Statistics for Stochastic Processes

Lecture given in Spring Semester 2020

Jakob Söhl
j.soehl@tudelft.nl
Delft University of Technology

Version: 27 January 2022
Comments are welcome

Contents

1 Diffusion processes

Definition 1.1. A (time-inhomogeneous) diffusion process on $\mathbb{R}$ is a stochastic process $(X_t)_{t\in\mathbb{R}_+}$ solving the stochastic differential equation (SDE)

$$\mathrm{d}X_t = b(X_t, t)\,\mathrm{d}t + \sigma(X_t, t)\,\mathrm{d}W_t, \qquad t \geqslant 0, \tag{1.1}$$

with initial condition $X_0 = X^{(0)}$, where $b : \mathbb{R}\times\mathbb{R}_+ \to \mathbb{R}$, $\sigma : \mathbb{R}\times\mathbb{R}_+ \to \mathbb{R}_+$ and $(W_t)_{t\in\mathbb{R}_+}$ is an one-dimensional Brownian motion.

We call b the drift coefficient and σ the diffusion coefficient (or the volatility). The intuition is that

$$\dot{X}_t = \frac{\mathrm{d}X_t}{\mathrm{d}t} = b(X_t, t) + \sigma(X_t, t)\dot{W}_t,$$

where $\dot{W}_t$ is Gaussian white noise.

The rigorous interpretation of (1.1) is given by integration:

X is a strong solution of the SDE (1.1), where W is defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and $X^{(0)}$ is independent of W on $(\Omega, \mathcal{F}, \mathbb{P})$ if

(a) $(X_t)_{t\in\mathbb{R}_+}$ is adapted to the completion by null sets of $\mathcal{F}_t^0 = \sigma((W_s)_{0\leqslant s\leqslant t}, X^{(0)})$

(b) X is a continuous process

(c) $\mathbb{P}(X_0 = X^{(0)}) = 1$

(d) $\mathbb{P}(\int_0^t(|b(X_s, s)| + |\sigma(X_s, s)|^2)\,\mathrm{d}s < \infty) = 1$ for all $t > 0$

(e) With probability one

$$\forall t \geqslant 0 \quad X_t = X_0 + \int_0^t b(X_s, s)\,\mathrm{d}s + \int_0^t \sigma(X_s, s)\,\mathrm{d}W_s$$

The stochastic integral is to be understood in the Itô sense as the limit in probability of sums

$$\sum_{j=1}^{m} \sigma(X_{t_{j-1}}, t_{j-1})(W_{t_j} - W_{t_{j-1}}),$$

where $0 = t_0 < t_1 < \cdots < t_m = t$ and $\Delta := \max_j |t_j - t_{j-1}| \to 0$.

Theorem 1.2. *Grant the following global Lipschitz and linear growth conditions*

(a) $|b(x,t) - b(y,t)| + |\sigma(x,t) - \sigma(y,t)| \leqslant K|x-y|$

(b) $|b(x,t)| + |\sigma(x,t)| \leqslant K(1 + |x|)$

for all $x, y \in \mathbb{R}$, $t \geqslant 0$ *and some constant* K. *Let* $X^{(0)} \in L^2$. *Then the SDE* (1.1) *has a strong solution, which is unique.*

If we observe the path $(X_t)_{t\in[0,T]}$ (continuous-time observations), then by taking refined partitions we can determine the quadratic variation

$$\int_0^t \sigma(X_s, s)^2 \,\mathrm{d}s$$

for all $t \in [0, T]$,

$$\sum_{j=1}^m (X_{t_j} - X_{t_{j-1}})^2 \to \int_0^t \sigma(X_s, s)^2 \,\mathrm{d}s$$

almost surely as $\Delta \to 0$ (see Theorem I.2.4 and the remarks thereafter in [20]). Thus $\sigma(X_t, t)^2$ can be identified by taking the derivative at time $t \in [0, T]$. If X does not visit x at time t, then there is no direct information on $\sigma(x, t)^2$ contained in the sample path. Continuous-time observations identify the diffusion coefficient as far as possible and the main interest is in the drift estimation. The main tool for identifying the drift is the Girsanov theorem.

Theorem 1.3. *(Girsanov theorem, Theorem 7.19 in [19]) Let $(X_t)_{t\in[0,T]}$ and $(Y_t)_{t\in[0,T]}$ be two diffusion processes with*

$$\begin{aligned}
\mathrm{d}X_t &= b_X(X_t, t)\,\mathrm{d}t + \sigma(X_t, t)\,\mathrm{d}W_t\\
\mathrm{d}Y_t &= b_Y(Y_t, t)\,\mathrm{d}t + \sigma(Y_t, t)\,\mathrm{d}W_t
\end{aligned}$$

and $X_0 = Y_0$ a.s. Let the coefficients of Y satisfy the global Lipschitz and linear growth conditions from Theorem 1.2 and let $b_X(x, t) = b_Y(x, t)$ for x and t such that $\sigma(x, t) = 0$. If

$$\begin{aligned}
\mathbb{P}\left(\int_0^T \frac{b_X(X_t, t)^2 + b_Y(X_t, t)^2}{\sigma(X_t, t)^2} \mathbb{1}_{\{\sigma(X_t,t)>0\}} \,\mathrm{d}t < \infty\right) &= 1,\\
\mathbb{P}\left(\int_0^T \frac{b_X(Y_t, t)^2 + b_Y(Y_t, t)^2}{\sigma(Y_t, t)^2} \mathbb{1}_{\{\sigma(Y_t,t)>0\}} \,\mathrm{d}t < \infty\right) &= 1,
\end{aligned}$$

then $\mathbb{P}_T^X$ and $\mathbb{P}_T^Y$ are equivalent and the Radon–Nikodym derivative is given by

$$\begin{aligned}
&\frac{\mathrm{d}\,\mathbb{P}_T^Y}{\mathrm{d}\,\mathbb{P}_T^X}((X_t)_{t\in[0,T]})\\
&\quad= \exp\left(\int_0^T \frac{(b_Y - b_X)(X_t, t)}{\sigma(X_t, t)^2} \mathbb{1}_{\{\sigma(X_t,t)>0\}} \,\mathrm{d}X_t - \frac{1}{2}\int_0^T \frac{(b_Y^2 - b_X^2)(X_t, t)}{\sigma(X_t, t)^2} \mathbb{1}_{\{\sigma(X_t,t)>0\}} \,\mathrm{d}t\right).
\end{aligned}$$

Examples. (a) Brownian motion with drift:

Let $b_X(x, t) = b_X(t)$, $b_Y(x, t) = b_Y(t)$, $\sigma(x, t) = \sigma > 0$ and $X^{(0)} = 0$. Then

$$X_t = \int_0^t b_X(s)\,\mathrm{d}s + \sigma W_t, \quad Y_t = \int_0^t b_Y(s)\,\mathrm{d}s + \sigma W_t$$

and the formula for the Radon–Nikodym derivative gives

$$\frac{\mathrm{d}\,\mathbb{P}_T^Y}{\mathrm{d}\,\mathbb{P}_T^X}((X_t)_{t\in[0,T]}) = \exp\left(\int_0^T \frac{(b_Y - b_X)(t)}{\sigma^2}\,\mathrm{d}X_t - \frac{1}{2}\int_0^T \frac{(b_Y^2 - b_X^2)(t)}{\sigma^2}\,\mathrm{d}t\right).$$

If we further specialise to $Y_t = \vartheta t + \sigma W_t$ and $X_t = \sigma W_t$, then

$$\frac{\mathrm{d}\,\mathbb{P}_T^Y}{\mathrm{d}\,\mathbb{P}_T^X}((X_t)_{t\in[0,T]}) = \exp\left(\frac{\vartheta}{\sigma^2}X_T - \frac{\vartheta^2 T}{2\sigma^2}\right) = \exp\left(-\frac{T}{2\sigma^2}\left(\frac{X_T}{T} - \vartheta\right)^2 + \frac{X_T^2}{2\sigma^2 T}\right).$$

We see that X_T is a sufficient statistic, i.e., for all statistical purposes it suffices to use X_T instead of the whole sample path $(X_t)_{t\in[0,T]}$. The maximum likelihood estimator (MLE) of $X_t = \vartheta t + \sigma W_t$ with ϑ unknown is given by $\widehat{\vartheta}_{\mathrm{MLE}} = X_T/T \sim N(\vartheta, \sigma^2/T)$. We have $\widehat{\vartheta}_{\mathrm{MLE}} \xrightarrow{d} \vartheta$ if and only if $T \to \infty$.

(b) Ornstein–Uhlenbeck process:
The Ornstein–Uhlenbeck process is the solution of the SDE

$$\begin{aligned} \mathrm{d}X_t &= aX_t\,\mathrm{d}t + \sigma\,\mathrm{d}W_t,\\ X_0 &= X^{(0)}. \end{aligned}$$

The SDE can be solved by variation of constants

$$X_t = e^{at}X^{(0)} + \int_0^t e^{a(t-s)}\sigma\,\mathrm{d}W_s. \tag{1.2}$$

Remark. Integrals of the form $\int_a^b f(s)\,\mathrm{d}W_s$, $f \in L^2([a,b])$, are called *Wiener integrals.* We have

$$\begin{aligned} \int_a^b f(s)\,\mathrm{d}W_s &\sim N(0, \|f\|^2_{L^2([a,b])}),\\ \mathbb{E}\left[\int_a^b f(s)\,\mathrm{d}W_s \int_a^b g(s)\,\mathrm{d}W_s\right] &= \int_a^b f(s)g(s)\,\mathrm{d}s, \quad f,g \in L^2([a,b]). \end{aligned}$$

If $a < 0$, then it follows from (1.2) that $X_t \xrightarrow{d} N(0, -\sigma^2/2a)$ as $t \to \infty$. If $X^{(0)}$ is Gaussian or deterministic, then (X_t) is a Gaussian process. Take $b_Y(x,t) = ax$, $b_X(x,t) = 0$. For $X^{(0)} \in L^2$ and $\sigma > 0$ the conditions of the Girsanov theorem are satisfied and it yields

$$\frac{\mathrm{d}\,\mathbb{P}_T^a}{\mathrm{d}\,\mathbb{P}_T^0}((X_t)_{t\in[0,T]}) := \frac{\mathrm{d}\,\mathbb{P}_T^Y}{\mathrm{d}\,\mathbb{P}_T^X}((X_t)_{t\in[0,T]}) = \exp\left(\int_0^T \frac{aX_s}{\sigma^2}\,\mathrm{d}X_s - \frac{1}{2}\int_0^T \frac{a^2X_s^2}{\sigma^2}\,\mathrm{d}s\right).$$

By taking the derivative of the log-likelihood

$$\frac{\mathrm{d}}{\mathrm{d}a}\log\left(\frac{\mathrm{d}\,\mathbb{P}_T^a}{\mathrm{d}\,\mathbb{P}_T^0}((X_t)_{t\in[0,T]})\right) = \int_0^T \frac{X_s}{\sigma^2}\,\mathrm{d}X_s - a\int_0^T \frac{X_s^2}{\sigma^2}\,\mathrm{d}s$$

we determine the MLE to be

$$\widehat{a}_T = \frac{\int_0^T X_s\,\mathrm{d}X_s}{\int_0^T X_s^2\,\mathrm{d}s}.$$

Under $\mathbb{P}_T^a$

$$\widehat{a}_T = \frac{\int_0^T X_s(aX_s\,\mathrm{d}s + \sigma\,\mathrm{d}W_s)}{\int_0^T X_s^2\,\mathrm{d}s} = a + \frac{\int_0^T X_s\sigma\,\mathrm{d}W_s}{\int_0^T X_s^2\,\mathrm{d}s}.$$

For $a < 0$ it can be shown $\sqrt{T}(\widehat{a}_T - a) \xrightarrow{d} N(0, -2a)$, see Example 5.2.5 in [17].

(c) Linear factor model:
We consider the SDE

$$\mathrm{d}X_t = \vartheta b(X_t, t)\,\mathrm{d}t + \sigma(X_t, t)\,\mathrm{d}W_t,$$
$$X_0 = X^{(0)},$$

with $\sigma(x,t) > 0$ for all x and t. The unknown parameter is $\vartheta \in \Theta$ and we assume $0 \in \Theta$. Let $X^{(0)} \in L^2$ and b, σ be such that the conditions of the Girsanov theorem are satisfied. Then we have

$$\frac{\mathrm{d}\,\mathbb{P}_T^{\vartheta}}{\mathrm{d}\,\mathbb{P}_T^{0}}((X_t)_{t\in[0,T]}) = \exp\left(\int_0^T \frac{\vartheta b(X_t,t)}{\sigma(X_t,t)^2}\,\mathrm{d}X_t - \frac{1}{2}\int_0^T \frac{\vartheta^2 b(X_t,t)^2}{\sigma(X_t,t)^2}\,\mathrm{d}t\right).$$

We take the derivative of the log-likelihood

$$\frac{\mathrm{d}}{\mathrm{d}\vartheta}\log\left(\frac{\mathrm{d}\,\mathbb{P}_T^{\vartheta}}{\mathrm{d}\,\mathbb{P}_T^{0}}((X_t)_{t\in[0,T]})\right) = \int_0^T \frac{b(X_t,t)}{\sigma(X_t,t)^2}\,\mathrm{d}X_t - \vartheta\int_0^T \frac{b(X_t,t)^2}{\sigma(X_t,t)^2}\,\mathrm{d}t.$$

The MLE is given by

$$\widehat{\vartheta}_T = \left(\int_0^T \frac{b(X_t,t)}{\sigma(X_t,t)^2}\,\mathrm{d}X_t\right)\Big/\left(\int_0^T \frac{b(X_t,t)^2}{\sigma(X_t,t)^2}\,\mathrm{d}t\right).$$

Under $\mathbb{P}_T^{\vartheta}$

$$\begin{aligned}\widehat{\vartheta}_T &= \left(\int_0^T \vartheta\frac{b(X_t,t)^2}{\sigma(X_t,t)^2}\,\mathrm{d}t + \int_0^T \frac{b(X_t,t)}{\sigma(X_t,t)}\,\mathrm{d}W_t\right)\Big/\left(\int_0^T \frac{b(X_t,t)^2}{\sigma(X_t,t)^2}\,\mathrm{d}t\right)\\ &= \vartheta + \left(\int_0^T \frac{b(X_t,t)}{\sigma(X_t,t)}\,\mathrm{d}W_t\right)\Big/\left(\int_0^T \frac{b(X_t,t)^2}{\sigma(X_t,t)^2}\,\mathrm{d}t\right).\end{aligned}$$

On appropriate assumptions the estimation error decays with a $\sqrt{T}$-rate or even a CLT holds for the estimator.

Remark. Let X be a solution of $\mathrm{d}X_t = b(X_t,t)\,\mathrm{d}t + \sigma(X_t,t)\,\mathrm{d}W_t$ and $f : \mathbb{R}\times[0,\infty) \to \mathbb{R}$ such that $\partial f/\partial x$, $\partial^2 f/\partial x^2$, $\partial f/\partial t$ exist and are continuous. Then the *Itô formula* holds

$$f(X_t,t) = f(X_0,0) + \int_0^t \frac{\partial}{\partial t}f(X_s,s)\,\mathrm{d}s + \int_0^t \frac{\partial}{\partial x}f(X_s,s)\,\mathrm{d}X_s + \frac{1}{2}\int_0^t \frac{\partial^2}{\partial x^2}f(X_s,s)\sigma(X_s,s)^2\,\mathrm{d}s.$$

2 Nonparametric drift estimation with continuous-time observations

We consider the SDE

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t, \qquad t \geqslant 0, \tag{2.1}$$

and our aim is the nonparametric estimation of b. We suppose that we observe the whole sample path X_t, $t \in [0,T]$, up to time T (continuous-time observations). To get an intuition we analyse rescaled increments

$$\frac{X_\Delta - X_0}{\Delta} = \underbrace{\frac{1}{\Delta}\int_0^\Delta b(X_s)\,\mathrm{d}s}_{\sim b(X_0) \text{ if } b \text{ cts.}} + \underbrace{\frac{1}{\Delta}\int_0^\Delta \sigma(X_s)\,\mathrm{d}W_s}_{\mathbb{E}[\dots]=0 \text{ if } \sigma \text{ bounded}}.$$

We see

$$\mathbb{E}\left[\tfrac{1}{\Delta}(X_{t+\Delta}-X_t)\,\middle|\,X_t=x\right]\sim b(x)$$

for $\Delta>0$ small. The same holds if we condition on a small neighbourhood

$$\mathbb{E}\left[\tfrac{1}{\Delta}(X_{t+\Delta}-X_t)\,\middle|\,x-h\leqslant X_t\leqslant x+h\right]\sim b(x).$$

Letting $\Delta\to 0$ we obtain heuristically

$$\frac{\int_0^T \frac{\mathrm{d}X_t}{\mathrm{d}t}\,\mathbb{1}_{[x-h,x+h]}(X_t)\,\mathrm{d}t}{\int_0^T \mathbb{1}_{[x-h,x+h]}(X_t)\,\mathrm{d}t}\sim b(x).$$

This motivates the estimator

$$\widehat{b}_T(x,h)=\frac{\int_0^T \mathbb{1}_{[x-h,x+h]}(X_t)\,\mathrm{d}X_t}{\int_0^T \mathbb{1}_{[x-h,x+h]}(X_t)\,\mathrm{d}t}\sim b(x).$$

We decompose the error

$$|\widehat{b}_T(x,h)-b(x)|\leqslant\underbrace{\left|\frac{\int_0^T \mathbb{1}_{[x-h,x+h]}(X_t)(b(X_t)-b(x))\,\mathrm{d}t}{\int_0^T \mathbb{1}_{[x-h,x+h]}(X_t)\,\mathrm{d}t}\right|}_{\text{bias part } B_{x,h}}+\underbrace{\left|\frac{\int_0^T \mathbb{1}_{[x-h,x+h]}(X_t)\sigma(X_t)\,\mathrm{d}W_t}{\int_0^T \mathbb{1}_{[x-h,x+h]}(X_t)\,\mathrm{d}t}\right|}_{\text{variance part } V_{x,h}}.$$

In order to control the bias part $B_{x,h}$ we assume Hölder continuity of b. Let there be $\alpha\in(0,1]$ and $R>0$ such that for all $x,y\in\mathbb{R}$

$$|b(x)-b(y)|\leqslant R|x-y|^\alpha.$$

For all $x\in\mathbb{R}$ this yields the bound

$$B_{x,h}\leqslant Rh^\alpha.$$

We simplify the analysis of the variance part $V_{x,h}$ by assuming that X is stationary.

Definition 2.1. Let $\mathcal{T}\subseteq\mathbb{R}$ be such that $s,t\in\mathcal{T}$ implies $s+t\in\mathcal{T}$. A stochastic process $(X_t)_{t\in\mathcal{T}}$ is called *stationary* if

$$\forall n\in\mathbb{N},\ t_1,\dots,t_n,t\in\mathcal{T}:\quad (X_{t_1},\dots,X_{t_n})\overset{d}{=}(X_{t_1+t},\dots,X_{t_n+t}).$$

If X is a stationary solution* of an SDE, then the distribution of any X_t, $t\in\mathcal{T}$, (and thus of all X_t) is called an *invariant measure* of the SDE.

Remark. Let $f(X_t,t)$ be adapted. Then we have the *Itô isometry*

$$\mathbb{E}\left[\left(\int_a^b f(X_t,t)\,\mathrm{d}W_t\right)^2\right]=\mathbb{E}\left[\int_a^b f(X_t,t)^2\,\mathrm{d}t\right]$$

provided the right hand side is finite.

*Solution can be read throughout as either strong or weak solution.

For analysing the variance part we suppose that X is a stationary solution. Furthermore, we assume that a Lebesgue density μ of the corresponding invariant measure exists. For the numerator of the variance part we have by the Itô isometry

$$\begin{aligned}\mathbb{E}\left[\left(\int_0^T \mathbb{1}_{[x-h,x+h]}(X_t)\sigma(X_t)\,\mathrm{d}W_t\right)^2\right] &= \int_0^T \mathbb{E}\left[\mathbb{1}_{[x-h,x+h]}(X_t)\sigma(X_t)^2\right]\,\mathrm{d}t \\ &= T\,\mathbb{E}\left[\mathbb{1}_{[x-h,x+h]}(X_0)\sigma(X_0)^2\right] \\ &= T\int_{x-h}^{x+h} \sigma(y)^2\mu(y)\,\mathrm{d}y \\ &\leqslant 2Th\|\sigma^2\mu\|_\infty \sim Th,\end{aligned}$$

where finiteness of $\|\sigma^2\mu\|_\infty$ was assumed. Turning to the denominator we see

$$\begin{aligned}\mathbb{E}\left[\int_0^T \mathbb{1}_{[x-h,x+h]}(X_t)\,\mathrm{d}t\right] &= T\,\mathbb{E}\left[\mathbb{1}_{[x-h,x+h]}(X_0)\right] \\ &= 2Th\frac{1}{2h}\int_{x-h}^{x+h}\mu(y)\,\mathrm{d}y \sim Th\end{aligned}$$

if μ and $1/\mu$ are locally bounded. We hope that the denominator concentrates around its expectation such that the variance part is of order $O_{\mathbb{P}}\left(\frac{\sqrt{Th}}{Th}\right) = O_{\mathbb{P}}\left(\frac{1}{\sqrt{Th}}\right)$.

Remark. For random variables $(X_\alpha)_{\alpha\in A}$ we write $X_\alpha = O_{\mathbb{P}}(1)$ if for all $\varepsilon > 0$ there exists $M > 0$ such that $\sup_{\alpha\in A}\mathbb{P}(|X_\alpha| > M) < \varepsilon$. Given random variables $(R_\alpha)_{\alpha\in A}$ we further introduce the notation $X_\alpha = O_{\mathbb{P}}(R_\alpha)$ if $X_\alpha = R_\alpha Y_\alpha$ and $Y_\alpha = O_{\mathbb{P}}(1)$.

Proposition 2.2. *(See Lemma 9 and Theorem 18 in [23, Chapter I]) Let b, σ and $1/\sigma$ be measurable and locally bounded functions. Let*

$$\int_0^x \exp\left(-\int_0^y \frac{2b(z)}{\sigma^2(z)}\,\mathrm{d}z\right)\mathrm{d}y \to \pm\infty$$

as $x \to \pm\infty$ and

$$G := \int_{-\infty}^{\infty}\frac{1}{\sigma^2(y)}\exp\left(\int_0^y \frac{2b(z)}{\sigma^2(z)}\,\mathrm{d}z\right)\mathrm{d}y < \infty.$$

(a) If the SDE (2.1) *has a solution for every initial distribution,*† *then there exists a stationary solution of the SDE.*

(b) Let X be a stationary solution of the SDE (2.1)*. Then the invariant measure of the SDE is unique and absolutely continuous with respect to the Lebesgue measure. Its density is given by*

$$\mu(x) = \frac{1}{G\sigma^2(x)}\exp\left(\int_0^x \frac{2b(y)}{\sigma^2(y)}\,\mathrm{d}y\right).$$

Proposition 2.3. *Let b and σ be measurable and locally bounded and let $\inf_{x\in\mathbb{R}}\sigma^2(x) \geqslant \underline{\sigma}^2 > 0$. Let there be $M, \gamma > 0$ such that $\operatorname{sign}(x)\frac{2b}{\sigma^2}(x) \leqslant -\gamma$ for all x with $|x| \geqslant M$. Let X be a stationary*

†The assumptions of the proposition ensure that for every initial distribution there exists a weak solution that is unique in the sense of probability in law, see [16, Section 5.5.B].

solution of the SDE (2.1)*. Then the invariant measure* μ *is unique and there exists a constant* C *such that for all functions* $f \in L^1(\mu)$ *with* $\mathbb{E}[f(X_0)] = 0$ *we have*

$$\mathbb{E}\left[\left(\int_0^T f(X_t)\,\mathrm{d}t\right)^2\right] \leqslant C(1+T)\left(\|f\|_{L^1(\mu)}^2 + \sup_{|x|\geqslant M} |f(x)|^2\right).$$

The constant C *depends only on* M*,* γ*,* G*,* $\underline{\sigma}^2$ *and* $\sup_{|x|\leqslant M} |b(x)|$*.*

Proof. (a) (invariant density) We are in the setting of Proposition 2.2(b).

