Lévy process, Lévy driven SDE, and quasi-likelihood estimation *

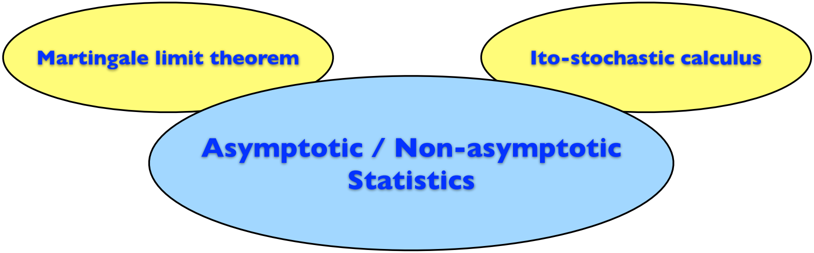
Hiroki Masuda

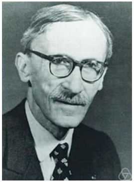
Kyushu University
JST CREST

YUIMA Summer School 2019
Brixen-Bressanone, Italy
June 25–28, 2019

* This version: June 27, 2019

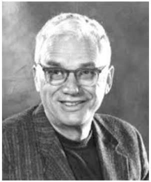
Contents, June 27 a.m.



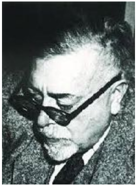


Paul Lévy
(1886-1971)

Kiyosi Itô
(1915-2008)



Joseph L. Doob
(1910-2004)



Norbert Wiener
(1894-1964)

- Standard references:
  - [Applebaum, 2009]
  - [Bertoin, 1996]
  - [Protter, 2005, (2nd.) Chapter I.4]
  - [Sato, 1999]
  - [Iacus and Yoshida, 2018] for many YUIMA examples

Discrete-time random walk $S_1, S_2, \ldots$

$$S_n := \sum_{j=1}^{n} \epsilon_j, \qquad S_0 := 0$$

- $\epsilon_1, \epsilon_2, \ldots$: i.i.d. random variables
- **Independent and stationary increments**

$$S_k - S_l = \sum_{j=l+1}^{k} \epsilon_j$$

1. $S_{j_1} - S_{j_0}, S_{j_2} - S_{j_1}, \ldots, S_{j_n} - S_{j_{n-1}}$ independent $(n \in \mathbb{N})$
2. $S_{j_k} - S_{j_{k-1}} \sim S_{j_k - j_{k-1}}$ $(k \in \mathbb{N})$

- Natural *continuous-time* counterpart?

Lévy process: Continuous-time random walk

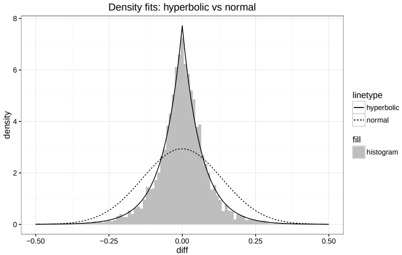
$$X_t = \sum_{j=1}^{n}(X_{t_j} - X_{t_{j-1}}) =: \sum_{j=1}^{n} \Delta_j X \quad (X_0 = 0 \text{ a.s.})$$

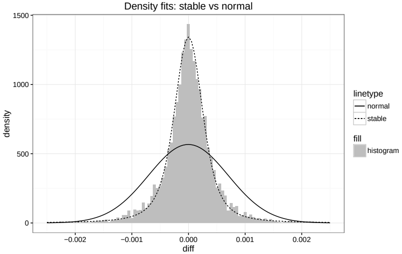
Definition

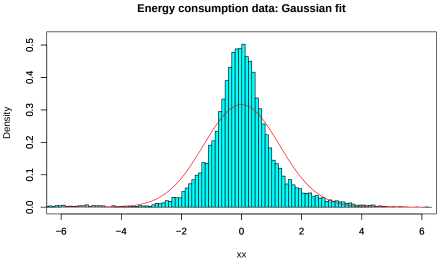
1. **Independent and stationary increments** $(0 = t_0 < t_1 < \cdots < t_n;\, n \in \mathbb{N})$
   - $X_{t_1} - X_{t_0},\ X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}}$ are independent.
   - $X_{t_j} - X_{t_{j-1}} \sim X_{t_j - t_{j-1}}$
2. Continuity in probability: $X_s \xrightarrow{p} X_t$ as $s \to t$.

- No pre-assigned jump time: $\mathbb{P}(|\Delta X_t| > 0) = 0$ for each $t > 0$.
- W.l.g. we may suppose that $t \mapsto X_t(\omega)$ is càdlàg.
- $\exists$Lévy process X s.t. $X_1 \sim F \iff F$ is **infinitely divisible**
  - F is infinitely divisible $\overset{\text{def.}}{\iff} \forall n \, \exists F_n,\ F = F_n^{*n} (:= F_n * \cdots * F_n)$

- Real data may be often leptokurtic: higher kurtosis than the normal (NYSE minutes data); maybe also skewed (energy consumption data).







Two prominent cases

- *If X is a counting process*:

$$X_t = \sum_{j \in \mathbb{N}} I(s \leq \tau_j) \text{ where } 0 < \tau_1 < \tau_2 < \dots \text{ are event-occurrence times,}$$

then X is necessarily a **Poisson process** with intensity λ:

$$\exists \lambda > 0, \quad X_t \sim \mathrm{Pois}(\lambda t), \quad t \in \mathbb{R}_+.$$

- *If X has continuous sample paths*, then X is necessarily a **Wiener process**:

$$\exists \mu \in \mathbb{R}\ \exists \sigma \geq 0, \quad X_t \sim N(\mu t, \sigma^2 t), \quad t \in \mathbb{R}_+,$$

i.e. we may write

$$X_t = \mu t + \sigma w_t$$

for a standard Wiener process w ($w_t \sim N(0, t)$).

