} \right) \frac{\tilde{X}}{\sqrt{2 \pi y}} p_{\gamma}(y) dy = \frac{\sqrt{\nu}}{\pi (\tilde{X}^2 + \nu)}
\]
yields
\[
\frac{1}{h_n^2} \left( \frac{\partial}{\partial \beta} p_{\theta_n}(X_{h_n} | \theta) \right)^2 \xrightarrow{\beta} \left( \frac{\sqrt{\nu \tilde{X}}}{\pi (\tilde{X}^2 + \nu)} \right)^2 = \sigma^2 \tilde{X}^2.
\]
This means that for rate \(A_n^{(33)} = \frac{1}{h_n^{\nu/\mu}}\) we have
\[
E_\theta \left[\sigma^2 \tilde{X}^2 \right] = \infty,
\]
implying that, by Fatou’s Lemma, the LAN property does not hold true.

The issue of having no LAN result cannot be solved by simply choosing another rate \(A_n^{(33)}\). For example, if \(A_n^{(33)} = (ah_n)^{-1}\) this would cause a singular Fisher information matrix making the LAN result not meaningful (see Masuda 2015).

This means that joint asymptotic normality (and optimality) for the MLE is not available. Fortunately, the result of Theorem 1 remains valid if we treat the non-zero skewness parameter \(\beta\) as known.

**Corollary 1.** Let \(\{X_t\}\) be the Skew Student-Lévy process such that \(\mathcal{L}(X_t) = t(\nu, \mu, \sigma^2, \beta)\) (with known \(\nu\) and \(\beta\)). The LAN holds true for \(\theta = (\mu, \sigma)\) with Fisher information \((\tilde{\theta})\) and rate \((\tilde{h})\) of Theorem 2.

**Proof.** The proof is analogous to the proof of Theorem 1 but using the local Cauchy property of Proposition 1. Note that \(N(X_{h_n} | h_n \mu + \beta \sigma^2 Y_{h_n}, \sigma^2 Y_{h_n})\) and \(\frac{\partial}{\partial \beta} N(X_{h_n} | h_n \mu + \beta \sigma^2 Y_{h_n}, \sigma^2 Y_{h_n})\) have the same limiting behavior as for \(\beta = 0\) because \(\frac{X_{h_n} - h_n \mu - \beta \sigma^2 Y_{h_n}}{h_n \sigma} \overset{\beta}{\sim} \frac{X_{h_n} - h_n \mu}{h_n \sigma} \overset{\nu}{\sim} \frac{Y_{h_n}}{h_n \nu}\). Moreover, for the computation of \(\mathcal{I}^{(22)}(\theta)\) and \(\mathcal{I}^{(12)}(\theta)\), we use \(X_{h_n} = h_n \mu - Y_{h_n} \beta \sigma^2 \) to eliminate.

**Remark 3.** We here found a special case of the GH Lévy process for which the LAN does not hold. This means that it cannot be true for all parameter constellations of the GH process. More research is needed to find the conditions under which the LAN property holds.

We end the discussion of discretely sampled Student-Lévy processes with a short remark concerning low-frequency sampling.

**Remark 4.** Lemma 1 is also valid for low-frequency sampling with the difference that \(h_n \to h\) for \(n \to \infty\) implying that \(p_{\theta_n}(x | \theta)\) is the \(h\) transition density of the \(X_h\) of the Student-Lévy process and is not available in closed form. This carries over to the Fisher information and is therefore omitted here. There is one exception, namely if \(h = 1\). Then the transition density and the Fisher information are known explicitly, but we refer to Lange et al. (1989) for this standard case.
2.1 Continuous sampling

We now take a little detour and discuss the case of continuous data where the full path \( \{X_t\}_{t \in [0,T]} \) is observed. We are interested in estimation of parameters and the asymptotics when \( T \to \infty \). Although this setting is unrealistic, it is interesting to spell out the differences to high-frequency sampling. It may be the case that some parameters can be estimated without error when we observe the whole path. To identify these parameters \( \text{Raible} (2000) \) (see also \( \text{Masuda} (2015) \) and \( \text{Sato} (1999) \)) proved the following proposition, which provides a criterion when the likelihood ratio \( \frac{P_{\theta_1}|_{\mathcal{F}_T}}{P_{\theta_2}|_{\mathcal{F}_T}} \) for parameters \( \theta_1 \) and \( \theta_2 \) is well-defined, where \( P_\theta|_{\mathcal{F}_T} \) denotes the restriction of \( P_\theta \) to the natural filtration \( \mathcal{F}_T \) generated by \( \{X_t\}_{t \leq T} \).

**Proposition 2.** Let \( \{X_t\} \) be a (one-dimensional) Lévy process with characteristics \((\gamma(\theta), A(\theta), \Pi(dx; \theta))\) for \( \theta \in \Theta \subset \mathbb{R}^p \). Let \( T > 0 \) and \( \theta_1, \theta_2 \in \Theta \). Then the measures \( P_{\theta_1}|_{\mathcal{F}_T} \) and \( P_{\theta_2}|_{\mathcal{F}_T} \) are equivalent iff the following conditions hold true.

(i) \( \Pi(dx; \theta_2) = k(x; \theta_1, \theta_2)\Pi(dx; \theta_2) \) for some Borel function \( k(\cdot; \theta_1, \theta_2) : \mathbb{R} \to (0, \infty) \),

(ii) \( \gamma(\theta_2) = \gamma(\theta_1) + \int_k x(k(x; \theta_1, \theta_2) - 1)\Pi(dx; \theta_1) + \sqrt{A(\theta_1)}b \) for some \( b \),

(iii) \( A(\theta_2) = A(\theta_1) \),

(iv) \( \int_{B} \left( 1 - \sqrt{k(x; \theta_1, \theta_2)} \right)^2 \Pi(dx; \theta) < \infty \).

Using this criterion we obtain the following

**Corollary 2.** Let \( T > 0 \) and let \( P_{\theta_k}, k = 1, 2 \) denote the distribution of the Student-Lévy process with parameters \( \theta_k = (\nu_k, \mu_k, \sigma_k, \beta_k) \). Then \( P_{\theta_1}|_{\mathcal{F}_T} \) and \( P_{\theta_2}|_{\mathcal{F}_T} \) are equivalent iff \( \mu_1 = \mu_2 \) and \( \sqrt{\nu_1}\sigma_1 = \sqrt{\nu_2}\sigma_2 \).

**Proof.** \( \text{Raible} (2000) \) proved this for the more general GH process with 1-increments distributed as \( GH(\lambda_k, \alpha_k, \beta_k, \theta_k, \mu_k) \). The measures are equivalent iff \( \delta_1 = \delta_2 \) and \( \mu_1 = \mu_2 \). The Student \( t \) distribution \( t(\nu_k, \mu_k, \sigma_k) \) is the GH distribution limiting case \( GH(-\frac{\nu_k}{\sqrt{\nu_k}}, |\beta_k|, \sqrt{\nu_k}\sigma_k, \mu_k) \).