(b) (initial bound) We start by considering the *Itô formula* (Itô–Tanaka formula)

$$\begin{aligned} \mathrm{d}F(X_t) &= F'(X_t)\,\mathrm{d}X_t + \frac{1}{2}F''(X_t)\sigma^2(X_t)\,\mathrm{d}t \\ &= \underbrace{\left(F'(X_t)b(X_t) + \frac{1}{2}F''(X_t)\sigma^2(X_t)\right)}_{:=AF(X_t)}\mathrm{d}t + F'(X_t)\sigma(X_t)\,\mathrm{d}W_t. \end{aligned}$$

Let $S(x) = \frac{1}{2}\sigma^2(x)\mu(x) = \frac{1}{2G}\exp\left(\int_0^x \frac{2b(y)}{\sigma^2(y)}\,\mathrm{d}y\right)$. This yields

$$AF(x) = b(x)F'(x) + \frac{1}{2}\sigma^2(x)F''(x) = \frac{1}{\mu(x)}\left(S(x)F'(x)\right)'.$$

We call A *infinitesimal generator*. We obtain $\int_0^T AF(X_t)\,\mathrm{d}t = F(X_T) - F(X_0) - \int_0^T F'(X_t)\sigma(X_t)\,\mathrm{d}W_t$. Suppose we can find F such that $AF = f$. Then

$$\begin{aligned} \mathbb{E}\left[\left(\int_0^T f(X_t)\,\mathrm{d}t\right)^2\right] &\leqslant 3\,\mathbb{E}[F(X_T)^2] + 3\,\mathbb{E}[F(X_0)^2] + 3\,\mathbb{E}\left[\left(\int_0^T F'(X_t)\sigma(X_t)\,\mathrm{d}W_t\right)^2\right] \\ &= 6\,\mathbb{E}[F(X_0)^2] + 3T\,\mathbb{E}[F'(X_0)^2\sigma(X_0)^2]. \end{aligned} \tag{2.2}$$

(c) (finding F) We define

$$F(x) := \int_0^x \frac{2}{\sigma^2(y)\mu(y)}\left(\int_{-\infty}^y f(z)\mu(z)\,\mathrm{d}z\right)\mathrm{d}y,$$

where μ denotes the Lebesgue density of the invariant measure. To check that $AF = f$ we calculate the first two derivatives

$$\begin{aligned} F'(x) &= \frac{2}{\sigma^2(x)\mu(x)}\int_{-\infty}^x f(z)\mu(z)\,\mathrm{d}z \\ &= 2\int_{-\infty}^x \frac{f(z)}{\sigma^2(z)}\exp\left(-\int_z^x \frac{2b}{\sigma^2}(y)\,\mathrm{d}y\right)\mathrm{d}z, \\ F''(x) &= \frac{2f(x)}{\sigma^2(x)} + 2\int_{-\infty}^x \frac{f(z)}{\sigma^2(z)}\left(-\frac{2b}{\sigma^2}(x)\right)\exp\left(-\int_z^x \frac{2b}{\sigma^2}(y)\,\mathrm{d}y\right)\mathrm{d}z. \end{aligned}$$

We verify

$$AF(x) = \left(\frac{\sigma^2}{2}F'' + bF'\right)(x) = f(x) - b(x)F'(x) + b(x)F'(x) = f(x).$$

(d) (bounding $\mathbb{E}[F'(X_0)^2\sigma(X_0)^2]$) For $x \leqslant -M$ we obtain

$$\begin{aligned}
|F'(x)| &= 2\left|\int_{-\infty}^{x} \frac{f(z)}{\sigma^2(z)} \exp\left(-\int_z^x \frac{2b}{\sigma^2}(y)\,\mathrm{d}y\right)\,\mathrm{d}z\right| \\
&\leqslant 2\int_{-\infty}^{x} \frac{|f(z)|}{\sigma^2(z)} \exp\left(-(x-z)\gamma\right)\,\mathrm{d}z \\
&\leqslant C \sup_{x\leqslant -M} |f(x)|.
\end{aligned}$$

Using $\int_{-\infty}^{x} f(z)\mu(z)\,\mathrm{d}z = -\int_x^\infty f(z)\mu(z)\,\mathrm{d}z$ we likewise obtain for $x \geqslant M$

$$\begin{aligned}
|F'(x)| &= 2\left|\int_{x}^{\infty} \frac{f(z)}{\sigma^2(z)} \exp\left(\int_x^z \frac{2b}{\sigma^2}(y)\,\mathrm{d}y\right)\,\mathrm{d}z\right| \\
&\leqslant 2\int_{x}^{\infty} \frac{|f(z)|}{\sigma^2(z)} \exp\left(-(z-x)\gamma\right)\,\mathrm{d}z \\
&\leqslant C \sup_{x\geqslant M} |f(x)|.
\end{aligned}$$

We conclude that

$$\sup_{|x|\geqslant M} |F'(x)| \leqslant C \sup_{|x|\geqslant M} |f(x)|.$$

With this preparation we bound

$$\begin{aligned}
\mathbb{E}[F'(X_0)^2\sigma(X_0)^2] &= \int_{\mathbb{R}} F'(x)^2\sigma(x)^2\mu(x)\,\mathrm{d}x \\
&\leqslant \int_{-M}^{M} \frac{4}{\sigma(x)^2\mu(x)} \left(\int_{-\infty}^{x} f(z)\mu(z)\,\mathrm{d}z\right)^2 \mathrm{d}x \\
&\quad + C^2 \sup_{|x|\geqslant M} |f(x)|^2 \int_{|x|\geqslant M} \sigma(x)^2\mu(x)\,\mathrm{d}x \\
&\leqslant \|f\|_{L^1(\mu)}^2 \int_{-M}^{M} 4G \exp\left(-\int_0^x \frac{2b}{\sigma^2}(y)\,\mathrm{d}y\right)\mathrm{d}x \\
&\quad + C^2 \sup_{|x|\geqslant M} |f(x)|^2 \int_{|x|\geqslant M} \frac{1}{G} \exp\left(\int_0^x \frac{2b}{\sigma^2}(y)\,\mathrm{d}y\right)\mathrm{d}x \\
&\leqslant C'\left(\|f\|_{L^1(\mu)}^2 + \sup_{|x|\geqslant M} |f(x)|^2\right).
\end{aligned} \tag{2.3}$$

(e) (bounding $\mathbb{E}[F(X_0)^2]$) We can bound $|F(x)|$ by

$$\begin{aligned}
|F(x)| &\leqslant \sup_{x\in[-M,M]} |F(x)| + \max(|x|-M,0) \sup_{|x|\geqslant M} |F'(x)| \\
&\leqslant M \sup_{x\in[-M,M]} \frac{2}{\sigma^2(x)\mu(x)} \left|\int_{-\infty}^{x} f(z)\mu(z)\,\mathrm{d}z\right| + |x| \sup_{|x|\geqslant M} |F'(x)| \\
&\leqslant 2M\|f\|_{L^1(\mu)} \sup_{x\in[-M,M]} G\exp\left(-\int_0^x \frac{2b}{\sigma^2}(y)\,\mathrm{d}y\right) + C|x| \sup_{|x|\geqslant M} |f(x)| \\
&\leqslant C''\|f\|_{L^1(\mu)} + C|x| \sup_{|x|\geqslant M} |f(x)|.
\end{aligned}$$

By the exponential decay of μ we see that $\mathbb{E}[X_0^2]$ is bounded and obtain

$$\begin{aligned}\mathbb{E}[F(X_0)^2] &\leqslant 2C''^2\|f\|_{L^1(\mu)}^2 + 2C^2\,\mathbb{E}[X_0^2]\sup_{|x|\geqslant M}|f(x)|^2\\ &\leqslant C'''\left(\|f\|_{L^1(\mu)}^2 + \sup_{|x|\geqslant M}|f(x)|^2\right).\end{aligned} \tag{2.4}$$

The proposition follows by combining (2.2), (2.3) and (2.4). □

Let σ, b and X be as in the previous proposition. Then μ is bounded and the proposition applies to

$$f := \mathbb{1}_{[x-h,x+h]} - \mathbb{E}[\mathbb{1}_{[x-h,x+h]}(X_0)]$$

since

$$\begin{aligned}\mathbb{E}[|f(X_0)|] &= \mathbb{E}\left[\left|\mathbb{1}_{[x-h,x+h]}(X_0) - \mathbb{E}[\mathbb{1}_{[x-h,x+h]}(X_0)]\right|\right]\\ &\leqslant 2\,\mathbb{E}[\mathbb{1}_{[x-h,x+h]}(X_0)] \leqslant 4h\|\mu\|_\infty\end{aligned}$$

and $\mathbb{E}[f(X_0)] = 0$. Let I be a closed interval in $(-M, M)$. For $x \in I$ and $h > 0$ small enough

$$\sup_{|y|\geqslant M}|f(y)| = \mathbb{E}[\mathbb{1}_{[x-h,x+h]}(X_0)] \leqslant 2h\|\mu\|_\infty.$$

For $h > 0$ small enough we obtain

$$\operatorname{Var}\left(\int_0^T \mathbb{1}_{[x-h,x+h]}(X_t)\,\mathrm{d}t\right) = \mathbb{E}\left[\left(\int_0^T f(X_t)\,\mathrm{d}t\right)^2\right] \leqslant C(1+T)\cdot 20h^2\|\mu\|_\infty^2$$

It follows for $T \geqslant 1$ and for some constant $C' > 0$

$$\operatorname{Var}\left(\frac{1}{Th}\int_0^T \mathbb{1}_{[x-h,x+h]}(X_t)\,\mathrm{d}t\right) \leqslant \frac{C'}{T} \to 0 \tag{2.5}$$

as $T \to \infty$. Furthermore, $1/\mu$ is locally bounded such that for some $C'' > 0$

$$\mathbb{E}\left[\int_0^T \mathbb{1}_{[x-h,x+h]}(X_t)\,\mathrm{d}t\right] \geqslant C''Th \quad\Longrightarrow\quad \mathbb{E}\left[\frac{1}{Th}\int_0^T \mathbb{1}_{[x-h,x+h]}(X_t)\,\mathrm{d}t\right] \geqslant C'' > 0.$$

Consequently

$$\mathbb{P}\left(\frac{1}{Th}\int_0^T \mathbb{1}_{[x-h,x+h]}(X_t)\,\mathrm{d}t \geqslant \frac{C''}{2}\right) \to 1.$$

We conclude $V_{x,h} = O_{\mathbb{P}}\left(\frac{\sqrt{Th}}{Th}\right) = O_{\mathbb{P}}\left(\frac{1}{\sqrt{Th}}\right)$ and obtain the following theorem.

Theorem 2.4. *Let b be Hölder continuous of exponent $\alpha \in (0,1]$ and σ be measurable and locally bounded with $\inf_{x\in\mathbb{R}}\sigma^2(x) \geqslant \underline{\sigma}^2 > 0$. Let there be $M, \gamma > 0$ such that $\operatorname{sign}(x)\frac{2b}{\sigma^2}(x) \leqslant -\gamma$ for all x with $|x| \geqslant M$. Let X be a stationary solution and I a compact interval. Then uniformly for $x \in I$ we have*

$$|\widehat{b}_T(x,h) - b(x)| \leqslant Rh^\alpha + O_{\mathbb{P}}\left(\frac{1}{\sqrt{Th}}\right).$$

In particular, $\widehat{b}_T(x,h)$ is a consistent estimator of $b(x)$ if $h \to 0$ and $Th \to \infty$.

Corollary 2.5. *The choice $h \sim T^{-\frac{1}{2\alpha+1}}$ yields*

$$|\widehat{b}_T(x,h) - b(x)| = O_{\mathbb{P}}\left(T^{-\frac{\alpha}{2\alpha+1}}\right).$$

3 Nonparametric estimation of the invariant density with continuous-time observations

We consider

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t, \quad t \geqslant 0,$$

where b and σ are as in Proposition 2.3.

Definition 3.1. For a Borel set A define $\mu_T(A) = \int_0^T \mathbb{1}_A(X_t)\,\mathrm{d}t$. The Lebesgue density L_T of μ_T is called local time of X at time T (see [3, 20]). For all positive Borel measurable f we have $\int_0^T f(X_t)\,\mathrm{d}t = \int_{\mathbb{R}} f(x)L_T(x)\,\mathrm{d}x$.

This definition differs from the usual definition in the above and in other literature, where it is common to call $\sigma(x)^2 L_T(x)$ the local time.

There exists a version of the local time $L_T(x)$ such that almost surely

$$L_T(x) = \lim_{h\to 0} \frac{1}{h}\int_0^T \mathbb{1}_{[x,x+h)}(X_t)\,\mathrm{d}t$$

for every x and T (Corollary VI.1.9 in [20]).

Let σ be a càdlàg function (right-continuous with left limits). Then the invariant density μ is càdlàg, too. We estimate the invariant density by the normalised local time

$$\widehat{\mu}_T(x) := \frac{1}{T} L_T(x).$$

Let X be a stationary solution. We rewrite

$$\begin{aligned} |\widehat{\mu}_T(x) - \mu(x)| &= \left| \widehat{\mu}_T(x) - \lim_{h\to 0}\frac{1}{h}\int_x^{x+h} \mu(y)\,\mathrm{d}y \right| \\ &= \left| \lim_{h\to 0} \underbrace{\frac{1}{Th}\int_0^T \left(\mathbb{1}_{[x,x+h)}(X_t) - \mathbb{E}[\mathbb{1}_{[x,x+h)}(X_t)]\right)\mathrm{d}t}_{:=\mathcal{E}_{x,h,T}} \right|. \end{aligned}$$

As in (2.5) in the last section we deduce as $T\to\infty$ and for $h>0$ small enough

$$\operatorname{Var}(\mathcal{E}_{x,h,T}) \leqslant \frac{C}{T}$$

for some constant $C>0$. We obtain the following theorem.

Theorem 3.2. *Let b be a measurable, locally bounded function and σ a càdlàg function with $\inf_{x\in\mathbb{R}} \sigma^2(x) \geqslant \underline{\sigma}^2 > 0$. Let there be $M, \gamma > 0$ such that $\operatorname{sign}(x)\frac{2b}{\sigma^2}(x) \leqslant -\gamma$ for all x with $|x| \geqslant M$. Let X be a stationary solution and I a compact interval. Then uniformly for $x \in I$ we have*

$$|\widehat{\mu}_T(x) - \mu(x)| = O_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\right).$$

The invariant density can be estimated nonparametrically with a $\sqrt{T}$-rate.

4 Nonparametric volatility estimation with high-frequency data

4.1 Introduction

We consider the diffusion process

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t.$$

The observations are given by

$$X_0, X_\Delta, X_{2\Delta}, \dots, X_{N\Delta}.$$

We will base our estimator on the increments. To get an intuition we will analyse the approximate size of the different terms in the rescaled increments

$$\frac{X_\Delta - X_0}{\Delta} = \underbrace{\frac{1}{\Delta}\int_0^\Delta b(X_s)\,\mathrm{d}s}_{\sim b(X_0) \text{ if } b \text{ cts.}} + \underbrace{\frac{1}{\Delta}\int_0^\Delta \sigma(X_s)\,\mathrm{d}W_s}_{\substack{\mathbb{E}[\dots]=0 \text{ if } \mathbb{E}[\int_0^\Delta \sigma(X_s)^2\,\mathrm{d}s]<\infty,\\ \text{in particular if } \sigma \text{ is bounded}}}. \tag{4.1}$$

For the estimation of σ^2 we consider squared increments

$$\frac{(X_\Delta - X_0)^2}{\Delta} = \underbrace{\frac{1}{\Delta}\Big(\int_0^\Delta b(X_s)\,\mathrm{d}s\Big)^2}_{\sim\Delta} + 2\underbrace{\frac{1}{\Delta}\int_0^\Delta b(X_s)\,\mathrm{d}s}_{\sim 1}\underbrace{\int_0^\Delta \sigma(X_s)\,\mathrm{d}W_s}_{\sim\sqrt{\Delta}}$$
$$+ \underbrace{\frac{1}{\Delta}\Big(\int_0^\Delta \sigma(X_s)\,\mathrm{d}W_s\Big)^2}_{\substack{\mathbb{E}[\dots] = \frac{1}{\Delta}\mathbb{E}[\int_0^\Delta \sigma(X_s)^2\,\mathrm{d}s]\\ \sim \sigma(X_0)^2, \text{ by Itô isometry}}}.$$

As an example we consider $\mathrm{d}B_t = \sigma\,\mathrm{d}W_t$. We observe $B_0, B_\Delta, B_{2\Delta}, \dots, B_{N\Delta}$ with $N \to \infty$, $N\Delta = T$ fixed. The analysis of the increments motivates the estimator

$$\widehat{\sigma}^2 = \frac{1}{N}\sum_{n=0}^{N-1}\frac{(B_{(n+1)\Delta} - B_{n\Delta})^2}{\Delta} = \frac{1}{N}\sum_{n=0}^{N-1}\sigma^2 Y_n^2,$$

where (Y_n) are iid with distribution $N(0,1)$. Then the estimator is unbiased, $\mathbb{E}[\widehat{\sigma}^2] = \sigma^2$, and the quadratic risk is given by

$$\mathbb{E}[(\widehat{\sigma}^2 - \sigma^2)^2] = \mathbb{E}\left[\Big(\frac{1}{N}\sum_{n=0}^{N-1}\sigma^2(Y_n^2 - 1)\Big)^2\right] = \frac{\sigma^4}{N}\,\mathbb{E}[(Y_0^2-1)^2] = \frac{2\sigma^4}{N}.$$

We see $\mathbb{E}[(\widehat{\sigma}^2 - \sigma^2)^2]^{1/2} \sim N^{-1/2}$. By the CLT we even obtain $\sqrt{N}(\widehat{\sigma}^2 - \sigma^2) \xrightarrow{d} N(0, 2\sigma^4)$.

What makes this calculation easy?

- independent increments
- σ is constant

Remark. (a) By the *Burkholder–Davis–Gundy inequality* (BDG inequality) there is for all $p \in (0, \infty)$ a constant $C_p > 0$ such that for all $f(X_t, t)$ adapted

$$\mathbb{E}\left[\left|\int_a^b f(X_t, t)\,\mathrm{d}W_t\right|^p\right] \leqslant C_p\,\mathbb{E}\left[\left(\int_a^b f(X_t, t)^2\,\mathrm{d}t\right)^{p/2}\right].$$

(b) Let X be a solution of $\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t$. The *Tanaka formula* states

$$|X_t - x| = |X_0 - x| + \int_0^t \operatorname{sign}(X_s - x)\,\mathrm{d}X_s + \sigma^2(x)L_t(x),$$

where L_t is the local time at t, $\operatorname{sign}(x) = 1$ for $x > 0$ and $\operatorname{sign}(x) = -1$ for $x \leqslant 0$. (The Tanaka formula can be viewed as a generalisation of the Itô formula for $f(y) = |y - x|$.)

4.2 Error bounds for the Florens-Zmirou estimator

Definition 4.1. Let $0 < m < M$ and define

$$\Theta(m, M) = \left\{\sigma \in C^1(\mathbb{R})\,\middle|\, m \leqslant \inf_{x\in\mathbb{R}} \sigma(x) \leqslant \sup_{x\in\mathbb{R}} \sigma(x) \leqslant M, \quad \sup_{x\in\mathbb{R}} |\sigma'(x)| \leqslant M\right\}$$

Each $\sigma \in \Theta(m, M)$ satisfies the Lipschitz and the linear growth conditions and thus

$$\begin{aligned} \mathrm{d}X_t &= \sigma(X_t)\,\mathrm{d}W_t, \\ X_0 &= X^{(0)} \in L^2(\Omega), \end{aligned}$$

has a unique strong solution. We observe

$$X_0, X_\Delta, X_{2\Delta}, \ldots, X_{N\Delta}$$

as $N \to \infty$ and with $N\Delta = 1$ fixed. We define the *Florens-Zmirou estimator* [11] by

$$\sigma^2_{FZ}(x, h_\Delta) = \frac{\sum_{n=0}^{N-1} \mathbb{1}_{\{|X_{n\Delta} - x| < h_\Delta\}} \frac{1}{\Delta}(X_{(n+1)\Delta} - X_{n\Delta})^2}{\sum_{n=0}^{N-1} \mathbb{1}_{\{|X_{n\Delta} - x| < h_\Delta\}}}$$

if $\sum_{n=0}^{N-1} \mathbb{1}_{\{|X_{n\Delta} - x| < h_\Delta\}} > 0$. This estimator is of *Nadaraya–Watson type.*

Lemma 4.2. *For every $p > 0$ holds $\sup_{\sigma\in\Theta, x\in\mathbb{R}} \mathbb{E}[L(x)^p] \leqslant K_p$ for $L(x) = L_1(x)$.*

Proof. By the Tanaka formula

$$\begin{aligned} L(x) &= \frac{1}{\sigma(x)^2}\left(|X_1 - x| - |X_0 - x| - \int_0^1 \operatorname{sign}(X_t - x)\,\mathrm{d}X_t\right) \\ &\leqslant \frac{1}{m^2}\left(|X_1 - X_0| + \left|\int_0^1 \operatorname{sign}(X_t - x)\,\mathrm{d}X_t\right|\right), \end{aligned}$$

where $\operatorname{sign}(x) = 1$ for $x > 0$ and $\operatorname{sign}(x) = -1$ for $x \leqslant 0$. By the BDG inequality we have

$$\mathbb{E}\left[|X_1 - X_0|^p\right] = \mathbb{E}\left[\left|\int_0^1 \sigma(X_t)\,\mathrm{d}W_t\right|^p\right] \leqslant C_p\,\mathbb{E}\left[\left(\int_0^1 \sigma(X_s)^2\,\mathrm{d}s\right)^{p/2}\right] \leqslant C_p M^p,$$

$$\mathbb{E}\left[\left|\int_0^1 \operatorname{sign}(X_t - x)\,\mathrm{d}X_t\right|^p\right] \leqslant C_p\,\mathbb{E}\left[\left(\int_0^1 \operatorname{sign}(X_t - x)^2\sigma(X_t)^2\,\mathrm{d}t\right)^{p/2}\right] \leqslant C_p M^p.$$

$\square$

Theorem 4.3. *Let I be an open interval, $\nu > 0$ and $\mathcal{L} = \{\omega \in \Omega \,|\inf_{x\in I} L(x) \geqslant \nu\}$. Let $h_\Delta \sim \Delta^{1/3}$. Then there exists $C > 0$ such that for all $x \in I$*

$$\sup_{\sigma\in\Theta}\left(\mathbb{E}\left[\left|\sigma^2_{FZ}(x,h_\Delta)\wedge M^2 - \sigma^2(x)\right|^2 \mathbb{1}_{\mathcal{L}}\right]\right)^{1/2} \leqslant C\Delta^{1/3}.$$

Notation:
$f_\sigma \lesssim g_\sigma$ (or $g_\sigma \gtrsim f_\sigma$) means that there exists $C > 0$ such that $f_\sigma \leqslant C g_\sigma$ for all $\sigma \in \Theta$, $x \in I$. We write $f_\sigma \sim g_\sigma$ if $f_\sigma \lesssim g_\sigma$ and $f_\sigma \gtrsim g_\sigma$.

