Invariant adaptive density estimation for ergodic SDE with jumps over anisotropic classes

Chiara Amorino
From a joint work with Arnaud Gloter

Université d’Evry Val d’Essonne - Paris Saclay

March 19, 2020
École Polytechnique, France.
Non-parametric adaptive estimation of the invariant density $p$ associated to the $d$-dimensional diffusion process $(X_t)_{t \geq 0}$, solution of

$$X_t = X_0 + \int_0^t b(X_s)ds + \int_0^t a(X_s)dW_s + \int_0^t \int_{\mathbb{R}^d \setminus \{0\}} \gamma(X_{s-}) z \tilde{\mu}(ds, dz),$$

t \in \mathbb{R}_+, W = (W_t)_{t \geq 0}$ d-dimensional Brownian motion, $\tilde{\mu} = \mu - \bar{\mu}$ compensated Poisson random measure on $[0, T] \times \mathbb{R}^d$ with a possible infinity jump activity, associated to the Lévy process $L = (L_t)_{t \geq 0}$.

We assume that a continuous record $X^T = \{X_t, 0 \leq t \leq T\}$ up to time $T$ is observed.

The estimation of the invariant density in various frameworks is a widely studied problem through the literature (see for example Comte and Merlevede (2005) [4], Comte, Prieur and Samson (2017) [5], Schmisser (2013) [16] and Bosq and Blancke (2007) [2]).


Recent works on the recursive approximation of the invariant measure for a continuous diffusion in Honoré, Menozzi (2016) [12] and for a Poisson compound process in Gloter, Honoré, Loukianova (2018) [9].
In (Dalalyan, Reiss (2007)) [6], a continuous record of a $d$ dimensional diffusion process $X$ is observed up to $T$.

$$dX_t = b(X_t)dt + dW_t \quad X_0 = \zeta, \quad t \in [0, T].$$

**Definition**

For any multi-index $\alpha \in \mathbb{N}^d$ and $x \in \mathbb{R}^d$ we set $|\alpha| = \alpha_1 + \ldots + \alpha_d$ and $x^\alpha = x_1^{\alpha_1} \cdot \ldots \cdot x_d^{\alpha_d}$.

$$\mathcal{H}(\beta, L) = \left\{ f \in C^{[\beta]}(\mathbb{R}^d; \mathbb{R}) : \left| D^\alpha f(x) - D^\alpha f(y) \right| \leq |x - y|^{\beta - [\beta]} \right\}$$

for $|\alpha| = [\beta]$ and where $D^\alpha f = \frac{\partial |\alpha| f}{\partial x_1^{\alpha_1} \ldots \partial x_d^{\alpha_d}}$.

They assume $b = -\nabla V$, where $V \in C^2(\mathbb{R}^d)$ is referred to as potential, and $V \in \mathcal{H}(\beta + 1, L) \rightarrow b_i \in \mathcal{H}(\beta, L), \forall i = 1, \ldots, d$. 
To estimate the invariant density $\mu_b \in \mathcal{H}(\beta + 1, L)$ natural kernel estimator:

$$\hat{\mu}_{h,T}(x) := \frac{1}{T} \int_0^T K_h(x - X_t) dt, \quad x \in \mathbb{R}^d,$$

where $K_h(x) = h^{-d}K(h^{-1}x)$ and $K : \mathbb{R}^d \rightarrow \mathbb{R}$ is a smooth kernel function of compact support, satisfying $\int_{\mathbb{R}^d} K(dx) = 1$ and $\int_{\mathbb{R}^d} x^\alpha K(dx) = 0$ whenever $1 \leq |\alpha| \leq \lceil \beta \rceil$.

The usual bias-variance decomposition and approximation inequality yield (Efroymovich(1999) [7], § 8.9)

$$\mathbb{E}_b[|\hat{\mu}_{h,T}(x) - \mu_b(x)|^2] \leq h^{2(\beta+1)} + \frac{1}{T^2} \text{Var}_b(\int_0^T K_h((x - X_t)dt)).$$
Let $P_{b,t}$ be the transition semi-group of the process $X$ and $p_{b,t}$ the transition density:

$$P_{b,t}f(x) = \mathbb{E}_b[f(X_t)|X_0 = x] = \int_{\mathbb{R}^d} f(y)p_{b,t}(x,y)dy.$$ 

To study the variance, Assumptions required:

1. There is $\rho > 0$ such that, for any $f : \mathbb{R}^d \to \mathbb{R}$:
   $$\int_{\mathbb{R}^d} |f(x)|^2 \mu_b(dx) < \infty$$
   and for any $t > 0$:
   $$\|P_{b,t}f - \mu_b(f)\|_{\mu_b} \leq e^{-t\rho} \|f\|_{\mu_b}.$$ 

2. There is $C_0 > 0$ such that for any $t > 0$ and for any pair of points $x, y \in \mathbb{R}^d$ satisfying $|x - y|^2 < t$, it is
   $$p_{b,t}(x,y) \leq C_0(t^{-\frac{d}{2}} + t^{\frac{3d}{2}}).$$
They show that

$$\frac{1}{T^2} \text{Var}_b(\int_0^T K_h((x - X_t)dt)) \leq \frac{1}{T} \psi_d^2(h^d),$$

where

$$\psi_d(x) = \begin{cases} \max(1, \log^2(\frac{1}{x})), & d = 2 \\ x^{\frac{1}{d}-\frac{1}{2}}, & d \geq 3 \end{cases}.$$ 

The optimal choice $h = h(T) \sim T^{-\frac{1}{2\beta+d}}$ provides the convergence rate

$$\mathbb{E}_b[|\hat{\mu}_h, T(x) - \mu_b(x)|^2]^{\frac{1}{2}} \leq \begin{cases} T^{-\frac{1}{2}} (\log T)^2, & d = 2 \\ T^{-\frac{\beta+1}{2\beta+d}}, & d \geq 3. \end{cases}$$
In Strauch (2018) [17] a continuous record of a \(d\) dimensional diffusion process \(X\) is still observed up to \(T\). As in Dalalyan, Reiss (2007) [6], \(X\) strong solution of

\[
dX_t = b(X_t)dt + dW_t \quad X_0 = \zeta, \quad t \in [0, T].
\]

Again, \(b = -\nabla V\), where \(V \in C^2(\mathbb{R}^d)\) is referred to as potential. Now \(V\) belongs to the anisotropic Hölder class \(H_d(\beta + 1, \mathcal{L})\):

**Definition**

Let \(\beta = (\beta_1, ..., \beta_d), \ \beta_i > 0, \ \mathcal{L} = (\mathcal{L}_1, ..., \mathcal{L}_d), \ \mathcal{L}_i > 0\). A function \(g : \mathbb{R}^d \to \mathbb{R}\) is said to belong to the anisotropic Hölder class \(H_d(\beta, \mathcal{L})\) of functions if, for all \(i \in \{1, ..., d\}\),

\[
\left\|D_i^{\lfloor \beta_i \rfloor} g(. + te_i) - D_i^{\lfloor \beta_i \rfloor} g(.) \right\|_{\infty} \leq \mathcal{L}_i |t|^{\beta_i - \lfloor \beta_i \rfloor} \quad \forall t \in \mathbb{R}.
\]
To estimate the invariant density $\mu_b \in \mathcal{H}_d(\beta + 1, L)$:

$$
\hat{\mu}_{h, T}(x) = \frac{1}{T \prod_{l=1}^{d} h_l} \int_0^T \prod_{m=1}^{d} K \left( \frac{x_m - X_u^m}{h_m} \right) du,
$$

$h = (h_1, \ldots, h_d)$ is the multi-dimensional bandwidth. Nash and Poincaré inequalities lead Strauch to bounds analogous to the Assumptions 1 and 2 of Dalalyan, Reiss (2007).

$$
\Downarrow
$$

$$
\mathbb{E}_b[|\hat{\mu}_{h, T}(x) - \mu_b(x)|^2]^{\frac{1}{2}} \lesssim \begin{cases} 
T^{-\frac{1}{2}}(\log T)^2, & d = 2 \\
T^{-\frac{\beta+1}{2\beta+1+d-2}}, & d \geq 3,
\end{cases}
$$

where $\beta + 1$ is the mean smoothness over the $d$ different dimensions:

$$
\frac{1}{\beta + 1} := \frac{1}{d} \sum_{r=1}^{d} \frac{1}{\beta_r + 1}.
$$
We consider the question of non-parametric adaptive estimation of the invariant density $\mu$ associated to the $d$-dimensional diffusion process $(X_t)_{t \geq 0}$, solution of

$$X_t = X_0 + \int_0^t b(X_s) ds + \int_0^t a(X_s) dW_s + \int_0^t \int_{\mathbb{R}^d \setminus \{0\}} \gamma(X_s^-) z \tilde{\mu}(ds, dz),$$

where $b : \mathbb{R}^d \to \mathbb{R}^d$, $a : \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^d$ and $\gamma : \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^d$. $W = (W_t, t \geq 0)$ is a $d$-dimensional Brownian motion and $\tilde{\mu} = \mu - \bar{\mu}$ compensated Poisson random measure on $[0, T] \times \mathbb{R}^d$ with a possible infinity jump activity, associated to the Lévy process $L = (L_t)_{t \geq 0}$.