Where \( \nu \) is known, this reduces to \( \sigma_1 = \sigma_2 \). This means we can find \((\mu, \sigma)\) by observing the path. \( \text{Raible} (2000) \) proved that the statistics

\[
\hat{\sigma}_{T,n} := \frac{\pi}{\sqrt{nT}} \#\{t \leq T : \Delta X_t \geq 1/n\},
\]

\[
\hat{\mu}_{T,n} = \frac{1}{T} \left( X_T - \sum_{0 < t \leq T} \Delta X_t 1_{[1/n, \infty)}(\Delta X_t) \right)
\]

are strongly consistent estimators of \( \sigma \) and \( \mu \), as \( n \to \infty \). If we observe the path in continuous time we can compute \( \sigma \overset{a.s.}{=} \lim_{n \to \infty} \hat{\sigma}_{T,n} \) and \( \mu \overset{a.s.}{=} \lim_{n \to \infty} \hat{\mu}_{T,n} \). Section 4 compares these estimators (where the data is obviously not continuously available but in high-frequency) with the high-frequency MLE.

3 Numerical methods

Of course, the theory from the previous section is not directly informative about how to actually compute the MLE. As is the case for the Student \( t \) distribution (i.e., the Student-Lévy process observing 1-increments) the MLE does not exist in closed form. This carries over to the Student-Lévy process when observing \( \Delta^h_k X \) with \( h_n \neq 1 \). Moreover, the density function of \( \Delta^h_k X \) is not given explicitly. We discuss two approaches to tackling this issue. First, we numerically maximize the Fourier inversion of the characteristic function. Second, we use a Monte Carlo Expectation-Maximization (MCEM) algorithm. The first approach is less elegant and substantially slower to execute than the second one but involves no randomness.

Let \( \varphi_{X_t} = \varphi_{\nu, \mu, \sigma^2} \) be the characteristic function of the 1-increment, i.e., the characteristic function of \( t(\nu, \mu, \sigma^2) \). Then, the transition density of \( \Delta^h_k X \) can be numerically found via

\[
p_{h_n}(\Delta^h_k X | \theta) = \frac{1}{2\pi} \int \exp(-iu\Delta^h_k X)\varphi_{X_1}(u)^{h_n} du.
\]

There are multiple ways to numerically compute this integral. For example, by a suitable discretization and subsequent application of the Fast Fourier Transform algorithm; see \( \text{Walker} (1996) \) among many...
The log-likelihood function $\ell(\theta)$ can be numerically maximized in $\theta$ by the Nelder & Mead [1965] method. One issue with this method is that Fourier inversion needs to be executed extensively, which is highly time-consuming. Thus, we will use this approach only for comparison. We call it the Characteristic Function–Maximum Likelihood Estimator (CF-MLE).

As an alternative, we discuss an MCEM approach. The EM algorithm was developed by Dempster et al. (1977) and a Monte Carlo extension was proposed by Wei & Tanner (1990). We first sketch the details of how the EM algorithm works. Then we apply it to the present Student-Lévy scenario and explain why the MC extension is used. The resulting ML estimation routine is summarized in Algorithm 1.

We follow McLachlan & Krishnan (2007) for the details of the EM algorithm. The idea behind EM is to assume that besides the observed data $x$ with density function $p(x|\theta)$, there are missing values $y$ which we cannot observe. If we could observe them, ML estimation using the joint density $p(x, y|\theta)$ would be easy.

Denote by $\ell(\theta|x)$ the incomplete log-likelihood and by $\ell(\theta|x, y)$ the complete log-likelihood. Take some initial value $\theta_0$. The following E- and M-steps are repeated alternately. On the $(j + 1)$-th iteration we have:

- **E-Step.** Calculate
  \[ Q(\theta|\theta_j) := E_{\theta_j}[\ell(\theta|x, y)|x] = \int \log p(x, y|\theta)p(y|x, \theta_j)d\theta. \]

- **M-Step.** Find a value $\theta_{j+1}$ that maximizes $Q(\theta|\theta_j)$:
  \[ \theta_{j+1} = \arg \max Q(\theta|\theta_j). \]

In practice, we repeat the E- and M-steps until the sequence $\{\theta_j\}$ converges. We omit here the proof that the EM algorithm indeed finds the MLE and refer to Dempster et al. (1977) or McLachlan & Krishnan (2007), but note that the M-step implies that the incomplete-data log-likelihood function is non-decreasing, i.e.

$\ell(\theta_{j+1}|x) \geq \ell(\theta_j|x)$

for any $j = 0, 1, \ldots$.

In case the E-step is difficult to compute, i.e., the expectation has no closed form as in the present Student-Lévy case, we replace the E-step with the following MCE-step. Assume that the missing data $y$ can be sampled from the posterior latent distribution $p(y|x, \theta_j)$. Then we have

- **MCE-Step.**
  \[ \hat{Q}(\theta|\theta_j) := \frac{1}{B} \sum_{b=1}^{B} \log p(x, y_b|\theta) \rightarrow Q(\theta|\theta_j), \]

a.s., for $y_b \sim p(y|x, \theta_j)$, $b = 1, \ldots, B$, i.i.d., as $B \rightarrow \infty$.

Next, we apply the MCEM algorithm to sample paths of the Student-Lévy process. The procedure is similar to the standard Student $t$ distribution McLachlan & Krishnan (2007), but differs in some of the details. Let

$x = \{ \Delta^n_k X \}_{k=1, \ldots, n}, \quad y = \{ \Delta^n_k Y \}_{k=1, \ldots, n}, \quad \theta = \{ \mu, \sigma \},$

where $\{ X_t \}$ denotes a Student-Lévy process of which we observe a sample path and $\{ Y_t \}$ is the corresponding unobserved inverse gamma subordinator. $p_{h_n}(x|\theta)$ denotes the density of $X_{h_n}$ and $p_{h_n}(y)$ denotes the density of $Y_{h_n}$, which is independent of $\theta$. Then the joint density is given by

$p_{h_n}(x, y|\theta) = p_{h_n}(x|y, \theta)p_{h_n}(y) = N(x|h_n\mu, \sigma^2 Y)p_{h_n}(y).$

The complete log-likelihood function is given by

$\ell(\theta|\{ \Delta^n_k X \}, \{ \Delta^n_k Y \})$

$= \sum_{k=1}^{n} \log N(\Delta^n_k X|h_n\mu, \sigma^2 \Delta^n_k Y) + \log p_{h_n}(\Delta^n_k Y)$
The density $p_{h_n}(y)$ of the inverse gamma subordinator has no closed form. However, since it only depends on $\nu$, which we assume to be known, it is irrelevant for maximization in $(\mu, \sigma)$. The normal part of the complete likelihood is independent of $\nu$. This is indeed the reason why we assume $\nu$ to be fixed. As we do not explicitly know $p_{h_n}(\Delta_k^n Y)$, except that it depends solely on $\nu$, we cannot perform likelihood-based estimation.

