Limit theorems and statistical inference for ergodic solutions of Lévy driven SDE's

Alexei M.Kulik

Institute of mathematics, Kyiv, Ukraine, *kulik.alex.m@gmail.com*

Tokyo, 3 September 2013

I. A statistical model based on discrete observations of a Lévy driven SDE

This part of the talk is based on the joint research with D.Ivanenko.

Consider a solution X^θ to an SDE driven by a Lévy process Z:

$$dX_t^\theta = a_\theta(X_t^\theta)dt + dZ_t, \quad X_0 = x_0. \tag{1}$$

Denote by P_n^θ the distribution of the sample $(X_h, \dots, X_{nh})$, and consider the statistical experiments

$$\mathcal{E}^n = \Big(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathsf{P}_n^\theta, \theta \in \Theta\Big), \quad n \geq 1.$$

The state space for X, Z is $\mathbb{R}$, the parameter set Θ is an open interval in $\mathbb{R}$.
In our model:

- the noise is an *infinite intensity* Lévy process *without a diffusion part*;
- we consider the *fixed frequency case*: on the contrary to *high frequency models*, where $h_n \to 0$, we assume $h > 0$ to be fixed;
- we are mainly focused on the asymptotic properties of the *MLE* because we are aiming to get an *asymptotically efficient* estimator.

The likelihood function of the model

The likelihood function of our Markov model has the form

$$L_n(\theta; x_1, \dots, x_n) = \prod_{k=1}^{n} p_h^\theta(x_{k-1}, x_k), \quad (x_1, \dots, x_n) \in \mathbb{R}^n$$

(the initial value x_0 is assumed to be known), where $p_t^\theta(x, y)$ is the transition probability density of X^θ. Both the likelihood function and likelihood ratio

$$Z_n(\theta_0, \theta; x_1, \dots, x_n) = \frac{L_n(\theta; x_1, \dots, x_n)}{L_n(\theta_0; x_1, \dots, x_n)}$$

are implicit, because analytical expressions for $p_t^\theta(x, y)$ or their ratio are not available.

Specific feature of the model: likelihood function may be non-trivially degenerated

In general, $L_n(\theta;\cdot)$ may equal zero on a non-empty set $N_n^\theta \subset \mathbb{R}^n$. Moreover, this set *can depend non-trivially* on θ. To see that, consider an example of an Ornstein-Uhlenbeck process driven by a one-sided α-stable process with $\alpha < 1$:

$$dX_t^\theta = -\theta X_t^\theta\, dt + dZ_t, \quad Z_t = \int_0^t \int_0^\infty u\nu(ds, du).$$

Then by a support theorem for Lévy driven SDE's (Simon 2000), the (topological) support of P_n^θ is

$$S_n^\theta = \Big\{(x_1, \dots, x_n) : x_k \geq e^{-\theta h} x_{k-1}, k = 1, \dots, n\Big\},$$

which depends non-trivially on θ. Because

$$S_n^\theta = \mathrm{closure}(\mathbb{R}^n \setminus N_n^\theta),$$

this indicates that N_n^θ depends non-trivially on θ, as well.

Henceforth, our model *can not be considered as a model with a C^1 log-likelihood function.*

Main result: conditions on the noise

- **smoothness near the origin:** for some $u_0 > 0$, the restriction of μ on $[-u_0, u_0]$ has a positive density $\sigma \in C^2\left([-u_0, 0) \cup (0, u_0]\right)$ and there exists C_0 such that

$$|\sigma'(u)| \leq C_0 |u|^{-1} \sigma(u), \quad |\sigma''(u)| \leq C_0 u^{-2} \sigma(u), \quad |u| \in (0, u_0];$$

- **sufficiently high intensity of "small jumps":**

$$\left(\log \frac{1}{\varepsilon}\right)^{-1} \mu\Big(\{u : |u| \geq \varepsilon\}\Big) \to \infty, \quad \varepsilon \to 0;$$

- **moment bound for "large jumps":** for some $\varepsilon > 0$,

$$\int_{|u| \geq 1} u^{4+\varepsilon} \mu(du) < \infty.$$

An exapmle: tempered α-stable measure $\mu(du) = r(u) u^{-\alpha-1} du$.