Proof. (a) (error decomposition) For $n = 0, \ldots, N-1$ we define

$$\eta_n = \frac{1}{\Delta}\left(\int_{n\Delta}^{(n+1)\Delta}\sigma(X_s)\,\mathrm{d}W_s\right)^2 - \frac{1}{\Delta}\int_{n\Delta}^{(n+1)\Delta}\sigma(X_s)^2\,\mathrm{d}s.$$

- $\mathbb{E}[\eta_n|\,\mathcal{F}_{n\Delta}] = 0$ and for $m < n$ we have $\mathbb{E}[\eta_m\eta_n] = \mathbb{E}[\eta_m\,\mathbb{E}[\eta_n|\,\mathcal{F}_{n\Delta}]] = 0$.
- $\mathbb{E}[\eta_n^2|\,\mathcal{F}_{n\Delta}] \lesssim 1$ since by the BDG inequality

$$\begin{aligned}\Delta^2\,\mathbb{E}[\eta_n^2|\,\mathcal{F}_{n\Delta}] &\lesssim \mathbb{E}\left[\left(\int_{n\Delta}^{(n+1)\Delta}\sigma(X_s)\,\mathrm{d}W_s\right)^4\middle|\,\mathcal{F}_{n\Delta}\right] + \mathbb{E}\left[\left(\int_{n\Delta}^{(n+1)\Delta}\sigma(X_s)^2\,\mathrm{d}s\right)^2\middle|\,\mathcal{F}_{n\Delta}\right]\\ &\lesssim \mathbb{E}\left[\left(\int_{n\Delta}^{(n+1)\Delta}\sigma(X_s)^2\,\mathrm{d}s\right)^2\middle|\,\mathcal{F}_{n\Delta}\right] \lesssim \Delta^2.\end{aligned}$$

We decompose

$$\begin{aligned}&|\sigma^2_{\mathrm{FZ}}(x,h_\Delta) - \sigma^2(x)|\\ &= \left|\frac{\sum_{n=0}^{N-1}\mathbb{1}_{\{|X_{n\Delta}-x|<h_\Delta\}}\left(\frac{1}{\Delta}\left(\int_{n\Delta}^{(n+1)\Delta}\sigma(X_t)\,\mathrm{d}W_t\right)^2 - \sigma^2(x)\right)}{\sum_{n=0}^{N-1}\mathbb{1}_{\{|X_{n\Delta}-x|<h_\Delta\}}}\right|\\ &\leqslant \underbrace{\left|\frac{\sum_{n=0}^{N-1}\mathbb{1}_{\{|X_{n\Delta}-x|<h_\Delta\}}\eta_n}{\sum_{n=0}^{N-1}\mathbb{1}_{\{|X_{n\Delta}-x|<h_\Delta\}}}\right|}_{\text{martingale part } M_{x,\Delta}} + \underbrace{\left|\frac{\sum_{n=0}^{N-1}\mathbb{1}_{\{|X_{n\Delta}-x|<h_\Delta\}}\left(\frac{1}{\Delta}\int_{n\Delta}^{(n+1)\Delta}\sigma^2(X_t)\,\mathrm{d}t - \sigma^2(x)\right)}{\sum_{n=0}^{N-1}\mathbb{1}_{\{|X_{n\Delta}-x|<h_\Delta\}}}\right|}_{\text{bias part } B_{x,\Delta}}.\end{aligned}$$

(b) (good event of high probability) Define the modulus of continuity as the random variable

$$W^X(\Delta)_T := \sup_{\substack{0\leqslant s,t\leqslant T\\ |t-s|<\Delta}}|X_t - X_s|, \qquad W(\Delta) := W^X(\Delta)_1.$$

Let $0 < \varepsilon < 1/6$ and $\alpha = 3/2 - 3\varepsilon > 1$. We define $\mathcal{R} = \{\omega \in \Omega\,|W(\Delta) < h_\Delta^\alpha\}$. By Markov's inequality we have for all $p > 0$

$$\mathbb{P}(\mathcal{R}^c) \leqslant h_\Delta^{-p\alpha}\,\mathbb{E}[W(\Delta)^p]. \tag{4.2}$$

Claim:

$$\mathbb{E}[W^X(\Delta)_T^p] \leqslant C_p \left(\Delta \log\left(\frac{2T}{\Delta}\right)\right)^{p/2} \tag{4.3}$$

Reason:

- (4.3) is true for Brownian motion, see [10].
- Let $\mathrm{d}X_t = \sigma(X_t)\,\mathrm{d}W_t$. By the Dambis–Dubins–Schwarz theorem $X_t = B_{\int_0^t \sigma^2(X_u)\,\mathrm{d}u}$ for some Brownian motion B. Consequently for $0 \leq s, t \leq T$

$$|X_t - X_s| = \left|B_{\int_0^t \sigma^2(X_u)\,\mathrm{d}u} - B_{\int_0^s \sigma^2(X_u)\,\mathrm{d}u}\right| \leqslant W^B\left(|t-s|M^2\right)_{TM^2}.$$

We bound (4.2) by

$$\begin{aligned}\mathbb{P}(\mathcal{R}^c) &\lesssim \Delta^{-p\alpha/3}\left(\Delta\log\left(\frac{2}{\Delta}\right)\right)^{p/2}\\ &= \Delta^{p\varepsilon}\left(\log\left(\frac{2}{\Delta}\right)\right)^{p/2}\end{aligned}$$

and conclude that $\mathbb{P}(\mathcal{R}^c) \lesssim \Delta^{2/3}$ for p large enough.

(c) (bias part) If $|X_{n\Delta} - x| < h_\Delta$, then

$$\begin{aligned}\left|\frac{1}{\Delta}\int_{n\Delta}^{(n+1)\Delta} \sigma^2(X_t)\,\mathrm{d}t - \sigma^2(x)\right| &\lesssim \frac{1}{\Delta}\int_{n\Delta}^{(n+1)\Delta} |X_t - x|\,\mathrm{d}t\\ &\lesssim \frac{1}{\Delta}\int_{n\Delta}^{(n+1)\Delta} |X_t - X_{n\Delta}|\,\mathrm{d}t + |X_{n\Delta} - x|\\ &\lesssim W(\Delta) + h_\Delta.\end{aligned}$$

So we have $B_{x,\Delta}\mathbb{1}_{\mathcal{R}} \lesssim h_\Delta$.

(d) (martingale part) We define $N(x, h_\Delta) := \sum_{n=0}^{N-1} \mathbb{1}_{\{|X_{n\Delta}-x|<h_\Delta\}}$.
Claim: On $\mathcal{R}$ we have

$$\left|\frac{N(x,h_\Delta)}{Nh_\Delta} - \frac{1}{h_\Delta}\int_{x-h_\Delta}^{x+h_\Delta} L(z)\,\mathrm{d}z\right| \leqslant \frac{1}{h_\Delta}\int_{\{h_\Delta - h_\Delta^\alpha \leqslant |z-x| < h_\Delta + h_\Delta^\alpha\}} L(z)\,\mathrm{d}z$$

Proof of claim:

$$\begin{aligned}&\left|\frac{1}{N}\sum_{n=0}^{N-1} \mathbb{1}_{\{|X_{n\Delta}-x|<h_\Delta\}} - \int_0^1 \mathbb{1}_{\{|X_s-x|<h_\Delta\}}\,\mathrm{d}s\right|\\ &\leqslant \sum_{n=0}^{N-1}\int_{n\Delta}^{(n+1)\Delta} \left|\mathbb{1}_{\{|X_{n\Delta}-x|<h_\Delta\}} - \mathbb{1}_{\{|X_s-x|<h_\Delta\}}\right|\,\mathrm{d}s\\ &\leqslant \sum_{n=0}^{N-1}\int_{n\Delta}^{(n+1)\Delta} \mathbb{1}_{\{h_\Delta \leqslant |X_s-x| < h_\Delta + W(\Delta)\}}\,\mathrm{d}s + \sum_{n=0}^{N-1}\int_{n\Delta}^{(n+1)\Delta} \mathbb{1}_{\{h_\Delta - W(\Delta) \leqslant |X_s-x| < h_\Delta\}}\,\mathrm{d}s\\ &\leqslant \int_0^1 \mathbb{1}_{\{h_\Delta - h_\Delta^\alpha \leqslant |X_s-x| < h_\Delta + h_\Delta^\alpha\}}\,\mathrm{d}s\\ &= \int_{\{h_\Delta - h_\Delta^\alpha \leqslant |z-x| < h_\Delta + h_\Delta^\alpha\}} L(z)\,\mathrm{d}z\end{aligned}$$

For simplicity we define $A := \{z | h_\Delta - h_\Delta^\alpha \leqslant |z-x| < h_\Delta + h_\Delta^\alpha\}$ and observe that A has Lebesgue measure $4h_\Delta^\alpha$. Using Markov's and Jensen's inequalities we obtain for $p > 1$

$$\begin{aligned}&\mathbb{P}\left(\frac{1}{h_\Delta}\int_A L(z)\,\mathrm{d}z \geqslant \nu\right) \lesssim \mathbb{E}\left[\frac{1}{h_\Delta^p}\left(\int_A L(z)\,\mathrm{d}z\right)^p\right]\\ \lesssim\ & \frac{h_\Delta^{\alpha(p-1)}}{h_\Delta^p}\int_A \mathbb{E}\left[L(z)^p\right]\mathrm{d}z \lesssim h_\Delta^{(\alpha-1)p} \lesssim \Delta^{2/3}\end{aligned}$$

for p large enough. So there is an event $\mathcal{Q} \subseteq \mathcal{R}$ with $\mathbb{P}(\mathcal{Q}^c) \lesssim \Delta^{2/3}$ such that $N(x,h_\Delta)/(Nh_\Delta)$ is bounded from below on $\mathcal{Q} \cap \mathcal{L}$. Using the martingale properties of η_n we obtain

$$\begin{aligned}\mathbb{E}\left[M_{x,\Delta}^2 \mathbb{1}_{\mathcal{Q}\cap\mathcal{L}}\right] &= \mathbb{E}\left[\left(\frac{1}{N(x,h_\Delta)}\sum_{n=0}^{N-1}\mathbb{1}_{\{|X_{n\Delta}-x|<h_\Delta\}}\eta_n\right)^2 \mathbb{1}_{\mathcal{Q}\cap\mathcal{L}}\right]\\ &\lesssim \frac{1}{N^2h_\Delta^2}\,\mathbb{E}\left[\left(\sum_{n=0}^{N-1}\mathbb{1}_{\{|X_{n\Delta}-x|<h_\Delta\}}\eta_n\right)^2 \mathbb{1}_{\mathcal{Q}\cap\mathcal{L}}\right]\\ &\lesssim \frac{1}{N^2h_\Delta^2}\,\mathbb{E}\left[\sum_{n,m=0}^{N-1}\mathbb{1}_{\{|X_{n\Delta}-x|<h_\Delta\}}\mathbb{1}_{\{|X_{m\Delta}-x|<h_\Delta\}}\eta_n\eta_m\right]\\ &= \frac{1}{N^2h_\Delta^2}\,\mathbb{E}\left[\sum_{n=0}^{N-1}\mathbb{1}_{\{|X_{n\Delta}-x|<h_\Delta\}}\,\mathbb{E}[\eta_n^2|\,\mathcal{F}_{n\Delta}]\right]\\ &\lesssim \frac{1}{N^2h_\Delta^2}\,\mathbb{E}[N(x,h_\Delta)].\end{aligned}$$

Finally

$$\begin{aligned}\frac{1}{Nh_\Delta}\,\mathbb{E}\left[N(x,h_\Delta)\right] &\lesssim \frac{1}{Nh_\Delta}\,\mathbb{E}\left[N(x,h_\Delta)\mathbb{1}_{\mathcal{R}}\right] + \frac{1}{Nh_\Delta}\,\mathbb{E}[N(x,h_\Delta)\mathbb{1}_{\mathcal{R}^c}]\\ &\lesssim \mathbb{E}\left[\frac{1}{h_\Delta}\int_{x-h_\Delta}^{x+h_\Delta} L(z)\,\mathrm{d}z + \frac{1}{h_\Delta}\int_A L(z)\,\mathrm{d}z\right] + h_\Delta^{-1}\,\mathbb{P}\left(\mathcal{R}^c\right)\\ &\lesssim \frac{1}{h_\Delta}\int_{(x-h_\Delta,x+h_\Delta)\cup A}\mathbb{E}[L(z)]\,\mathrm{d}z + h_\Delta^{-1}\Delta^{2/3}\\ &\lesssim 1.\end{aligned}$$

(e) (conclusion) We have shown

$$\begin{aligned}\mathbb{E}\left[\left|\sigma_{\mathrm{FZ}}^2(x,h_\Delta) - \sigma^2(x)\right|^2 \mathbb{1}_{\mathcal{L}\cap\mathcal{Q}}\right] &\lesssim \mathbb{E}\left[M_{x,\Delta}^2\mathbb{1}_{\mathcal{L}\cap\mathcal{Q}} + B_{x,\Delta}^2\mathbb{1}_{\mathcal{R}}\right]\\ &\lesssim \frac{1}{Nh_\Delta} + h_\Delta^2 \sim \Delta^{2/3}.\end{aligned}$$

Furthermore,

$$\mathbb{E}\left[\left|\sigma_{\mathrm{FZ}}^2(x,h_\Delta)\wedge M^2 - \sigma^2(x)\right|^2 \mathbb{1}_{\mathcal{L}\cap\mathcal{Q}^c}\right] \lesssim \mathbb{P}(\mathcal{Q}^c) \lesssim \Delta^{2/3}.$$

□

Corollary 4.4. *Let* $\Theta^* = \Theta(m, M) \times \{b \in C(\mathbb{R}) \,|\, b \text{ is Lipschitz and } \sup_{x\in\mathbb{R}} |b(x)| \leqslant M\}$*. Let* $(\sigma, b) \in \Theta^*$ *and define* $dY_t = b(Y_t)\,\mathrm{d}t + \sigma(Y_t)\,\mathrm{d}W_t$*,* $Y_0 = X_0$*. For* $h_\Delta \sim \Delta^{1/3}$ *and* $\mathcal{L}$ *as before there exists* $C > 0$ *such that for all* $x \in I$

$$\sup_{(\sigma,b)\in\Theta^*} \mathbb{E}\left[\left|\sigma^2_{FZ}(x, h_\Delta) \wedge M^2 - \sigma^2(x)\right| \mathbb{1}_{\mathcal{L}}\right] \leqslant C\Delta^{1/3}.$$

Proof. The assumptions of the Girsanov theorem are satisfied. The laws of X and Y on $C([0, 1])$ are equivalent and

$$\begin{aligned}\frac{\mathrm{d}\,\mathbb{P}_Y}{\mathrm{d}\,\mathbb{P}_X}(X) &= \exp\left(\int_0^1 \frac{b}{\sigma^2}(X_s)\,\mathrm{d}X_s - \frac{1}{2}\int_0^1 \frac{b^2}{\sigma^2}(X_s)\,\mathrm{d}s\right)\\ &= \exp\left(\int_0^1 \frac{b}{\sigma}(X_s)\,\mathrm{d}W_s - \frac{1}{2}\int_0^1 \frac{b^2}{\sigma^2}(X_s)\,\mathrm{d}s\right).\end{aligned}$$

We define $\mathcal{E}_{x,\Delta} := \left|\sigma^2_{\mathrm{FZ}}(x, h_\Delta) \wedge M^2 - \sigma^2(x)\right| \mathbb{1}_{\mathcal{L}}$. By the Cauchy–Schwarz inequality

$$\begin{aligned}\mathbb{E}_Y\left[\mathcal{E}_{x,\Delta}\right] &= \mathbb{E}_X\left[\mathcal{E}_{x,\Delta}\frac{\mathrm{d}\,\mathbb{P}_Y}{\mathrm{d}\,\mathbb{P}_X}(X)\right]\\ &= \mathbb{E}_X\left[\mathcal{E}_{x,\Delta}\exp\left(\int_0^1 \frac{b}{\sigma}(X_s)\,\mathrm{d}W_s - \frac{1}{2}\int_0^1 \frac{b^2}{\sigma^2}(X_s)\,\mathrm{d}s\right)\right]\\ &\leqslant \mathbb{E}_X\left[\mathcal{E}_{x,\Delta}\exp\left(\int_0^1 \frac{b}{\sigma}(X_s)\,\mathrm{d}W_s\right)\right]\\ &\leqslant \mathbb{E}_X\left[\mathcal{E}^2_{x,\Delta}\right]^{1/2}\mathbb{E}_X\left[\exp\left(2\int_0^1 \frac{b}{\sigma}(X_s)\,\mathrm{d}W_s\right)\right]^{1/2}.\end{aligned}$$

It remains to show that

$$\mathbb{E}_X\left[\exp\left(\int_0^1 \frac{2b}{\sigma}(X_s)\,\mathrm{d}W_s\right)\right]$$

is uniformly bounded. Since $\mathbb{E}_X\left[\exp\left(\int_0^1 \frac{2b^2}{\sigma^2}(X_s)\,\mathrm{d}s\right)\right] < \infty$, by Novikov's condition the process

$$M_t := \exp\left(\int_0^t \frac{2b}{\sigma}(X_s)\,\mathrm{d}W_s - \int_0^t \frac{2b^2}{\sigma^2}(X_s)\,\mathrm{d}s\right), \quad t \in [0, 1],$$

is a martingale so that $\mathbb{E}_X[M_1] = \mathbb{E}_X[M_0] = 1$. We conclude

$$\mathbb{E}_X\left[\exp\left(\int_0^1 \frac{2b}{\sigma}(X_s)\,\mathrm{d}W_s\right)\right] \leqslant \exp\left(\frac{2M^2}{m^2}\right).$$

□

Theorem 4.5. *(Florens-Zmirou, 1993) Let* X *satisfy*

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t, \quad t \in [0, 1],$$

where b *is bounded with two continuous and bounded derivatives,* σ *has three continuous and bounded derivatives and* $m \leqslant \sigma \leqslant M$ *for some* $0 < m < M$*. If* Nh^3_Δ *tends to zero, then*

$$\sqrt{Nh_\Delta}\left(\frac{\sigma^2_{FZ}(x, h_\Delta)}{\sigma^2(x)} - 1\right) \xrightarrow{d} L(x)^{-1/2} Z,$$

where Z *is a standard normal random variable independent of* $L(x)$*.*

5 Nonparametric estimation with low-frequency data

We consider the SDE

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t, \quad t \geqslant 0.$$

For $\Delta > 0$ fixed we observe $X_0, X_\Delta, \dots, X_{N\Delta}$ as $N \to \infty$. We define the *transition operator*

$$P_\Delta f(x) := \mathbb{E}\left[\, f(X_\Delta) \middle|\, X_0 = x\right].$$

We recall the infinitesimal generator

$$Af(x) = \frac{1}{2}\sigma^2(x)f''(x) + b(x)f'(x).$$

We have $P_\Delta = \exp(\Delta A)$ in the operator sense. The estimation method can be summarised by

$$X_0, X_\Delta, \dots, X_{N\Delta} \overset{\text{estimation}}{\longrightarrow} P_\Delta \overset{\text{identification}}{\longrightarrow} A \quad \longrightarrow \quad (\sigma^2, b).$$

We simplify the statistical problem by considering a diffusion with boundary reflections

$$\begin{aligned} \mathrm{d}X_t &= b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t + v(X_t)\,\mathrm{d}L(X), \\ X_0 &= x_0 \quad \text{and} \quad X_t \in [0,1], \qquad t \geqslant 0, \end{aligned}$$

where $v(0) = 1$, $v(1) = -1$ and $L(X)$ is a continuous nondecreasing process that increases only when $X_t \in \{0, 1\}$.

For $s \geqslant 0$ we define the *Sobolev space*

$$H^s(\mathbb{R}) := \left\{ f \in L^2(\mathbb{R}) \middle|\, \|f\|^2_{H^s(\mathbb{R})} := \int_{\mathbb{R}} (u^2+1)^s |\mathcal{F}f(u)|^2\,\mathrm{d}u < \infty \right\},$$

where $\mathcal{F}f(u) = \int_{-\infty}^{\infty} e^{iux} f(x)\,\mathrm{d}x$ denotes the Fourier transform of f. We define

$$H^s([0,1]) := \left\{ f \in L^2([0,1]) \,\middle|\, \exists g \in H^s(\mathbb{R}) \text{ with } g|_{[0,1]} = f \right\},$$

and

$$\|f\|_{H^s([0,1])} := \inf\left\{ \|g\|_{H^s(\mathbb{R})} \,\middle|\, g \in H^s(\mathbb{R}), g|_{[0,1]} = f \right\}.$$

Definition 5.1. For $s > 1$ and given constants $C \geqslant c > 0$ we consider the class $\Theta_s = \Theta(s, C, c)$ defined by

$$\left\{ (\sigma, b) \in H^s([0,1]) \times H^{s-1}([0,1]) \middle|\, \|\sigma\|_{H^s([0,1])} \leqslant C, \|b\|_{H^{s-1}([0,1])} \leqslant C, \inf_{x\in[0,1]} \sigma(x) \geqslant c \right\}.$$

The invariant density has the form

$$\mu(x) = \frac{1}{G\sigma^2(x)} \exp\left(\int_0^x \frac{2b}{\sigma^2}(y)\,\mathrm{d}y \right).$$

We further define

$$S(x) = \frac{1}{2G} \exp\left(\int_0^x \frac{2b}{\sigma^2}(y)\,\mathrm{d}y \right).$$

The infinitesimal generator can be expressed by

$$Af(x) = \frac{1}{2}\sigma^2(x)f''(x) + b(x)f'(x) = \frac{S(x)}{\mu(x)}f''(x) + \frac{S'(x)}{\mu(x)}f'(x) = \frac{1}{\mu(x)}(S(x)f'(x))'.$$

The domain of this unbounded operator in $L^2(\mu)$ is given by

$$\mathrm{dom}(A) = \left\{f \in H^2([0,1]) | f'(0) = f'(1) = 0\right\}.$$

The operator A has a discrete point spectrum $\{\nu_k | k = 0, 1, \dots\}$. The largest eigenvalue is 0 with constant eigenfunction. Let ν_1 be the second largest eigenvalue with corresponding eigenfunction u_1. By the reflecting boundary $u_1'(0) = u_1'(1) = 0$ and thus we obtain from

$$Au_1(x) = \frac{1}{\mu(x)}(S(x)u_1'(x))' = \nu_1 u_1(x)$$

by integration

$$S(x)u_1'(x) = \nu_1 \int_0^x u_1(y)\mu(y)\,\mathrm{d}y.$$

We can choose u_1 such that $u_1'(x) > 0$ for all $x \in (0,1)$. Furthermore, u_1 is eigenfunction of P_Δ with eigenvalue $\kappa_1 = e^{\Delta\nu_1}$. We derive

$$S(x) = \frac{\Delta^{-1}\log(\kappa_1)\int_0^x u_1(y)\mu(y)\,\mathrm{d}y}{u_1'(x)}, \quad x \in (0,1),$$

so that

$$\sigma^2(x) = \frac{2S(x)}{\mu(x)} = \frac{2\Delta^{-1}\log(\kappa_1)\int_0^x u_1(y)\mu(y)\,\mathrm{d}y}{u_1'(x)\mu(x)}$$

and

$$b(x) = \frac{S'(x)}{\mu(x)} = \Delta^{-1}\log(\kappa_1)\frac{u_1(x)u_1'(x)\mu(x) - u_1''(x)\int_0^x u_1(y)\mu(y)\,\mathrm{d}y}{u_1'(x)^2\mu(x)}.$$

The estimation method can be summarised in more detail by

$$X_0, X_\Delta, \dots, X_{N\Delta} \overset{\text{estimation}}{\longrightarrow} (\mu, P_\Delta) \quad\longrightarrow\quad (\mu, u_1, \kappa_1) \quad\longrightarrow\quad (\mu, S) \quad\longrightarrow\quad (\sigma^2, b).$$

With this method estimators $\widehat{\sigma}^2$ and $\widehat{b}$ can be defined such that we have the following theorem.

Theorem 5.2. *(Gobet, Hoffmann, Reiß, 2004, [13]) For all $s > 1$, $C \geqslant c > 0$ and $0 < a < b < 1$ we have*

$$\sup_{(\sigma,b)\in\Theta_s} \mathbb{E}_{\sigma,b}\left[\|\widehat{\sigma}^2 - \sigma^2\|^2_{L^2([a,b])}\right]^{1/2} \lesssim N^{-s/(2s+3)}$$

$$\sup_{(\sigma,b)\in\Theta_s} \mathbb{E}_{\sigma,b}\left[\|\widehat{b} - b\|^2_{L^2([a,b])}\right]^{1/2} \lesssim N^{-(s-1)/(2s+3)}.$$

They also show that these rates are minimax optimal. Let $s_1 = s - 1$ be the regularity of the drift b and let $s_2 = s$ the regularity of the volatility σ. Then b can be estimated with rate $N^{-s_1/(2s_1+5)}$ and σ^2 with rate $N^{-s_2/(2s_2+3)}$.

The following table shows minimax convergence rates for the diffusion model with continuous, high-frequency and low-frequency observations.

	Parametric		Nonparametric	
	Volatility	Drift	Volatility	Drift
Continuous	known	$T^{-1/2}$	known	$T^{-s/(2s+1)}$
High-frequency	$N^{-1/2}$	$(N\Delta)^{-1/2}$	$N^{-s/(2s+1)}$	$(N\Delta)^{-s/(2s+1)}$
Low-frequency	$N^{-1/2}$	$N^{-1/2}$	$N^{-s/(2s+3)}$	$N^{-s/(2s+5)}$

6 Lévy processes

Definition 6.1. An $\mathbb{R}^d$-valued process $X = (X_t)_{t\geqslant 0}$ defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F})_{t\geqslant 0}, \mathbb{P})$ is called a *Lévy process* if it is $(\mathcal{F}_t)$-adapted and has the following properties

(a) X is continuous in probability, i.e., for fixed $u \geqslant 0$, $\mathbb{P}(|X_t - X_u| > \varepsilon) \to 0$ holds as $t \to u$ for all $\varepsilon > 0$.

(b) $\mathbb{P}(X_0 = 0) = 1$.

(c) For $0 \leqslant s \leqslant t$, $X_t - X_s$ is equal in distribution to X_{t-s}.

(d) For $0 \leqslant s \leqslant t$, $X_t - X_s$ is independent of $\mathcal{F}_s$.