We assume that a continuous record $X^T = \{X_t, 0 \leq t \leq T\}$ up to time $T$ is observed.
Our goal is to provide a non-parametric estimator of the invariant density $\mu$ with a fully data-driven procedure of the bandwidth.

Extension previous literature:

- Generalization process: presence of jumps.
- We no longer assume $b = -\nabla V$.
- $a$ is no longer $I$.
- Assumptions on the model instead of on the transition semi-group and density.
The functions $b(x)$, $\gamma(x)$ and $a(x)$ are globally Lipschitz and bounded.

Drift condition: there exist $c > 0$ and $\tilde{\rho} > 0$ such that $\langle x, b(x) \rangle \leq -c|x|$, $\forall x : |x| \geq \tilde{\rho}$.

Jumps:

1. Lévy measure $F$ absolutely continuous wrt Lebesgue.
2. To bound the transition density:
   $$F(z) \leq \frac{c}{|z|^{d+\alpha}}, \text{ with } \alpha \in (0, 2).$$
   If $\alpha = 1$, for any $0 < r < R < \infty$ $\int_{r < |z| < R} zF(z)dz = 0$.
3. Irreducibility: $\text{supp}(F) = \mathbb{R}^d$, $\forall x \in \mathbb{R}^d$, $\text{Det}(\gamma(x)) \neq 0$.
4. There exists $\epsilon > 0$ and $c > 0$ s.t $\int_{\mathbb{R}^d} |z|^2 e^{\epsilon|z|} F(z)dz \leq c$. 
To estimate the invariant density $\mu \in \mathcal{H}_d(\beta, \mathcal{L})$ we introduce some kernel function $K : \mathbb{R} \to \mathbb{R}$ satisfying

$$\int_{\mathbb{R}} K(x)dx = 1, \quad \|K\|_\infty < \infty, \quad \text{supp}(K) \subset [-1, 1], \quad \int_{\mathbb{R}} K(x)x^l dx = 0,$$

for all $l \in \{0, \ldots, M\}$ with $M \geq \max_i \beta_i$.

$$\hat{\mu}_{h,T}(x) = \frac{1}{T \prod_{l=1}^d h_l} \int_0^T \prod_{m=1}^d K\left(\frac{x_m - X_m^t}{h_m}\right) du.$$

Bias - variance decomposition:

$$\mathbb{E}[|\hat{\mu}_{h,T}(x) - \mu(x)|^2] < \sum_{l=1}^d h_l^{2\beta_l} +$$

$$+ T^{-2} \text{Var}\left(\frac{1}{\prod_{l=1}^d h_l} \int_0^T \prod_{m=1}^d K\left(\frac{x_m - X_t^m}{h_m}\right) dt\right).$$
Assumptions required in Dalalyan, Reiss (2007) [6] (analogous to those in Strauch (2018) [17]) to study the variance:

1. There is $\rho > 0$ such that, for any $f : \mathbb{R}^d \to \mathbb{R}$:
   \[
   \int_{\mathbb{R}^d} |f(x)|^2 \mu_b(dx) < \infty \quad \text{and for any } t > 0:
   \]
   \[
   \|P_{b,t}f - \mu_b(f)\|_{\mu_b} \leq e^{-t \rho} \|f\|_{\mu_b}.
   \]

2. There is $C_0 > 0$ such that for any $t > 0$ and for any pair of points $x, y \in \mathbb{R}^d$ satisfying $|x - y|^2 < t$, it is
   \[
   p_{b,t}(x, y) \leq C_0(t^{-\frac{d}{2}} + t^{\frac{3d}{2}}).
   \]
Result on transition semi-group

**Definition**

We say that \( X \) is exponentially ergodic if it admits a unique invariant distribution \( \pi \) and additionally if there exist positive constants \( c \) and \( \rho \) for which, for each \( f \) centered under \( \mu \),

\[
\| P_t f \|_{L^1(\mu)} \leq ce^{-\rho t} \| f \|_\infty.
\]