Main result: conditions on the coefficients

- **regularity and bounds:** $a \in C^{3,2}(\mathbb{R} \times \Theta)$ have bounded derivatives

  $$\partial_x a, \quad \partial^2_{xx} a, \quad \partial^2_{x\theta} a, \quad \partial^3_{xxx} a, \quad \partial^3_{xx\theta} a, \quad \partial^3_{x\theta\theta} a, \quad \partial^4_{xxx\theta} a, \quad \partial^4_{xx\theta\theta} a, \quad \partial^5_{xxx\theta\theta} a$$

  and

  $$|a_\theta(x)| + |\partial_\theta a_\theta(x)| + |\partial^2_{\theta\theta} a_\theta(x)| \leq C(1 + |x|);$$

- **"drift condition":** for any compact set $K \subset \Theta$,

  $$\limsup_{|x| \to +\infty} \frac{a_\theta(x)}{x} < 0 \quad \text{uniformly w.r.t. } \theta \in K.$$

An example: perturbed OU process,

$$a_\theta(x) = -\theta x + \alpha_\theta(x), \quad \alpha \in C^{3,2}_b(\mathbb{R} \times \Theta), \quad \Theta = (\theta_1, \theta_2), \quad \theta_1 > 0.$$

Main result

Theorem

Every experiment $\mathcal{E}_n, n \geq 1$ is *regular* (see below), and there exists

$$\lim_{n\to\infty} \frac{I_n(\theta)}{n} = \sigma^2(\theta) = \mathsf{E}\Big(g_h^\theta(X_0^{\theta,st}, X_h^{\theta,st})\Big)^2, \quad g_h^\theta = \frac{\partial_\theta p_h^\theta}{p_h^\theta}.$$

In addition, if the model is *locally identifiable* in the sense that

$$\sigma^2(\theta) > 0, \quad \theta \in \Theta,$$

and is *globally identifiable*, i.e. for every $\theta_1 \neq \theta_2$ there exists $x = x(\theta_1, \theta_2)$:

$$P_h^{\theta_1}(x, \cdot) \neq P_h^{\theta_2}(x, \cdot),$$

then the MLE $\hat{\theta}_n$ is *consistent*, *asymptotically normal* with $\mathcal{N}(0, \sigma^2(\theta))$ limit distribution, and is *asymptotically efficient* w.r.t. any loss function $w \in W_p$, i.e.

$$w(x, y) = v(|x - y|)$$

with convex v of at most polynomial growths at ∞.

The method

Because of lack of C^1-smoothness of the log-likelihood function, it was almost inevitable for us to choose as the main tool the Ibragimov-Khas'minskii approach (Ibragimov-Khas'minskii 1981), which basically consists of three following stages.

Ground stage	Regularity property	$\Rightarrow$ Rao-Cramer inequality
1-st stage	LAN property	$\Rightarrow$ Lower bounds for efficiency w.r.t. cost functions from W_p
2-st stage	Uniform LAN property; Hölder continuity and growth bounds for associated Hellinger processes	$\Rightarrow$ Asymptotic normality and efficiency of MLE

Malliavin-calculus based integral representations for transition densities and their derivatives

It is well known that in the framework of the Malliavin calculus a representation

$$p_t^\theta(x,y) = \mathsf{E}_x^\theta \delta(\Xi_t) 1\!\!1_{X_t>y}, \quad \Xi_t = \frac{DX_t}{\|DX_t\|_H^2}$$

can be obtained via an integration-by-parts procedure from the formal relation

$$p_t^\theta(x,y) = -\partial_y \mathsf{E}_x^\theta 1\!\!1_{X_t>y}.$$

Nualart 1995.
Similar heuristics leads to integral representations for the derivatives of $p_t^\theta(x,y)$.

$$\frac{\partial_\theta p_t^\theta(x,y)}{p_t^\theta(x,y)} = \mathsf{E}_{x,y}^{t,\theta} \delta(\Xi_t^1), \quad \Xi_t^1 = \frac{(\partial_\theta X_t^1) DX_t}{\|DX_t\|_H^2}$$

Gobet 2001, 2002; Corcuera, Kohatsu-Higa 2011.
Yoshida 1992, 1996.