Lévy-Khintchine representation[‡]

- Form of the Fourier transform:

$$\varphi_{X_t}(u) := \mathbb{E}(e^{iuX_t}) = \exp\{t\psi(u)\}, \quad u \in \mathbb{R},$$

with the characteristic exponent function

$$\psi(u) := iu\mu_1 - \frac{1}{2}\sigma^2 u^2 + \int \left(e^{iuz} - 1 - iuzI(|z| \leq 1)\right) \nu(dz).$$

- The element (μ_1, σ^2, ν) is called the **generating triplet** of Z:
  1. μ_1 is the **drift (location-shift)**,
  2. $\sigma^2 \geq 0$ is the **Gaussian variance**,
  3. ν is the **Lévy measure** (roughly, expected jump frequency).
     - $\int (1 \wedge |z|^2)\nu(dz) < \infty$ necessarily holds. [†]

[†] Can be of infinite mass: $\nu(\{|z| \leq 1\}) = +\infty$ in general, yet not too much.
[‡] Of dimension $d = 1$, while completely analogous for $d \geq 2$.

Remarks

- $\forall q > 0$: "$\mathbb{E}(|X_t|^q) < \infty \iff \int_{|z|>1} |z|^q \nu(dz) < \infty$"
- Mean and variance of X_t: if φ_{X_t} is of $\mathcal{C}^2$-class, then

$$\mathbb{E}(X_t) = i^{-1} t\psi'(0) = t\mu_1 + t\int_{|z|>1} z\nu(dz),$$
$$\mathrm{var}(X_t) = i^{-2} t\psi''(0) = t\sigma^2 + t\int z^2\nu(dz).$$

- kth cumulant of X_t: if φ_{X_t} is of $\mathcal{C}^k$-class ($k \geq 3$), then

$$i^{-k}\partial_u^k \log\varphi_{X_t}(0) = t\int z^k\nu(dz).$$

$$\frac{1}{t}\log\varphi_{X_t}(u) = iu\mu_1 - \frac{1}{2}\sigma^2 u^2 + \int \left(e^{iuz} - 1 - iuzI(|z|\le 1)\right)\nu(dz)$$

- In general, we should **not** do something like

$$\int \left(e^{iuz} - 1 - iuzI(|z|\le 1)\right)\nu(dz) \\ = \int \left(e^{iuz} - 1\right)\nu(dz) - iu\int_{|z|\le 1} z\nu(dz).$$

- The generating triplet uniquely determines the law of the **process** X, so that it determines e.g.

$$\mathcal{L}\Big(\sup_{t\in[0,1]}|X_t|\Big),\quad \mathcal{L}\big(\inf\{t\ge 0:\, |X_t|>1\}\big),\quad \mathcal{L}\Big(\int_0^1 X_s ds\Big).$$

Lévy-Itô decomposition of sample path

- Sum of independent Gaussian and non-Gaussian factors:

$$X_t = \mu_1 t + \sigma w_t + J_t$$

- More formally [Applebaum, 2009]:

$$X_t = t\mu_1 + \sigma w_t + \int_0^t \int_{|z|>1} z\mu(ds, dz) + \int_0^t \int_{|z|\le 1} z(\mu - \nu)(ds, dz). \qquad (1)$$

  - Poisson random measure $\mu((0,t], A) := \sum_{0<s\le t} I(\Delta X_s \in A \setminus \{0\})$
  - Semimartingale decomposition: local martingale plus finite-variation process.
  - The truncation level "1" could be any positive real instead.
- The jump-part form is mainly of theoretical importance; sometimes it may not be essential for simulation.

Toward generating discrete-time sample

Want to generate sample $X_{t_1}, X_{t_2}, \ldots, X_{t_n}$ from

$$X_t = \mu t + \sigma w_t + J_t,$$

where $0 = t_0 < t_1 < \cdots < t_n = t$ are (fine) sampling time points.

- Enough to be able to generate X_t for *any* $t > 0$:

$$X_t = \sum_{j=1}^{n} (X_{t_j} - X_{t_{j-1}}) = \sum_{j=1}^{n} \Delta_j X, \qquad \Delta_j X \sim X_{t_j - t_{j-1}}$$

- Simulator list in YUIMA [Brouste et al., 2014]:
  - Help documents of `rng` function in YUIMA.
  - [Iacus and Yoshida, 2018, Chapter 4]

Example: Compound Poisson process

$$X_t = \sum_{j=1}^{N_t} \epsilon_j$$

- $N_t \sim \mathrm{Pois}(\lambda t)$: Poisson process with intensity $\lambda > 0$.
- $\epsilon_1, \epsilon_2, \cdots \sim$ i.i.d. with $\mathbb{P}(\epsilon_1 = 0) = 0$, independent of N.
- *Any Lévy process can be a weak limit of a compound Poisson process.*
  - [Sato, 1999, Corollary 8.8]

▷ yss2019_hm_demo.html

Example: Inverse-Gaussian subordinator

- The density of $X_t \sim \mathrm{IG}(\delta t, \gamma)$, $\delta, \gamma > 0$, is

$$x \mapsto \frac{\delta t e^{\delta t \gamma}}{\sqrt{2\pi}} x^{-3/2} \exp\left\{ -\frac{1}{2}\left(\frac{(\delta t)^2}{x} + \gamma^2 x \right) \right\}, \quad x > 0.$$

▷ yss2019_hm_demo.html

Example: Normal inverse Gaussian Lévy process

- Normal variance-mean mixture, i.e. **subordination** §:

$$X_t = \mu t + \beta\tau_t + w_{\tau_t}$$

  - $\tau_t \sim \mathrm{IG}(\delta t, \gamma)$,
  - Standard Wiener process w independent of τ.

- The density of $X_t \sim NIG(\alpha, \beta, \delta t, \mu t)$ is

$$x \mapsto \frac{\alpha\delta t \exp\{\delta t\sqrt{\alpha^2 - \beta^2} + \beta(x - \mu t)\}K_1(\alpha\psi(x;\delta t, \mu t))}{\pi\psi(x;\delta t, \mu t)}$$

  - $\alpha^2 := \gamma^2 + \beta^2$
  - $\psi(x;\delta t, \mu t) := \sqrt{(\delta t)^2 + (x - \mu t)^2}$

▷ yss2019_hm_demo.html

§ General subordination formulae for probability and Lévy densities are available ($\tau \to X$); see [Iacus and Yoshida, 2018, Sect 4.8.3]; [Sato, 1999, chap 6] for general account.