Remark. Every Lévy process has a càdlàg modification. Without loss of generality we will assume that all sample paths of Lévy processes are càdlàg.

Definition 6.2. A *Lévy measure* on $\mathbb{R}^d$ is a σ-finite measure ν on $\mathbb{R}^d$ such that $\nu(\{0\}) = 0$ and

$$\int_{\mathbb{R}^d} (1 \wedge |x|^2)\,\mathrm{d}\nu(x) < \infty.$$

Proposition 6.3. *(Lévy–Khintchine Representation) Let X be a Lévy process taking values in $\mathbb{R}^d$. Then for each $t \geqslant 0$ the characteristic function φ_t of X_t satisfies*

$$\varphi_t(u) := \mathbb{E}\left[e^{i\langle u, X_t\rangle}\right] = e^{t\psi(u)}, \qquad u \in \mathbb{R}^d,$$

with characteristic exponent $\psi(u)$ given by

$$\psi(u) = i\langle u, \gamma\rangle - \frac{1}{2}\langle u, \Sigma u\rangle + \int_{\mathbb{R}^d} \left(e^{i\langle u,x\rangle} - 1 - i\langle u, x\rangle \mathbb{1}_{\{|x|\leqslant 1\}}\right) \mathrm{d}\nu(x), \tag{6.1}$$

where $\gamma \in \mathbb{R}^d$, $\Sigma \in \mathbb{R}^{d\times d}$ is a positive semi-definite matrix and ν is a Lévy measure on $\mathbb{R}^d$.

The quantity (γ, Σ, ν) is called the *characteristic triplet* of X. If $d = 1$, we also write σ^2 instead of Σ. Under additional assumptions on ν (6.1) has a simpler form:

(a) If $\int_{\mathbb{R}^d} |x| \mathbb{1}_{\{|x| \leqslant 1\}} \, \mathrm{d}\nu(x) < \infty$ holds, then (6.1) reduces to

$$\psi(u) = i\langle u, \gamma_0 \rangle - \frac{1}{2}\langle u, \Sigma u \rangle + \int_{R^d} \left(e^{i\langle u, x\rangle} - 1 \right) \mathrm{d}\nu(x)$$

with $\gamma_0 = \gamma - \int_{\mathbb{R}^d} x \mathbb{1}_{\{|x| \leqslant 1\}} \, \mathrm{d}\nu(x)$.

(b) If $\int_{\mathbb{R}^d} |x| \mathbb{1}_{\{|x| > 1\}} \, \mathrm{d}\nu(x) < \infty$ holds, then we can write (6.1) as

$$\psi(u) = i\langle u, \gamma_1 \rangle - \frac{1}{2}\langle u, \Sigma u \rangle + \int_{R^d} \left(e^{i\langle u, x\rangle} - 1 - i\langle u, x \rangle \right) \mathrm{d}\nu(x)$$

with $\gamma_1 = \gamma + \int_{\mathbb{R}^d} x \mathbb{1}_{\{|x| > 1\}} \, \mathrm{d}\nu(x)$ and we have $\mathbb{E}[X_t] = \gamma_1 t$.

(c) If $d = 1$ and $\int_{-\infty}^{\infty} x^2 \, \mathrm{d}\nu(x) < \infty$ holds, then we have the Kolmogorov representation

$$\begin{aligned}\psi(u) &= iu\gamma_1 - \frac{\sigma^2 u^2}{2} + \int_{-\infty}^{\infty} \frac{e^{iux} - 1 - iux}{x^2} \, \mathrm{d}\widetilde{\nu}(x) \\ &= iu\gamma_1 + \int_{-\infty}^{\infty} \frac{e^{iux} - 1 - iux}{x^2} \, \mathrm{d}\nu_\sigma(x)\end{aligned}$$

with $\mathrm{d}\widetilde{\nu}(x) = x^2 \, \mathrm{d}\nu(x)$ and $\mathrm{d}\nu_\sigma(x) = \mathrm{d}\widetilde{\nu}(x) + \sigma^2 \, \mathrm{d}\delta_0(x)$, using at $x = 0$ the continuous extension of the integrand to $-u^2/2$ in the second representation. We have $\mathbb{E}[X_t] = \gamma_1 t$ and $\mathrm{Var}(X_t) = (\sigma^2 + \widetilde{\nu}(\mathbb{R}))t = \nu_\sigma(\mathbb{R})t$.

Proposition 6.4. *(Corollary 25.8, [22]) Let X be a Lévy process and $p > 0$. Then $\mathbb{E}[|X_t|^p] < \infty$ for one $t > 0$ implies $\mathbb{E}[|X_t|^p] < \infty$ for all $t > 0$. We have $\mathbb{E}[|X_t|^p] < \infty$ if and only if $\int_{\mathbb{R}^d} |x|^p \mathbb{1}_{\{|x| > 1\}} \, \mathrm{d}\nu(x) < \infty$.*

7 Empirical characteristic functions and processes

Definition 7.1. The *empirical characteristic function* (ecf) of i.i.d. $\mathbb{R}^d$-valued random variables $X_1, \ldots, X_n$ is given by

$$\varphi_n(u) = \frac{1}{n} \sum_{k=1}^{n} e^{i\langle u, X_k \rangle}, \qquad u \in \mathbb{R}^d,$$

and the *empirical characteristic process* (ecp) is given by

$$u \mapsto \mathcal{C}_n(u) = \sqrt{n}(\varphi_n(u) - \varphi(u))$$

with $\varphi(u) = \mathbb{E}[e^{i\langle u, X_1 \rangle}]$.

It holds $\mathcal{C}_n \xrightarrow{fidi} \Gamma$ as $n \to \infty$ for a centred complex-valued Gaussian process $\Gamma(u)$ satisfying $\overline{\Gamma(u)} = \Gamma(-u)$ and $\mathbb{E}[\Gamma(u)\Gamma(v)] = \varphi(u+v) - \varphi(u)\varphi(v)$, i.e., for all $k \in \mathbb{N}$ and $u_1, \ldots, u_k$ we have $(\mathcal{C}_n(u_1), \ldots, \mathcal{C}_n(u_k)) \xrightarrow{d} (\Gamma(u_1), \ldots, \Gamma(u_k))$.

Proposition 7.2. *(Hoeffding's Inequality) Suppose the real-valued and centred random variables $Y_1, \ldots, Y_n$ are i.i.d. and set $S_n = \sum_{k=1}^n Y_k$. If there exists a deterministic number R with $|Y_1| \leqslant R$ almost surely, then for all $\tau > 0$*

$$\mathbb{P}(|S_n| \geqslant \tau) \leqslant 2 \exp\left(-\frac{\tau^2}{2nR^2} \right)$$

Proposition 7.3. *For i.i.d. random vectors $(X_k)_{k\geqslant 1}$ in $\mathbb{R}^d$ with $X_k \in L^1$ and any constant $R > 8\sqrt{d}$ the empirical characteristic process satisfies uniformly in $n \in \mathbb{N}$ and $K \geqslant 2$*

$$\mathbb{P}\left(\max_{u\in[-K,K]^d} |\mathcal{C}_n(u)| \geqslant R\sqrt{\log(nK^2)}\right) \leqslant C(\sqrt{n}K)^{(64d-R^2)/(64d+64)}$$

for some constant C depending on d and $\mathbb{E}[|X_1|]$ only.

Proof. First we treat the real part and define

$$S_n(u) := \sum_{k=1}^{n} (\cos(\langle u, X_k\rangle) - \mathbb{E}[\cos(\langle u, X_k\rangle)]) .$$

For each $u \in \mathbb{R}^d$, $S_n(u)$ is the sum of centred i.i.d. random variables bounded by 2 so that Hoeffding's inequality yields

$$\mathbb{P}\left(|S_n(u)| \geqslant \frac{\tau}{2}\right) \leqslant 2\exp\left(-\frac{(\tau/2)^2}{8n}\right).$$

For $J = J(n)$ we consider the grid on the cube $[-K, K]^d$ given by the $(2J)^d$ points $u_j = jK/J$, $j \in G_J^d := \{-J+1, -J+2, \dots, 0, 1, \dots, J\}^d$ and obtain

$$\mathbb{P}\left(\max_{j\in G_J^d} |S_n(u_j)| \geqslant \frac{\tau}{2}\right) \leqslant \sum_{j\in G_J^d} 2\exp\left(-\frac{(\tau/2)^2}{8n}\right) = 2(2J)^d \exp\left(-\frac{\tau^2}{32n}\right).$$

For all $u, v \in \mathbb{R}^d$ we have $|\cos(\langle u, X_k\rangle) - \cos(\langle v, X_k\rangle)| \leqslant |u-v||X_k|$. Since $\mathbb{E}[|X_k|] < \infty$, we have $|S_n(u) - S_n(v)| \leqslant |u-v| \sum_{k=1}^{n}(|X_k| - \mathbb{E}[|X_k|])$. Further $\max_{u\in[-K,K]^d} \min_j |u - u_j| \leqslant \sqrt{d}K/J$ so that

$$\mathbb{P}\left(\max_{u\in[-K,K]^d} |S_n(u)| \geqslant \tau\right) \leqslant \mathbb{P}\left(\max_{j\in G_J^d} |S_n(u_j)| + \sqrt{d}KJ^{-1}\sum_{k=1}^{n}(|X_k| + \mathbb{E}[|X_k|]) \geqslant \tau\right).$$

By Markov's inequality we obtain

$$\begin{aligned}
&\mathbb{P}\left(\max_{u\in[-K,K]^d} |S_n(u)| \geqslant \tau\right) \\
&\leqslant \mathbb{P}\left(\max_{j\in G_J^d} |S_n(u_j)| \geqslant \frac{\tau}{2}\right) + \mathbb{P}\left(\sqrt{d}KJ^{-1}\sum_{k=1}^{n}(|X_k| + \mathbb{E}[|X_k|]) \geqslant \frac{\tau}{2}\right) \\
&\leqslant 2(2J)^d \exp\left(-\frac{\tau^2}{32n}\right) + \sqrt{d}KJ^{-1}(\tau/2)^{-1}\sum_{k=1}^{n}\mathbb{E}[|X_k| + \mathbb{E}[|X_k|]] \\
&= 2^{d+1}J^d \exp\left(-\frac{\tau^2}{32n}\right) + 4\sqrt{d}nKJ^{-1}\tau^{-1}\,\mathbb{E}[|X_1|].
\end{aligned}$$

The choice $J = (nK/\tau)^{1/(d+1)} \exp\left(\tau^2/(32(d+1)n)\right)$ yields

$$\mathbb{P}\left(\max_{u\in[-K,K]^d} |S_n(u)| \geqslant \tau\right) \leqslant C\left(\frac{nK}{\tau}\right)^{d/(d+1)} \exp\left(-\frac{\tau^2}{32(d+1)n}\right)$$

with $C = 2^{d+1} + 4\sqrt{d}\,\mathbb{E}[|X_1|]$. Since $R > 8\sqrt{d}$ and $nK^2 \geqslant 4$, we obtain

$$\begin{aligned}\mathbb{P}\left(\max_{u\in[-K,K]^d} |S_n(u)| \geqslant \frac{R}{2}\sqrt{n\log(nK^2)}\right) &\leqslant C(\sqrt{n}K)^{d/(d+1)} \exp\left(-\frac{R^2\log(nK^2)}{128(d+1)}\right)\\ &\leqslant C(\sqrt{n}K)^{d/(d+1)-R^2/(64(d+1))}.\end{aligned}$$

An analogous result holds for the imaginary part. The statement follows by

$$\begin{aligned}&\mathbb{P}\left(\max_{u\in[-K,K]^d} |\varphi_n(u) - \varphi(u)| \geqslant \rho\right)\\ &\leqslant \mathbb{P}\left(\max_{u\in[-K,K]^d} |\operatorname{Re}(\varphi_n(u) - \varphi(u))| \geqslant \frac{\rho}{2}\right) + \mathbb{P}\left(\max_{u\in[-K,K]^d} |\operatorname{Im}(\varphi_n(u) - \varphi(u))| \geqslant \frac{\rho}{2}\right).\end{aligned}$$

□

Proposition 7.3 implies that the empirical characteristic function converges uniformly on compact sets in L^p, $p \geqslant 1$, to the true characteristic function with rate $(\log(n)/n)^{1/2}$. Using empirical processes, in particular bracketing entropy arguments, it is possible to improve to a $1/n^{1/2}$-rate and to bound any derivative on the whole real axis.

Theorem 7.4. *(Kappus and Reiß, 2012, [15]) Let X be a one-dimensional Lévy process with finite $(2k+\gamma)$-th moment and choose $w(u) = (\log(e + |u|))^{-1/2-\delta}$ for some constants $\gamma, \delta > 0$ and an integer $k \geqslant 0$. Then for the k-th derivative $\mathcal{C}_{n,\Delta}^{(k)}$ of the empirical characteristic process*

$$\mathcal{C}_{n,\Delta}(u) = \sqrt{n}\left(\frac{1}{n}\sum_{k=1}^{n} e^{iu(X_{k\Delta}-X_{(k-1)\Delta})} - \mathbb{E}\left[e^{iuX_\Delta}\right]\right), \qquad u \in \mathbb{R},\ \Delta > 0,$$

we have

$$\sup_{n\geqslant 1, \Delta\leqslant 1} \Delta^{-(k\wedge 1)/2}\, \mathbb{E}\left[\sup_{u\in\mathbb{R}} \left|\mathcal{C}_{n,\Delta}^{(k)}(u)\right| w(u)\right] < \infty.$$

8 Spectral estimation of the Lévy triplet in the finite intensity case

8.1 Estimation method

Consider a Lévy process X on $\mathbb{R}$, where the Lévy measure ν is absolutely continuous with respect to the Lebesgue measure and with $\lambda = \nu(\mathbb{R}) < \infty$. We observe $X_0, X_\Delta, \ldots, X_{n\Delta}$ for $n \to \infty$, and with $\Delta > 0$ fixed. Our aim is to estimate σ^2, γ, λ and ν. By the Lévy–Khintchine representation we have $\varphi_t(u) = e^{t\psi(u)}$ with

$$\psi(u) = -\frac{1}{2}\sigma^2u^2 + i\gamma u - \lambda + \mathcal{F}\nu(u), \tag{8.1}$$

where $\mathcal{F}\nu(u) = \int_{-\infty}^{\infty} e^{iux}\,\mathrm{d}\nu(x)$ denotes the Fourier transform of ν. By the Riemann–Lebesgue lemma $\mathcal{F}\nu(u) \to 0$ as $|u| \to \infty$. We view ψ as quadratic polynomial in u plus $\mathcal{F}\nu$. We consider the optimisation problem

$$\inf_{(\sigma^2,\gamma,\lambda)} \int_{\{|u|>A\}} w(u)\left|\psi(u) + \frac{1}{2}\sigma^2u^2 - i\gamma u + \lambda\right|^2 \mathrm{d}u$$

for some nonnegative function w and $A > 0$. Let $\varphi_n(u) = \frac{1}{n}\sum_{j=1}^n e^{iu(X_{j\Delta}-X_{(j-1)\Delta})}$ and define $\psi_n(u) = \Delta^{-1}\log(\varphi_n(u))$, where the complex logarithm is taken such that ψ_n is continuous on $(-u_{0,n}, u_{0,n})$ with $\psi_n(0) = 0$ and $u_{0,n}$ being the smallest positive zero of φ_n. Using that φ does not vanish on $\mathbb{R}$ one can show that $u_{0,n} \to \infty$ almost surely [24, Thm 3.2.1, p.165].

We have

$$\psi_n(u) - \psi(u) = \Delta^{-1}\left(\log(\varphi_n(u)) - \log(\varphi(u))\right) \approx \Delta^{-1}\frac{\varphi_n(u) - \varphi(u)}{\varphi(u)}. \tag{8.2}$$

For $\sigma^2 > 0$, $|\varphi(u)|$ decreases exponentially in u so that ψ_n is only a good approximation of ψ for u not too large. So we restrict to $u \in [0, U_n]$ with $U_n \to \infty$ as $n \to \infty$. Let

$$\widetilde{w}^{U_n}(u) := \frac{1}{U_n}\widetilde{w}\left(\frac{u}{U_n}\right),$$

where $\widetilde{w}(u)$ is continuous, $\operatorname{supp}\widetilde{w} \subseteq [0,1]$ and $\widetilde{w}(u) > 0$ on $(0,1)$. We consider the optimisation problem

$$(\sigma_n^2, \lambda_n) := \underset{(\sigma^2,\lambda)}{\operatorname{argmin}} \int_0^\infty \widetilde{w}^{U_n}(u)\left(\operatorname{Re}\psi_n(u) + \tfrac{1}{2}\sigma^2 u^2 + \lambda\right)^2 \mathrm{d}u.$$

The solution is given by

$$\begin{aligned}\sigma_n^2 &= \int_0^\infty w_\sigma^{U_n}(u)\operatorname{Re}\psi_n(u)\,\mathrm{d}u \qquad \text{and}\\ \lambda_n &= \int_0^\infty w_\lambda^{U_n}(u)\operatorname{Re}\psi_n(u)\,\mathrm{d}u\end{aligned}$$

for some $w_\sigma^{U_n}$ and $w_\lambda^{U_n}$. We have

$$\begin{aligned}&\int_0^{U_n}(-u^2/2)w_\sigma^{U_n}(u)\,\mathrm{d}u = 1, && \int_0^{U_n} w_\sigma^{U_n}(u)\,\mathrm{d}u = 0,\\ &\int_0^{U_n}(-1)w_\lambda^{U_n}(u)\,\mathrm{d}u = 1 \qquad \text{and} && \int_0^{U_n}(-u^2/2)w_\lambda^{U_n}(u)\,\mathrm{d}u = 0.\end{aligned} \tag{8.3}$$

Further $w_\sigma^{U_n}(u) = U_n^{-3}w_\sigma^1(u/U_n)$ and $w_\lambda^{U_n}(u) = U_n^{-1}w_\lambda^1(u/U_n)$. The optimisation problem

$$\gamma_n := \underset{\gamma}{\operatorname{argmin}} \int_0^\infty \widetilde{w}^{U_n}(u)\left(\operatorname{Im}\psi_n(u) - \gamma u\right)^2 \mathrm{d}u$$

is solved by $\gamma_n = \int_0^\infty w_\gamma^{U_n}(u)\operatorname{Im}\psi_n(u)\,\mathrm{d}u$ for some $w_\gamma^{U_n}$. We have $\int_0^{U_n} u w_\gamma^{U_n}(u)\,\mathrm{d}u = 1$ and $w_\gamma^{U_n}(u) = U_n^{-2}w_\gamma^1(u/U_n)$. All functions w_σ^1, w_γ^1, w_λ^1 are bounded and supported on $[0,1]$. We denote by ν both the Lévy measure and its density. We define the inverse Fourier transform by $\mathcal{F}^{-1}f(u) = \frac{1}{2\pi}\int_{-\infty}^\infty e^{-iux}f(x)\,\mathrm{d}x$ and estimate the Lévy density by

$$\nu_n(x) = \mathcal{F}^{-1}\left[\left(\psi_n(\cdot) + \frac{\sigma_n^2}{2}(\cdot)^2 - i\gamma_n(\cdot) + \lambda_n\right)w_\nu\left(\frac{\cdot}{U_n}\right)\right](x), \qquad x \in \mathbb{R},$$

where w_ν is a symmetric weight function supported on $[-1,1]$. The estimated Lévy density ν_n might take negative values. One could modify the estimator to ensure nonnegative values.

8.2 Error decomposition

We will exemplify the error analysis by considering $\sigma_n^2 - \sigma^2$. By (8.1) and (8.3) we have

$$\begin{aligned}\sigma_n^2 - \sigma^2 &= \int_0^{U_n} w_\sigma^{U_n}(u)\,\mathrm{Re}(\psi_n(u) - \psi(u))\,\mathrm{d}u + \int_0^{U_n} w_\sigma^{U_n}(u)\,\mathrm{Re}(\psi(u))\,\mathrm{d}u - \sigma^2 \\ &= \underbrace{\int_0^{U_n} w_\sigma^{U_n}(u)\,\mathrm{Re}(\psi_n(u) - \psi(u))\,\mathrm{d}u}_{\text{Stochastic error}} + \underbrace{\int_0^{U_n} w_\sigma^{U_n}(u)\,\mathrm{Re}(\mathcal{F}\,\nu(u))\,\mathrm{d}u}_{\text{Bias}}.\end{aligned}$$

The approximation (8.2) motivates the decomposition

$$\int_0^{U_n} w_\sigma^{U_n}(u)\,\mathrm{Re}(\psi_n(u) - \psi(u))\,\mathrm{d}u = \underbrace{\frac{1}{\Delta}\int_0^{U_n} w_\sigma^{U_n}(u)\,\mathrm{Re}\left(\frac{\varphi_n(u) - \varphi(u)}{\varphi(u)}\right)\mathrm{d}u}_{=:L_n \text{ Linear term}} + \underbrace{R_n}_{\text{Remainder}}.$$

Linear term

By the exercise we know $\mathbb{E}[L_n] = 0$ and

$$\begin{aligned}\mathrm{Cov}_{\mathbb{C}}(\varphi_n(u), \varphi_n(v)) &= \mathbb{E}\left[\varphi_n(u)\overline{\varphi_n(v)}\right] - \mathbb{E}\left[\varphi_n(u)\right]\mathbb{E}\left[\overline{\varphi_n(v)}\right] \\ &= \frac{1}{n}\left(\varphi(u-v) - \varphi(u)\varphi(-v)\right).\end{aligned}$$

Using $|\varphi(u)| \leqslant 1$ for all $u \in \mathbb{R}$ we obtain

$$\begin{aligned}\mathrm{Var}(L_n) &\leqslant \frac{1}{\Delta^2}\int_0^{U_n}\int_0^{U_n} w_\sigma^{U_n}(u) w_\sigma^{U_n}(v)\,\mathrm{Cov}_{\mathbb{C}}\left(\frac{\varphi_n(u)}{\varphi(u)}, \frac{\varphi_n(v)}{\varphi(v)}\right)\mathrm{d}u\,\mathrm{d}v \\ &= \frac{1}{n\Delta^2}\int_0^{U_n}\int_0^{U_n} w_\sigma^{U_n}(u) w_\sigma^{U_n}(v)\varphi^{-1}(u)\varphi^{-1}(-v)(\varphi(u-v) - \varphi(u)\varphi(-v))\,\mathrm{d}u\,\mathrm{d}v \\ &\leqslant \frac{2}{n\Delta^2}\left(\int_0^{U_n}\left|w_\sigma^{U_n}(u)/\varphi(u)\right|\,\mathrm{d}u\right)^2 \\ &= \frac{2}{nU_n^4\Delta^2}\left(\int_0^1\left|w_\sigma^1(u)/\varphi(uU_n)\right|\,\mathrm{d}u\right)^2 =: \varepsilon_{1,n}^2/\Delta^2.\end{aligned}$$

By Markov's inequality

$$\mathbb{P}\left(|L_n| > \frac{A}{\Delta}\varepsilon_{1,n}\right) \leqslant A^{-2}. \tag{8.4}$$

Remainder term

We define the good event

$$\mathcal{G}_n := \left\{\left\|\frac{\varphi_n - \varphi}{\varphi}\right\|_{U_n} \leqslant \frac{1}{2}\right\} \qquad \text{with } \|f\|_{U_n} := \sup_{|u| \leqslant U_n} |f(u)|.$$

It holds $|\log(1+z)-z|\leqslant 2|z|^2$ for $|z|<1/2$. This yields on $\mathcal{G}_n$

$$\psi_n(u)-\psi(u)=\frac{1}{\Delta}(\log\varphi_n(u)-\log\varphi(u))$$
$$=\frac{1}{\Delta}\log\left(1+\frac{\varphi_n(u)-\varphi(u)}{\varphi(u)}\right)=\frac{1}{\Delta}\left(\frac{\varphi_n(u)-\varphi(u)}{\varphi(u)}+O\left(\left|\frac{\varphi_n(u)-\varphi(u)}{\varphi(u)}\right|^2\right)\right).$$

By Proposition 7.3 for $R>8$, $n\in\mathbb{N}$ and $U_n\geqslant 2$

$$\mathbb{P}\left(\sqrt{n}\|\varphi_n-\varphi\|_{U_n}\geqslant R\sqrt{\log(nU_n^2)}\right)\leqslant C(\sqrt{n}U_n)^{(64-R^2)/128}.$$