Integral representations (continued)

To get the integration-by-part framework on the Poisson probability space, we use the approach close to the one introduced in Bismut 1981, modified and simplified in order to give integral representations explicitly.
Let ν be the Poisson point measure involved into Itô-Lévy representation for the Lévy process Z:

$$Z_t = \int_0^t \int_{|u|\le 1} u\Big(\nu(ds,du) - ds\mu(du)\Big) + \int_0^t \int_{|u|>1} u\nu(ds,du).$$

Then

$$D\int_0^t \int_{\mathbb{R}} f(u)\nu(ds,du) = \int_0^t \int_{\mathbb{R}} f'(u)\varrho(u)\nu(ds,du),$$

where $\varrho \in C_0^\infty$ is a function which equals $\varrho(u) = u^2$ in some neighbourhood of the point $u = 0$.

$$(\tau, u) \rightsquigarrow (\tau, Q_\varepsilon(u)), \quad \partial_\varepsilon Q_\varepsilon(u)|_{\varepsilon=0} = \varrho(u).$$

Integral representations (continued)

Theorem

There exists continuous and bounded $p_h^\theta(x,y), \partial_\theta p_h^\theta(x,y), \partial^2_{\theta\theta} p_h^\theta(x,y)$, and

$$p_h^\theta(x,y) = \mathsf{E}_x^\theta \delta(\Xi_h) 1\!\!1_{X_h>y}, \quad \frac{\partial_\theta p_h^\theta(x,y)}{p_h^\theta(x,y)} =: g_h^\theta(x,y) = \mathsf{E}_{x,y}^{h,\theta} \delta(\Xi_h^1),$$

$$\frac{\partial^2_{\theta\theta} p_h^\theta(x,y)}{p_h^\theta(x,y)} =: f_h^\theta(x,y) = \mathsf{E}_{x,y}^{h,\theta} \delta(\Xi_h^2)$$

with explicitly given Ξ_h, Ξ_h^1, Ξ_h^2 such that

$$\mathsf{E}_x \Big(|\delta(\Xi_h)|^p + |\delta(\Xi_h^1)|^p + |\delta(\Xi_h^2)|^p \Big) \leq C(1+|x|^p), \quad p < 4+\varepsilon.$$

Consequently, for every $p < 4+\varepsilon$

$$p_t^\theta(x,y) \leq C(1+|x-y|)^{-p}, \quad \mathsf{E}_x^\theta \Big(\Big| g_t^\theta(x, X_t) \Big|^p + \Big| f_t^\theta(x, X_t) \Big|^p \Big) \leq C(1+|x|)^p.$$

Regularity of the model

Recall that an experiment is said to be *regular*, if

- for λ^d-a.a. $(x_1, \dots, x_n) \in \mathbb{R}^n$ the mapping $\theta \mapsto L_n(\theta; x_1, \dots, x_n)$ is continuous;
- the mapping $\theta \mapsto \sqrt{L_n(\theta; \cdot)} \in L_2(\mathbb{R}^n)$ is continuously differentiable.

For a regular experiment, the Fisher information is given by

$$I_n(\theta) = 4 \int_{\mathbb{R}^n} \left(\partial_\theta \sqrt{L_n(\theta; \mathbf{x})} \right)^2 d\mathbf{x} = \mathsf{E} G_n^2(\theta; X_h^\theta, \dots, X_{nh}^\theta), \quad G_n = 2 \frac{\partial_\theta \sqrt{L_n}}{\sqrt{L_n}}.$$

Using the above bounds and approximating the function $x \mapsto \sqrt{x}$ by C^1-functions properly, we get that the model is regular and

$$\partial_\theta \sqrt{L_n(\theta; \cdot)} = \frac{1}{2} G_n(\theta; \cdot) \partial_\theta \sqrt{L_n(\theta; \cdot)}, \quad G_n(\theta; x_1, \dots, x_n) = \sum_{k=1}^n g_h^\theta(x_{k-1}, x_k).$$