**Result 1**: Suppose that Assumptions hold. Then, the process \( X \) is exponentially ergodic.

**Idea of the proof**: Requiring \( <x, b(x)> \leq -c|x| \) and

\[
\int_{\mathbb{R}^d} |z|^2 e^{\epsilon|z|} F(z)dz \leq c
\]

we are able to show the existence of a Lyapunov function. We use then Proposition 3.8 in Masuda (2007) [14].
Result 2: Suppose that Assumptions \( \diamond \) hold. Then, for \( T \geq 0 \), there exists \( p_t(x, y) \) for which for any \( t \in [0, T] \) there are a \( c_0 > 0 \) and a \( \lambda_0 > 0 \) such that, for any pair of points \( x, y \in \mathbb{R}^d \), we have

\[
p_t(x, y) \leq c_0 \left( t^{-\frac{d}{2}} e^{-\lambda_0 \frac{|y-x|^2}{t}} + \frac{t}{(t^{\frac{1}{2}} + |y-x|)^{d+\alpha}} \right).
\]

Idea of the proof:
It relies on the first point of Theorem 1.1 in Chen (2017) [3]. We show that assumptions required on \( a, b, \gamma \) and on the jumps hold true.
Suppose that Assumptions hold and let $f : \mathbb{R}^d \to \mathbb{R}$ be a bounded, measurable function with support $S$ satisfying $|S| < 1$. Then, there exists a constant $C$ independent of $f$ such that

- for $d = 1$
  \[
  \text{Var}(\int_0^T f(X_t)dt) \leq CT \|f\|_\infty^2 |S|^2 \left(1 + \log\left(\frac{1}{|S|}\right)\right)^2 - \frac{(1+\alpha)}{2} + \log\left(\frac{1}{|S|}\right)),
  \]

- for $d = 2$
  \[
  \text{Var}(\int_0^T f(X_t)dt) \leq CT \|f\|_\infty^2 |S|^2 \left(1 + \log\left(\frac{1}{|S|}\right)\right)
  \]

- for $d \geq 3$
  \[
  \text{Var}(\int_0^T f(X_t)dt) \leq CT \|f\|_\infty^2 |S|^{1+\frac{2}{d}}
  \]
Sketch of the proof, $d \geq 3$

We define $f_c := f - \mu(f)$.

$$\text{Var} (\int_0^T f(X_s)ds) = 2 \int_0^T (T - u) \mathbb{E}[f_c(X_0)f_c(X_u)]du =$$

$$= 2(\int_0^\delta + \int_\delta^D + \int_D^T) (T - u) \mathbb{E}[f_c(X_0)f_c(X_u)]du,$$

with $0 < \delta < D \leq T$, the specific choice of $\delta$ and $D$ will be given later. Useful control we will use on $[0, D]$: $\forall a, b \in \mathbb{R},$

$$\int_a^b (T - u) \mathbb{E}[f_c(X_0)f_c(X_u)]du =$$

$$= \int_a^b (T - u)(\mathbb{E}[f(X_0)f(X_u)]-(\mu(f))^2)du \leq cT \int_a^b |< P_uf, f >_\mu | du.$$
Sketch of the proof, \( d \geq 3 \)

On \([0, \delta]\) main argument: \( P_uf \) contraction \( \Rightarrow \)

\[ \int_0^\delta | < P_uf, f >_\mu | \, du \leq c \| f \|_\infty^2 |S|\delta. \]

On \([\delta, D]\) main argument: Result 2 (bound on the transition density) \( \Rightarrow \)

\[ \int_\delta^D | < P_uf, f >_\mu | \, du \leq \int_\delta^D \int \mathbb{R}^d |f(x)| \int \mathbb{R}^d |f(y)| p_u(x, y) dy \mu(x) dx \, du \leq \]

\[ c \| f \|_\infty^2 |S|^2 \int_\delta^D (u^{-\frac{d}{2}} + u^{1-\frac{(d+\alpha)}{2}} + 1) \, du \leq c \| f \|_\infty^2 |S|^2 (\delta^{1-\frac{d}{2}} + D). \]