Next, we seek the posterior of the latent subordinator in order to take the expectation w.r.t. this posterior. In the case of the Student-$t$ distribution, the latent variables are inverse gamma distributed and therefore the posterior is also inverse gamma distributed because it is a conjugate prior for the normal distribution. This is not the case for the inverse gamma subordinator and general $h_n \neq 1$. By Bayes’ law,

$$p_{h_n}(y|\Delta_k^n X, \theta) = \frac{p_{h_n}(\Delta_k^n X|y, \theta)p_{h_n}(y)}{p_{h_n}(\Delta_k^n X|\theta)} = \frac{N(\Delta_k^n X|h_n \mu, \sigma^2 y)p_{h_n}(y)}{p_{h_n}(\Delta_k^n X|\theta)},$$  

(27)

To find $Q(\theta|\theta_j)$ we integrate the log-likelihood with respect to this posterior.

$$Q(\theta|\theta_j) = E_{\theta_j} \left[ \ell(\theta|\{\Delta_k^n X\}, \{\Delta_k^n Y\}) |\Delta_k^n X\right] = \sum_{k=1}^{n} -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log \sigma^2 - \frac{1}{2} \log \Delta_k^n Y - \frac{(\Delta_k^n X - h_n \mu)^2}{2\sigma^2 \Delta_k^n Y} + \log p_{h_n}(\Delta_k^n Y).$$

$$= \sum_{k=1}^{n} -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log \sigma^2 - \frac{1}{2} \log \Delta_k^n Y - \frac{(\Delta_k^n X - h_n \mu)^2}{2\sigma^2 \Delta_k^n Y} + \log p_{h_n}(\Delta_k^n Y) + \log \Delta_k^n Y.$$

(27)

We do not need to find $E_{\theta_j} [\log \Delta_k^n Y | \Delta_k^n X]$ and $E_{\theta_j} [\log p_{h_n}(\Delta_k^n Y) | \Delta_k^n X]$ because when we take the derivative w.r.t. $\mu$ or $\sigma$ these terms vanish. It only remains to find $E_{\theta_j} \left[ \frac{1}{\Delta_k^n Y} | \Delta_k^n X \right]$. By (27),

$$E_{\theta_j} \left[ \frac{1}{\Delta_k^n Y} | \Delta_k^n X \right] = \int_0^\infty \frac{1}{y} \frac{N(\Delta_k^n X|h_n \mu, \sigma^2 y)}{p_{h_n}(\Delta_k^n X|\theta)} p_{h_n}(y) dy.$$

(28)

At this point, Monte Carlo integration is useful. We approximate (28) by

$$E_{\theta_j} \left[ \frac{1}{\Delta_k^n Y} | \Delta_k^n X \right] = \frac{1}{B} \sum_{b=1}^{B} \frac{N(\Delta_k^n X|h_n \mu, \sigma^2 Y_{h_n,b})}{p_{h_n}(\Delta_k^n X|\theta)}.$$

where $Y_{h_n,b}, b = 1, \ldots, B$, are i.i.d. draws from $p_{h_n}(y)$. They are simulated by series representations with the inverse Lévy measure method using the numerical inversion algorithms of Massing [2018], which make proper use of numerical integration in combination with Newton interpolation instead of slow root finding. Then $E_{\theta_j} \left[ \frac{1}{\Delta_k^n Y} | \Delta_k^n X \right] \to E_{\theta_j} \left[ \frac{1}{\Delta_k^n Y} | \Delta_k^n X \right]$ a.s. for $B \to \infty$. Note that we have to use the Fourier inversion [26] to compute $p_{h_n}(\Delta_k^n X|\theta)$. Although this is the most time-consuming step in the proposed MCEM algorithm, it is still much faster than the Nelder-Mead approach mentioned above.

We conclude the $(j+1)$-th MCE-Step.

$$\hat{Q}(\theta|\theta_j) = \sum_{k=1}^{n} -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log \sigma^2 - \frac{1}{2} \log \Delta_k^n Y - \frac{(\Delta_k^n X - h_n \mu)^2}{2\sigma^2 \Delta_k^n Y} + C,$$

(29)

where we collect the terms vanishing during maximization in the constant $C$.

In order to maximize (29), we set $\frac{\partial Q}{\partial \mu} = 0$ and $\frac{\partial Q}{\partial \sigma} = 0$. We obtain the $(j+1)$-th M-Step.

$$h_n \sum_{k=1}^{n} \hat{E}_{\theta_j} \left[ \frac{1}{\Delta_k^n Y} | \Delta_k^n X \right]$$

(30)
and
\[ \sigma_j^2 + 1 = \frac{1}{n} \sum_{k=1}^{n} (\Delta_k^n X - h_n \mu_j + 1)^2 \hat{E}_{\theta_j} \left[ \frac{1}{\Delta_k^n X} \right] \Delta_k^n X. \] (31)

Note that we need to find \( \mu_{j+1} \) first, since it is needed to update \( \sigma_{j+1} \). The MCE-step and the M-step are repeated iteratively until we observe convergence of \((\mu_j, \sigma_j)\). To speed up the MCEM, we draw \( Y_{h_n,b} \), \( b = 1, \ldots, B \), only once and reuse them in any MCE-step as recommended by Levine & Casella (2001).

For the initial \((\mu_0, \sigma_0)\) we take the raw moment estimates if \( \nu > 2 \). For \( 1 < \nu \leq 2 \) we take some other initial values. The whole MCEM routine is summarized in compact form in Algorithm 1.

### Algorithm 1 MCEM Algorithm for the Student-Lévy process with known \( \nu \)

**Input:** Sample path observed at \((X_{kh_n})_{1 \leq k \leq n} \), such that \( \Delta_k^n X \sim p_{h_n}(x|\theta); \)

**Output:** Maximum likelihood estimates \( \hat{\mu} \) and \( \hat{\sigma}^2 \).

1: \( \mu_0 \leftarrow \frac{1}{h_n n} \sum_{k=1}^{n} \Delta_k^n X; \)
2: \( \sigma_0^2 \leftarrow \frac{\nu}{\nu-2} \frac{1}{h_n n} \sum_{k=1}^{n} (\Delta_k^n X)^2 - \frac{\nu-2}{\nu-2} h_n \mu_0^2; \) \( \triangleright \) Start with moment estimation.
3: Draw \( B \) i.i.d. random variates \( Y_{h_n,1}, \ldots, Y_{h_n,B} \sim p_{h_n}(y); \)
4: \( j \leftarrow 0; \)
5: repeat
6: for \( k = 1 \) to \( n \) do
7: \( \hat{E}_{\theta_j} \left[ \frac{1}{\Delta_k^n X} \right] \Delta_k^n X \left[ \frac{1}{\Delta_k^n X} \right] \Delta_k^n X \equiv \frac{1}{\pi} \sum_{b=1}^{B} \frac{1}{Y_{h_n,b} - (\Delta_k^n X) \sigma_j; \}
8: end for
9: M step:
10: for \( k = 1 \) to \( n \) do
11: \( \mu_{j+1} \leftarrow \frac{\sum_{k=1}^{n} \Delta_k^n X}{h_n \sum_{k=1}^{n} \Delta_k^n X} \); \( \hat{E}_{\theta_j} \left[ \frac{1}{\Delta_k^n X} \right] \Delta_k^n X; \)
12: \( \sigma_{j+1}^2 \leftarrow \frac{1}{\pi} \sum_{k=1}^{n} (\Delta_k^n X - h_n \mu_{j+1})^2 \hat{E}_{\theta_j} \left[ \frac{1}{\Delta_k^n X} \right] \Delta_k^n X; \)
13: end for
14: until convergence.