Since $g_h^\theta(X_{(k-1)h}^\theta, X_{kh}^\theta), k = 1, \dots, n$ is a martingale-difference sequence w.r.t. P_n^θ, the Fisher information of the model equals

$$I_n(\theta) = \sum_{k=1}^n \mathsf{E} \Big(g_h^\theta(X_{(k-1)h}^\theta, X_{kh}^\theta) \Big)^2.$$

LAN property of the model

Theorem

$$Z_{n,\theta}(u) := \frac{d\mathsf{P}_n^{\theta+\varphi(n)u}}{d\mathsf{P}_n^{\theta}}(X^n) = \exp\left\{\Delta_n(\theta)u - \frac{1}{2}u^2 + \Psi_n(u,\theta)\right\},$$

with $\varphi(n) = I_n^{-1/2}(\theta)$,

$$\Delta_n(\theta) \stackrel{\mathsf{P}^\theta}{\Rightarrow} \mathcal{N}(0,1), \quad \Psi_n(u,\theta) \xrightarrow{\mathsf{P}^\theta} 0, \quad n \to \infty.$$

The proof is an extension to the Markov setting of the proof of that property for a regular experiment based on i.i.d. observations, given in Ibragimov, Khas'minskii 1981 , Chapter II; (Le Cam 1970).

$$\log Z_{n,\theta}(u) \approx 2\sum_{j=1}^n \eta_{jn}^\theta(u) - \sum_{j=1}^n \Big(\eta_{jn}^\theta(u)\Big)^2,$$

$$\eta_{jn}^\theta(u) = \left(\left(\frac{p_h^{\theta+\varphi(n)u}(X_{h(j-1)}, X_{hj})}{p_h^\theta(X_{h(j-1)}, X_{hj})}\right)^{1/2} - 1\right) \mathbb{I}_{p_h^\theta(X_{h(j-1)}, X_{hj}) \neq 0}.$$

A key point in the whole proof is that "elementary increments" $\eta_{jn}(u)$ can be "linearized w.r.t. u":

$$\eta^\theta_{jn}(u) \approx \frac{1}{2}\varphi(n) u g^\theta_h(X_{h(j-1)}, X_{hj}), \quad g^\theta_h = \frac{\partial_\theta p^\theta_h}{p^\theta_h}.$$

Note that the *drift condition* above and smoothness of transition probabilities of X yield that the process X^θ is *exponentially ergodic*:

$$\|P^\theta_T(x,dy) - \pi^\theta(dy)\|_{TV} \leq Ce^{-\beta t} V(x), \quad V(x) = (1+|x|^2),$$

e.g. Masuda 2007. Then using a typical "perturbation of stationary limit theorems" trick, e.g. Bhattacharya 1982, one can show that

$$\frac{1}{n}\sum_{k=1}^n \Big(g^\theta_h(X^\theta_{(k-1)h}, X^\theta_{kh})\Big)^2 \stackrel{L_1(\mathsf{P}^\theta)}{\to} \sigma^2(\theta);$$

$$\frac{1}{\sqrt{n}}\sum_{k=1}^n g^\theta_h(X^\theta_{(k-1)h}, X^\theta_{kh}) \stackrel{\mathsf{P}^\theta}{\Rightarrow} \mathcal{N}(0, \sigma^2(\theta)).$$

LAN property of the model: proof (continued)

We prove the "linearization w.r.t. u" relation using regularity of the model, the LLN and CLT given above, and the following integral version of uniform continuity type condition for $q_h(\theta, x, y) = \partial_\theta \sqrt{p_h^\theta(x, y)}$: for every $N > 0$

$$\sup_{|v|<N} \varphi^2(n) \mathsf{E} \sum_{j=1}^{n} \int_{\mathbb{R}} \left(q_h \left(\theta + \varphi(n) v, X^\theta_{h(j-1)}, y \right) - q_h(\theta, X^\theta_{h(j-1)}, y) \right)^2 dy \to 0.$$