On \([D, T]\) main argument: Result 1 (exponential ergodicity) \( \Rightarrow \)

\[ 2 \int_D^T (T - u) \mathbb{E}[f_c(X_0)f_c(X_u)] \, du \leq c \, T \| f \|_\infty^2 e^{-\rho D}. \]
In all,

\[ \text{Var}(\int_0^T f(X_s) ds) \leq c \, T \, \| f \|_\infty^2 (|S| \delta + |S|^2 \delta^{1 - \frac{d}{2}} + |S|^2 D + e^{-\rho D}). \]

To balance we choose \( \delta := |S|^{\frac{2}{d}} \) and \( D := \max(-\frac{2}{\rho} \log(|S|), 1). \)

\[ \Downarrow \]

\[ \text{Var}(\int_0^T f(X_s) ds) \leq c \, T \, \| f \|_\infty^2 (|S|^{\frac{2}{d} + 1} + |S|^{\frac{2}{d} + 1} + |S|^2 \log(|S|) + |S|^2) = \]

\[ = c \, T \, \| f \|_\infty^2 |S|^{\frac{2}{d} + 1}. \]
Proposition

Suppose that Assumptions hold. If \( \mu \in \mathcal{H}_d(\beta, \mathcal{L}) \), then

\[
\mathbb{E}[|\hat{\mu}_{h,T}(x) - \mu(x)|^2] \lesssim \sum_{l=1}^{d} h_l^{2\beta_l} + T^{-1} \left( \prod_{l=1}^{d} h_l \right)^{2/d - 1}.
\]

Defined \( \frac{1}{\beta} := \frac{1}{d} \sum_{l=1}^{d} \frac{1}{\beta_l} \), the rate optimal choice \( h_l(T) = \left( \frac{1}{T} \right)^{\frac{\beta_l}{\beta_l(2\beta + d - 2)}} \) yields

\[
\mathbb{E}[|\hat{\mu}_{h,T}(x) - \mu(x)|^2] \lesssim T^{-\frac{2\beta}{2\beta + d - 2}}.
\]

Convergence rate found by Strauch (2018) in [17] for the estimation of \( \mu \in \mathcal{H}_d(\beta, \mathcal{L}) \) was \( T^{-\frac{2\beta}{2\beta + d - 2}} \).

\[\downarrow\]

The convergence rate we obtain is the same it was in the case without jumps.
Proposition

Suppose that Assumptions hold. If $\mu \in H_d(\beta, \mathcal{L})$, then

$$\mathbb{E}[|\hat{\mu}_{h,T}(x) - \mu(x)|^2] \lesssim h^{2\beta} + \frac{1}{T} (1 + (\log(\frac{1}{h}))^{2-(\frac{1+\alpha}{2})} + \log(\frac{1}{h})).$$

The rate optimal choice for $h$ yields to the convergence rate

$$\mathbb{E}[|\hat{\mu}_{h,T}(x) - \mu(x)|^2] \lesssim \frac{(\log T)^{2-(\frac{1+\alpha}{2})} \vee 1}{T}.$$

Without jumps the convergence rate was $\frac{1}{T}$, now it depends on the degree of the jumps $\alpha$ and it is between $\frac{\log T}{T}$ and $\frac{(\log T)^{\frac{3}{2}}}{T}$.

$\Rightarrow$ Sharper bound using other approaches?
Proposition

Suppose that Assumptions hold. If $\mu \in \mathcal{H}_d(\beta, \mathcal{L})$, then

$$
\mathbb{E}[|\hat{\mu}_{h,T}(x) - \mu(x)|^2] \lesssim h_1^{2\beta_1} + h_2^{2\beta_2} + \frac{1}{T}(1 + \log(\frac{1}{h_1 h_2})).
$$

The rate optimal choice for $h$ yields to the convergence rate

$$
\mathbb{E}[|\hat{\mu}_{h,T}(x) - \mu(x)|^2] \lesssim \frac{\log T}{T}.
$$

Continuous convergence rate: $\frac{(\log T)^4}{T} \Rightarrow$ it seems faster in presence of jumps. Why?