Put simply

$$\varphi_{X_t}(u) := \mathbb{E}(e^{iuX_t}) = \exp\{t\psi(u)\}, \quad u \in \mathbb{R},$$
$$\psi(u) := iu\mu_1 - \frac{1}{2}\sigma^2 u^2 + \int \left(e^{iuz} - 1 - iuzI(|z| \le 1)\right) \nu(dz).$$

- Lévy process is completely characterized by the **generating triplet**, which
  - sometimes crucial in calculations,
  - while sometimes does not matter at all.
- Given any **infinitely divisible** distribution F, there essentially uniquely corresponds a Lévy process X such that $X_1 \sim F$.
- `rng` has several slots for generating specific Lévy process (see help file)
  - Approximate "inputting-$\nu(dz)$" way, yet to be implemented.

- A reader-friendly and comprehensive monograph is [Applebaum, 2009].

- Diffusion process is an SDE driven by a Wiener process:

$$dX_t = a(X_t)dt + b(X_t)dw_t,$$

a strong solution X realized as a functional form

$$X = F(X_0, w).$$

  - e.g. Geometric Brownian motion:

$$dX_t = X_t(\mu dt + \sigma dw_t),$$
$$X_t = X_0 \exp\left\{\sigma w_t + \left(\mu - \frac{\sigma^2}{2}\right) t\right\}.$$

- The driving Wiener process w could be replaced by a **Lévy process**.

Lévy driven Stochastic Differential Equation (SDE)

$$dX_t = a(X_t)dt + b(X_t)dw_t + c(X_{t-})dJ_t$$

$$\left(X_t = x_0 + \int_0^t a(X_s)ds + \int_0^t b(X_s)dw_s + \int_0^t c(X_{s-})dJ_s \right)$$

- Initial variable $X_0 \in \mathbb{R}^d$, possibly random.
- Driving noises:
  - d'-dimensional standard Wiener process $w = (w^j)_{j=1}^{d'}$;
  - d''-dimensional pure-jump Lévy process $J = (J^j)_{j=1}^{d''}$ of the form

$$J_t := \int_0^t \int_{|z|\le 1} z(\mu - \nu)(ds, dz) + \int_0^t \int_{|z|>1} z\mu(ds, dz).$$

- Coefficient functions:
  - Drift coefficient $a(x) = \{a_k(x)\}_{k\le d} : \mathbb{R}^d \to \mathbb{R}^d$
  - Diffusion coefficient $b(x) = \{b_{kl}(x)\}_{k\le d;\, l\le d'} : \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^{d'}$
  - Jump coefficient $c(x) = \{c_{kl}(x)\}_{k\le d;\, l\le d''} : \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^{d''}$

$$dX_t = a(X_t)dt + b(X_t)dw_t + c(X_{t-})dJ_t$$
$$\left(X_t = x_0 + \int_0^t a(X_s)ds + \int_0^t b(X_s)dw_s + \int_0^t c(X_{s-})dJ_s \right)$$

- The stochastic integrations:

$$\int_0^t Y_{s-}dJ_s := \lim_{n\to\infty} \sum_{j=1}^n Y_{(j-1)t/n} \left(J_{jt/n} - J_{(j-1)t/n} \right)$$
$$= \int_0^t \int_{|z|>1} Y_s z\mu(ds, dz) + \int_0^t \int_{|z|\le 1} Y_s z(\mu - \nu)(ds, dz)$$

with the notation in (1), [Applebaum, 2009, Section 6.2];

- L^2-stochastic integration theory for small-jump part,
- Pathwise interlacing of large-jump component for large-jump part.

$$dX_t = a(X_t)dt + b(X_t)dw_t + c(X_{t-})dJ_t$$
$$\left(X_t = x_0 + \int_0^t a(X_s)ds + \int_0^t b(X_s)dw_s + \int_0^t c(X_{s-})dJ_s \right)$$

- Globally Lipschitz (a, b, c): $\exists K > 0,\ \forall x_1, x_2 \in \mathbb{R}^d$,

$$|a(x_1) - a(x_2)| + |b(x_1) - b(x_2)| + |c(x_1) - c(x_2)| \le K|x_1 - x_2|,$$

  leads to existence of unique strong solution (a (w, J)-Lévy functional)

$$X =: F(x_0, w, J).$$

  - [Applebaum, 2009, Theorems 6.2.9 and 6.4.6].

- The simplest but widely applicable device to approximate a solution process to is the **Euler(-Maruyama) scheme** [Platen and Bruti-Liberati, 2010].

Euler-discretization scheme

- As in the case of diffusions, for $t_j - t_{j-1}$ small enough,

$$\begin{aligned} X_{t_{j+1}} &= X_{t_j} + \int_{t_j}^{t_{j+1}} a(X_s)ds + \int_{t_j}^{t_{j+1}} b(X_s)dw_s + \int_{t_j}^{t_{j+1}} c(X_{s-})dJ_s \\ &\approx X_{t_j} + \int_{t_j}^{t_{j+1}} a(X_{t_{j-1}})ds + \int_{t_j}^{t_{j+1}} b(X_{t_{j-1}})dw_s + \int_{t_j}^{t_{j+1}} c(X_{t_{j-1}})dJ_s \\ &\approx X_{t_{j-1}} + a(X_{t_{j-1}})(t_j - t_{j-1}) + b(X_{t_{j-1}})(w_{t_j} - w_{t_{j-1}}) \\ &\quad + c(X_{t_{j-1}})(J_{t_j} - J_{t_{j-1}}) \end{aligned} \tag{2}$$

Need to generate $J_{t_j} - J_{t_{j-1}} (\sim \Delta_j J)$-random number at each step.

- For this, YUIMA internally use the **rng** slots.