To prove this condition, we use the L_2-bound for

$$\partial_\theta q_h = \partial^2_{\theta\theta} \sqrt{p_h^\theta} = \left(\frac{\partial^2_{\theta\theta} p_h^\theta}{2 p_h^\theta} - \left(\frac{\partial_\theta p_h^\theta}{4 p_h^\theta} \right)^2 \right) \sqrt{p_h^\theta},$$

which follow from the Malliavin-type integral representations, and based on them L_p-bounds ($p = 4$) for

$$g_h^\theta = \frac{\partial_\theta p_h^\theta}{p_h^\theta}, \quad f_h^\theta = \frac{\partial^2_{\theta\theta} p_h^\theta}{p_h^\theta}.$$

Asymptotic properties of the MLE

General theorem by Ibragimov and Khas'minskii (Ibragimov, Khas'minskii 1981, Chapter III.1) makes it possible to prove asymptotic normality of the MLE and its asymptotic efficiency w.r.t. a wide class of loss functions under the following principal assumptions.

- Uniform LAN condition: $Z_{n,\theta_n}(u_n)$ instead of $Z_{n,\theta}(u)$, with $\theta_n \to \theta, u_n \to u$.
- Integral Hölder continuity and growth conditions on the Hellinger process $H^{n,\theta}_{1/2}(u) = (Z_{n,\theta}(u))^{1/2}, \quad u \in \mathbb{R}$.

1-st *assumption* can be verified, e.g., following the next scheme:

Uniform bounds for the α-mixing coefficitent for $X^\theta, \theta \in \Theta$

$\Rightarrow$ LLN and CLT for a sequence of strictly stationary processes

$\Rightarrow$ uniform LLN and CLT required for the uniform LAN

2-nd *assumption* follows from the identifiability conditions.

II. "Martingale problem approach" for proving diffusion approximation theorems: an outline

Let $X_k, k \geq 0$ be a Markov process with the state space $\mathbb{X}$, which is *ergodic*, i.e. has unique invariant distribution π, and the rate of convergence of the transition probabilities $P_k(x, \cdot)$ to π has the following bounds.
Consider a *distance-like* function $d : \mathbb{X} \times \mathbb{X} \to \mathbb{R}^+$, i.e. d is symmetric, lower semicontinuous, and $d(x, y) = 0 \Leftrightarrow x = y$. Denote by $C(\mu, \nu)$ the class of measures on $\mathbb{X} \times \mathbb{X}$ which μ, ν as their projections, and define the *coupling distance* on the set $\mathcal{P}(\mathbb{X})$ of all probability measures on $\mathbb{X}$ by

$$d(\mu, \nu) = \inf_{\chi \in C(\mu,\nu)} \int d(x, y)\chi(dx, dy).$$

- if $d(x, y) = \mathbb{I}_{x \neq y}$ (the discrete metric), then $d(\mu, \nu) = (1/2)\|\mu - \nu\|_{TV}$;
- if $d(x, y) = \rho^p(x, y)$, when $d^{1/p}(\mu, \nu)$ is the (Wasserstein-) Kantorovich - Rubinshtein metric of the power p, associated with the metric ρ.

In what follows, we assume, for some properly chosen r, V,

$$d(P_k(x, \cdot), \pi) \leq r(k)V(x), \quad k \geq 0, x \in \mathbb{X}.$$

Constructing the potential

We will explain the method on a particular case of the CLT for a sequence $\xi_{n,n}$,

$$\xi_{k,n} = \frac{1}{\sqrt{n}} \sum_{j=1}^{k} A(X_j), \quad k = 1, \dots, n;$$

in more generality, *diffusion approximation* theorems deal with

$$\xi_{k,n} = \xi_{0,n} + \frac{1}{n} \sum_{j=1}^{k} a(X_j, \xi_{j-1,n}) + \frac{1}{\sqrt{n}} \sum_{j=1}^{k} A(X_j, \xi_{j-1,n}), \quad k = 1, \dots, n.$$

Define the (extended) potential of A by

$$\mathcal{R}A(x) = \sum_{k=1}^{\infty} \mathsf{E}_x^X A(X_k).$$

- A is centered, i.e. $\int, d\pi = 0$;
- A is d-Hölder with the index γ: $|A(x) - A(y)| \leq d^\gamma(x, y)$.