Dalalyan and Reiss (2007) assume $p_{b,t}(x,y) \leq C_0(t^{-\frac{d}{2}} + t^{\frac{3d}{2}})$, which is a bound different from the one we get in Lemma 1.
Convergence rates for the pointwise estimation of the invariant density $\mu \in \mathcal{H}_d(\beta, \mathcal{L})$ of a continuous diffusion:

$$\mathbb{E}[|\hat{\mu}_{h,T}(x) - \mu(x)|^2] \sim \begin{cases} \frac{1}{T} & \text{for } d = 1, \\
\frac{(\log T)^4}{T} & \text{for } d = 2, \\
T^{-\frac{2\beta}{2\beta+d-2}} & \text{for } d \geq 3.\end{cases}$$

Convergence rates for the pointwise estimation of the invariant density $\mu \in \mathcal{H}_d(\beta, \mathcal{L})$ of our diffusion with jumps:

$$\mathbb{E}[|\hat{\mu}_{h,T}(x) - \mu(x)|^2] \sim V_d(T) := \begin{cases} \frac{(\log T)^{2-\frac{(1+\alpha)}{2}}}{T} & \text{for } d = 1, \\
\frac{\log T}{T} & \text{for } d = 2, \\
T^{-\frac{2\beta}{2\beta+d-2}} & \text{for } d \geq 3.\end{cases}$$

Estimation on $L^2(A)$, $A \subset \mathbb{R}^d$, compact:

$$\mathbb{E}[\|\hat{\mu}_{h,T} - \mu\|^2_A] \sim V_d(T).$$
Adaptive procedure, continuous case, \(d \geq 3\)

Let \(\mathcal{H}_t\) be the set of candidate bandwidths. For \(h = (h_1, ..., h_d) \in \mathcal{H}_t\)

\[
\mathcal{F}(\mathcal{H}_t) := \left\{ \hat{\mu}_{h, T}(x) = \frac{1}{T \prod_{l=1}^{d} h_l} \int_{0}^{T} \prod_{m=1}^{d} K\left(\frac{x_m - X^m_{u}}{h_m}\right)du, \quad x \in \mathbb{R}^d \right\}
\]

is the associate set of candidate estimators. Strauch proposes a fully data-driven selection procedure of the bandwidth inspired by Goldenshluger and Lespki (2011) [10]. The selected estimator \(\hat{\mu}_{\hat{h}} \in \mathcal{F}(\mathcal{H}_t)\) satisfies, for any \(T > 0\),

\[
\mathbb{E}[\|\hat{\mu}_{\hat{h}} - \mu(x)\|_{\infty}^q]^{\frac{1}{q}} \leq c_1 \inf_{h \in \mathcal{H}_T} \left\{ B(h) + \prod_{j=1}^{d} h_j^{\frac{1}{2}} \sqrt{\frac{\log T}{T}} \right\} + \frac{c_2}{\sqrt{T}},
\]

where \(\mathcal{H}_T \subset \mathcal{H}_t\) denotes a dyadic grid and \(B(\cdot)\) can be viewed as the approximation error of \(\mu\) measured in the supremum norm.
Adaptive procedure, jump framework, $d \geq 3$

$\mathcal{H}_t$ and $\mathcal{F}(\mathcal{H}_t)$ defined as before. We want to select $\hat{\mu}_{\hat{h}} \in \mathcal{F}(\mathcal{H}_t)$.

How?

For any bandwidths $h = (h_1, ..., h_d)^T$, $\eta = (\eta_1, ..., \eta_d)^T \in \mathcal{H}_T$ and $x \in \mathbb{R}^d$, we define

$$K_h \ast K_\eta(x) := \prod_{j=1}^{d} (K_{h_j} \ast K_{\eta_j})(x_j) = \prod_{j=1}^{d} \int_{\mathbb{R}} K_{h_j}(u - x_j)K_{\eta_j}(u)du,$$

$$\hat{\mu}_{h,\eta}(x) := \frac{1}{T} \int_{0}^{T} (K_h \ast K_\eta)(X_u - x)du, \quad x \in \mathbb{R}^d.$$

The proposed selection procedure relies on comparing the differences $\hat{\mu}_{h,\eta} - \hat{\mu}_\eta$. 
Adaptive procedure, jump framework, $d \geq 3$