Generation of discretized process

The Euler-discretized process $X^\Delta =: (X_t^\Delta)$, for $\Delta > 0$ small enough:

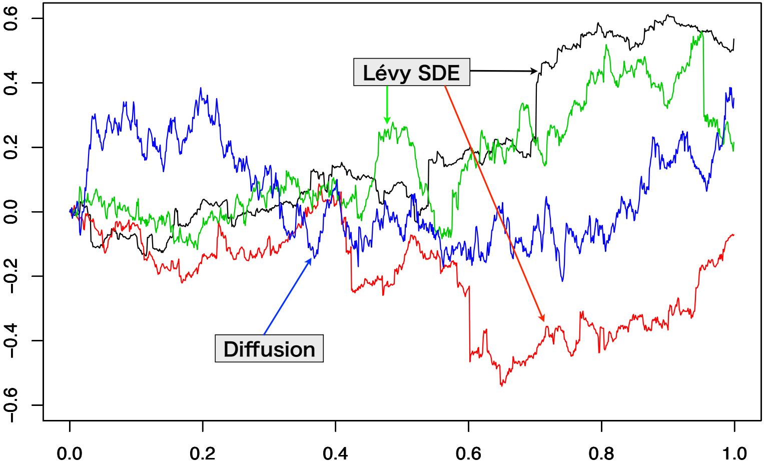
1. $X_t^\Delta := X_0$ for $t \in [0, \Delta)$.
2. For $t \in [j\Delta, (j+1)\Delta)$, $j \in \mathbb{N}$,

$$X_t^\Delta := X_{(j-1)\Delta}^\Delta + a_{j-1}\Delta + b_{j-1}\Delta_j w + c_{j-1}\Delta_j J.$$

- $f_{j-1} := f(X_{(j-1)\Delta})$
- $\Delta_j x = \Delta_j^n x := x_{t_j} - x_{t_{j-1}}$: the jth increment of a process x

- Then, we approximate as $X_t \approx X_t^\Delta$ over a period $[0, T]$;
  - Having generated a finest-approximating process X^Δ,
  - we can extract any thinned process, say $X^{k\Delta}$ for some $k \geq 2$,
  - which plays a role of discretely observed sample from X.
- Strong and weak approximation errors are defined as in diffusion cases.

- May seem like:



Example: Diffusion with compound-Poisson jumps

- For J begin a compound Poisson process with $\Gamma(3,3)$-distributed jumps,

$$dX_t = \{\sin(X_t) - X_t\}\, dt + 2dw_t - dJ_t.$$

  - Downward jumps only.

- The recipe YUIMA uses for $(X_t)_{t\in[0,T]}$:

$$dX_t = a(X_t)dt + b(X_t)dw_t + c(X_{t-})dJ_t$$

0. Generate X_0 and $N_T \leftarrow \mathrm{Pois}(\lambda T)$ independently, and set $j = 1$.
    - 0.1. If $N_T = 0$, then $(X_t)_{t\in[0,T]}$ is a diffusion $dX_t = a(X_t)dt + b(Z_t)dw_t$.
    - 0.2. Otherwise, generate $U_1, \dots, U_{N_T} \sim$ i.i.d.$U(0,T)$;
      Sort them as $U_{(1)} < U_{(2)} < \cdots < U_{(N_T)}$;
      For $k \le N_T$, pick a $j_k \in \{1, 2, \dots, [T/\Delta]\}$ s.t. $U_{(k)} \in ((j_k - 1)\Delta, j_k\Delta]$. ¶
1. Generate $\eta_j \sim N_{d'}(0, I_{d'})$ and then
    - 1.1. If $j = j_k$ for some k, then generate $\zeta_k \sim F(dz)$ (jump law) and deliver

$$X_{j\Delta} \leftarrow X_{(j-1)\Delta} + a_{j-1}\Delta + b_{j-1}\sqrt{\Delta}\,\eta_j + c_{j-1}\zeta_k.$$

    - 1.2. Otherwise, $X_{j\Delta} = X_{(j-1)\Delta} + a_{j-1}\Delta + b_{j-1}\sqrt{\Delta}\,\eta_j$.
2. Update $j = j + 1$ and return to step 1: repeat step 1 until $j = [T/\Delta]$.

▷ yss2019_hm_demo.html

¶Ignores the possibility of multiple i_j: that'll be negligoble for Δ small enough.

Example: Geometric Lévy process

- SDE driven by a general Lévy process $X_t = \mu t + \sigma w_t + J_t$:

$$dY_t = Y_{t-} dX_t, \qquad Y_0 = 1,$$

$$Y_t = \exp\left(X_t - \frac{\sigma^2}{2} t\right) \prod_{0<s\le t} (1 + \Delta X_s) e^{-\Delta X_s}$$

  - Can be checked by **Itô's formula** (in the next page).
  - If in particular $J \equiv 0$, then Y is the geometric Brownian motion:

$$Y_t = \exp\left\{\sigma w_t + \left(\mu - \frac{\sigma^2}{2}\right) t\right\}$$

- Yet, we apply Euler-Maruyama scheme directly.

Itô's formula (Univariate case)

$$dX_t = a(X_t)dt + b(X_t)dw_t + c(X_{t-})dJ_t$$

$$\begin{aligned}
f(X_t) &= f(X_0) + \int_0^t f'(X_{s-})dX_s + \frac{1}{2}\int_0^t f''(X_{s-})d\langle X^c\rangle_s \\
&\quad + \sum_{0<s\le t}\{f(X_s) - f(X_{s-}) - f'(X_{s-})\Delta X_s\} \\
&= f(X_0) + \int_0^t \left(a(X_s)f'(X_{s-}) + \frac{1}{2}b^2(X_s)f''(X_{s-})\right) ds \\
&\quad + \int_0^t f'(X_{s-})b(X_s)dw_s + \int_0^t f'(X_{s-})c(X_{s-})dJ_s \\
&\quad + \sum_{0<s\le t}\{f(X_{s-} + c(X_{s-})\Delta J_s) - f(X_{s-}) - f'(X_{s-})c(X_{s-})\Delta J_s\}.
\end{aligned}$$

▷ yss2019_hm_demo.html

Example: Heavy-tailed SDE

- Non-Gaussian infinite-variance Lévy process $J_t \sim S_\alpha(\beta, \sigma, \mu)$:

$$\varphi_{J_t}(u) = \begin{cases} -(t^{1/\alpha}\sigma)^\alpha |u|^\alpha \left(1 - i\beta \mathrm{sign}(u) \tan\frac{\alpha\pi}{2}\right) + i\mu t u, & \alpha \neq 1 \\ -t\sigma|u|\left(1 + i\frac{2\beta}{\pi}\mathrm{sign}(u)\log|u|\right) + i\mu t u, & \alpha = 1 \end{cases}$$

  - $(\alpha, \beta, \sigma, \mu) \in (0,2) \times [-1,1] \times (0,\infty) \times \mathbb{R}$:
    - $\alpha > 1 \Rightarrow$ Finite mean and infinite variance
    - $\alpha = 1 \Rightarrow$ Cauchy (possibly skewed)
    - $\alpha < 1 \Rightarrow$ Infinite mean
  - The index $\alpha \in (0,2)$ controls tail heaviness and small-jump activity.