### 4 Monte Carlo study

In this section we briefly present some experimental evidence for the above methods. The section is split into three experiments. First, we test the MCEM algorithm and verify that a higher frequency leads to a better estimation result. A second experiment compares the MCEM algorithm with the Nelder-Mead maximization of the Fourier inversion. Third, we investigate the estimators \( \hat{\mu}_{T,n} \) and \( \hat{\sigma}_{T,n} \) for continuous sampling.

The first experiment tests the MCEM algorithm. We sample Student-Lévy paths for different degrees of freedom \( \nu = 4, 12, 39 \) with \( h_n = 0.01, 0.1, 0.5, 1 \) increments until \( T = T_n = 100 \). We consider 5 constellations for \( \theta = (\mu, \sigma) \). For each sampled path and each \( h_n \) we compute the ML estimate \( (\hat{\mu}_{ML,\sigma}, \hat{\sigma}_{ML}) \) with the MCEM algorithm and the method of moments (MoM) estimate \( (\hat{\mu}_{MoM}, \hat{\sigma}_{MoM}) \). We repeat this 10,000 times for each constellation and compute the empirical bias and the empirical root mean squared error (RMSE). Tables 1, 2 and 3 report the results. In order to reduce computing time we only estimate the parameters for \( h_n = 0.01 \) in the setting \( \theta = (0, 1) \).

Unsurprisingly, the estimates are closer to the true parameters for smaller step sizes in all constellations. In almost all setups, the ML estimates are better than the MoM estimates. Moreover, this pattern is more clearly visible for smaller degrees of freedom. This is due to the fact that, for high degrees of freedom, the increments are approximately normally distributed and the MoM estimator and the MLE coincide for Gaussian increments. Thus, the MLE performs better for low degrees of freedom.

Further, the bias for the scaling parameter is typically negative while the bias for the location parameter is positive in some cases. Next, the RMSEs for the ML estimates of both parameters only depend on the true \( \sigma \) (and not on \( \mu \)). This is supported by Theorem 1 as the Fisher information matrix does not depend on \( \mu \). As \( \nu \) increases, the differences in RMSEs for \( \mu \) tend to vanish along \( h_n \). Again, this is reasonable as the MLE and the MoM estimator are numerically close for high degrees of freedom. Although the Fisher
This is because Student-Lévy increments for high degrees of freedom are approximately normal. However, for different true \( \nu \) and step size \( h_n \), there are (if at all) only small differences. All figures share the same true parameter setup \((\mu, \sigma) = (0, 1)\).

Note that the moment estimator for \( \mu \) is numerically equal to any digit among different times \( h_n \) since the moment estimator is not consistent if \( T_n \to \infty \).

Figures 1 and 2 show kernel density estimates of the realizations of \( \sqrt{n} \left( \hat{\mu}_{ML} - \mu \right) / \sqrt{\nu \sigma^2} \) (panel (a)) and \( \sqrt{n} \left( \hat{\mu}_{MoM} - \mu \right) / \sqrt{2 \sigma^2} \) (panel (b)) for different \( h_n \) compared with the theoretical standard normal density. Figure 1 considers \( \nu = 4 \), Figure 2 is for \( \nu = 12 \) and Figure 3 for \( \nu = 39 \). All figures share the same true parameter setup \((\mu, \sigma) = (0, 1)\).

Figure 3 illustrates asymptotic normality for both estimators, and the density estimates for \( h_n = 0.01 \) are not too far from the standard normal density. Figure 2 and 3 show that this is now less valid for large \( \nu \). In Figure 3(b) all kernel density estimates are closer to each other and obviously not standard normal. This is because Student-Lévy increments for high degrees of freedom are approximately normal. However, for Brownian motions the LAN holds true, but with Fisher information \( \mathcal{I}^{(22)} = \frac{\nu}{2\nu} \) instead of \( \frac{1}{2\nu} \) for Student-Lévy processes (yet with the same rate). If we do not interpret the high degrees of freedom for \( \nu \).

### Table 1: Empirical bias and RMSE (in parenthesis) for \( \nu = 4 \) comparing the MLE and the MoM estimator for different true \( \theta \) and step size \( h_n \).