We define

$$A(h) := \sup_{\eta \in \mathcal{H}_T} (\| \hat{\mu}_{h,\eta} - \hat{\mu}_\eta \|_A^2 - V(\eta))_+,$$

with

$$V(h) := \frac{k}{T} (\prod_{l=1}^{d} h_l)^{\frac{2d}{d-1}}.$$ 

Heuristically, $A(h)$ is an estimate of the squared bias and $V(h)$ of the variance bound. Thus, the selection is done by setting

$$\hat{h} := \arg \min_{h \in \mathcal{H}_T} (A(h) + V(h)).$$

We show that

$$\mathbb{E}[\| \hat{\mu}_{\hat{h}} - \mu \|_A^2] \leq c_1 \inf_{h \in \mathcal{H}_T} (\sum_{l=1}^{d} h_l^{2\beta_l} + V(h)) + c_1 e^{-c_2 (\log T)^2}.$$
We recall that Proposition 2 provides, for $d \geq 3$, the rate optimal choice $h_l(T) = \left(\frac{1}{T}\right)^{\frac{\beta}{\beta_l(2\beta+d-2)}}$. As $h(T) \in \mathcal{H}_T$, we obtain

$$\mathbb{E}[\|\hat{\mu}_{\tilde{h}} - \mu\|_A^2] \leq c_1 T^{-\frac{2\beta}{2\beta+d-2}} + c_1 e^{-c_2(\log T)^2}.$$ 

\[\downarrow\]

The risk estimates we get using the selected bandwidth $\tilde{h}$ converges to zero fast: its convergence rate coincides to the one provided by both Dalalyan and Reiss (2007) [6] and Strauch (2018) [17] in the case without jumps.

Is it optimal?
For the computation of lower bounds, we introduce the family of SDE with jumps:

\[
X_t = X_0 + \int_0^t b(X_s) \, ds + \int_0^t a \, dW_s + \int_0^t \int_{\mathbb{R}^d \setminus \{0\}} \gamma z \tilde{\mu}(ds, dz).
\]

The model still satisfies Assumptions ✿ and \( \int_{\mathbb{R}^d} F(z) \, dz = \lambda_1 \). The Lyapounov function exists and the process admits a unique stationary measure \( \mu_b \). \( \mathbb{P}_{\mu_b} \) law of the stationary solution \( (X_t)_{t \geq 0} \) and \( \mathbb{E}_{\mu_b} \) the corresponding expectation.

We introduce the minimax risk for the estimation at some point \( x_0 \in \mathbb{R}^d \).

\[
R_T(\beta, \mathcal{L}) := \inf_{\tilde{\mu}_T} \sup_{\mu_b \in \mathcal{H}_d(\beta, \mathcal{L})} \mathbb{E}_{\mu_b}[(\tilde{\mu}_T(x_0) - \mu_b(x_0))^2].
\]

We want to show that \( R_T(\beta, \mathcal{L}) \geq c\psi^2(T) \).
Lower bound (ongoing work)

Method: lower bound based on two hypotheses

\[ R(\tilde{\mu}_T(x_0)) := \sup_{\mu_b \in \mathcal{H}_d(\beta, \mathcal{L})} \mathbb{E}_{\mu_b}[(\tilde{\mu}_T(x_0) - \mu_b(x_0))^2] \geq \]

\[ \geq \max_{\mu_b \in \{\mu_{b_0}, \mu_{b_1}\} \subset \mathcal{H}_d(\beta, \mathcal{L})} \mathbb{E}_{\mu_b}[(\tilde{\mu}_T(x_0) - \mu_b(x_0))^2] \geq \]

\[ \geq \frac{1}{2} \mathbb{E}_{\mu_{b_0}}[(\tilde{\mu}_T(x_0) - \mu_{b_0}(x_0))^2] + \frac{1}{2} \mathbb{E}_{\mu_{b_1}}[(\tilde{\mu}_T(x_0) - \mu_{b_1}(x_0))^2]. \]

We need the laws \( \mathbb{P}_{\mu_{b_0}} \) and \( \mathbb{P}_{\mu_{b_1}} \) to be close. Link between \( b \) and \( \mu_b \) needed.

Without jumps it was \( b_0(x) = -\nabla V_0(x) = \frac{1}{2} \nabla (\log \mu_{b_0})(x) \), now it is no longer true.

\[ \Downarrow \]

With jumps, how can we get a relation between \( b \) and \( \mu_b \)?
We introduce $A$, the generator of the diffusion. $A = A_c + A_d$

$$A_c f(x) := \frac{1}{2} \sum_{i,j=1}^{d} a_{ij}^2 \partial_{i,j}^2 f(x) + \sum_{i=1}^{d} b_i(x) \partial_i f(x)$$

$$A_d f(x) := \int_{\mathbb{R}^d} (f(x + \gamma \cdot z) - f(x) - \gamma \cdot z \cdot \nabla f(x)) F(z) dz,$$