- SDE driven by a Lévy process $J_1 \sim \mathrm{Stable}(1.3, 0, 1, 0)$:

$$dX_t = -\frac{X_t}{\sqrt{1 + X_t^2}} dt + dJ_t$$

▷ yss2019_hm_demo.html

Example: Multidimensional nonlinear SDE

- The same recipe as in the one-dimensional case (2):

$$
\begin{aligned}
\underset{d\times 1}{X_{t_{j+1}}} &= X_{t_j} + \int_{t_j}^{t_{j+1}} \underset{d\times 1}{a(X_s)}ds + \int_{t_j}^{t_{j+1}} \underset{d\times r}{b(X_s)}\underset{r\times 1}{dw_s} + \int_{t_j}^{t_{j+1}} \underset{d\times m}{c(X_{s-})}\underset{m\times 1}{dJ_s} \\
&\approx X_{t_{j-1}} + a(X_{t_{j-1}})(t_j - t_{j-1}) + b(X_{t_{j-1}})(w_{t_j} - w_{t_{j-1}}) \\
&\qquad + c(X_{t_{j-1}})(J_{t_j} - J_{t_{j-1}})
\end{aligned}
$$

  - Just matrix multiplications, keeping the form:

(Predictable coefficient) $\times$ **(Noise increment)**

- YUIMA has slots for exact generation of multidimensional $J_{t_j} - J_{t_{j-1}}$.

- Two-dim. SDE $X = (X^1, X^2)$ driven by a 2-dim. NIG Lévy process:

$$d\begin{pmatrix} X_t^1 \\ X_t^2 \end{pmatrix} = \begin{pmatrix} -2X_t^1 \\ 0.3X_t^1 - 1/\sqrt{1+(X_t^2)^2} \end{pmatrix} dt + \begin{pmatrix} 1/\sqrt{1+(X_t^1)^2} & -0.5 \\ 0 & 1 \end{pmatrix} dJ_t$$

▷ yss2019_hm_demo.html

Put simply

$$dX_t = a(X_t)dt + b(X_t)dw_t + c(X_{t-})dJ_t$$

$$J_t = \int_0^t \int_{|z|\le 1} z(\mu - \nu)(ds, dz) + \int_0^t \int_{|z|>1} z\mu(ds, dz)$$

- Good (a, b, c) leads to the existence of unique strong solution.
- `simulate` in YUIMA can generate X, as soon as YUIMA can generate J_h.
- YUIMA has several options for $\mathcal{L}(J_h)$-random numbers.

- Discrete-time location-scale time series model:

$$X_n = b(X_{n-1}, \beta) + a(X_{n-1}, \alpha)\epsilon_n$$

  - $\epsilon_1, \epsilon_2, \cdots \sim$ i.i.d. $(0, 1)$
  - $\theta = (\alpha, \beta)$: Statistical parameter, to be estimated from $(X_1, \dots, X_n)$.

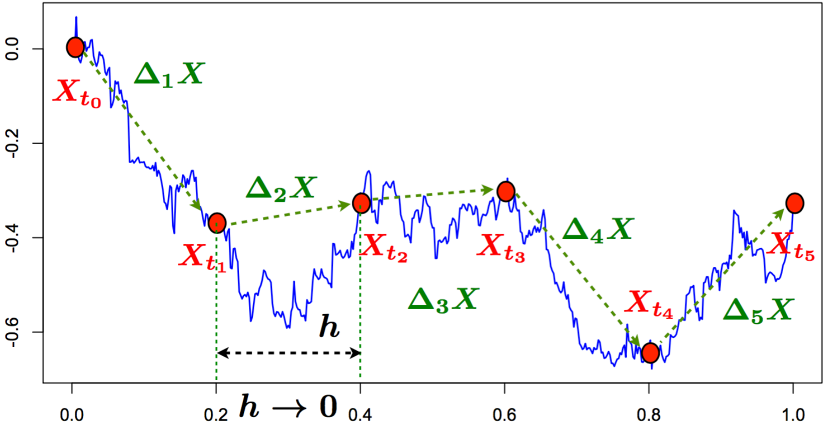
- A natural continuous-time counterpart is [a]

$$dX_t = a(X_{t-}, \alpha)dZ_t + b(X_t, \beta)dt, \qquad \theta = (\alpha, \beta)$$

  - Z is a **standard Lévy process**: $\mathbb{E}(Z_t) = 0$ and $\mathrm{var}(Z_t) = t$.
  - Estimate θ from $(X_{t_0}, X_{t_1}, \dots, X_{t_n})$.

[a] Note: the coefficient notation got changed! $(a \leftrightarrow b)$

- *High-frequency sampling* can provide us with *unified* inference strategies, which generally cannot be shared with the discrete-time framework.



Setup: joint asymptotics

- Univariate parametric Stochastic differential equation (SDE):

$$dX_t = a(X_{t-}, \alpha)dZ_t + b(X_t, \beta)dt, \qquad \theta = (\alpha, \beta)$$

- Available data $(X_{t_j})_{j=0}^n$; $t_j = jh_n = jh$, $T_n := nh \to \infty$, $nh^2 \to 0$.
- Driving Lévy process s.t. $\mathbb{E}(Z_t) = 0$, $\mathrm{var}(Z_t) = t$:

$$Z_t = \sigma W_t + \int_0^t \int z\,(\mu - \nu)(ds, dz),$$

  - A nuisance element.