Then

$$|\mathsf{E}_x^X A(X_k)| \leq r^\gamma(k) V^\gamma(x),$$

and $\mathcal{R}A$ is well defined provided that $\sum_k r^\gamma(k) < \infty$.

The corrector term method

Consider the processes $Y_n(t), t \in [0,1]$ such that

$$Y_n(t) = \xi_{k-1,n}, \quad t \in [(k-1)/n, k/n), \quad Y_n(1) = \xi_{n,n}.$$

For a given test function $f \in C^2(\mathbb{R})$ with bounded derivatives, consider a *corrector term process*

$$U_n(t) = (\mathcal{R}A)(X_{[nt]}) f'(Y_n(t)), \quad t \in [0,1],$$

then we have the following representation for the *corrected* value of the test function applied to Y_n:

$$f(Y_n(t)) + \frac{1}{\sqrt{n}} U_n(t) = \frac{1}{2n} \sum_{j \le [nt]} \Big[A^2(X_j) + 2A(X_{j-1}) \mathcal{R}A(X_{j-1}) \Big] f''(Y_n((j-1)/r$$

$$+ f(y_0) + \text{(martingale)} + \text{(remainder term)}.$$

The corrector term method (continued)

Given this representation, we are able to prove the "large ball containment" property:

$$\sup_n \sup_t P(|Y_n| > R) \to 0, \quad R \to \infty,$$

and the bound for the increments for processes Y_n:

$$\limsup_{n\to\infty} \sup_{|t-s|\leq\delta} E\Big(|Y_n(s) - Y_n(t)| \wedge 1\Big) \to 0, \quad \delta \to 0.$$

Then $\{Y_n\}$ is compact in the sense of weak convergence of finite-dimensional distributions. Using the above representation once more, we prove that any limiting point Y is a solution to the martingale problem

$$Lf = \sigma^2/2f'', \quad f \in C_0^\infty(\mathbb{R}), \quad \sigma^2 := \int \Big[A^2 + 2A\mathcal{R}A\Big]\, d\pi.$$

Because this martingale problem is well posed,

$$Y_n \Rightarrow \sigma W, \quad \text{in particular} \quad \xi_{n,n} \Rightarrow \mathcal{N}(0, \sigma^2).$$

“Martingale problem approach”: summary

We have proved the above (functional) CLT under the following set of assumptions:

- ergodicity bound for X with the distance function d, rate function r, and state-dependent weight V;
- A is centered, A, A^2, and $A\mathcal{R}A$ are π-integrable and d-Hölder with the index γ;
- $\sum_k r^\gamma(k) < \infty$;
- $(1/n)\max_{k\leq n} \mathsf{E}_x^X V^\gamma(X_k) \leq C$.

The proof can be modified easily if X, A depend on θ, and under the uniform version of these conditions the *uniform CLT* is available.
The proof well applies to Markov systems “with extinct memory”, for which e.g. α-mixing coefficients does not vanish when $t \to \infty$; e.g. fBm, solutions to SDDE's, SPDE's.

"Martingale problem approach": milestones

Feller 1956: "Distributions as operators" (Volume II, Chapter VIII Section 3).

Papanicolau, Stroock, Varadhan 1977: "Martingale approach" under *a priori* assumptions on existence, smoothness, and growth bounds for potentials RA.

Koroliuk, Limnious 2005: These *a priori* assumptions can be *verified* in the terms of the semigroup theory for X, if the process X is *uniformly ergodic.*

Pardoux, Veretennikov 2001, 2003, 2005: For *a diffusion* process X, potential RA is replaced by the weak solution to the *Poisson equation* $L^X u = -A$. The Itô formula for the corrector term can be applied because of analytical results about solutions to elliptic 2-nd order PDE's.

Kulik, Veretennikov 2011, 2012, 2013: The Itô formula in the whole approach is systematically replaced by the (extended) Dynkin's formula. This makes the approach to be completely insensitive w.r.t. the structure of the process X, and to involve into the domain of applications weakly ergodic Markov processes.