$A^*$ is the adjoint operator on $L^2(\mathbb{R}^d)$. $\mu_b$ can be computed as solution of $A^* \mu_b = 0$, ie

$$\frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} (a a^T)_{ij} \partial^2 \partial_{x_i} \partial_{x_j} \mu_b(x) - \left( \sum_{i=1}^{d} \frac{\partial b_i}{\partial x_i} \mu_b(x) + b_i \frac{\partial \mu_b}{\partial x_i}(x) \right) +$$

$$+ \int_{\mathbb{R}^d} |\gamma^{-1}| \mu_b(y) F(\gamma^{-1}(x - y)) dy - \lambda_1 \mu_b(x) +$$

$$+ \int_{\mathbb{R}^d} \sum_{k=1}^{d} z_k \left( \sum_{j=1}^{d} \gamma_{kj} \frac{\partial \mu_b(x)}{\partial x_j} \right) F(z) dz = 0.$$
⇒ For a given $\mu_b$ we find the associated $b$.
We define

$$\mu_{b_0}(x) := e^{-\epsilon \sum_{i=1}^{d} |x_i|}.$$  

For $\epsilon$ small enough $\mu_{b_0} \in H_d(\beta, \mathcal{L})$.
We show that, for such a $\mu_{b_0}$, the associated $b_0$ satisfies Assumptions ◻.

We define

$$\mu_{b_1}(x) := \mu_{b_0}(x) + \frac{1}{M_T} \prod_{i=1}^{d} K\left(\frac{x_i - x_i^0}{h_i(T)}\right),$$  

where $M_T$ and $h_i(T)$ will be calibrated later and satisfy,
$\forall i \in \{1, ..., d\}$ $M_T \to \infty$, $h_i(T) \to 0$ as $T \to \infty$. 
Lower bound (ongoing work)

- **Result 1**
  If, for any $i \in \{1, \ldots, d\}$,
  \[
  \frac{1}{M_T} \leq C_k h_i(T)_{\beta_i},
  \]
  then $\mu_{b_1} \in \mathcal{H}_d(\beta, \mathcal{L})$.

- **Result 2** (Required to get $\mathbb{P}_{\mu_{b_0}}$ next to $\mathbb{P}_{\mu_{b_1}}$)
  If
  \[
  T \prod_{i=1}^{d} h_i(T) \sum_{i=1}^{d} h_i(T)^{-2} < c_1,
  \]
  then there exist $\lambda_0$ and $c > 0$ such that
  \[
  \sup_{T \geq 0} T \int_{\mathbb{R}^d} |b_0(x) - b_1(x)|^2 dx < c.
  \]
If the Results 1 and 2 hold true, we show that $R(\tilde{\mu}_T(x_0)) \geq \frac{c}{M_T^2}$.

\[\Downarrow\]

We look for the larger choice of $\left(\frac{1}{M_T}\right)^2$ such that

1. \[T \frac{\prod_{l=1}^{d} h_l(T)}{M_T^2} \sum_{i=1}^{d} h_i(T)^{-2} < c,\]

2. \[\frac{1}{M_T} \leq C_k h_i(T)^{\beta_i}, \; \forall i \in \{1, ..., d\} .\]

After computation, it follows

\[M_T = cT^{d + 2\beta - 2\beta_{\min}}.\]

We can now state the main result:
There exists $c > 0$ such that

$$\inf_{\tilde{\mu}_T} \sup_{\mu_b \in \mathcal{H}_d(\beta, \mathcal{L})} \mathbb{E}_{\mu_b}[(\tilde{\mu}_T(x_0) - \mu_b(x_0))^2] > T^{-\frac{2\bar{\beta}}{d + 2\bar{\beta} - 2\frac{\beta}{\beta_{\text{min}}}}}.$$ 

We recall the upper bound for $d \geq 3$:

$$\mathbb{E}[|\hat{\mu}_h, T(x_0) - \mu_b(x_0)|^2] \leq T^{-\frac{2\bar{\beta}}{2\bar{\beta} + d - 2}}.$$ 

$$\frac{2\bar{\beta}}{d + 2\bar{\beta} - 2\frac{\beta}{\beta_{\text{min}}}} > \frac{2\bar{\beta}}{2\bar{\beta} + d - 2}, \text{ in general.}$$

$$\beta_{\text{min}} = \bar{\beta} \text{ in the isotropic case } \Rightarrow \text{ the convergence rate is optimal in the isotropic context.}$$


Thank you for your attention!