<table>
<thead>
<tr>
<th>True ( \theta )</th>
<th>( h_n )</th>
<th>( \hat{\mu}_{ML} )</th>
<th>( \hat{\sigma}_{ML} )</th>
<th>( \hat{\mu}_{MoM} )</th>
<th>( \hat{\sigma}_{MoM} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0, 1)</td>
<td>1</td>
<td>-1.3 \times 10^{-3}</td>
<td>-5.5 \times 10^{-3}</td>
<td>-10^{-3}</td>
<td>-0.019</td>
</tr>
<tr>
<td></td>
<td>(.118)</td>
<td>(.094)</td>
<td>(.141)</td>
<td>(.162)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>-9.1 \times 10^{-4}</td>
<td>-3.2 \times 10^{-3}</td>
<td>-10^{-3}</td>
<td>-0.016</td>
</tr>
<tr>
<td></td>
<td>(.106)</td>
<td>(.07)</td>
<td>(.141)</td>
<td>(.156)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>5 \times 10^{-5}</td>
<td>-1.5 \times 10^{-4}</td>
<td>-10^{-3}</td>
<td>-0.013</td>
</tr>
<tr>
<td></td>
<td>(.069)</td>
<td>(.037)</td>
<td>(.141)</td>
<td>(.149)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.01</td>
<td>1.9 \times 10^{-4}</td>
<td>-3.2 \times 10^{-4}</td>
<td>-10^{-3}</td>
<td>-0.012</td>
</tr>
<tr>
<td></td>
<td>(.027)</td>
<td>(.013)</td>
<td>(.141)</td>
<td>(.148)</td>
<td></td>
</tr>
<tr>
<td>(0, 3)</td>
<td>1</td>
<td>-9 \times 10^{-3}</td>
<td>-1.7 \times 10^{-2}</td>
<td>-8.6 \times 10^{-3}</td>
<td>-0.066</td>
</tr>
<tr>
<td></td>
<td>(.361)</td>
<td>(.278)</td>
<td>(.428)</td>
<td>(.459)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>-7.2 \times 10^{-3}</td>
<td>-9.9 \times 10^{-3}</td>
<td>-8.6 \times 10^{-3}</td>
<td>-0.056</td>
</tr>
<tr>
<td></td>
<td>(.324)</td>
<td>(.211)</td>
<td>(.428)</td>
<td>(.435)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>-2 \times 10^{-3}</td>
<td>-3.1 \times 10^{-3}</td>
<td>-8.6 \times 10^{-3}</td>
<td>-0.046</td>
</tr>
<tr>
<td></td>
<td>(.211)</td>
<td>(.111)</td>
<td>(.428)</td>
<td>(.416)</td>
<td></td>
</tr>
<tr>
<td>(0, 0.1)</td>
<td>1</td>
<td>4.4 \times 10^{-5}</td>
<td>-5.9 \times 10^{-4}</td>
<td>2.4 \times 10^{-5}</td>
<td>-1.9 \times 10^{-3}</td>
</tr>
<tr>
<td></td>
<td>(.0118)</td>
<td>(.0094)</td>
<td>(.0142)</td>
<td>(.0179)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>3.7 \times 10^{-5}</td>
<td>-4.1 \times 10^{-4}</td>
<td>2.4 \times 10^{-5}</td>
<td>-1.5 \times 10^{-3}</td>
</tr>
<tr>
<td></td>
<td>(.0106)</td>
<td>(.007)</td>
<td>(.0142)</td>
<td>(.0172)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>3.3 \times 10^{-5}</td>
<td>-1.5 \times 10^{-4}</td>
<td>2.4 \times 10^{-5}</td>
<td>-1.1 \times 10^{-3}</td>
</tr>
<tr>
<td></td>
<td>(.0069)</td>
<td>(.0036)</td>
<td>(.0142)</td>
<td>(.0166)</td>
<td></td>
</tr>
<tr>
<td>(2, 1)</td>
<td>1</td>
<td>-8 \times 10^{-4}</td>
<td>-4.7 \times 10^{-3}</td>
<td>-2.1 \times 10^{-3}</td>
<td>-0.019</td>
</tr>
<tr>
<td></td>
<td>(.118)</td>
<td>(.093)</td>
<td>(.141)</td>
<td>(.16)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>-9.9 \times 10^{-5}</td>
<td>-1.9 \times 10^{-3}</td>
<td>-2.1 \times 10^{-3}</td>
<td>-0.015</td>
</tr>
<tr>
<td></td>
<td>(.106)</td>
<td>(.07)</td>
<td>(.141)</td>
<td>(.152)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>-1.3 \times 10^{-3}</td>
<td>-8.7 \times 10^{-4}</td>
<td>-2.1 \times 10^{-3}</td>
<td>-0.013</td>
</tr>
<tr>
<td></td>
<td>(.07)</td>
<td>(.036)</td>
<td>(.141)</td>
<td>(.147)</td>
<td></td>
</tr>
<tr>
<td>(-0.5, 1)</td>
<td>1</td>
<td>-6.8 \times 10^{-4}</td>
<td>-3.7 \times 10^{-3}</td>
<td>-1.7 \times 10^{-3}</td>
<td>-0.016</td>
</tr>
<tr>
<td></td>
<td>(.118)</td>
<td>(.094)</td>
<td>(.142)</td>
<td>(.166)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>-1.3 \times 10^{-3}</td>
<td>-1.8 \times 10^{-4}</td>
<td>-1.7 \times 10^{-3}</td>
<td>-0.013</td>
</tr>
<tr>
<td></td>
<td>(.106)</td>
<td>(.07)</td>
<td>(.142)</td>
<td>(.158)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>-1.2 \times 10^{-3}</td>
<td>-3.3 \times 10^{-4}</td>
<td>-1.7 \times 10^{-3}</td>
<td>-0.01</td>
</tr>
<tr>
<td></td>
<td>(.07)</td>
<td>(.036)</td>
<td>(.142)</td>
<td>(.152)</td>
<td></td>
</tr>
<tr>
<td>True $\theta$</td>
<td>$h_n$</td>
<td>$\hat{\mu}_{ML}$</td>
<td>$\hat{\sigma}_{ML}$</td>
<td>$\hat{\mu}_{MoM}$</td>
<td>$\hat{\sigma}_{MoM}$</td>
</tr>
<tr>
<td>--------------</td>
<td>-------</td>
<td>-----------------</td>
<td>-------------------</td>
<td>-----------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>(0, 1)</td>
<td>1</td>
<td>$5.7 \cdot 10^{-4}$</td>
<td>$-6.3 \cdot 10^{-3}$</td>
<td>$4 \cdot 10^{-4}$</td>
<td>$-9 \cdot 10^{-3}$</td>
</tr>
<tr>
<td></td>
<td>(.106)</td>
<td>(.08)</td>
<td>(.108)</td>
<td>(.082)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>$8.9 \cdot 10^{-4}$</td>
<td>$-3.4 \cdot 10^{-3}$</td>
<td>$4 \cdot 10^{-4}$</td>
<td>$-5.7 \cdot 10^{-3}$</td>
</tr>
<tr>
<td></td>
<td>(.102)</td>
<td>(.059)</td>
<td>(.108)</td>
<td>(.065)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>$10^{-3}$</td>
<td>$-1.3 \cdot 10^{-3}$</td>
<td>$4 \cdot 10^{-4}$</td>
<td>$-2.1 \cdot 10^{-3}$</td>
</tr>
<tr>
<td></td>
<td>(.086)</td>
<td>(.031)</td>
<td>(.108)</td>
<td>(.048)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.01</td>
<td>$3.5 \cdot 10^{-5}$</td>
<td>$-3.2 \cdot 10^{-3}$</td>
<td>$4 \cdot 10^{-4}$</td>
<td>$-1.4 \cdot 10^{-3}$</td>
</tr>
<tr>
<td></td>
<td>(.042)</td>
<td>(.013)</td>
<td>(.108)</td>
<td>(.043)</td>
<td></td>
</tr>
<tr>
<td>(0, 3)</td>
<td>1</td>
<td>$-6.8 \cdot 10^{-4}$</td>
<td>-.023</td>
<td>$-6.6 \cdot 10^{-4}$</td>
<td>$-0.29$</td>
</tr>
<tr>
<td></td>
<td>(.319)</td>
<td>(.24)</td>
<td>(.326)</td>
<td>(.25)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>$-2.5 \cdot 10^{-4}$</td>
<td>-.013</td>
<td>$-6.6 \cdot 10^{-4}$</td>
<td>$-0.19$</td>
</tr>
<tr>
<td></td>
<td>(.31)</td>
<td>(.179)</td>
<td>(.326)</td>
<td>(.2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>$-7.1 \cdot 10^{-5}$</td>
<td>$-4.9 \cdot 10^{-3}$</td>
<td>$-6.6 \cdot 10^{-4}$</td>
<td>$-8.5 \cdot 10^{-3}$</td>
</tr>
<tr>
<td></td>
<td>(.259)</td>
<td>(.093)</td>
<td>(.326)</td>
<td>(.145)</td>
<td></td>
</tr>
<tr>
<td>(0, 0.1)</td>
<td>1</td>
<td>$1.2 \cdot 10^{-4}$</td>
<td>$-6.1 \cdot 10^{-4}$</td>
<td>$1.4 \cdot 10^{-4}$</td>
<td>$-8.1 \cdot 10^{-4}$</td>
</tr>
<tr>
<td></td>
<td>(.0108)</td>
<td>(.008)</td>
<td>(.0111)</td>
<td>(.0083)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>$1.4 \cdot 10^{-4}$</td>
<td>$-3.2 \cdot 10^{-4}$</td>
<td>$1.4 \cdot 10^{-4}$</td>
<td>$-4.7 \cdot 10^{-4}$</td>
</tr>
<tr>
<td></td>
<td>(.0106)</td>
<td>(.006)</td>
<td>(.0111)</td>
<td>(.0066)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>$5.4 \cdot 10^{-5}$</td>
<td>$-9.8 \cdot 10^{-5}$</td>
<td>$1.4 \cdot 10^{-4}$</td>
<td>$-1.4 \cdot 10^{-4}$</td>
</tr>
<tr>
<td></td>
<td>(.0088)</td>
<td>(.0031)</td>
<td>(.0111)</td>
<td>(.0049)</td>
<td></td>
</tr>
<tr>
<td>(2, 1)</td>
<td>1</td>
<td>$-1.8 \cdot 10^{-3}$</td>
<td>$-7.3 \cdot 10^{-3}$</td>
<td>$-1.8 \cdot 10^{-3}$</td>
<td>$-9.3 \cdot 10^{-3}$</td>
</tr>
<tr>
<td></td>
<td>(.107)</td>
<td>(.079)</td>
<td>(.109)</td>
<td>(.082)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>$-1.8 \cdot 10^{-3}$</td>
<td>$-4.1 \cdot 10^{-3}$</td>
<td>$-1.8 \cdot 10^{-3}$</td>
<td>$-5.5 \cdot 10^{-3}$</td>
</tr>
<tr>
<td></td>
<td>(.104)</td>
<td>(.059)</td>
<td>(.109)</td>
<td>(.065)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>$-7.2 \cdot 10^{-4}$</td>
<td>$-9.4 \cdot 10^{-4}$</td>
<td>$-1.8 \cdot 10^{-3}$</td>
<td>$-1.8 \cdot 10^{-3}$</td>
</tr>
<tr>
<td></td>
<td>(.087)</td>
<td>(.031)</td>
<td>(.109)</td>
<td>(.048)</td>
<td></td>
</tr>
<tr>
<td>(-0.5, 1)</td>
<td>1</td>
<td>$1.3 \cdot 10^{-4}$</td>
<td>$-7.4 \cdot 10^{3}$</td>
<td>$5 \cdot 10^{-4}$</td>
<td>$-9.5 \cdot 10^{5}$</td>
</tr>
<tr>
<td></td>
<td>(.108)</td>
<td>(.08)</td>
<td>(.11)</td>
<td>(.083)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>$1.9 \cdot 10^{-4}$</td>
<td>$-3.1 \cdot 10^{3}$</td>
<td>$5 \cdot 10^{-4}$</td>
<td>$-5 \cdot 10^{3}$</td>
</tr>
<tr>
<td></td>
<td>(.105)</td>
<td>(.059)</td>
<td>(.11)</td>
<td>(.066)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>$8.6 \cdot 10^{-4}$</td>
<td>$-1.1 \cdot 10^{3}$</td>
<td>$5 \cdot 10^{-4}$</td>
<td>$-2 \cdot 10^{3}$</td>
</tr>
<tr>
<td></td>
<td>(.087)</td>
<td>(.031)</td>
<td>(.11)</td>
<td>(.048)</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Empirical bias and RMSE (in parenthesis) for $\nu = 12$ comparing the MLE and the MoM estimator for different true $\theta$ and step size $h_n$.