Regularity conditions

$$dX_t = a(X_{t-}, \alpha)dZ_t + b(X_t, \beta)dt$$

- Correctly specified [‖] smooth coefficients, known up to $\theta := (\alpha, \beta)$.
- Stability ((Exponential) Ergodicity and bounded moments) [**]:

$$\frac{1}{T}\int_0^T f(X_s)ds \xrightarrow{p} \int f(x)\pi(dx), \quad T \to \infty, \tag{3}$$

$$\forall q > 0, \quad \sup_{t \in \mathbb{R}_+} \mathbb{E}\left(|X_t|^q\right) < \infty. \tag{4}$$

- Identifiability:

$$\text{“}(a(\cdot, \alpha), b(\cdot, \beta)) = (a(\cdot, \alpha_0), b(\cdot, \beta_0)),\ \pi\text{-a.e.”} \Rightarrow \theta = \theta_0$$

[‖] i.e. True data-generating parameter $\theta_0 = (\alpha_0, \beta_0)$ is assumed to exist.

[**] A reader-friendly and comprehensive monograph is [Kulik, 2018].

Remark on the stability assumption

$$dX_t = a(X_t)dt + b(X_t)dw_t + c(X_{t-})dJ_t.$$

- X is "exponentially" ergodic (hence (3)) and (4) if:
  1. (a, b, c) is of class $\mathcal{C}^1(\mathbb{R})$ and globally Lipschitz, and (b, c) is bounded.
  2. Either one of the following conditions holds true:
     (i) $b(x') \neq 0$ for some x', $c(x'') \neq 0$ for every x'', and there exists a constant $\overline{\epsilon} > 0$ such that $\nu(-\epsilon, 0) \wedge \nu(0, \epsilon) > 0$ for every $\epsilon \in (0, \overline{\epsilon})$;
     (ii) $b \equiv 0$, $c(x'') \neq 0$ for every x'', and we have the decomposition
     $$\nu = \nu_\star + \nu_\natural$$
     for two Lévy measures $\nu_\star$ and $\nu_\natural$, where the restriction of $\nu_\star$ to some open set of the form $(-\overline{\epsilon}, 0) \cup (0, \overline{\epsilon})$ admits a continuously differentiable positive density $g_\star$.
  3. $\mathbb{E}(J_1) = 0$ and $\int_{|z|>1} |z|^q \nu(dz) < \infty$ for every $q > 0$, and
     $$\limsup_{|x|\to\infty} \frac{a(x)}{x} < 0.$$
- See [Masuda, 2013, Sect 5] for details; another conditions are possible.

GQLF and GQMLE

$$dX_t = a(X_{t-}, \alpha)dZ_t + b(X_t, \beta)dt, \qquad \theta = (\alpha, \beta)$$

- Gaussian approximation in small time ($f_{j-1}(\theta) := f(X_{t_{j-1}}, \theta)$):

$$X_{t_j} \overset{\mathbb{P}_\theta}{\approx} X_{t_{j-1}} + a_{j-1}(\alpha)\Delta_j Z + hb_{j-1}(\beta)$$
$$\overset{\mathcal{L}(\mathbb{P}_\theta)}{\approx} X_{t_{j-1}} + a_{j-1}(\alpha)N(0,h) + hb_{j-1}(\beta)$$

Gaussian quasi-likelihood function (GQLF) and Gaussian QMLE (GQMLE)

$$\mathbb{H}_n(\theta) = \sum_{j=1}^{n} \log\phi\left(X_{t_j};\, X_{t_{j-1}} + hb_{j-1}(\beta),\, ha^2_{j-1}(\alpha)\right), \qquad (5)$$

$$\hat{\theta}_n = (\hat{\alpha}_n, \hat{\beta}_n) \in \operatorname{argmax} \mathbb{H}_n.$$

- User's input:
    - Function forms of scale $a(x, \alpha)$ and drift $b(x, \beta)$.
    - Small sampling stepsize value $h = h_n$.

Asymptotic normality: joint asymptotics

$$dX_t = a(X_{t-},\alpha)dZ_t + b(X_t,\beta)dt, \qquad \theta = (\alpha,\beta)$$

- Diffusion case $Z = w$ [Kessler, 1997]:

$$\left(\sqrt{n}(\hat{\alpha}_n - \alpha_0),\ \sqrt{T_n}(\hat{\beta}_n - \beta_0)\right) \xrightarrow{\mathcal{L}} N_{p_\alpha+p_\beta}\left(0,\ \mathrm{diag}\{\mathcal{I}_\alpha^{-1}(\alpha_0), \mathcal{I}_\beta^{-1}(\theta_0)\}\right)$$

- In the presence of jumps [Masuda, 2013]

$$\left(\sqrt{T_n}(\hat{\alpha}_n - \alpha_0),\ \sqrt{T_n}(\hat{\beta}_n - \beta_0)\right) \xrightarrow{\mathcal{L}} N_{p_\alpha+p_\beta}\left(0,\ \begin{pmatrix} \nu_4 \mathcal{I}_\alpha^{-1}(\theta_0) & \text{sym.} \\ \nu_3 \mathcal{J}_{\alpha\beta}(\theta_0) & \mathcal{I}_\beta^{-1}(\beta_0) \end{pmatrix}\right)$$

- Straightforward to estimate $\mathcal{I}_\alpha(\theta_0)$, $\mathcal{I}_\beta(\beta_0)$, and $\mathcal{J}_{\alpha\beta}(\theta_0)$ empirically.
- Non-Gaussian structure of Z appears in the asymptotic covariance:

$$\nu_k := \int z^k \nu(dz), \qquad k = 3, 4.$$

- Difference in magnitude in small time:

$$X_{t_{j+1}} = X_{t_j} + \underbrace{\int_{t_j}^{t_{j+1}} b(X_s, \beta)ds}_{\approx O_p(h)} + \underbrace{\int_{t_j}^{t_{j+1}} a(X_{s-}, \alpha)dZ_s}_{\approx O_p(\sqrt{h})}$$

- Suggests:
  - First estimate α with ignoring $b(x, \beta)$;
  - Then estimate β with plugging in $\hat{\alpha}_n$,

  even for general standard Lévy process Z;
  - See [Kamatani and Uchida, 2015] and the ref's therein for the diffusion case.

Setup: stepwise asymptotics

Objective

- Estimate true parameter $\theta_0 = (\alpha_0, \gamma_0)$ of

$$dX_t = a(X_t, \alpha, \gamma)dt + c(X_{t-}, \gamma)dJ_t.$$

from discrete-time sample $(X_{t_j})_{j=1}^n$ for $t_j = jh_n$ with $h = h_n$ s.t.