We have

$$\mathbb{P}(\mathcal{G}_n^c)\leqslant\mathbb{P}\left(\sqrt{n/\log(nU_n^2)}\|\varphi_n-\varphi\|_{U_n}>\frac{1}{2}\sqrt{n/\log(nU_n^2)}\inf_{|u|\leqslant U_n}|\varphi(u)|\right)$$
$$=\mathbb{P}\left(\sqrt{n/\log(nU_n)}\|\varphi_n-\varphi\|_{U_n}>\kappa_n\right)$$
$$=O\left((\sqrt{n}U_n)^{(64-\kappa_n^2)/128}\right)$$

provided that U_n is chosen such that

$$\kappa_n:=\frac{1}{2}\sqrt{n/\log(nU_n^2)}\inf_{|u|\leqslant U_n}|\varphi(u)|>8.$$

This means that U_n should not increase too fast. We define $\varepsilon_{2,n}:=1/\kappa_n$ and using again Proposition 7.3 we obtain

$$\mathbb{P}\left(\|(\varphi_n-\varphi)/\varphi\|_{U_n}^2>A\varepsilon_{2,n}^2\right)\leqslant\mathbb{P}\left(n\|\varphi_n-\varphi\|_{U_n}^2>4A\log(nU_n^2)\right)$$
$$=O\left((\sqrt{n}U_n)^{(64-4A)/128}\right)\tag{8.5}$$

for $A>16$. On $\mathcal{G}_n$ we have

$$|R_n|\lesssim\Delta^{-1}\left\|\frac{\varphi_n-\varphi}{\varphi}\right\|_{U_n}^2\int_0^{U_n}|w_\sigma^{U_n}(u)|\,\mathrm{d}u\lesssim\Delta^{-1}\left\|\frac{\varphi_n-\varphi}{\varphi}\right\|_{U_n}^2U_n^{-2}.\tag{8.6}$$

Remark. (a) The definition of the Fourier transform can be extended from $L^1(\mathbb{R})$ to $L^1(\mathbb{R})\cup L^2(\mathbb{R})$ and the *Plancherel identity* states for all $f,g\in L^2(\mathbb{R})$

$$\int_{-\infty}^{\infty}f(x)\overline{g(x)}\,\mathrm{d}x=\frac{1}{2\pi}\int_{-\infty}^{\infty}\mathcal{F}f(u)\overline{\mathcal{F}g(u)}\,\mathrm{d}u.$$

(b) Let $f\in L^2(\mathbb{R})$ be such that for all $k\in\{0,1,\dots,s\}$ the (weak) derivative $f^{(k)}$ satisfies $f^{(k)}\in L^2(\mathbb{R})$. Then for all $k\in\{0,1,\dots,s\}$

$$\mathcal{F}\left[f^{(k)}\right](u)=(iu)^k\,\mathcal{F}f(u).$$

(c) For $U>0$ we have

$$\mathcal{F}f(u)=U\,\mathcal{F}[f(U\bullet)](Uu),$$
$$\mathcal{F}^{-1}f(u)=U\,\mathcal{F}^{-1}[f(U\bullet)](Uu).$$

Bias

Let ν satisfy for an integer $s \geqslant 0$ that $\max_{k=0,\dots,s} \|\nu^{(k)}\|_{L^2(\mathbb{R})} \leqslant C$ and $\|\nu^{(s)}\|_\infty \leqslant C$ for some $C > 0$. Let $w_\sigma^1(u)/u^s \in L^2(\mathbb{R})$ and $\mathcal{F}\left[w_\sigma^1(u)/u^s\right] \in L^1(\mathbb{R})$. By the Plancherel identity we have

$$\begin{aligned}\left|\int_0^\infty w_\sigma^{U_n}(u)\operatorname{Re}(\mathcal{F}\nu(u))\,\mathrm{d}u\right| &\leqslant \left|\int_0^\infty w_\sigma^{U_n}(u)\,\mathcal{F}\nu(u)\,\mathrm{d}u\right| \\ &= 2\pi\left|\int_{-\infty}^\infty \nu^{(s)}(x)\overline{\mathcal{F}^{-1}[w_\sigma^{U_n}(u)/(iu)^s](x)}\,\mathrm{d}x\right| \\ &= 2\pi U_n^{-(s+3)}\left|\int_{-\infty}^\infty \nu^{(s)}(x)\overline{\mathcal{F}^{-1}[w_\sigma^1(u/U_n)/(u/U_n)^s](x)}\,\mathrm{d}x\right| \\ &\leqslant U_n^{-(s+3)}\|\nu^{(s)}\|_\infty \|\,\mathcal{F}[w_\sigma^1(u)/u^s]\|_{L^1(\mathbb{R})}.\end{aligned}$$

So we obtain

$$\left|\int_0^\infty w_\sigma^{U_n}(u)\operatorname{Re}(\mathcal{F}\nu(u))\,\mathrm{d}u\right| \lesssim U_n^{-(s+3)}. \tag{8.7}$$

8.3 Convergence rates

Definition 8.1. For an integer $s \geqslant 0$ and $R, \sigma_{\max} > 0$ let $\mathcal{G}_s(R, \sigma_{\max})$ denote the set of all Lévy triplets $\tau = (\gamma, \sigma^2, \nu)$ such that ν is s-times (weakly) differentiable and

$$\sigma \in [0, \sigma_{\max}], \quad |\gamma|, \lambda \in [0, R], \quad \max_{k=0,1,\dots,s} \|\nu^{(k)}\|_{L^2(\mathbb{R})} \leqslant R \quad \text{and} \quad \|\nu^{(s)}\|_\infty \leqslant R.$$

Definition 8.2. Let $\{\mathbb{P}_\vartheta, \vartheta \in \Theta\}$ be a family of probability measures on $(\Omega, \mathcal{F})$. Assume that $\xi_n = \xi_n(\vartheta)$ is a sequence of random variables on $(\Omega, \mathcal{F})$. We write $\xi_n = O_{\mathbb{P},\Theta}(r_n)$ for a sequence of positive numbers r_n if

$$\lim_{A\to\infty} \limsup_{n\to\infty} \sup_{\vartheta\in\Theta} \mathbb{P}_\vartheta(|\xi_n(\vartheta)| \geqslant A r_n) = 0.$$

Theorem 8.3. *Suppose that the weight functions w_σ^1, w_γ^1, w_λ^1 and w_ν^1 satisfy*

$$\begin{aligned}&w_\sigma^1(u)/u^s, w_\gamma^1(u)/u^s, w_\lambda^1(u)/u^s, (1 - w_\nu^1(u))/u^s \in L^2(\mathbb{R}), \\ &\mathcal{F}[w_\sigma^1(u)/u^s], \mathcal{F}[w_\gamma^1(u)/u^s], \mathcal{F}[w_\lambda^1(u)/u^s], \mathcal{F}[(1 - w_\nu^1(u))/u^s] \in L^1(\mathbb{R}).\end{aligned}$$

Choosing for some $\bar\sigma > \sigma_{\max}$ the cut-off value $U_n := \bar\sigma^{-1}(\log(n)/\Delta)^{1/2}$, we obtain the convergence rates

$$\begin{aligned}\sigma_n^2 - \sigma^2 &= O_{\mathbb{P},\mathcal{G}_s}((\log n)^{-(s+3)/2}), &&\text{for } s \geqslant 0,\\ \gamma_n - \gamma &= O_{\mathbb{P},\mathcal{G}_s}((\log n)^{-(s+2)/2}), &&\text{for } s \geqslant 0,\\ \lambda_n - \lambda &= O_{\mathbb{P},\mathcal{G}_s}((\log n)^{-(s+1)/2}), &&\text{for } s \geqslant 0,\\ \|\nu_n - \nu\|_\infty &= O_{\mathbb{P},\mathcal{G}_s}((\log n)^{-s/2}), &&\text{for } s \geqslant 1.\end{aligned}$$

Proof for σ_n, sketch of proof for γ_n, λ_n, ν_n. We recall the error decomposition

$$\sigma_n^2 - \sigma^2 = \underbrace{\int_0^{U_n} w_\sigma^{U_n}(u)\operatorname{Re}(\mathcal{F}\nu(u))\,\mathrm{d}u}_{=:B_n\ \text{Bias}} + \underbrace{\frac{1}{\Delta}\int_0^{U_n} w_\sigma^{U_n}(u)\operatorname{Re}\left(\frac{\varphi_n(u) - \varphi(u)}{\varphi(u)}\right)\mathrm{d}u}_{=:L_n\ \text{Linear term}} + \underbrace{R_n}_{\text{Remainder}}.$$

By (8.4) and (8.7) we have

$$|B_n| \lesssim U_n^{-(s+3)} = \left(\frac{\Delta\bar{\sigma}^2}{\log(n)}\right)^{\frac{s+3}{2}},$$
$$\mathbb{P}\left(|L_n| > \frac{A}{\Delta}\varepsilon_{1,n}\right) \leqslant A^{-2}.$$

For n large enough

$$\begin{aligned}\varepsilon_{1,n} &= \frac{\sqrt{2}}{\sqrt{n}U_n^2}\int_0^1 |w_\sigma^1(u)/\varphi(uU_n)|\,\mathrm{d}u \\ &\lesssim \frac{1}{\sqrt{n}U_n^2}\left\|\frac{1}{\varphi}\right\|_{U_n}\int_0^1 |w_\sigma^1(u)|\,\mathrm{d}u \\ &\lesssim \frac{1}{\sqrt{n}\log(n)}n^{\sigma^2/(2\bar{\sigma}^2)} = O(n^{-(1-\sigma_{\max}^2/\bar{\sigma}^2)/2}).\end{aligned}$$

We have by (8.5) and (8.6)

$$|R_n| \lesssim \Delta^{-1}\left\|\frac{\varphi_n-\varphi}{\varphi}\right\|_{U_n}^2 U_n^{-2} \qquad \text{on } \mathcal{G}_n := \left\{\left\|\frac{\varphi_n-\varphi}{\varphi}\right\|_{U_n} \leqslant \frac{1}{2}\right\}$$

and

$$\mathbb{P}\left(\|(\varphi_n-\varphi)/\varphi\|_{U_n}^2 > A\varepsilon_{2,n}^2\right) = O\left((\sqrt{n}U_n)^{(64-4A)/128}\right)$$

with $A > 16$. Furthermore,

$$\begin{aligned}\varepsilon_{2,n} &= 2\sqrt{\log(nU_n^2)/n}\left\|\frac{1}{\varphi}\right\|_{U_n} \\ &\lesssim \sqrt{\frac{\log n}{n}}n^{\sigma^2/(2\bar{\sigma}^2)} = O\left(\sqrt{\log n}\,n^{-(1-\sigma_{\max}^2/(\bar{\sigma}^2))/2}\right).\end{aligned}$$

So $\mathbb{P}(\mathcal{G}_n) \to 1$ as $n \to \infty$. The above bounds yield

$$\lim_{A\to\infty}\limsup_{n\to\infty}\sup_{(\sigma^2,\gamma,\nu)\in\mathcal{G}_s}\mathbb{P}_{(\sigma^2,\gamma,\nu)}\left(|\sigma_n^2-\sigma^2| > A\left(\frac{\Delta\bar{\sigma}^2}{\log n}\right)^{(s+3)/2}\right) = 0.$$

The bounds for the error terms of γ_n and λ_n are larger than the error terms of σ_n^2 by a factor U_n and U_n^2, respectively. Otherwise the convergence rates for γ_n and λ_n follow similarly.

For ν_n we have

$$\begin{aligned}\nu_n(x)-\nu(x) = \mathcal{F}^{-1}&\left[\left((\psi_n-\psi)(u)+\frac{\sigma_n^2-\sigma^2}{2}u^2 - i(\gamma_n-\gamma)u+\lambda_n-\lambda\right)w_\nu\left(\frac{u}{U_n}\right)\right](x) \\ &-\mathcal{F}^{-1}\left[\left(1-w_\nu\left(\frac{u}{U_n}\right)\right)\mathcal{F}\,\nu(u)\right](x).\end{aligned}$$

By the exercises we know

$$\|\,\mathcal{F}^{-1}[(1-w_\nu(u/U_n))\,\mathcal{F}\,\nu(u)]\|_\infty \lesssim U_n^{-s}.$$

The term $\mathcal{F}^{-1}[(\psi_n - \psi)(u)w_\nu(u/U_n)]$ is treated similarly to the stochastic error of σ_n^2. The following terms remain

$$\frac{\sigma_n^2 - \sigma^2}{2} U_n^3 \mathcal{F}^{-1}[u^2 w_\nu(u)](U_n x) - i(\gamma_n - \gamma)U_n^2 \mathcal{F}^{-1}[u w_\nu(u)](U_n x) + (\lambda_n - \lambda)U_n \mathcal{F}^{-1} w_\nu(U_n x).$$

Since $(1-w_\nu(u))/u^s \in L^2(\mathbb{R})$ and $\mathcal{F}[(1-w_\nu(u))/u^s] \in L^1(\mathbb{R})$, we have $(1-w_\nu(u))/u^s \in L^\infty(\mathbb{R})$. By the bounded support of w_ν we infer $w_\nu \in L^\infty(\mathbb{R})$, so that $u^2 w_\nu(u), u w_\nu(u), w_\nu \in L^1(\mathbb{R})$. This yields $\mathcal{F}^{-1}[u^2 w_\nu(u)]$, $\mathcal{F}^{-1}[u^1 w_\nu(u)]$, $\mathcal{F}^{-1} w_\nu \in L^\infty(\mathbb{R})$. The result follows by

$$\left|\frac{\sigma_n^2 - \sigma^2}{2}\right| U_n^3 + |\gamma_n - \gamma| U_n^2 + |\lambda_n - \lambda| U_n = O_{\mathbb{P},\mathcal{G}_s}((\log n)^{-s/2}).$$

□

These rates of σ_n^2, γ_n and λ_n are minimax optimal over the class $\mathcal{G}_s(R, \sigma_{\max})$ [2].

9 Extension to the infinite intensity case

The estimators σ_n, λ_n are designed for the finite intensity case. We want to analyse their behaviour in the infinite intensity case, i.e., under model misspecification. In the infinite intensity case $\operatorname{Re}(\psi(u)) \to -\infty$ even if $\sigma = 0$. Since the jump part of $\operatorname{Re}(\psi(u))$ diverges slower than $-u^2$, adding an additional infinite intensity jump part leads to larger σ_n^2 and larger λ_n when fitting $-\sigma_n^2 u^2/2 - \lambda_n$ to $\operatorname{Re}(\psi(u))$. For $d = 1$ symmetric stable Lévy processes ($\sigma^2 = 0$, $\gamma = 0$, $\nu(x) = -c|x|^{-\alpha-1}$) have the characteristic exponent $\psi(u) = -c'|u|^\alpha$, $\alpha \in (0, 2)$, $c' > 0$. We restrict the analysis to stable like behaviour.

Proposition 9.1. *Suppose the Lévy triplet of the Lévy process X satisfies $\sigma > 0$ as well as $\int_{-\infty}^{\infty}(1 - \cos(ux))\,\mathrm{d}\nu(x) = c_\alpha u^\alpha + O(u^\beta)$ for $0 \leqslant \beta < \alpha < 2$ and $c_\alpha > 0$ with the asymptotics $u \to \infty$. Then for any $\bar\sigma > \sigma$*

$$\sigma_n^2 = \sigma^2 + O_{\mathbb{P}}\left(U_n^{-(2-\alpha)} + n^{-1/2} U_n^{-2} e^{\Delta\bar\sigma^2 U_n^2/2}\right),$$
$$\lambda_n \gtrsim U_n^\alpha + O_{\mathbb{P}}\left(n^{-1/2} e^{\Delta\bar\sigma^2 U_n^2/2}\right).$$

In particular, for U_n as in Theorem 8.3 the estimator σ_n^2 is consistent with rate $(\log n)^{-(2-\alpha)/2}$.

Proof. The bias term of σ_n^2 can be expressed using the general formula (6.1) for ψ:

$$\int_0^{U_n} w_\sigma^{U_n}(u) \operatorname{Re}\psi(u)\,\mathrm{d}u - \sigma^2 = \int_0^{U_n} w_\sigma^{U_n}(u) \int_{-\infty}^{\infty} (\cos(ux) - 1)\,\mathrm{d}\nu(x)\,\mathrm{d}u.$$

Substituting $s = u/U_n$ and using the assumption on ν we obtain

$$\begin{aligned}\left|\int_0^{U_n} w_\sigma^{U_n}(u) \operatorname{Re}\psi(u)\,\mathrm{d}u - \sigma^2\right| &= \left|U_n^{-2} \int_0^1 w_\sigma^1(s) \int_{-\infty}^{\infty} (1 - \cos(U_n s x))\,\mathrm{d}\nu(x)\,\mathrm{d}s\right| \\ &\lesssim U_n^{-2} \int_0^1 \left|w_\sigma^1(s)\right| U_n^\alpha s^\alpha\,\mathrm{d}s + U_n^{-2} \int_0^1 \left|w_\sigma^1(s)\right| U_n^\beta s^\beta\,\mathrm{d}s \\ &\lesssim U_n^{\alpha-2}.\end{aligned}$$

λ_n decomposes into stochastic error and

$$\int_0^{U_n} w_\lambda^{U_n}(u)\operatorname{Re}(\psi(u))\,\mathrm{d}u = \int_0^1 w_\lambda^1(s)\int_{-\infty}^{\infty}(\cos(U_n sx)-1)\,\mathrm{d}\nu(x)\,\mathrm{d}s$$
$$= -c_\alpha U_n^\alpha \int_0^1 w_\lambda^1(s)s^\alpha\,\mathrm{d}s + O(U_n^\beta).$$

By the exercises we know

$$w_\lambda^1(u) = \widetilde{w}(u)\frac{\int_0^1 \widetilde{w}(s)s^2\,\mathrm{d}s\,u^2 - \int_0^1 \widetilde{w}(s)s^4\,\mathrm{d}s}{\int_0^1 \widetilde{w}(s)s^4\,\mathrm{d}s\int_0^1 \widetilde{w}(s)\,\mathrm{d}s - (\int_0^1 \widetilde{w}(s)s^2\,\mathrm{d}s)^2}$$

so that

$$\int_0^1 w_\lambda^1(u)u^\alpha\,\mathrm{d}u = C\left(\int_0^1 \widetilde{w}\,s^2\int_0^1 \widetilde{w}\,s^{2+\alpha} - \int_0^1 \widetilde{w}\,s^4\int_0^1 \widetilde{w}\,s^\alpha\right), \qquad C>0.$$

By the Hölder inequality in $L^1(\widetilde{w})$ with $p=(4-\alpha)/(2-\alpha)$, $q=(4-\alpha)/2$ we obtain

$$\int_0^1 \widetilde{w}\,s^2 = \int_0^1 \widetilde{w}\,s^{\frac{8-4\alpha}{4-\alpha}}s^{\frac{2\alpha}{4-\alpha}} < \left(\int_0^1 \widetilde{w}\,s^4\right)^{1/p}\left(\int_0^1 \widetilde{w}\,s^\alpha\right)^{1/q},$$
$$\int_0^1 \widetilde{w}\,s^{2+\alpha} = \int_0^1 \widetilde{w}\,s^{\frac{8}{4-\alpha}}s^{\frac{2\alpha-\alpha^2}{4-\alpha}} < \left(\int_0^1 \widetilde{w}\,s^4\right)^{1/q}\left(\int_0^1 \widetilde{w}\,s^\alpha\right)^{1/p}.$$

This shows $\int_0^1 w_\lambda^1(u)u^\alpha\,\mathrm{d}u < 0$. Consequently, $\int_0^{U_n} w_\lambda^{U_n}(u)\operatorname{Re}(\psi(u))\,\mathrm{d}u \gtrsim U_n^\alpha$. The analysis of the stochastic errors is as before. □

σ_n^2 achieves the rate $(\log n)^{-(2-\alpha)/2}$, which can be shown to be minimax optimal with respect to jump components whose characteristic function decays at most like $e^{-c|u|^\alpha}$ as $|u|\to\infty$, $c>0$.

10 Spectral estimation for general Lévy measures

Assume $\int_{-\infty}^{\infty} x^2\,\mathrm{d}\nu(x) < \infty$. Then

$$\mathrm{d}\nu_\sigma(x) := \sigma^2\,\mathrm{d}\delta_0(x) + x^2\,\mathrm{d}\nu(x)$$

is a finite measure. The measure ν_σ is a natural object of the Lévy process X since $\operatorname{Var}(X_t) = \nu_\sigma(\mathbb{R})t$, $\psi''(u) = -\sigma^2 + \int_{-\infty}^{\infty}(ix)^2e^{ixu}\,\mathrm{d}\nu(x) = -\mathcal{F}\,\nu_\sigma(u)$ and by the Kolmogorov representation $\varphi_t(u) = e^{t\psi(u)}$ with $\psi(u) = i\gamma u + \int_{-\infty}^{\infty}(e^{iux}-1-iux)x^{-2}\,\mathrm{d}\nu_\sigma(x)$, where the integrand is continuously extended to $-u^2/2$ at $x=0$. Define the reweighted measure $\bar\nu_\sigma$ of ν_σ by

$$\mathrm{d}\bar\nu_\sigma(x) := \sigma^2\,\mathrm{d}\delta_0(x) + \frac{x^2}{1+x^2}\,\mathrm{d}\nu(x).$$

Let $\bar\gamma$ be such that

$$\psi(u) = iu\bar\gamma - \frac{\sigma^2}{2}u^2 + \int_{-\infty}^{\infty}\left(e^{iux}-1-\frac{iux}{1+x^2}\right)\mathrm{d}\nu(x)$$
$$= iu\bar\gamma + \int_{-\infty}^{\infty}\frac{(e^{iux}-1)(1+x^2)-iux}{x^2}\,\mathrm{d}\bar\nu_\sigma(x).$$

The pair $(\bar\gamma,\bar\nu_\sigma)$ characterises weak convergence of $\mathbb{P}_{(\bar\gamma,\bar\nu_\sigma)}$, the law of X_1. By Theorem 19.1 in [12] we have

Proposition 10.1. *The convergence* $\mathbb{P}_{(\bar\gamma_m,\bar\nu_{\sigma,m})} \xrightarrow{w} \mathbb{P}_{(\bar\gamma,\bar\nu_\sigma)}$ *for a sequence of pairs* $(\bar\gamma_m,\bar\nu_{\sigma,m})_{m\geqslant 1}$ *takes place if and only if* $\bar\gamma_m \to \bar\gamma$ *and* $\bar\nu_{\sigma,m}\to\bar\nu_\sigma$ *(weak convergence of finite measures).*

We introduce the Sobolev norm and Sobolev space by

$$\|f\|_{H^1} := \frac{1}{\sqrt{2\pi}}\left\|(1+u^2)^{1/2}\,\mathcal{F} f(u)\right\|_{L^2}$$
$$H^1 := H^1(\mathbb{R}) := \left\{f\in L^2(\mathbb{R}) \mid \|f\|_{H^1}<\infty\right\}.$$

An equivalent norm of H^1 is given by $\|f\|_{L^2}+\|f'\|_{L^2}$, where f' denotes the weak derivative of f. We estimate ν_σ and analyse the performance in H^{-1}, the dual space of H^1. In the spectral domain we shall use

$$\|\mu\|_{H^{-1}} = \frac{1}{\sqrt{2\pi}}\left\|(1+u^2)^{-1/2}\,\mathcal{F}\mu(u)\right\|_{L^2}.$$

We will also use $|\int_{-\infty}^{\infty} f\,\mathrm{d}\mu| \leqslant \|f\|_{H^1}\|\mu\|_{H^{-1}}$ and $\|\mu\|_{H^{-1}} = \sup_{\|f\|_{H^1}=1}|\int_{-\infty}^{\infty} f\,\mathrm{d}\mu|$. We base the estimation on the identity

$$\nu_\sigma = -\mathcal{F}^{-1}[\psi''] = -\frac{1}{\Delta}\mathcal{F}^{-1}\left[(\log\varphi)''\right] = -\frac{1}{\Delta}\mathcal{F}^{-1}\left[\frac{\varphi''}{\varphi}-\left(\frac{\varphi'}{\varphi}\right)^2\right]$$

and a plug-in approach. Let $K\in L^1(\mathbb{R})$ be such that $\int_{-\infty}^{\infty}K(x)\,\mathrm{d}x=1$ and $\operatorname{supp}(\mathcal{F}K)\subseteq[-1,1]$. We define $K_h(x) := \frac{1}{h}K(\frac{x}{h})$ for $h>0$ and

$$\nu_{\sigma,n} := -\mathcal{F}^{-1}[\psi_n''\,\mathcal{F}K_h] := -\frac{1}{\Delta}\mathcal{F}^{-1}\left[\left(\frac{\varphi_n''}{\varphi_n}-\left(\frac{\varphi_n'}{\varphi_n}\right)^2\right)\mathcal{F}K_h\right].$$

We obtain the following error decomposition for ν_σ

$$\nu_{\sigma,n}-\nu_\sigma := \underbrace{-\mathcal{F}^{-1}[(\psi_n''-\psi'')\,\mathcal{F}K_h]}_{\text{stochastic error}}\underbrace{-\mathcal{F}^{-1}[\psi''(\mathcal{F}K_h-1)]}_{\text{approximation error}}.$$

The approximation error can be represented by $-\mathcal{F}^{-1}[\psi''(\mathcal{F}K_h-1)] = K_h*\nu_\sigma-\nu_\sigma$.