freedom as approximately normal but follow the Student-Lévy LAN theory, the asymptotic normality hence “occurs later”.

### 4.1 Comparison between ML methods

The second experiment aims to specify the numerical error between the two different methods, MCEM-MLE and CF-MLE. We again test $\nu = 4, 12, 39$ and $h_n = 0.1, 0.5, 1$ but restrict ourselves to the setting $\mu = 0, \sigma = 1$. Since the execution time of CF-ML is too long, we only perform 100 iterations instead of 10,000 as before. In each iteration we simulate a Student-Lévy path and estimate parameters for $h_n$-increments both with MCEM and CF-ML. We then estimate the root mean squared deviance between both: $ \left( \frac{1}{100} \sum_j (\mu_{n,MCEM}^{(j)} - \mu_{n,CF-ML}^{(j)})^2 \right)^{1/2} $. For $h_n = 1$ (Student $t$ random numbers) we also compare each with the standard EM-MLE. Table 4 shows the results.

Apparently, the randomness caused by the Monte Carlo integration has little impact on the estimation results. This seems to be true for all degrees of freedom considered. There are some exceptions for the CF-ML, viz. the Nelder-Mead maximization occasionally fails to find the optimum. These outliers have
Figure 1: Panel (a) compares kernel density estimates of $\sqrt{n} \left( \hat{\mu}_{ML} - \mu \right)$ for different $h_n$ with the standard normal density. Analogously, panel (b) for $\sqrt{n} \left( \hat{\sigma}_{ML} - \sigma \right)$.

Figure 2: Panel (a) compares kernel density estimates of $\sqrt{n} \left( \hat{\mu}_{ML} - \mu \right)$ for different $h_n$ with the standard normal density. Analogously, panel (b) for $\sqrt{n} \left( \hat{\sigma}_{ML} - \sigma \right)$.