  - $\exists \epsilon_0 \in (0, 1),\ nh^{1+\epsilon_0} \to \infty$
  - $nh^2 \to 0$

Assumed

- Smooth parametric coefficients known up to $\theta = (\alpha, \gamma) \in \mathbb{R}^p$.
- Standardized pure-jump Lévy process J s.t. $\forall q > 0,\ \mathbb{E}(|J_t|^q) < \infty$.
- Exponentially ergodicity and bounded moments.
- Identifiability

$$dX_t = a(X_t, \alpha, \gamma)dt + c(X_{t-}, \gamma)dZ_t$$

$$X_{t_j} \overset{\mathbb{P}_\theta}{\approx} X_{t_{j-1}} + ha_{j-1}(\alpha, \gamma) + c_{j-1}(\gamma)\Delta_j Z$$

Stepwise-estimation recipe [Uehara and Masuda, 2017], with $\mathbb{H}_n(\alpha, \gamma)$ of (5)

1. $\mathcal{L}(X_{t_j} | X_{t_{j-1}} = x) \overset{\mathbb{P}_\theta}{\approx} N\left(x,\ hc^2(x, \gamma)\right)$: $\quad \hat{\gamma}_n \in \operatorname{argmin}_\gamma \mathbb{H}_n^1(\gamma)$,

$$\mathbb{H}_n^1(\gamma) := \sum_{j=1}^{n} \log \phi\left(X_{t_j};\ X_{t_{j-1}},\ hc_{j-1}^2(\gamma)\right)$$

2. $\mathcal{L}(X_{t_j} | X_{t_{j-1}} = x) \overset{\mathbb{P}_\theta}{\approx} N\left(x + ha(x, \alpha, \gamma),\ hc^2(x, \hat{\gamma}_n)\right)$: $\hat{\alpha}_n \in \operatorname{argmin}_\alpha \mathbb{H}_n(\alpha, \hat{\gamma}_n)$,

$$\mathbb{H}_n(\alpha, \gamma) := \sum_{j=1}^{n} \log \phi\left(X_{t_j};\ X_{t_{j-1}} + ha_{j-1}(\alpha, \gamma),\ hc_{j-1}^2(\gamma)\right)$$

Result: [Masuda and Uehara, 2017] & [Uehara and Masuda, 2017]

Joint asymptotic normality of $(\hat{\gamma}_n, \hat{\alpha}_n)$ at speed $\sqrt{T_n}$ $\quad (T_n := nh)$

Some technical details

- Asymptotics of $\hat{\gamma}_n$ is the same as in the joint estimation, and as for $\hat{\alpha}_n$:

Explicit stochastic expansions

$$\sqrt{T_n}(\hat{\alpha}_n - \alpha_0) = \hat{\mathcal{I}}_{\alpha,n}^{-1} \left(\hat{A}_n \sqrt{T_n}(\hat{\gamma}_n - \gamma_0) + v_n \right) + o_p(1)$$

$$v_n := \frac{1}{\sqrt{T_n}} \sum_{j=1}^{n} \frac{\partial_\alpha a_{j-1}(\alpha_0, \gamma_0)}{c_{j-1}^2(\gamma_0)} (\Delta_j X - h a_{j-1}(\alpha_0, \gamma_0)),$$

$$\hat{\mathcal{I}}_{\alpha,n} := \frac{1}{n} \sum_{j=1}^{n} \frac{(\partial_\alpha \hat{a}_{j-1})^{\otimes 2}}{\hat{c}_{j-1}^2},$$

$$\hat{A}_n := \frac{1}{n} \sum_{j=1}^{n} \frac{\partial_\alpha \hat{a}_{j-1} \otimes \partial_\gamma \hat{a}_{j-1}}{\hat{c}_{j-1}^2} \xrightarrow{p} \int \frac{\partial_\alpha a(x, \alpha_0, \gamma_0) \otimes \partial_\gamma a(x, \alpha_0, \gamma_0)}{c^2(x, \gamma_0)} \pi_0(dx),$$

Asymptotic normality

$$\sqrt{T_n}\hat{\Sigma}_n^{-1/2}\begin{pmatrix}\hat{\mathcal{I}}_{\alpha,n}^{-1} & \hat{\mathcal{I}}_{\alpha,n}^{-1}\hat{A}_n \\ O & \hat{\mathcal{I}}_{\gamma,n}^{-1}\end{pmatrix}\begin{pmatrix}\hat{\alpha}_n-\alpha_0 \\ \hat{\gamma}_n-\gamma_0\end{pmatrix}\xrightarrow{\mathcal{L}} N_p(0,I)$$

- Clarifies effect of simultaneous presence of γ in the coefficients.
- $\hat{\mathcal{I}}_{\gamma,n}^{-1} := \frac{1}{n}\sum_{j=1}^{n}\frac{(\partial_\gamma \hat{c}_{j-1})^{\otimes 2}}{\hat{c}_{j-1}^2}$, and $\hat{\Sigma}_n$ is also given explicitly.
- Readily provides us with an approximate $(1-s)$-confidence set:

$$\left\{(\alpha,\gamma):\ \left|\sqrt{T_n}\hat{\Sigma}_n^{-1/2}\begin{pmatrix}\hat{\mathcal{I}}_{\alpha,n}^{-1} & \hat{\mathcal{I}}_{\alpha,n}^{-1}\hat{A}_n \\ O & \hat{\mathcal{I}}_{\gamma,n}^{-1}\end{pmatrix}\begin{pmatrix}\hat{\alpha}_n-\alpha \\ \hat{\gamma}_n-\gamma\end{pmatrix}\right|^2 \le \chi^2(p;s)\right\}$$

Put simply

- For the univariate parametric Stochastic differential equation (SDE):

$$dX_t = a(X_{t-}, \alpha)dZ_t + b(X_t, \beta)dt, \qquad \theta = (\alpha, \beta)$$

or

$$dX_t = c(X_{t-}, \gamma)dJ_t + a(X_t, \alpha, \gamma)dt, \qquad \theta = (\alpha, \gamma).$$

for stepwise estimation, we can make use of the explicit GQLF

$$\mathbb{H}_n(\theta) = \sum_{j=1}^{n} \log \phi\left(X_{t_j};\, X_{t_{j-1}} + hb_{j-1}(\beta),\, ha_{j-1}^2(\alpha)\right).$$

  - Available data $(X_{t_j})_{j=0}^{n}$; $t_j = jh_n = jh$, $T_n := nh \to \infty$, $nh^2 \to 0$.
  - Driving Lévy process s.t. $\mathbb{E}(Z_t) = 0$, $\mathrm{var}(Z_t) = t$.