Lemma 10.2. *Suppose that the kernel* K *satisfies* $\int_{-\infty}^{\infty}|\eta|^{1/2}|K(\eta)|\,\mathrm{d}\eta<\infty$. *Then we have as* $h\to 0$

$$\|K_h*\nu_\sigma-\nu_\sigma\|_{H^{-1}} \lesssim h^{1/2}.$$

Proof. We calculate by the dual definition of H^{-1}, $\int_{-\infty}^{\infty}K=1$ and by the Cauchy–Schwarz inequality:

$$\begin{aligned}
\|K_h*\nu_\sigma-\nu_\sigma\|_{H^{-1}} &= \sup_{\|f\|_{H^1}=1}\left|\int_{-\infty}^{\infty} f\,\mathrm{d}(K_h*\nu_\sigma-\nu_\sigma)\right|\\
&= \sup_{\|f\|_{H^1}=1}\left|\int_{-\infty}^{\infty}(K_h(-\bullet)*f-f)\,\mathrm{d}\nu_\sigma\right|\\
&\leqslant \sup_{\|f\|_{H^1}=1}\sup_{x\in\mathbb{R}}|(K_h(-\bullet)*f-f)(x)|\,\nu_\sigma(\mathbb{R})\\
&\lesssim \sup_{\|f\|_{H^1}=1}\sup_{x\in\mathbb{R}}\left|\int_{-\infty}^{\infty}(f(x+y)-f(x))K_h(y)\,\mathrm{d}y\right|
\end{aligned}$$

$$
\begin{aligned}
&\lesssim \sup_{\|f'\|_{L^2}=1} \sup_{x\in\mathbb{R}} \left| \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f'(z) \mathbb{1}_{[x,x+y]}(z)\, \mathrm{d}z \right) K_h(y)\, \mathrm{d}y \right| \\
&\leqslant \int_{-\infty}^{\infty} |y|^{1/2} \, |K_h(y)| \, \mathrm{d}y = h^{1/2} \int_{-\infty}^{\infty} |\eta|^{1/2} |K(\eta)| \, \mathrm{d}\eta \lesssim h^{1/2}.
\end{aligned}
$$

□

For the stochastic error we have

Lemma 10.3. *Let X be a one-dimensional Lévy process with finite $(4+\gamma)$-th moment for some $\gamma > 0$. Let $M_h := \max_{k=0,1,2} \sup_{|u|\leqslant 1/h} |(1/\varphi)^{(k)}(u)|$. If $M_h = o(n^{1/2} \log(h_n^{-1})^{-(1+\delta)/2})$ holds for a sequence $h_n \to 0$ and some $\delta > 0$ then we have*

$$
\mathcal{F}^{-1}[\mathcal{F} K_{h_n} \Delta(\psi_n'' - \psi'')](x) = \mathcal{F}^{-1}[\mathcal{F} K_{h_n}((\varphi_n - \varphi)/\varphi)''](x) + R_n(x)
$$

with a second order term R_n satisfying

$$
\|R_n\|_{H^{-1}} = O_{\mathbb{P}}\left(M_{h_n}^2 n^{-1} \log(h_n^{-1})^{1+\delta} \right).
$$

Proof. To linearise $\psi_n'' - \psi'' = \Delta^{-1}(\log(\varphi_n/\varphi))''$, we set $F(y) = \log(1+y)$, $\eta = (\varphi_n - \varphi)/\varphi$ and use

$$
\begin{aligned}
(F \circ \eta)''(u) &= F'(\eta(u))\eta''(u) + F''(\eta(u))\eta'(u)^2 \\
&= F'(0)\eta''(u) + O(\|F''\|_\infty(\|\eta\|_\infty \|\eta''\|_\infty + \|\eta'\|_\infty^2)),
\end{aligned}
$$

where the supremum norms are taken over the ranges of u and $\eta(u)$, respectively. On the event $\Omega_n := \left\{ \|(\varphi_n - \varphi)/\varphi\|_{L^\infty([-1/h,1/h])} \leqslant 1/2 \right\}$ the values of η are in $[-1/2, 1/2]$ and we obtain the error estimate

$$
\begin{aligned}
\sup_{|u|\leqslant h^{-1}} |(\log(\varphi_n/\varphi))''(u) - ((\varphi_n - \varphi)/\varphi)''(u)| &= O\left(\max_{k=0,1,2} \|((\varphi_n - \varphi)/\varphi)^{(k)}\|_{L^\infty([-1/h,1/h])}^2 \right) \\
&= O\left(M_h^2 \max_{k=0,1,2} \|(\varphi_n - \varphi)^{(k)}\|_{L^\infty([-1/h,1/h])}^2 \right).
\end{aligned}
$$

By the moment assumption and by Theorem 7.4 we have for $k = 0, 1, 2$ and any $\delta > 0$

$$
\|(\varphi_n - \varphi)^{(k)}\|_{L^\infty([-1/h,1/h])} = O_{\mathbb{P}}\left(n^{-1/2} \Delta^{(k\wedge 1)/2} \log(h^{-1})^{(1+\delta)/2} \right).
$$

Combining this with the growth assumption on M_h yields $\mathbb{P}(\Omega_n) \to 1$ and then

$$
\sup_{|u|\leqslant h_n^{-1}} \left| \Delta(\psi_n''(u) - \psi''(u)) - ((\varphi_n - \varphi)/\varphi)''(u) \right| = O_{\mathbb{P}}\left(M_{h_n}^2 n^{-1} \log(h_n^{-1})^{1+\delta} \right).
$$

We conclude

$$
\begin{aligned}
\|R_n\|_{H^{-1}} &= \frac{1}{\sqrt{2\pi}} \left\| (1+u^2)^{-1/2} \mathcal{F} R_n(u) \right\|_{L^2} \\
&\leqslant \frac{1}{\sqrt{2\pi}} \left\| (1+u^2)^{-1/2} \right\|_{L^2} \|\mathcal{F} R_n\|_\infty \\
&= O_{\mathbb{P}}\left(M_{h_n}^2 n^{-1} \log(h_n^{-1})^{1+\delta} \right).
\end{aligned}
$$

□

By the exercises $\mathrm{Var}_{\mathbb{C}}\left(\varphi_n^{(k)}(u)\right) \leqslant \frac{1}{n}\mathbb{E}\left[X_\Delta^{2k}\right]$ for $k=0,1,2$. We bound the main stochastic error:

$$\mathbb{E}\left[\left\|\mathcal{F}^{-1}[\mathcal{F}K_h((\varphi_n-\varphi)/\varphi)'']\right\|_{H^{-1}}^2\right] = \frac{1}{2\pi}\mathbb{E}\left[\left\|(1+u^2)^{-1/2}\mathcal{F}K_h((\varphi_n-\varphi)/\varphi)''\right\|_{L^2}^2\right]$$
$$\lesssim M_h^2\int_{-1/h}^{1/h}(1+u^2)^{-1}\sum_{k=0}^{2}\mathrm{Var}_{\mathbb{C}}(\varphi_n^{(k)}(u))\,\mathrm{d}u \lesssim n^{-1}M_h^2.$$

We have proved the following result, where the condition on M_h ensures that R_n is of appropriate order.

Proposition 10.4. *Let X be a one-dimensional Lévy process with finite $(4+\gamma)$-th moment for some $\gamma>0$. Let $K\in L^1(\mathbb{R})$, $\int_{-\infty}^{\infty}K(x)\,\mathrm{d}x=1$, $\mathrm{supp}(\mathcal{F}K)\subseteq[-1,1]$ and $\int_{-\infty}^{\infty}|\eta|^{1/2}|K(\eta)|\,\mathrm{d}\eta<\infty$. Suppose that $h\to 0$ as $n\to\infty$ such that $M_h=O(n^{1/2}\log(h^{-1})^{-(1+\delta)})$ holds for some $\delta>0$. Then the estimator $\nu_{\sigma,n}$ of ν_σ satisfies*

$$\|\nu_{\sigma,n}-\nu_\sigma\|_{H^{-1}} = O_{\mathbb{P}}\left(h^{1/2}+n^{-1/2}M_h\right).$$

Depending on the growth of M_h this result leads to rates ranging from $O_{\mathbb{P}}((\log n)^{-1/4})$ to $O_{\mathbb{P}}(n^{-1/2})$.

11 More on Lévy processes

11.1 Lévy–Itô decomposition

Theorem 11.1. *(See Theorem 2.1 in [18]) Given any $\gamma\in\mathbb{R}$, $\sigma\geqslant 0$ and a Lévy measure ν on $\mathbb{R}$, there exists a probability space on which three independent Lévy processes exist, $X^{(1)}$, $X^{(2)}$ and $X^{(3)}$:*

- *$X^{(1)}$ is a Brownian motion with drift,*

$$X_t^{(1)} = \gamma t + \sigma W_t, \quad t\geqslant 0.$$

- *$X^{(2)}$ is a square integrable martingale with characteristic exponent*

$$\psi^{(2)}(u) = \int_{\mathbb{R}}(e^{iux}-1-iux)\mathbb{1}_{\{|x|\leqslant 1\}}\,\mathrm{d}\nu(x).$$

- *$X^{(3)}$ is a compound Poisson process,*

$$X_t^{(3)} = \sum_{i=1}^{N_t} Y_i, \quad t\geqslant 0,$$

where $N=(N_t)_{t\geqslant 0}$ is a Poisson process with intensity $\lambda := \nu(\mathbb{R}\setminus[-1,1])$ independent of the i.i.d. sequence $(Y_i)_{i\geqslant 1}$ with distribution concentrated on the set $\{x\,|\,|x|>1\}$ and given by $\mathrm{d}\nu/\lambda$ (unless $\lambda=0$ in which case $X^{(3)}$ is identically zero).

By taking $X := X^{(1)} + X^{(2)} + X^{(3)}$ we see that there exists a probability space on which a Lévy process is defined with characteristic exponent

$$\psi(u) = iu\gamma - \frac{\sigma^2 u^2}{2} + \int_{\mathbb{R}} (e^{iux} - 1 - iux\mathbb{1}_{\{|x|\leqslant 1\}})\,\mathrm{d}\nu(x).$$

In other words, the Lévy–Itô decomposition tells us that X is a Lévy process with characteristic triplet (γ, σ^2, ν) if and only if it can be written as the sum of three independent Lévy processes:

$$X_t = \gamma t + \sigma W_t + \lim_{\eta\to 0}\left(\sum_{s\leqslant t}\Delta X_s\mathbb{1}_{\eta<|\Delta X_s|\leqslant 1} - t\int_{\eta<|x|\leqslant 1} x\,\mathrm{d}\nu(x)\right) + \sum_{s\leqslant t}\Delta X_s\mathbb{1}_{|\Delta X_s|>1},$$

where:

- $W = (W_t)_{t\geqslant 0}$ is a standard Brownian motion.
- $(\sum_{s\leqslant t}\Delta X_s\mathbb{1}_{\eta<|\Delta X_s|\leqslant 1} - t\int_{\eta<|x|\leqslant 1} x\,\mathrm{d}\nu(x))_{t\geqslant 0}$ converges in L^2, as η tends to zero, to a martingale denoted by $M = (M_t)_{t\geqslant 0}$ with characteristic function given by

$$\mathbb{E}[e^{iuM_t}] = \exp\left(t\int_{|x|\leqslant 1}(e^{iux} - 1 - iux)\,\mathrm{d}\nu(x)\right).$$

- $(\sum_{s\leqslant t}\Delta X_s\mathbb{1}_{|\Delta X_s|>1})_{t\geqslant 0}$ is a Lévy process with finite Lévy measure, i.e., it is a compound Poisson process with intensity $\lambda := \nu(\{x|\,|x|>1\})$ and jump distribution concentrated on the set $\{x|\,|x|>1\}$ and given by $\mathrm{d}\nu/\lambda$. In particular, its characteristic function is given by

$$\exp\left(t\int_{|x|>1}(e^{iux} - 1)\,\mathrm{d}\nu(x)\right).$$

- The processes $(\gamma t + \sigma W_t)_{t\geqslant 0}$, $(M_t)_{t\geqslant 0}$ and $(\sum_{s\leqslant t}\Delta X_s\mathbb{1}_{|\Delta X_s|>1})_{t\geqslant 0}$ are three independent Lévy processes.

Definition 11.2. If the limit $\lim_{\eta\to 0}\int_{\eta<|x|\leqslant 1} x\,\mathrm{d}\nu(x)$ exists and is finite then we define $\gamma := \lim_{\eta\to 0}\int_{\eta<|x|\leqslant 1} x\,\mathrm{d}\nu(x)$ and call the Lévy process X with the characteristic triplet $(\gamma, 0, \nu)$ a *pure jump Lévy process* (also called purely discontinuous Lévy process).

The above limit γ exists for example if $\int_{-1}^{1}|x|\,\mathrm{d}\nu(x) < \infty$ or if ν is symmetric with respect to the origin that is $\nu([a,b]) = \nu([-b,-a])$ for all $0 < a < b$.

Nota Bene: In the general form of the Lévy–Itô decomposition one separates the large jumps $(\sum_{s\leqslant t}\Delta X_s\mathbb{1}_{|\Delta X_s|>1})_{t\geqslant 0}$ from the small jumps since the infinite sum

$$\sum_{s\leqslant t}\Delta X_s\mathbb{1}_{\Delta X_s\neq 0},\quad t\geqslant 0,$$

is almost surely not defined for Lévy measures ν such that $\int_{-1}^{1}|x|\,\mathrm{d}\nu(x) = \infty$. It can be shown that $\left|\sum_{s\leqslant t}\Delta X_s\right| < \infty$ a.s. whenever $\int_{-1}^{1}|x|\,\mathrm{d}\nu(x) < \infty$. In particular, a pure jump Lévy process X with a Lévy measure ν such that $\int_{-1}^{1}|x|\,\mathrm{d}\nu(x) < \infty$ can be written as the sum of all its jumps, i.e.,

$$X_t = \sum_{s\leqslant t}\Delta X_s\mathbb{1}_{\Delta X_s\neq 0},\quad t\geqslant 0.$$

Observe that the corresponding characteristic triplet is given by $(\int_{|x|\leqslant 1} x\,\mathrm{d}\nu(x), 0, \nu)$, that is its characteristic function is given by

$$\exp\left(t\int_{\mathbb{R}}(e^{iux}-1)\,\mathrm{d}\nu(x)\right).$$

Examples.

- Brownian motion with drift: $X_t = \gamma t + \sigma W_t$, $t \geqslant 0$. The characteristic triplet is given by $(\gamma, \sigma^2, 0)$.
- Poisson process: let N be a Poisson process with intensity λ, then its characteristic triplet is given by $(\lambda, 0, \lambda\delta_1)$.
- Compound Poisson process: $X_t = \sum_{i=1}^{N_t} Y_i$, where N is a Poisson process of intensity λ independent of the i.i.d. sequence $(Y_i)_{i\geqslant 1}$ with common law F. We call F the *jump measure* and λ the *intensity* of X. The characteristic triplet of X is given by $(\lambda\int_{|x|\leqslant 1} x\,\mathrm{d}F(x), 0, \lambda F)$.

11.2 Relationship between the Lévy measure of X and the law of X

Let X be a compound Poisson process with intensity λ and jump measure F. Denote by N_t the number of jumps of X on $[0,t]$. Then for any Borel set A,

$$\begin{aligned}\mathbb{P}(X_t \in A) &= \sum_{n=0}^{\infty}\mathbb{P}(X_t\in A|N_t=n)\,\mathbb{P}(N_t=n)\\ &= e^{-\lambda t}\delta_0(A) + \sum_{n=1}^{\infty}F^{*n}(A)\frac{e^{-\lambda t}(\lambda t)^n}{n!},\end{aligned}$$

where F^{*n} denotes the n-th convolution power of F and δ_0 stands for the Dirac measure at 0. Let ν be the Lévy measure of X, that is

$$\nu(A) = \lambda F(A) = \lambda\,\mathbb{P}(Y_1\in A), \quad \forall A \in \mathcal{B}(\mathbb{R}).$$

In particular, for any Borel set A that does not contain 0, we have

$$\lim_{t\to 0}\frac{\mathbb{P}(X_t\in A)}{t} = \lim_{t\to 0}\left(\lambda\,\mathbb{P}(Y_1\in A)e^{-\lambda t} + \lambda\sum_{n=2}^{\infty}\mathbb{P}(Y_1+\dots+Y_n\in A)\frac{e^{-\lambda t}(\lambda t)^{n-1}}{n!}\right) = \nu(A) \tag{11.1}$$

since

$$0 \leqslant \lambda\sum_{n=2}^{\infty}\mathbb{P}(Y_1+\dots+Y_n\in A)\frac{e^{-\lambda t}(\lambda t)^{n-1}}{n!} \leqslant \frac{e^{-\lambda t}}{t}\sum_{n=2}^{\infty}\frac{(\lambda t)^n}{n!} = \frac{e^{-\lambda t}}{t}(e^{\lambda t}-1-\lambda t)\to 0$$

as $t\to 0$. For general Lévy processes the following theorem holds.

Theorem 11.3. *([14], see also [7]) Let X be a Lévy process with characteristic triplet (γ, σ^2, ν).*

(a) If f is ν-a.e. continuous, bounded and satisfies $f(x) = o(x^2)$ as $x \to 0$ then

$$\lim_{t \to 0} \frac{1}{t} \mathbb{E}[f(X_t)] = \int_{-\infty}^{\infty} f(x) \, \mathrm{d}\nu(x).$$

(b) If f is ν-a.e. continuous, bounded and satisfies $f(x) \sim x^2$ as $x \to 0$ then

$$\lim_{t \to 0} \frac{1}{t} \mathbb{E}[f(X_t)] = \sigma^2 + \int_{-\infty}^{\infty} f(x) \, \mathrm{d}\nu(x).$$

In particular, we have for any point of continuity $s > 0$ of ν that

$$\lim_{t \to 0} \frac{1}{t} \mathbb{P}(X_t \geqslant s) = \nu([s, \infty)).$$

12 High-frequency intensity estimation for compound Poisson processes

Let X be a compound Poisson process, i.e.,

$$X_t = \sum_{i=1}^{N_t} Y_i, \quad t \geqslant 0,$$

where N is a Poisson process with intensity λ and $(Y_i)_{i \geqslant 1}$ is an independent sequence of i.i.d. random variables with common law F. We suppose that F is absolutely continuous with respect to the Lebesgue measure and denote its density by f. In particular, X is a Lévy process with Lévy measure $\nu = \lambda F$. We denote the density of ν by ρ. We observe $\lambda = \nu(\mathbb{R} \setminus \{0\})$.

Our aim is to estimate the intensity λ from discrete observations of X. We observe

$$X_0, X_\Delta, X_{2\Delta}, \ldots, X_{(n-1)\Delta}, X_{n\Delta} \qquad \text{with } n\Delta = T,$$

where $\Delta > 0$ is the observation distance and T the time horizon. We assume that $\Delta \to 0$ and $T \to \infty$ as $n \to \infty$. We set

$$Z_i := X_{i\Delta} - X_{(i-1)\Delta}, \quad i = 1, \ldots, n.$$

The random variables $Z_1, Z_2, \ldots, Z_n$ are i.i.d. with the same law as X_Δ.

By (11.1) we have

$$\lim_{\Delta \to 0} \frac{\mathbb{P}(X_\Delta \neq 0)}{\Delta} = \nu(\mathbb{R} \setminus \{0\}) = \lambda.$$

So for Δ small enough we have

$$\lambda \approx \frac{\mathbb{P}(X_\Delta \neq 0)}{\Delta}. \tag{12.1}$$

We define

$$\widehat{n}(0) := \sum_{i=1}^{n} \mathbb{1}_{Z_i \neq 0}.$$

Replacing $\mathbb{P}(X_\Delta \neq 0)$ by its empirical counterpart $\widehat{n}(0)/n$ in (12.1) leads to the estimator

$$\widehat{\lambda}_n := \frac{\widehat{n}(0)}{n\Delta}. \tag{12.2}$$

The following proposition says that the mean squared error of $\widehat{\lambda}_n$ is of order $\frac{1}{T} + \Delta^2$.

Proposition 12.1. *For $\lambda \in [0, \Lambda]$ the estimator $\widehat{\lambda}_n$ satisfies*

$$\mathbb{E}\left[|\widehat{\lambda}_n - \lambda|^2\right] = O\left(\frac{1}{T} + \Delta^2\right).$$

Proof. By the bias-variance decomposition we have

$$\mathbb{E}\left[|\widehat{\lambda}_n - \lambda|^2\right] = \left(\mathbb{E}\left[\widehat{\lambda}_n\right] - \lambda\right)^2 + \operatorname{Var}\left(\widehat{\lambda}_n\right).$$

We first analyse the bias. Since F is absolutely continuous with respect to the Lebesgue measure we have

$$\mathbb{P}(Z_i \neq 0) = \mathbb{P}(X_\Delta \neq 0) = \mathbb{P}(N_\Delta \neq 0) = 1 - e^{-\lambda\Delta}.$$

It follows

$$\mathbb{E}\left[\widehat{\lambda}_n\right] = \frac{1}{n\Delta}\mathbb{E}\left[\sum_{i=1}^n \mathbb{1}_{Z_i \neq 0}\right] = \frac{1 - e^{-\lambda\Delta}}{\Delta} = \lambda + O(\Delta).$$

Now we analyse the variance. From the previous computations we know $\mathbb{E}\left[\widehat{n}(0)\right] = n(1 - e^{-\lambda\Delta})$. Furthermore,

$$\begin{aligned}
\mathbb{E}\left[\widehat{n}(0)^2\right] &= \mathbb{E}\left[\sum_{i,j=1}^n \mathbb{1}_{Z_i \neq 0}\mathbb{1}_{Z_j \neq 0}\right] \\
&= n\,\mathbb{P}(Z_1 \neq 0) + n(n-1)(\mathbb{P}(Z_1 \neq 0))^2 \\
&= n(1 - e^{-\lambda\Delta}) + (n^2 - n)(1 - e^{-\lambda\Delta})^2.
\end{aligned}$$

This yields

$$\begin{aligned}
\operatorname{Var}\left(\widehat{n}(0)\right) &= \mathbb{E}\left[\widehat{n}(0)^2\right] - \mathbb{E}\left[\widehat{n}(0)\right]^2 = n(1 - e^{-\lambda\Delta}) - n(1 - e^{-\lambda\Delta})^2 \\
&= n(1 - e^{-\lambda\Delta})(1 - (1 - e^{-\lambda\Delta})) = n(1 - e^{-\lambda\Delta})e^{-\lambda\Delta}.
\end{aligned}$$

We recall $n\Delta = T$ and conclude

$$\operatorname{Var}(\widehat{\lambda}_n) = \frac{\operatorname{Var}\left(\widehat{n}(0)\right)}{n^2\Delta^2} = \frac{(1 - e^{-\lambda\Delta})e^{-\lambda\Delta}}{n\Delta^2} = O\left(\frac{1}{T}\right)$$

as $\Delta \to 0$. □

Remark. Another estimator of the intensity can be based on

$$\mathbb{P}(Z_i \neq 0) = 1 - e^{-\lambda\Delta}.$$

This leads to the alternative estimator

$$\widetilde{\lambda}_n := -\frac{1}{\Delta}\log\left(1 - \frac{\widehat{n}(0)}{n}\right).$$

Linearising the estimator $\widetilde{\lambda}_n$ for small Δ we recover the estimator $\widehat{\lambda}_n$ in (12.2). The advantage of $\widetilde{\lambda}_n$ is that it can be expected to work for large Δ as well.

The jump density can be estimated from the density of the nonzero increments (see e.g. [5]). Observe that the the number of nonzero increments and thus the sample size is random.