Figure 3: Panel (a) compares kernel density estimates of $\sqrt{n} \left( \hat{\mu}_{ML} - \mu \right)$ for different $h_n$ with the standard normal density. Analogously, panel (b) for $\sqrt{n} \left( \hat{\sigma}_{ML} - \sigma \right)$.
<table>
<thead>
<tr>
<th>True $\theta$</th>
<th>$h_n$</th>
<th>$\hat{\mu}_{ML}$</th>
<th>$\hat{\sigma}_{ML}$</th>
<th>$\hat{\mu}_{MoM}$</th>
<th>$\hat{\sigma}_{MoM}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0, 1)</td>
<td>1</td>
<td>$-1.1 \cdot 10^{-3}$</td>
<td>$-8.1 \cdot 10^{-3}$</td>
<td>$-10^{-3}$</td>
<td>$-8.7 \cdot 10^{-3}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.013)</td>
<td>(.074)</td>
<td>(.013)</td>
<td>(.074)</td>
</tr>
<tr>
<td></td>
<td>.5</td>
<td>$-9.2 \cdot 10^{-4}$</td>
<td>$-4.6 \cdot 10^{-3}$</td>
<td>$-10^{-3}$</td>
<td>$-5.2 \cdot 10^{-3}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.012)</td>
<td>(.054)</td>
<td>(.013)</td>
<td>(.054)</td>
</tr>
<tr>
<td></td>
<td>.1</td>
<td>$-7.3 \cdot 10^{-4}$</td>
<td>$-1.9 \cdot 10^{-3}$</td>
<td>$-10^{-3}$</td>
<td>$-2.3 \cdot 10^{-3}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.097)</td>
<td>(.027)</td>
<td>(.013)</td>
<td>(.031)</td>
</tr>
<tr>
<td></td>
<td>.01</td>
<td>$3.6 \cdot 10^{-4}$</td>
<td>$-3.9 \cdot 10^{-3}$</td>
<td>$-10^{-3}$</td>
<td>$-1.7 \cdot 10^{-3}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.065)</td>
<td>(.019)</td>
<td>(.013)</td>
<td>(.022)</td>
</tr>
<tr>
<td>(0, 3)</td>
<td>1</td>
<td>$-7.7 \cdot 10^{-4}$</td>
<td>$-2.8 \cdot 10^{-2}$</td>
<td>$-7.7 \cdot 10^{-4}$</td>
<td>$-3 \cdot 10^{-2}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.308)</td>
<td>(.221)</td>
<td>(.309)</td>
<td>(.222)</td>
</tr>
<tr>
<td></td>
<td>.5</td>
<td>$-1.1 \cdot 10^{-3}$</td>
<td>$-1.4 \cdot 10^{-2}$</td>
<td>$-7.7 \cdot 10^{-4}$</td>
<td>$-1.6 \cdot 10^{-2}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.307)</td>
<td>(.159)</td>
<td>(.309)</td>
<td>(.161)</td>
</tr>
<tr>
<td></td>
<td>.1</td>
<td>$5.3 \cdot 10^{-4}$</td>
<td>$-6.6 \cdot 10^{-3}$</td>
<td>$-7.2 \cdot 10^{-4}$</td>
<td>$-7.8 \cdot 10^{-3}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.292)</td>
<td>(.081)</td>
<td>(.309)</td>
<td>(.091)</td>
</tr>
<tr>
<td>(0, 0.1)</td>
<td>1</td>
<td>$9.5 \cdot 10^{-5}$</td>
<td>$-6.9 \cdot 10^{-4}$</td>
<td>$9.5 \cdot 10^{-5}$</td>
<td>$-2.5 \cdot 10^{-4}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.0103)</td>
<td>(.0073)</td>
<td>(.0103)</td>
<td>(.0074)</td>
</tr>
<tr>
<td></td>
<td>.5</td>
<td>$10^{-5}$</td>
<td>$-3.6 \cdot 10^{-4}$</td>
<td>$9.5 \cdot 10^{-5}$</td>
<td>$-1.7 \cdot 10^{-4}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.0103)</td>
<td>(.0053)</td>
<td>(.0103)</td>
<td>(.0054)</td>
</tr>
<tr>
<td></td>
<td>.1</td>
<td>$9.6 \cdot 10^{-5}$</td>
<td>$-1.7 \cdot 10^{-4}$</td>
<td>$9.5 \cdot 10^{-5}$</td>
<td>$-1.4 \cdot 10^{-4}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.0097)</td>
<td>(.0027)</td>
<td>(.0103)</td>
<td>(.003)</td>
</tr>
<tr>
<td>(2, 1)</td>
<td>1</td>
<td>$2.7 \cdot 10^{-4}$</td>
<td>$-7.9 \cdot 10^{-3}$</td>
<td>$2.6 \cdot 10^{-4}$</td>
<td>$-8.6 \cdot 10^{-3}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.101)</td>
<td>(.074)</td>
<td>(.101)</td>
<td>(.075)</td>
</tr>
<tr>
<td></td>
<td>.5</td>
<td>$2.7 \cdot 10^{-4}$</td>
<td>$-4.5 \cdot 10^{-3}$</td>
<td>$2.6 \cdot 10^{-4}$</td>
<td>$-5 \cdot 10^{-3}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.101)</td>
<td>(.054)</td>
<td>(.101)</td>
<td>(.055)</td>
</tr>
<tr>
<td></td>
<td>.1</td>
<td>$10^{-5}$</td>
<td>$-1.7 \cdot 10^{-3}$</td>
<td>$2.6 \cdot 10^{-4}$</td>
<td>$-1.9 \cdot 10^{-3}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.096)</td>
<td>(.027)</td>
<td>(.101)</td>
<td>(.031)</td>
</tr>
<tr>
<td>(-0.5, 1)</td>
<td>1</td>
<td>$-1.3 \cdot 10^{-3}$</td>
<td>$-8.5 \cdot 10^{-3}$</td>
<td>$-1.3 \cdot 10^{-3}$</td>
<td>$-9.1 \cdot 10^{-3}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.104)</td>
<td>(.074)</td>
<td>(.104)</td>
<td>(.075)</td>
</tr>
<tr>
<td></td>
<td>.5</td>
<td>$-1.1 \cdot 10^{-3}$</td>
<td>$-4.8 \cdot 10^{-3}$</td>
<td>$-1.3 \cdot 10^{-3}$</td>
<td>$-5.3 \cdot 10^{-3}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.103)</td>
<td>(.053)</td>
<td>(.104)</td>
<td>(.054)</td>
</tr>
<tr>
<td></td>
<td>.1</td>
<td>$-1.1 \cdot 10^{-3}$</td>
<td>$-2 \cdot 10^{-3}$</td>
<td>$-1.3 \cdot 10^{-3}$</td>
<td>$-2.1 \cdot 10^{-3}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.097)</td>
<td>(.027)</td>
<td>(.104)</td>
<td>(.03)</td>
</tr>
</tbody>
</table>

Table 3: Empirical bias and RMSE (in parenthesis) for $\nu = 39$ comparing the MLE and the MoM estimator for different true $\theta$ and step size $h_n$.

been excluded in the table.

## 4.2 Continuous sampling

Finally, we discuss the estimators (24) and (25) for the continuous sampling scheme. Of course, continuous sampling is physically impossible. However, if a path is generated by a series representation (according to Massing (2018)), we expect a sufficient number of jumps for estimation. See also Raible (2000) for the normal inverse Gaussian Lévy process.