- **Stability (Ergodicity)** is essential here.

- The YUIMA function `qmleLevy` was composed by Dr. Yuma Uehara.

YUIMA demo: qmleLevy

$$dX_t = a(X_t, \alpha, \gamma)dt + c(X_{t-}, \gamma)dZ_t$$

- **Usage**

```
qmleLevy(yuima, start, lower, upper, joint = FALSE, third = FALSE)
```

- **Arguments**

yuima a yuima object
lower a named list for specifying lower bounds of parameters.
upper a named list for specifying upper bounds of parameters.
start initial values to be passed to the optimizer.
joint perform joint estimation or two stage estimation?
  - by default joint=FALSE. If there exists an overlapping parameter, joint=TRUE currently does not work.

third perform third estimation?
  - by default third=FALSE. If there exists an overlapping parameter, third=TRUE currently does not work.

- **Value**

first estimated values of first estimation (scale parameters)
second estimated values of second estimation (drift parameters)
third estimated values of third estimation (scale parameters)

Example: bilateral gamma

$$dX_t = -\theta_0 X_t dt + \frac{\theta_1}{\sqrt{1+X_t^2}} dZ_t,$$

- $Z_t \sim$ Bilateral gamma$(t, \sqrt{2}, t, \sqrt{2})$, $\quad (\theta_{0,0}, \theta_{1,0}) = (1, 2)$

▷ yss2019_hm_demo.html

Example: Normal inverse Gaussian

$$dX_t = -\theta_0 X_t dt + \frac{\theta_1}{\sqrt{1+X_t^2}} dZ_t,$$

- $Z_t \sim NIG\,(\delta, 0, \delta t, 0, 1)$ with $\delta = 10$, $\quad (\theta_{0,0}, \theta_{1,0}) = (1, 2)$.

▷ yss2019_hm_demo.html

2-dim. Example: variance gamma

$$d\begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} = \begin{pmatrix} 1-\theta_0 X_{1,t} - X_{2,t} \\ -\theta_1 X_{2,t} \end{pmatrix} dt + \begin{pmatrix} \frac{\theta_2}{1+X_{1,t}^2}+1 & 0 \\ 1 & 1 \end{pmatrix} dZ_t,$$

- $Z_t \sim$ Variance gamma $\left(\frac{1}{2}t, 1, \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$, $\quad (\theta_0, \theta_1, \theta_2) = (1, 2, 3)$.

▷ yss2019_hm_demo.html

2-dim. Example: Normal inverse Gaussian

$$d\begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} = \begin{pmatrix} 1-\theta_0 X_{1,t} \\ -\theta_1 X_{2,t} \end{pmatrix} dt + \begin{pmatrix} \exp\left(-\frac{\theta_2}{1+X_{1,t}^2}\right) & 0 \\ 1 & \exp\left(-\frac{\theta_3}{\sqrt{1+X_{2,t}^2}}\right) \end{pmatrix} dZ_t,$$

- $Z_t \sim NIG_2\left(\frac{1}{\sqrt{\pi}}t, 1, \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right), \quad (\theta_0, \theta_1, \theta_2, \theta_3) = (1, 2, 3, 4).$

▷ yss2019_hm_demo.html

References I

Applebaum, D. (2009).
Lévy processes and stochastic calculus, volume 116 of *Cambridge Studies in Advanced Mathematics*.
Cambridge University Press, Cambridge, second edition.

Bertoin, J. (1996).
Lévy processes, volume 121 of *Cambridge Tracts in Mathematics*.
Cambridge University Press, Cambridge.

Brouste, A., Fukasawa, M., Hino, H., Iacus, S. M., Kamatani, K., Koike, Y., Masuda, H., Nomura, R., Ogihara, T., Shimizu, Y., Uchida, M., and Yoshida, N. (2014).
The yuima project: A computational framework for simulation and inference of stochastic differential equations.
Journal of Statistical Software, 57(4):1–51.

Iacus, S. M. and Yoshida, N. (2018).
Simulation and inference for stochastic processes with YUIMA.
Use R! Springer, Cham.
A comprehensive R framework for SDEs and other stochastic processes.

Kamatani, K. and Uchida, M. (2015).
Hybrid multi-step estimators for stochastic differential equations based on sampled data.
Stat. Inference Stoch. Process., 18(2):177–204.

Kessler, M. (1997).
Estimation of an ergodic diffusion from discrete observations.
Scand. J. Statist., 24(2):211–229.

References II

Kulik, A. (2018).
Ergodic behavior of Markov processes, volume 67 of *De Gruyter Studies in Mathematics*.
De Gruyter, Berlin.
With applications to limit theorems.

Masuda, H. (2013).
Convergence of Gaussian quasi-likelihood random fields for ergodic Lévy driven SDE observed at high frequency.
Ann. Statist., 41(3):1593–1641.

Masuda, H. and Uehara, Y. (2017).
Two-step estimation of ergodic Lévy driven SDE.
Stat. Inference Stoch. Process., 20(1):105–137.

Platen, E. and Bruti-Liberati, N. (2010).
Numerical solution of stochastic differential equations with jumps in finance, volume 64 of *Stochastic Modelling and Applied Probability*.
Springer-Verlag, Berlin.

Protter, P. E. (2005).
Stochastic integration and differential equations, volume 21 of *Stochastic Modelling and Applied Probability*.
Springer-Verlag, Berlin.
Second edition. Version 2.1, Corrected third printing.

References III

Sato, K.-i. (1999).
Lévy processes and infinitely divisible distributions, volume 68 of *Cambridge Studies in Advanced Mathematics*.
Cambridge University Press, Cambridge.
Translated from the 1990 Japanese original, Revised by the author.

Uehara, Y. and Masuda, H. (2017).
Stepwise estimation of a Lévy driven stochastic differential equation.
Proc. Inst. Statist. Math. (Japanese), 65(1):21–38.