13 High-frequency estimation of the intensity outside a zero neighbourhood

In the last section we estimated the intensity of compound Poisson processes. In this section we estimate the intensity of general Lévy processes outside of a zero neighbourhood. Let ν be a Lévy measure. If $\int_{|x|\leqslant 1} |x| \, \mathrm{d}\nu(x) < \infty$, the corresponding pure jump process has characteristic triplet $(\int_{|x|\leqslant 1} x \, \mathrm{d}\nu(x), 0, \nu)$ and can be written as

$$X_t = \sum_{s \leqslant t} \Delta X_s \mathbb{1}_{\Delta X_s \neq 0}.$$

Otherwise we will consider the Lévy process with characteristic triplet $(0, 0, \nu)$. So we will focus on the class $\mathscr{L}$ of Lévy processes with characteristic triplets $(\gamma_\nu, 0, \nu)$, where

$$\gamma_\nu := \begin{cases} \int_{|x|\leqslant 1} x \, \mathrm{d}\nu(x) & \text{if } \int_{|x|\leqslant 1} |x| \, \mathrm{d}\nu(x) < \infty, \\ 0 & \text{otherwise.} \end{cases}$$

Thanks to the Lévy–Itô decomposition any X in $\mathscr{L}$ can be written for any $0 < \varepsilon \leqslant 1$ as

$$X_t = B_t(\varepsilon) + M_t(\varepsilon) + t b_\nu(\varepsilon),$$

where:

- $B(\varepsilon) = (B_t(\varepsilon))_{t \geqslant 0}$ is a compound Poisson process with jumps larger than ε. We can write

  $$B_t(\varepsilon) = \sum_{s \leqslant t} \Delta X_s \mathbb{1}_{|\Delta X_s| > \varepsilon}.$$

  $B(\varepsilon)$ has intensity $\lambda_\varepsilon := \nu(\mathbb{R} \setminus [-\varepsilon, \varepsilon])$ and jump distribution $F_\varepsilon := \frac{\nu}{\lambda_\varepsilon} \mathbb{1}_{\mathbb{R} \setminus [-\varepsilon, \varepsilon]}$.

- $M(\varepsilon) = (M_t(\varepsilon))_{t \geqslant 0}$ is a martingale with jumps smaller than ε. We can write

  $$M_t(\varepsilon) = \lim_{\eta \to 0} \left(\sum_{s \leqslant t} \Delta X_s \mathbb{1}_{\eta < |\Delta X_s| \leqslant \varepsilon} - t \int_{\eta < |x| \leqslant \varepsilon} x \, \mathrm{d}\nu(x) \right).$$

- $b_\nu(\varepsilon)$ is given by

  $$b_\nu(\varepsilon) := \begin{cases} \int_{|x|\leqslant \varepsilon} x \, \mathrm{d}\nu(x) & \text{if } \int_{|x|\leqslant 1} |x| \, \mathrm{d}\nu(x) < \infty, \\ -\int_{\varepsilon < |x| \leqslant 1} x \, \mathrm{d}\nu(x) & \text{otherwise.} \end{cases}$$

Assume that ν is absolutely continuous with respect to the Lebesgue measure. We denote the densities of ν and F_ε by ρ and f_ε, respectively. Next we will briefly outline the role of intensity estimation when estimating ρ. Let $\widehat{\rho}$ be an estimator of ρ on a compact set A bounded away from zero. We consider the L^p-risk

$$\mathbb{E}\left[\int_A |\widehat{\rho}(x) - \rho(x)|^p \, \mathrm{d}x \right].$$

Let ε be small enough but fixed such that

$$\rho(x)\mathbb{1}_A(x) = \lambda_\varepsilon f_\varepsilon(x)\mathbb{1}_{|x|>\varepsilon}\mathbb{1}_A(x).$$

We can estimate ρ by

$$\widehat{\rho}(x) = \widehat{\lambda}_\varepsilon \widehat{f}_\varepsilon(x) \qquad \text{for all } x \in A,$$

where $\widehat{\lambda}_\varepsilon$ and $\widehat{f}_\varepsilon$ are estimators of λ_ε and f_ε, respectively. We observe that

$$\begin{aligned}\mathbb{E}\left[\int_A |\widehat{\rho}(x) - \rho(x)|^p \,\mathrm{d}x\right] &= \mathbb{E}\left[\int_A \left|\widehat{\lambda}_\varepsilon \widehat{f}_\varepsilon(x) - \widehat{\lambda}_\varepsilon f_\varepsilon(x) + \widehat{\lambda}_\varepsilon f_\varepsilon(x) - \lambda_\varepsilon f_\varepsilon(x)\right|^p \,\mathrm{d}x\right] \\ &\leqslant 2^{p-1}\,\mathbb{E}\left[|\widehat{\lambda}_\varepsilon|^p \int_A |\widehat{f}_\varepsilon(x) - f_\varepsilon(x)|^p \,\mathrm{d}x\right] + 2^{p-1}\,\mathbb{E}[|\widehat{\lambda}_\varepsilon - \lambda_\varepsilon|^p] \int_A |f_\varepsilon(x)|^p \,\mathrm{d}x.\end{aligned}$$

Furthermore, by the Cauchy–Schwarz inequality we have

$$\int_A \mathbb{E}\left[|\widehat{\lambda}_\varepsilon|^p |\widehat{f}_\varepsilon(x) - f_\varepsilon(x)|^p\right] \,\mathrm{d}x \leqslant \sqrt{\mathbb{E}\left[|\widehat{\lambda}_\varepsilon|^{2p}\right]} \int_A \sqrt{\mathbb{E}\left[|\widehat{f}_\varepsilon(x) - f_\varepsilon(x)|^{2p}\right]} \,\mathrm{d}x.$$

In particular, in order to control the L^p-risk of $\widehat{\rho}$ it is enough to control the L^p- and L^{2p}-risks of $\widehat{\lambda}_\varepsilon$ and $\widehat{f}_\varepsilon$. We will focus on the estimation of λ_ε only. The estimation of f_ε is more involved than in the compound Poisson case owing to the small jumps (see [6]).

Since ν is absolutely continuous with respect to the Lebesgue measure Theorem 11.3 yields

$$\lim_{\Delta\to 0} \frac{\mathbb{P}(|X_\Delta| > \varepsilon)}{\Delta} = \nu(\mathbb{R}\setminus[-\varepsilon,\varepsilon]) = \lambda_\varepsilon.$$

This motivates the estimator

$$\widehat{\lambda}_\varepsilon := \frac{n(\varepsilon)}{n\Delta}$$

with $n(\varepsilon) := \sum_{i=1}^n \mathbb{1}_{(\varepsilon,\infty)}(|X_{i\Delta} - X_{(i-1)\Delta}|)$.

In order to compute the L^p-risk of $\widehat{\lambda}_\varepsilon$ we use Rosenthal's inequality.

Theorem 13.1. *(Rosenthal's inequality [21]) Let $2 < p < \infty$. Then there exists a constant C_p depending only on p, so that if $\xi_1, \ldots, \xi_n$ are independent random variables with $\mathbb{E}[\xi_i] = 0$ and $\mathbb{E}[|\xi_i|^p] < \infty$ for all i, then*

$$\mathbb{E}\left[\left|\sum_{i=1}^n \xi_i\right|^p\right] \leqslant C_p \max\left(\sum_{i=1}^n \mathbb{E}\left[|\xi_i|^p\right], \left(\sum_{i=1}^n \mathbb{E}\left[\xi_i^2\right]\right)^{p/2}\right).$$

Using $(a+b)^p \leqslant 2^{p-1}a^p + 2^{p-1}b^p$ for all $p \geqslant 1$ and for all $a, b \geqslant 0$ we obtain

$$\begin{aligned}\mathbb{E}\left[|\lambda_\varepsilon - \widehat{\lambda}_\varepsilon|^p\right] &= \mathbb{E}\left[\left|\lambda_\varepsilon - \mathbb{E}\left[\frac{n(\varepsilon)}{n\Delta}\right] + \mathbb{E}\left[\frac{n(\varepsilon)}{n\Delta}\right] - \frac{n(\varepsilon)}{n\Delta}\right|^p\right] \\ &\leqslant 2^{p-1}\left|\lambda_\varepsilon - \frac{\mathbb{P}(|X_\Delta| > \varepsilon)}{\Delta}\right|^p + 2^{p-1}\frac{1}{\Delta^p}\,\mathbb{E}\left[\left|\mathbb{P}(|X_\Delta| > \varepsilon) - \frac{n(\varepsilon)}{n}\right|^p\right].\end{aligned}$$

Define

$$U_i := \frac{\mathbb{1}_{(\varepsilon,\infty)}(|X_{i\Delta} - X_{(i-1)\Delta}|) - \mathbb{P}(|X_\Delta| > \varepsilon)}{n} \qquad \text{for } i = 1, \ldots, n.$$

We observe that $U_1, \ldots, U_n$ are i.i.d. bounded centred random variables satisfying

$$\left|\sum_{i=1}^n U_i\right| = \left|\mathbb{P}(|X_\Delta| > \varepsilon) - \frac{n(\varepsilon)}{n}\right|.$$

Applying Rosenthal's inequality for $p > 2$ we obtain

$$\mathbb{E}\left[\left|\mathbb{P}(|X_\Delta| > \varepsilon) - \frac{n(\varepsilon)}{n}\right|^p\right] \leqslant C_p \max\left(\sum_{i=1}^n \mathbb{E}\left[|U_i|^p\right], \left(\sum_{i=1}^n \mathbb{E}\left[U_i^2\right]\right)^{p/2}\right).$$

By the variance of Bernoulli random variables we have

$$\mathbb{E}[U_1^2] = \frac{\mathbb{P}(|X_\Delta| > \varepsilon)(1 - \mathbb{P}(|X_\Delta| > \varepsilon))}{n^2} \leqslant \frac{\mathbb{P}(|X_\Delta| > \varepsilon)}{n^2}$$

and we derive

$$\left(\sum_{i=1}^n \mathbb{E}\left[U_i^2\right]\right)^{p/2} \leqslant \left(\frac{\mathbb{P}(|X_\Delta| > \varepsilon)}{n}\right)^{p/2}.$$

Furthermore, for $p > 2$

$$\begin{aligned}\mathbb{E}\left[\left|\mathbb{1}_{|X_\Delta|>\varepsilon} - \mathbb{P}(|X_\Delta| > \varepsilon)\right|^p\right] &= \mathbb{E}\left[\left|\mathbb{1}_{|X_\Delta|>\varepsilon} - \mathbb{P}(|X_\Delta| > \varepsilon)\right|^2 \left|\mathbb{1}_{|X_\Delta|>\varepsilon} - \mathbb{P}(|X_\Delta| > \varepsilon)\right|^{p-2}\right] \\ &\leqslant \mathbb{E}\left[\left|\mathbb{1}_{|X_\Delta|>\varepsilon} - \mathbb{P}(|X_\Delta| > \varepsilon)\right|^2\right] \leqslant \mathbb{P}(|X_\Delta| > \varepsilon)\end{aligned}$$

and thus $\mathbb{E}[|U_1|^p] \leqslant \mathbb{P}(|X_\Delta| > \varepsilon)/n^p$. Combing the above results we obtain for $p > 2$

$$\mathbb{E}\left[\left|\mathbb{P}(|X_\Delta| > \varepsilon) - \frac{n(\varepsilon)}{n}\right|^p\right] \leqslant C_p \max\left(\frac{\mathbb{P}(|X_\Delta| > \varepsilon)}{n^{p-1}}, \left(\frac{\mathbb{P}(|X_\Delta| > \varepsilon)}{n}\right)^{p/2}\right).$$

Let $n \geqslant 1$ and $\Delta > 0$ such that $n\,\mathbb{P}(|X_\Delta| > \varepsilon) \geqslant 1$. In [6] it is shown that

$$\frac{C_p}{\Delta^p} \max\left(\frac{\mathbb{P}(|X_\Delta| > \varepsilon)}{n^{p-1}}, \left(\frac{\mathbb{P}(|X_\Delta| > \varepsilon)}{n}\right)^{p/2}\right) = O\left(\left(\frac{\mathbb{P}(|X_\Delta| > \varepsilon)}{n\Delta^2}\right)^{p/2}\right).$$

For $p > 2$ we conclude that there exists C depending only on p such that

$$\mathbb{E}\left[|\lambda_\varepsilon - \widehat{\lambda}_\varepsilon|^p\right] \leqslant 2^{p-1}\left|\lambda_\varepsilon - \frac{\mathbb{P}(|X_\Delta| > \varepsilon)}{\Delta}\right|^p + C\left(\frac{\mathbb{P}(|X_\Delta| > \varepsilon)}{n\Delta^2}\right)^{p/2}.$$

For the case $p = 2$ we have

$$\begin{aligned}\mathbb{E}\left[|\lambda_\varepsilon - \widehat{\lambda}_\varepsilon|^2\right] &= (\lambda_\varepsilon - \mathbb{E}[\widehat{\lambda}_\varepsilon])^2 + \mathrm{Var}(\widehat{\lambda}_\varepsilon) \\ &= \left(\lambda_\varepsilon - \frac{\mathbb{P}(|X_\Delta| > \varepsilon)}{\Delta}\right)^2 + \frac{\mathbb{P}(|X_\Delta| > \varepsilon)(1 - \mathbb{P}(|X_\Delta| > \varepsilon))}{n\Delta^2} \\ &\leqslant \left(\lambda_\varepsilon - \frac{\mathbb{P}(|X_\Delta| > \varepsilon)}{\Delta}\right)^2 + \frac{\mathbb{P}(|X_\Delta| > \varepsilon)}{n\Delta^2}.\end{aligned}$$

Turning to the case $1 \leqslant p < 2$ we obtain by Jensen's inequality and the above bound

$$\begin{aligned}
\mathbb{E}\left[|\lambda_\varepsilon - \widehat{\lambda}_\varepsilon|^p\right] &\leqslant \left(\mathbb{E}\left[(\lambda_\varepsilon - \widehat{\lambda}_\varepsilon)^2\right]\right)^{p/2} \\
&\leqslant \left(\left(\lambda_\varepsilon - \frac{\mathbb{P}(|X_\Delta| > \varepsilon)}{\Delta}\right)^2 + \frac{\mathbb{P}(|X_\Delta| > \varepsilon)}{n\Delta^2}\right)^{p/2} \\
&\leqslant \left|\lambda_\varepsilon - \frac{\mathbb{P}(|X_\Delta| > \varepsilon)}{\Delta}\right|^p + \left(\frac{\mathbb{P}(|X_\Delta| > \varepsilon)}{n\Delta^2}\right)^{p/2}.
\end{aligned}$$

Let $n \geqslant 1$ and $\Delta > 0$ such that $n\,\mathbb{P}(|X_\Delta| > \varepsilon) \geqslant 1$. Then the above results yield Theorem 1 in [6], i.e., there exists a constant $C > 0$ depending only on p such that

$$\mathbb{E}\left[|\lambda_\varepsilon - \widehat{\lambda}_\varepsilon|^p\right] \leqslant 2^{p-1}\left|\lambda_\varepsilon - \frac{\mathbb{P}(|X_\Delta| > \varepsilon)}{\Delta}\right|^p + C\left(\frac{\mathbb{P}(|X_\Delta| > \varepsilon)}{n\Delta^2}\right)^{p/2} \qquad \text{for all } p \in [1, \infty).$$

We combine the above statement with the following proposition.

Proposition 13.2. *(Proposition 2.1 in [9]) Suppose that the Lévy density ρ of X is Lipschitz in an open set D_0 containing $D = [a, b] \subseteq \mathbb{R}\setminus\{0\}$ and that $\rho(x)$ is uniformly bounded on $|x| > \eta$ for any $\eta > 0$. Then there exist $k > 0$ and $\Delta_0 > 0$ such that for all $0 < \Delta < \Delta_0$*

$$\begin{aligned}
\sup_{y \in D}\left|\frac{1}{\Delta}\,\mathbb{P}(X_\Delta \geqslant y) - \nu([y, \infty))\right| &< k\Delta \qquad \text{if } D \subseteq \mathbb{R}_{>0}, \\
\sup_{y \in D}\left|\frac{1}{\Delta}\,\mathbb{P}(X_\Delta \leqslant y) - \nu((-\infty, y])\right| &< k\Delta \qquad \text{if } D \subseteq \mathbb{R}_{<0}.
\end{aligned}$$

If the statement of above propositions holds at $y = \varepsilon$ and $y = -\varepsilon$ then we obtain

$$\mathbb{E}\left[|\lambda_\varepsilon - \widehat{\lambda}_\varepsilon|^p\right] \leqslant \widetilde{C}\left(\Delta^p + \left(\frac{\lambda_\varepsilon + \Delta}{n\Delta}\right)^{\frac{p}{2}}\right),$$

where $\widetilde{C} > 0$ depends on p and k only.

14 High-frequency estimation of the Lévy density

We are interested in estimating the Lévy density ρ on an interval $D := [a, b] \subseteq \mathbb{R}\setminus\{0\}$ based on discrete observations up to time T. The interval D is bounded away from zero. We use the *method of sieves.* We consider finite dimensional linear models of functions

$$\mathcal{S} := \{\beta_1\varphi_1 + \cdots + \beta_d\varphi_d \,|\, \beta_1, \ldots, \beta_d \in \mathbb{R}\},$$

where $\varphi_1, \ldots, \varphi_d$ have support in D and are orthonormal with respect to the inner product $\langle p, q\rangle := \int_D p(x)q(x)\,\mathrm{d}x$. We denote by $\|\cdot\|$ the associated norm $\langle\cdot,\cdot\rangle^{1/2}$ on $L^2(D, \mathrm{d}x)$. Relative to the induced distance the element closest to ρ in $\mathcal{S}$ is given by the orthogonal projection

$$\rho^\perp(x) := \sum_{i=1}^{d} \beta(\varphi_i)\varphi_i(x),$$

where $\beta(\varphi_i) := \langle \varphi_i, \rho \rangle = \int_D \varphi_i(x)\rho(x)\,\mathrm{d}x$.

We will estimate ρ by an empirical version of $\rho^\perp$ with coefficients $\beta(\varphi_i)$ replaced by estimators $\widehat{\beta}_n(\varphi_i)$. We denote the observation times by $0 = t_0^n < t_1^n < \cdots < t_n^n = T$. Further we define $\pi^n := (t_k^n)_{k=0}^n$ and $\bar{\pi}^n := \max_k(t_k^n - t_{k-1}^n)$, where we will sometimes drop the superscript n. We suppose that $T \to \infty$ and $\bar{\pi}^n \to 0$ as $n \to \infty$. We estimate $\beta(\varphi)$ by

$$\widehat{\beta}^{\pi^n}(\varphi) := \frac{1}{t_n^n} \sum_{k=1}^n \varphi\left(X_{t_k^n} - X_{t_{k-1}^n}\right).$$

Let us motivate the estimator in the case of equidistant observations $t_k^n - t_{k-1}^n = T/n = \Delta_n$ for all k. We have

$$\begin{aligned}
\mathbb{E}[\widehat{\beta}^{\pi^n}(\varphi)] &= \frac{1}{\Delta_n}\,\mathbb{E}[\varphi(X_{\Delta_n})], \\
\operatorname{Var}\left(\widehat{\beta}^{\pi^n}(\varphi)\right) &= \frac{1}{T}\left(\frac{1}{\Delta_n}\,\mathbb{E}[\varphi^2(X_{\Delta_n})]\right) - \frac{1}{n}\left(\frac{1}{\Delta_n}\,\mathbb{E}[\varphi(X_{\Delta_n})]\right)^2.
\end{aligned}$$

If φ is ν-a.e. continuous, bounded and has support in D then by Theorem 11.3

$$\lim_{n\to\infty} \mathbb{E}[\widehat{\beta}^{\pi^n}(\varphi)] = \int_D \varphi(x)\rho(x)\,\mathrm{d}x \quad \text{and} \quad \lim_{n\to\infty} \operatorname{Var}\left(\widehat{\beta}^{\pi^n}(\varphi)\right) = 0.$$

So $\widehat{\beta}^{\pi^n}(\varphi)$ is an asymptotically unbiased estimator of $\beta(\varphi)$ and its mean squared error vanishes asymptotically. This justifies the estimator

$$\widehat{\rho}^{\pi^n}(x) := \sum_{i=1}^d \widehat{\beta}^{\pi^n}(\varphi_i)\varphi_i(x). \tag{14.1}$$

The estimator $\widehat{\rho}^{\pi^n}$is independent of the specific orthonormal basis of $\mathcal{S}$ since it can shown that $\widehat{\rho}^{\pi^n}$ is the unique solution of the minimisation problem

$$\min_{f\in\mathcal{S}} \gamma_D^{\pi^n}(f),$$

where $\gamma_D^{\pi^n} : L^2(D, \mathrm{d}x) \to \mathbb{R}$ is given by

$$\gamma_D^{\pi^n}(f) := -\frac{2}{t_n^n} \sum_{k=1}^n f(X_{t_k^n} - X_{t_{k-1}^n}) + \int_D f^2(x)\,\mathrm{d}x.$$

We call $\gamma_D^{\pi^n}$ the *contrast function.*

14.1 Properties of the estimators

We decompose the estimation error

$$\widehat{\beta}^{\pi}(\varphi) - \beta(\varphi) = \underbrace{\widehat{\beta}^{\pi}(\varphi) - \mathbb{E}\left[\widehat{\beta}^{\pi}(\varphi)\right]}_{\text{variance part}} + \underbrace{\mathbb{E}\left[\widehat{\beta}^{\pi}(\varphi)\right] - \beta(\varphi)}_{\text{bias part}},$$

where $\beta(\varphi) := \int_{-\infty}^{\infty} \varphi(x)\,\mathrm{d}\nu(x)$. We begin by studying the variance part.

Proposition 14.1. *(Proposition 2.1 in [8]) Let φ be ν-a.e. continuous, bounded and such that $\varphi(x) = o(|x|)$ as $x \to 0$. Let $t_n \to \infty$ and $\bar{\pi} \to 0$ as $n \to \infty$. Then*

$$\sqrt{t_n}\left(\widehat{\beta}^{\pi}(\varphi) - \mathbb{E}\left[\widehat{\beta}^{\pi}(\varphi)\right]\right) \xrightarrow{d} \nu(\varphi^2)^{1/2} Z \qquad \text{as } n \to \infty,$$

where $\nu(\varphi^2) = \int_{-\infty}^{\infty} \varphi^2(x)\,\mathrm{d}\nu(x)$ and Z is a standard normal random variable.

Proof. Let $\Gamma_t(\varphi) := \mathbb{E}[\varphi^2(X_t)] - (\mathbb{E}[\varphi(X_t)])^2$ and $\Delta_k := t_k - t_{k-1}$. We write

$$\sqrt{t_n}\left(\widehat{\beta}^{\pi}(\varphi) - \mathbb{E}\left[\widehat{\beta}^{\pi}(\varphi)\right]\right) = \sum_{k=1}^{n} \xi_k^{\pi},$$

where $\xi_k^{\pi} = \frac{1}{\sqrt{t_n}}(\varphi(X_{t_k} - X_{t_{k-1}}) - \mathbb{E}[\varphi(X_{t_k - t_{k-1}})])$. The assumptions of Lemma 5.5 (a) in [14] are satisfied and it yields $\limsup_{\Delta\to 0} |\frac{1}{\Delta}\Gamma_{\Delta}(\varphi) - \nu(\varphi^2)| = 0$. It follows

$$\sigma_{n,\pi}^2 := \operatorname{Var}\left(\sum_{k=1}^{n} \xi_k^{\pi}\right) = \frac{1}{t_n}\sum_{k=1}^{n} \Gamma_{\Delta_k}(\varphi) \tag{14.2}$$

and

$$\sigma_{n,\pi}^2 - \nu(\varphi^2) = \frac{1}{t_n}\sum_{k=1}^{n} \Delta_k \left(\frac{1}{\Delta_k}\Gamma_{\Delta_k}(\varphi) - \nu(\varphi^2)\right) \longrightarrow 0 \tag{14.3}$$

as $\bar{\pi} \to 0$. This shows the result for $\nu(\varphi^2) = 0$.

For $\nu(\varphi^2) > 0$ we use that φ is bounded and obtain

$$\frac{|\xi_k^{\pi}|}{\sigma_{n,\pi}} \leqslant C\frac{1}{\sqrt{t_n}} \to 0$$

as $n \to \infty$. The result follows by the Lindeberg central limit theorem. □

Next we consider the bias part. Let φ be ν-a.e. continuous, bounded and satisfy $\varphi(x) = o(x^2)$ as $x \to 0$. We define $\mu(f) := \int_{-\infty}^{\infty} f(x)\,\mathrm{d}\mu(x)$. We recall that by Theorem 11.3

$$\limsup_{\Delta\to 0}\left|\frac{1}{\Delta}\,\mathbb{E}[\varphi(X_{\Delta})] - \nu(\varphi)\right| = 0.$$

We obtain

$$\left|\mathbb{E}\left[\widehat{\beta}^{\pi}(\varphi)\right] - \beta(\varphi)\right| \leqslant \frac{1}{t_n}\sum_{k=1}^{n}\Delta_k\left|\frac{1}{\Delta_k}\,\mathbb{E}[\varphi(X_{\Delta_k})] - \nu(\varphi)\right| \to 0 \qquad \text{as } \bar{\pi} \to 0.$$

Combining this with Proposition 14.1 we obtain that $\widehat{\beta}^{\pi}(\varphi)$ is a consistent estimator of $\beta(\varphi)$ if $t_n \to \infty$ and $\bar{\pi} \to 0$.