Let $T = 1$, $\theta = (0, 1)$ and let $\tau = 50, 400$; $87, 300$; $157, 300$ for $\nu = 4, 12, 39$, respectively, be the truncation levels for the random truncated series representation. The different levels of truncation are chosen such that the series contain all jumps up to size $10^{-9}$ for each $\nu$. For each $\nu$ we generate 10,000 paths and compute the continuous sampling estimators $\hat{\mu}_{T,n}$ and $\hat{\sigma}_{T,n}$ for various $n$. Figure 4 plots the sample mean of the estimates versus $\log_{10} n$, both for $\hat{\mu}_{T,n}$ (panel (a)) and $\hat{\sigma}_{T,n}$ (panel (b)). The results illustrate the strong consistency of $\hat{\mu}_{T,n}$ for $\mu$. However, there is evidently a problem for $\hat{\sigma}_{T,n}$ causing the decay. This is due to the fact that on average there are $\tau$ jumps in each path (due to the random truncation). This bounds $\# \{ t \leq T : \Delta X_t \geq 1/n \}$ and, eventually, $\hat{\sigma}_{T,n} \to 0$ a.s. for $n \to \infty$. 
Table 4: Empirical root mean squared deviances between the different ML estimation approaches for different $\hat{h}_n$ and $\nu$. Exceptions occur where Nelder-Mead does not work. These outliers were exclude from analysis. The cases are highlighted with $\ast$.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>$\nu$ =</th>
<th>4</th>
<th>12</th>
<th>39</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCEM vs 1</td>
<td>0.5</td>
<td>2.4 · 10^{-4}</td>
<td>1.6 · 10^{-4}</td>
<td>7.8 · 10^{-5}</td>
</tr>
<tr>
<td>CF-ML vs 1</td>
<td>0.1</td>
<td>3.2 · 10^{-4}</td>
<td>7.7 · 10^{-5}</td>
<td>1.6 · 10^{-4}</td>
</tr>
<tr>
<td>EM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCEM vs 0.1</td>
<td>1</td>
<td>2.4 · 10^{-4}</td>
<td>1.6 · 10^{-4}</td>
<td>7.8 · 10^{-5}</td>
</tr>
<tr>
<td>CF-ML vs 0.1</td>
<td>0.1</td>
<td>6.1 · 10^{-7}</td>
<td>2.8 · 10^{-7}</td>
<td>6.8 · 10^{-8}</td>
</tr>
<tr>
<td>EM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 4: Empirical means of estimates of 10,000 trajectories using (24) and (25) for different $\nu$ and $\theta = (0, 1)$. If we pick $n = 10,000$ such that $\hat{\sigma}_{T,n}$ attains its maximum value, we find the RMSEs given in Table 5 outperforming the MCEM-MLE (see Tables 1 to 3, lines 1 to 3, values in parentheses). In practice, the series representation is not available but these estimators may be an alternative if a very high number of jumps is observed.

5 Conclusion and future work

This paper discusses and proves local asymptotic normality for the Student-Lévy process for high-frequency sampling. We find the rate of convergence and the Fisher information matrix. The LAN implies asymptotic normality and asymptotic efficiency for the maximum likelihood estimator. Additionally, we find that the
Table 5: RMSEs for estimators (24) and (25) for different $\nu$ and $\theta = (0, 1)$.

<table>
<thead>
<tr>
<th>$\nu$</th>
<th>$\hat{\mu}_{T,10^4}$</th>
<th>$\hat{\mu}_{T,10^6}$</th>
<th>$\hat{\sigma}_{T,10^4}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>$8.7 \cdot 10^{-3}$</td>
<td>$1.5 \cdot 10^{-5}$</td>
<td>$1.3 \cdot 10^{-2}$</td>
</tr>
<tr>
<td>12</td>
<td>$1.2 \cdot 10^{-2}$</td>
<td>$1.9 \cdot 10^{-5}$</td>
<td>$9.8 \cdot 10^{-3}$</td>
</tr>
<tr>
<td>39</td>
<td>$1.5 \cdot 10^{-2}$</td>
<td>$2.6 \cdot 10^{-5}$</td>
<td>$8.2 \cdot 10^{-3}$</td>
</tr>
</tbody>
</table>

LAN fails to hold for the Skew Student-Lévy process. We propose and test in simulations a Monte Carlo EM approach for numerical computations, which seems to work well.

In our future research we intend to further investigate estimation of GH Lévy processes. This involves classifying all special cases where a LAN does or does not hold. Furthermore, we aim to estimate the parameter $\nu$, since this is also possible for Student $t$ random numbers. Unfortunately, the density of $Y_t, t \neq 1$ is not available. In order to tackle this issue we plan to use appropriate approximations, e.g., Approximate Bayesian Computation.

Next, it is interesting to apply the discussed procedures to real world data, e.g., financial or physical, high-frequency data and compare how the resulting approach competes with existing ones.

**Acknowledgement**

The author is grateful to Christoph Hanck, Yannick Hoga and Paul Navas Alban as well as the anonymous reviewers for valuable comments which helped to substantially improve this paper. Last but not least, I thank Theresa Kemper for carefully reading my work. Full responsibility is taken for all remaining errors.

**References**


URL: https://academic.oup.com/jfec/article/4/2/275/788320


URL: http://stacks.iop.org/1751-8121/41/i=26/a=265004


URL: http://www.jstor.org/stable/2984875


URL: https://books.google.de/books?id=H0PhBwAAQBAJ


URL: http://scholarsmine.mst.edu/doctoral_dissertations/2274/


URL: http://alea.impa.br/articles/v12/12-32.pdf

URL: https://doi.org/10.1007/978-3-642-33134-3_87


URL: http://dx.doi.org/10.1080/02331888.2014.918980


URL: http://www.sciencedirect.com/science/article/pii/S0167715210003512


URL: https://doi.org/10.1051/ps/2011101


URL: https://amstat.tandfonline.com/doi/abs/10.1080/01621459.1989.10478852


URL: https://doi.org/10.1007/978-1-4684-0377-0_5


URL: https://www.tandfonline.com/doi/abs/10.1198/106186001317115045


URL: https://books.google.de/books?id=AyVeBAAAQBAJ


URL: https://academic.oup.com/biomet/article/81/4/633/224325


URL: http://www.jstor.org/stable/24305551


URL: https://doi.org/10.1007/s00180-018-0814-y


URL: https://www.jstage.jst.go.jp/article/jrss/39/1/39_1_49/-article/-char/ja/


URL: https://doi.org/10.1007/s10463-007-0131-7


URL: https://doi.org/10.1007/978-3-319-12373-8_3


URL: https://books.google.de/books?hl=en&lr=&id=NBawzWoW6C


**URL:** [https://academic.oup.com/comjnl/article/7/4/308/354237](https://academic.oup.com/comjnl/article/7/4/308/354237)


**URL:** [https://www.freidok.uni-freiburg.de/fedora/objects/freidok:51/datastreams/FILE1/content](https://www.freidok.uni-freiburg.de/fedora/objects/freidok:51/datastreams/FILE1/content)


**URL:** [https://books.google.de/books?id=tbZPLquJjSoC](https://books.google.de/books?id=tbZPLquJjSoC)


**URL:** [https://d-nb.info/1010445391/34](https://d-nb.info/1010445391/34)


**URL:** [https://books.google.de/books?id=cOA-vwKJfjIkC](https://books.google.de/books?id=cOA-vwKJfjIkC)


**URL:** [https://www.tandfonline.com/doi/abs/10.1080/01621459.1990.10474930](https://www.tandfonline.com/doi/abs/10.1080/01621459.1990.10474930)


**URL:** [https://freidok.uni-freiburg.de/dnb/download/295](https://freidok.uni-freiburg.de/dnb/download/295)