III. Ergodicity for Lévy driven SDE's: an outline

Typically, a *Harris-type theorem* gives for a Markov process X an ergodic bound of the form

$$\|P_t(x,\cdot)-\pi\|_{TV}\leq r(t)V(x),$$

provided two principal assumptions are verified:

- recurrence, e.g. $L^X V\leq -\alpha V+C$ and every set $\{V\leq c\}$ is compact;
- irreducibility.

The choice of the form of the irreducibility condition is non-trivial. If we adopt the strategy from Meyn, Tweedie '93, then we need to verify (some version of) the minorization condition:

$$P_t(x,dy)\geq c\kappa(dy),\quad x\in K.$$

This can be proved by means either of Bismut's approach/Malliavin calculus for SDE's with jumps, or of Picard's method/Ishikawa-Kunita's calculus on Wiener-Poisson space (Bichteler, Gravereaux, Jacod '87, Picard '96, Ishikawa, Kunita '05).
This is exactly approach from Masuda '07.

Ergodicity for Lévy driven SDE's: an outline (continued)

Applying the above strategy, when a diffusion noise is absent, requires the jump noise to have "sufficiently high" intensity of the small jumps:

$$\varphi(\rho) := \int_{|u|\leq\rho} u^2\Pi(du) \asymp \rho^{2-\alpha}, \quad \text{or at least} \quad \varphi(\rho) \gg \log\left(\frac{1}{\rho}\right), \quad \rho\to 0.$$

This limitation can be removed completely by considering the irreducibility condition in another form, called the *Dobrushin condition*:

$$\|P_t(x_1, dy) - P_t(x_2, dy)\|_{TV} \leq 2(1-c), \quad x_1, x_2 \in K.$$

The latter condition can be verified either by means of the *stratification method* by Davydov; Kulik '09, or by means of *stochastic control for Lévy driven SDE's*; Bodnarchuk, Kulik '12.

Ergodicity for Lévy driven SDE's: an outline (continued)

The *explicit* bounds for the ergodic rates in above theorems involving TV distance may be very poor, e.g. typically one gets $r(t) = Ce^{-\beta t}$ with large C and small $\beta > 0$. Such rates can be improved drastically when the TV-distance is replaced by some weaker distance, e.g. the (Wasserstein)-Kantorovich-Rubinshtein one. In that case applying the version of the above CLT with the weak ergodic bounds may lead to a significant *improvement of the accuracy* in statistical inference, simulation, etc.

An instructive example here is the OU process

$$dX(t) = -aX(t)\,dt + dZ(t),$$

which admits ergodic rates with $r(t) = e^{-\beta t}$ and explicitly given $\beta = \beta(a)$. This is a simplest example of a "dissipative" system, where respective weak ergodic bound comes from the Itô formula combined with the Gronwall lemma.

Dissipativity is a sort of a structurall assumption, which although can be reduced greatly, by using the machinery of *general Harris type theorems*, developed in Hairer, Mattingly, Scheutzow '11.

References

- Ivanenko, D.O., Kulik, A.M. (2013) *Malliavin calculus approach to statistical inference for Lévy driven SDE's.* arXiv:1301.5141
- Ivanenko, D.O., Kulik, A.M. (2013) *Asymptotic properties of MLE for a discretely observed Lévy driven SDE.* In progress.
- Veretennikov, A.Yu., Kulik, A.M. (2011) *On extended Poisson equation for weakly ergodic Markov processes.* Probab. Theory and Mat. Stat. **85**, 22 – 38.
- Veretennikov, A.Yu., Kulik, A.M. (2012, 2013) *Diffusion approximation for systems with weakly ergodic Markov perturbations I, II.* Prob. Theory and Math. Stat. **87, 88**.
- Kulik, A. M. (2009) *Exponential ergodicity of the solutions to SDEs with a jump noise*. Stochastic Processes and Appl. 119, 602 – 632.
- Bodnarchuk S. V., Kulik A. M. (2012) *Stochastic control based on time-change transformations for stochastic processes with Lévy noise,* Probab. Theory and Mat. Stat. **86**, 11 - 27.