For the convergence rate and for asymptotic normality we need stronger assumptions. For simplicity we assume that $[a,b] \subseteq \mathbb{R}_{>0}$.

Lemma 14.2. *(Lemma 3.2 in [8]) Suppose that φ has support in $[c,d] \subseteq \mathbb{R}_{>0}$ and that $\varphi|_{[c,d]}$ is continuous with continuous derivative. Then we have*

$$\left|\frac{\mathbb{E}[\varphi(X_{\Delta})]}{\Delta} - \nu(\varphi)\right| \leqslant \left(|\varphi(c)| + \int_c^d |\varphi'(u)|\,\mathrm{d}u\right) M_{\Delta}([c,d]),$$

where $M_{\Delta}([c,d]) := \sup_{y\in[c,d]} |\frac{1}{\Delta}\,\mathbb{P}(X_{\Delta} \geqslant y) - \nu([y,\infty))|$.

Let the Lévy density ρ of X be Lipschitz in an open set D_0 containing $D = [a, b] \subseteq \mathbb{R}_{>0}$ and let $\rho(x)$ be uniformly bounded on $|x| > \eta$ for any $\eta > 0$. Then by Proposition 13.2 there exist $C > 0$ and $\Delta_0 > 0$ such that for all $0 < \Delta < \Delta_0$ we have $M_\Delta([a, b]) < C\Delta$ and thus for $[c, d] \subseteq [a, b]$

$$\left|\frac{\mathbb{E}[\varphi(X_\Delta)]}{\Delta} - \nu(\varphi)\right| \leqslant C\left(|\varphi(c)| + \int_c^d |\varphi'(u)|\,\mathrm{d}u\right)\Delta. \tag{14.4}$$

Definition 14.3. Let Φ be the class of functions φ for which there exists $[c, d] \subseteq [a, b]$ such that φ has support in $[c, d]$ and such that $\varphi|_{[c,d]}$ is continuous with continuous derivative.

Assume $\varphi \in \Phi$. Writing $\Delta_k = t_k - t_{k-1}$ we bound the bias of the estimator by

$$\begin{aligned}\left|\mathbb{E}\left[\widehat{\beta}^\pi(\varphi)\right] - \beta(\varphi)\right| &\leqslant \frac{1}{t_n}\sum_{k=1}^n \Delta_k \left|\frac{1}{\Delta_k}\mathbb{E}[\varphi(X_{\Delta_k})] - \nu(\varphi)\right| \\ &< C\left(|\varphi(c)| + \int_c^d |\varphi'(u)|\,\mathrm{d}u\right)\frac{1}{t_n}\sum_{k=1}^n \Delta_k^2 \\ &\leqslant C\left(|\varphi(c)| + \int_c^d |\varphi'(u)|\,\mathrm{d}u\right)\bar{\pi}.\end{aligned} \tag{14.5}$$

We see that the bias is of order $O(\bar{\pi})$. We can extend the bias bound to linear combinations of functions in Φ. In the proof of Proposition 14.1 we have seen that $\mathrm{Var}(\widehat{\beta}^\pi(\varphi)) = O(t_n^{-1})$. Combining bias and variance bound yields

Theorem 14.4. *Let $t_n \to \infty$ and $\bar{\pi} \to 0$ as $n \to \infty$. If φ is a linear combination of functions in Φ then we have*

$$\mathbb{E}\left[\left(\widehat{\beta}^\pi(\varphi) - \beta(\varphi)\right)^2\right] = O\left(\frac{1}{t_n} + \bar{\pi}^2\right).$$

Under the undersmoothing condition $\bar{\pi}\sqrt{t_n} \to 0$ the bias is asymptotically negligible even after scaling with $\sqrt{t_n}$ and we obtain

Theorem 14.5. *(Theorem 2.3 in [8]) Let $t_n \to \infty$ and $\bar{\pi}\sqrt{t_n} \to 0$ as $n \to \infty$. If φ is a linear combination of functions in Φ then we have*

$$\sqrt{t_n}\left(\widehat{\beta}^\pi(\varphi) - \beta(\varphi)\right) \xrightarrow{d} \nu(\varphi^2)^{1/2} Z \qquad \text{as } n \to \infty.$$

Corollary 14.6. *(Corollary 2.5 in [8]) Suppose that $\varphi_1, \ldots, \varphi_d \in \Phi$ have support in D and are orthonormal with respect to the inner product $\langle p, q\rangle = \int_D p(x)q(x)\,\mathrm{d}x$. Let $t_n \to \infty$ and $\bar{\pi}\sqrt{t_n} \to 0$ as $n \to \infty$. Then the estimator $\widehat{\rho}^\pi$ defined in* (14.1) *satisfies*

$$\sqrt{t_n}\left(\widehat{\rho}^\pi(x) - \rho^\perp(x)\right) \xrightarrow{d} V(x)^{1/2} Z \qquad \text{as } n \to \infty,$$

where $V(x) := \nu(f_x^2) = \int_{-\infty}^\infty f_x^2(y)\,\mathrm{d}\nu(y)$ with $f_x(y) := \sum_{i=1}^d \varphi_i(x)\varphi_i(y)$.

Proof. By linearity of $\widehat{\beta}^\pi$ and β we derive

$$\begin{aligned}&\sqrt{t_n}\left(\widehat{\rho}^\pi(x) - \rho^\perp(x)\right) = \sqrt{t_n}\sum_{i=1}^d \left(\widehat{\beta}^\pi(\varphi_i) - \beta(\varphi_i)\right)\varphi_i(x) \\ &= \sqrt{t_n}\left(\widehat{\beta}^\pi\left(\sum_{i=1}^d \varphi_i(x)\varphi_i\right) - \beta\left(\sum_{i=1}^d \varphi_i(x)\varphi_i\right)\right) = \sqrt{t_n}\left(\widehat{\beta}^\pi(f_x) - \beta(f_x)\right) \xrightarrow{d} V(x)^{1/2} Z\end{aligned}$$

as $n \to \infty$ by Theorem 14.5. □

Remark. Notice that we have the following bound for the variance

$$V(x) \leqslant \|\rho\|_{\infty,D} \sum_{i=1}^{d} \varphi_i^2(x),$$

where $\|\rho\|_{\infty,D} := \sup_{y\in D} \rho(y)$.

14.2 The stochastic error on an interval

We decompose

$$\|\widehat{\rho}^\pi - \rho\|^2 = \underbrace{\|\widehat{\rho}^\pi - \rho^\perp\|^2}_{\text{stochastic error}} + \underbrace{\|\rho^\perp - \rho\|^2}_{\text{approximation error}},$$

where $\|f\|^2 = \int_D f^2(x)\,\mathrm{d}x$.

Standing Assumption 1. *The linear model $\mathcal{S}$ is generated by an orthonormal basis $\mathcal{G} := \{\varphi_1, \ldots, \varphi_d\}$ with $\varphi_i \in \Phi$ for $i = 1, \ldots, d$.*

We introduce the following notation:

$$D_1(\mathcal{S}) := \inf_{\mathcal{G}} \max_{\varphi\in\mathcal{G}} \|\varphi\|_\infty^2,$$
$$D_2(\mathcal{S}) := \inf_{\mathcal{G}} \max_{\varphi\in\mathcal{G}} \|\varphi'\|_1^2,$$

where the infimums are taken over all orthonormal bases $\mathcal{G}$ of $\mathcal{S}$. By Standing Assumption 1 we have that $D_1(\mathcal{S})$ and $D_2(\mathcal{S})$ are finite. They may grow as $\dim(\mathcal{S}) \to \infty$.

Proposition 14.7. *(Proposition 3.4 in [8]) Let the Lévy density ρ of X be Lipschitz on an open set D_0 containing $D = [a, b] \subseteq \mathbb{R}\setminus\{0\}$ and let $\rho(x)$ be uniformly bounded on $|x| > \eta$ for any $\eta > 0$. Then there exists a constant $K > 0$ such that*

$$\mathbb{E}\left[\|\widehat{\rho}^\pi - \rho^\perp\|^2\right] \leqslant K \frac{\dim(\mathcal{S})}{T}$$

for any linear model $\mathcal{S}$ satisfying Standing Assumption 1 and for any partition $\pi : 0 = t_0 < t_1 < \cdots < t_n = T$ such that $T > \max(D_1(\mathcal{S}), D_2(\mathcal{S}))$ and $\bar{\pi} \leqslant T^{-1}$.

Proof. Fix an orthonormal basis $\mathcal{G} := \{\varphi_1, \ldots, \varphi_d\}$ of $\mathcal{S}$ with $\varphi_i \in \Phi$ and corresponding intervals $[c_i, d_i]$ for $i = 1, \ldots, d$. Let $D_\Delta(\varphi) := \frac{1}{\Delta}\mathbb{E}[\varphi(X_\Delta)] - \nu(\varphi)$. For any φ_i we have

$$\mathbb{E}\left[\left(\widehat{\beta}^\pi(\varphi_i) - \beta(\varphi_i)\right)^2\right] = \operatorname{Var}\left(\widehat{\beta}^\pi(\varphi_i)\right) + \left(\mathbb{E}\left[\widehat{\beta}^\pi(\varphi_i)\right] - \beta(\varphi_i)\right)^2.$$

By (14.2), (14.3) and (14.4) we obtain

$$\operatorname{Var}\left(\widehat{\beta}^\pi(\varphi_i)\right) = \frac{\sigma_{n,\pi}^2}{t_n} \leqslant \frac{\nu(\varphi_i^2)}{t_n} + \frac{1}{t_n^2}\sum_{k=1}^{n} \Delta_k \left|D_{\Delta_k}(\varphi_i^2)\right|$$
$$\leqslant \frac{1}{t_n}\int_{c_i}^{d_i} \varphi_i^2(x)\,\mathrm{d}\nu(x) + \frac{C}{t_n^2}\left(|\varphi_i^2(c_i)| + \int_{c_i}^{d_i} |2\varphi_i(u)\varphi_i'(u)|\,\mathrm{d}u\right),$$

where we used $\sum_{k=1}^n \Delta_k^2 \leqslant \sum_{k=1}^n \Delta_k/t_n = 1$. By (14.5) we have

$$\left(\mathbb{E}\left[\widehat{\beta}^\pi(\varphi_i)\right] - \beta(\varphi_i)\right)^2 \leqslant \frac{C^2}{t_n^2}\left(|\varphi_i(c_i)| + \int_{c_i}^{d_i} |\varphi_i'(u)| \, \mathrm{d}u\right)^2.$$

Combining the above yields

$$\begin{aligned}\mathbb{E}\left[\left(\widehat{\beta}^\pi(\varphi_i) - \beta(\varphi_i)\right)^2\right] &\leqslant \frac{1}{T}\int_{c_i}^{d_i} \varphi_i^2(x)\,\mathrm{d}\nu(x) + \frac{C+C^2}{T^2}\left(\|\varphi_i\|_\infty + \|\varphi_i'\|_1\right)^2 \\ &\leqslant \frac{\|\rho\|_{\infty,D}}{T} + 2(C+C^2)\frac{\max_j\left(\|\varphi_j\|_\infty^2 + \|\varphi_j'\|_1^2\right)}{T^2}.\end{aligned}$$

Consequently

$$\mathbb{E}\left[\|\widehat{\rho}^\pi - \rho^\perp\|^2\right] \leqslant \frac{\dim(\mathcal{S})}{T}\left(\|\rho\|_{\infty,D} + 2(C+C^2)\frac{\max_j\left(\|\varphi_j\|_\infty^2 + \|\varphi_j'\|_1^2\right)}{T}\right).$$

The result follows by the assumption $T > \max(D_1(\mathcal{S}), D_2(\mathcal{S}))$. □

14.3 The approximation error on an interval

In order to bound the approximation error we will need smoothness assumptions on ρ. We assume that $\rho|_{[a,b]}$ belongs to the *Besov space* $\mathcal{B}^s_{p\infty}([a,b])$ for some $s > 0$ and $p \in [2,\infty]$ (see for example [4] for further information). Define the difference operator $\Delta_h(f,x) := f(x+h) - f(x)$ and inductively the higher order differences

$$\Delta_h^r(f,x) := \Delta_h(\Delta_h^{r-1}(f,\cdot),x)$$

for all $x \in [a,b]$ such that $x + rh \in [a,b]$ and $r \in \mathbb{N}$. The space $\mathcal{B}^s_{p\infty}([a,b])$ consists of the functions f belonging to $L^p([a,b])$ with $0 < p < \infty$ (or being uniformly continuous for $p = \infty$) such that

$$\|f\|_{\mathcal{B}^s_{p\infty}} := \sup_{\delta>0}\frac{1}{\delta^s}\sup_{0<h\leqslant\delta}\|\Delta_h^r(f,\cdot)\|_p < \infty,$$

where $r := \lfloor s \rfloor + 1$ with $\lfloor s \rfloor$ denoting the integer part of s.

The advantage of working with Besov-smooth functions is that we have bounds available for the approximation errors by polynomials, splines, trigonometric polynomials and wavelets (see [4] and [1]). For example, let $\mathcal{S}_{k,m}$ be the space of piecewise polynomials of degree at most k on a regular partition of $[a,b]$ into m subintervals of equal length. Let $\rho \in \mathcal{B}^s_{p\infty}([a,b])$ with $s < k+1$. Then there exists a constant $c_{\lfloor s \rfloor} < \infty$ such that

$$\inf_{f\in\mathcal{S}_{k,m}} \|\rho - f\|_p \leqslant c_{\lfloor s \rfloor}(b-a)^s\|\rho\|_{\mathcal{B}^s_{p\infty}} m^{-s}$$

and for $p \in [2,\infty]$

$$\|\rho - \rho_m^\perp\| \leqslant c_{\lfloor s \rfloor}(b-a)^{\frac{1}{2}-\frac{1}{p}+s}\|\rho\|_{\mathcal{B}^s_{p\infty}} m^{-s},$$

where $\rho_m^\perp$ denotes the orthogonal projection of ρ onto $\mathcal{S}_{k,m}$. Notice that the functions in $\mathcal{S}_{k,m}$ are not necessarily smooth (not even continuous). The above bounds can be extended to certain subsets of splines in $\mathcal{S}_{k,m}$.

Let us gives bounds on $D_1(\mathcal{S}_{k,m})$ and $D_2(\mathcal{S}_{k,m})$. We will use *Legendre polynomials.* For $j=0,1,\dots$ let P_j be a polynomial of degree j such that

$$\int_{-1}^{1} P_j(x)P_i(x)\,\mathrm{d}x = 0 \qquad \text{if } j \neq i.$$

This determines the Legendre polynomials up to their scale, which we fix by $P_j(1)=1$. The space $\mathcal{S}_{k,m}$ is generated by the orthonormal functions

$$\widehat{\varphi}_{i,j}(x) := \sqrt{\frac{2j+1}{x_i - x_{i-1}}} P_j\left(\frac{2x-(x_i+x_{i-1})}{x_i-x_{i-1}}\right)\mathbb{1}_{(x_{i-1},x_i)}(x),$$

where $i=1,\dots,m$, $j=0,\dots,k$, and $a=x_0<\dots<x_m=b$ are equally spaced points. It holds $|P_j(x)|\leqslant 1$ and $|P_j'(x)|\leqslant P_j'(1)=\frac{j(j+1)}{2}$. Denoting $\Delta_x := x_i - x_{i-1} = (b-a)/m$ we have

$$\widehat{\varphi}_{i,j}'(x) = 2\sqrt{2j+1}\Delta_x^{-3/2}P_j'\left(\frac{2x-(x_i+x_{i-1})}{x_i-x_{i-1}}\right)\mathbb{1}_{(x_{i-1},x_i)}(x),$$
$$\|\widehat{\varphi}_{i,j}'\|_1 \leqslant 2\sqrt{2j+1}\Delta_x^{-3/2}\int_{x_{i-1}}^{x_i}\sup_{u\in[-1,1]}\left|P_j'(u)\right|\,\mathrm{d}x \leqslant \sqrt{2j+1}\Delta_x^{-1/2}j(j+1).$$

It follows

$$D_2(\mathcal{S}_{k,m}) \leqslant \max_{i,j}\{\|\varphi_{i,j}'\|_1^2\} \leqslant \frac{(k+1)^2k^2(2k+1)}{b-a}m$$

and

$$D_1(\mathcal{S}_{k,m}) \leqslant \frac{2k+1}{b-a}m.$$

14.4 Convergence rate on an interval

Let $a,b\in\mathbb{R}$ and $\varepsilon>0$ be given such that $D_0=(a-\varepsilon,b+\varepsilon)\subseteq\mathbb{R}\backslash\{0\}$. Let $s,L>0$ and $M:\mathbb{R}_{>0}\to\mathbb{R}_{>0}$ such that $\liminf_{\eta\to 0}M(\eta)>0$. Define $\Theta^s(L,M)$ to be the class of Lévy densities ρ such that

- ρ is L-Lipschitz on D_0,
- for any $\eta>0$ we have $\rho(x)\leqslant M(\eta)$ for all x with $|x|>\eta$ and
- $\rho|_{[a,b]}$ belongs to $\mathcal{B}^s_{p\infty}([a,b])$ with $\|\rho\|_{\mathcal{B}^s_{p\infty}}<L$ for some $p\in[2,\infty]$.

Theorem 14.8. *(Proposition 3.5 in [8]) Let $m_T := \lfloor T^{1/(2s+1)}\rfloor$ and let $\bar{\pi}\leqslant T^{-1}$. Then*

$$\limsup_{T\to\infty} T^{s/(2s+1)} \sup_{\rho\in\Theta^s(L,M)} \left(\mathbb{E}\left[\|\widehat{\rho}_T-\rho\|^2\right]\right)^{1/2} < \infty,$$

where for each T the estimator $\widehat{\rho}_T = \widehat{\rho}^{\pi}_{m_T}$ is given by (14.1) *with $\mathcal{S}=\mathcal{S}_{k,m_T}$ and $k>s-1$.*

Proof. From the two previous sections we know that there exists a constant K (depending on $k,a,b,\varepsilon,s,p,L,M$) such that

$$\mathbb{E}\left[\|\widehat{\rho}^{\pi}_m-\rho^{\perp}_m\|^2\right] \leqslant K\frac{m}{T} \quad \text{and} \quad \|\rho^{\perp}_m-\rho\|\leqslant Km^{-s},$$

for $m\in\mathcal{M}_T:=\{m'|T>Km'\}$. So there exists a constant $C>0$ such that for T large enough

$$\sup_{\rho\in\Theta^s(L,M)} \mathbb{E}\left[\|\widehat{\rho}_T-\rho\|^2\right] \leqslant C\left(\lfloor T^{1/(2s+1)}\rfloor T^{-1} + \lfloor T^{1/(2s+1)}\rfloor^{-2s}\right).$$

This shows the statement of the theorem. □

14.5 Lower bound on an interval

In this section we state a lower bound result that ensures that no estimator can achieve a faster convergence rate than $T^{-s/(2s+1)}$ even under continuous-time observations. Inspection of the proofs of the lower bounds in [8] shows that they are also valid for the slightly smaller classes $\Theta^s(L, M)$ defined above. So we have

$$\liminf_{T\to\infty} T^{s/(2s+1)} \left(\inf_{\widehat{\rho}_T} \sup_{\rho\in\Theta^s(L,M)} \left(\mathbb{E}\left[\|\widehat{\rho}_T - \rho\|^2\right]\right)^{1/2} \right) > 0,$$

where the infimum is taken over all estimators $\widehat{\rho}_T$ based on continuous-time observations $(X_t)_{t\in[0,T]}$. This means that no estimator can achieve uniformly over the class $\Theta^s(L, M)$ a faster convergence rate than $T^{-s/(2s+1)}$. The estimator $\widehat{\rho}_T$ from the previous sections achieves this minimax optimal rate using only discrete-time observations.

Acknowledgement

Sections 1, 2 and 4 are based on the script for a course given by Markus Reiß in the winter semester 2014/2015. There is no direct reference for Section 3. Section 5 summarises the results of the paper by Emmanuel Gobet, Marc Hoffmann and Markus Reiß [13]. Sections 6-10 follow the exposition by Denis Belomestny and Markus Reiß in the book Lévy Matters IV [2]. Sections 11-13 are based on the lecture notes of the course Statistics for Stochastic Processes given by Ester Mariucci in the winter semester 2017/2018. Section 14 presents the results of a paper by José E. Figueroa-López [8] in the light of his improved small-time asymptotic result in [9].

References

[1] Andrew Barron, Lucien Birgé, and Pascal Massart. Risk bounds for model selection via penalization. *Probab. Theory Related Fields*, 113(3):301–413, 1999.

[2] Denis Belomestny and Markus Reiß. Estimation and calibration of Lévy models via Fourier methods. In *Lévy Matters IV*, volume 2128 of *Lecture Notes in Math.*, pages 1–76. Springer, Cham, 2015.

[3] Thomas Björk. The pedestrian's guide to local time. ArXiv:1512.08912, 2015.

[4] Ronald A. DeVore and George G. Lorentz. *Constructive approximation.* Springer–Verlag, Berlin, 1993.

[5] Céline Duval. Density estimation for compound Poisson processes from discrete data. *Stochastic Process. Appl.*, 123(11):3963–3986, 2013.

[6] Céline Duval and Ester Mariucci. Spectral-free estimation of Lévy densities in high-frequency regime. ArXiv:1702.08787, 2017.

[7] José E. Figueroa-López. Small-time moment asymptotics for Lévy processes. *Statist. Probab. Lett.*, 78(18):3355–3365, 2008.

[8] José E. Figueroa-López. Nonparametric estimation for Lévy models based on discrete-sampling. *IMS Lecture Notes-Monograph Series*, 57:117–146, 2009.

[9] José E. Figueroa-López. Sieve-based confidence intervals and bands for Lévy densities. *Bernoulli*, 17(2):643–670, 2011.

[10] Markus Fischer and Giovanna Nappo. On the moments of the modulus of continuity of Itô processes. *Stoch. Anal. Appl.*, 28(1):103–122, 2010.

[11] Daniéle Florens-Zmirou. On estimating the diffusion coefficient from discrete observations. *J. Appl. Probab.*, 30(4):790–804, 1993.

[12] B. V. Gnedenko and A. N. Kolmogorov. *Limit distributions for sums of independent random variables.* Addison-Wesley Publishing Company, Inc., Cambridge, Mass., 1954.

[13] Emmanuel Gobet, Marc Hoffmann, and Markus Reiß. Nonparametric estimation of scalar diffusions based on low frequency data. *Ann. Statist.*, 32(5):2223–2253, 2004.

[14] Jean Jacod. Asymptotic properties of power variations of Lévy processes. *ESAIM Probab. Stat.*, 11:173–196, 2007.

[15] Johanna Kappus and Markus Reiß. Estimation of the characteristics of a Lévy process observed at arbitrary frequency. *Stat. Neerl.*, 64(3):314–328, 2010.

[16] Ioannis Karatzas and Steven E. Shreve. *Brownian Motion and Stochastic Calculus.* Springer–Verlag, New York, second edition, 1991.

[17] Uwe Küchler and Michael Sørensen. *Exponential Families of Stochastic Processes.* Springer Series in Statistics. Springer–Verlag, New York, 1997.

[18] Andreas E. Kyprianou. *Introductory lectures on fluctuations of Lévy processes with applications.* Universitext. Springer–Verlag, Berlin, 2006.

[19] Robert S. Liptser and Albert N. Shiryaev. *Statistics of Random Processes. I General Theory.* Springer–Verlag, Berlin, 2001.

[20] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion.* Springer-Verlag, Berlin, third edition, 1999.

[21] Haskell P. Rosenthal. On the subspaces of L^p $(p > 2)$ spanned by sequences of independent random variables. *Israel J. Math.*, 8:273–303, 1970.

[22] Ken-iti Sato. *Lévy Processes and Infinitely Divisible Distributions.* Cambridge University Press, Cambridge, 2013.

[23] A. V. Skorokhod. *Asymptotic Methods in the Theory of Stochastic Differential Equations.* American Mathematical Society, Providence, RI, 1989.

[24] Nikolai G. Ushakov. *Selected Topics in Characteristic Functions.* Modern Probability and Statistics. VSP, Utrecht, 1999.