Penalized nonparametric drift estimation in a continuous time one-dimensional diffusion.

joint work with Eva Löcherbach and Oleg Loukianov, Paris 12.

1. Penalized nonparametric drift estimation
2. Polynomial deviation inequality in the ergodic theorem for one-dimensional diffusion and integrability of hitting times

Let $(X_t)_{t\geq 0}$ be one-dimensional positive recurrent diffusion

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad X_0 = x,$$

observed continuously during $[0, t]$.

We want to estimate non-parametrically the drift function b on $[0, 1]$. The estimator we propose is obtained by a penalized least square approach:

1. Consider families of finite-dimensional linear subspaces of $\mathbb{L}^2([0, 1])$. Compute for each sub-space an associated least-squares estimator (non-adaptive estimation).
2. Adaptive procedure chooses among the resulting collection of estimators the "best one" via some penalization device (adaptive estimation).

Assumptions on X.

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad X_0 = x,$$

- b and σ are locally Lipschitz and there is some $C > 0$ such that $|b(x)| \leq C(1+|x|)$.
- There exist $0 < \sigma_0^2 \leq \sigma_1^2 < \infty$ such that for all x,

$$\sigma_0^2 \leq \sigma^2(x) \leq \sigma_1^2.$$

- There exist γ, $2\gamma > 31\sigma_1^2$ and $M_0 > 0$, such that $[0,1] \subset [-M_0, M_0]$ and

$$xb(x) \leq -\gamma \quad \text{for} \quad |x| \geq M_0.$$

The third hypothesis means “polynomial β-mixing case”. The analogous problem was considered by Genon-Catalot, Compte, Rozenholc for X starting from stationary distribution and under the assumption of exponential β-mixing.

Assumptions on the collection of models

We aim to estimate b on $[0,1]$ using a data-driven procedure. For this purpose we consider a collection $(S_m)_{m\in\mathcal{M}_t}$ of finite dimensional sub-spaces of $\mathbb{L}^2([0,1],dx)$. Denote D_m dimension of S_m.

- $\forall m \in \mathcal{M}_t\ S_m \subseteq S_t$. (All models belong to a "biggest" model.)
- The dimension of "biggest space" verifies $D_t \leq t$ and $Card\mathcal{M}_t \leq D_t$.
- Let $\{\phi_1, \ldots \phi_{D_t}\}$ be an orthonormal basis of $S_t \subset \mathbb{L}^2([0,1])$. Then there exists a positive constant Φ_1 such that for all i,

$$\mathrm{Card}\{j:\ \|\phi_i\phi_j\|_\infty \neq 0\} \leq \Phi_1.$$

Non-adaptive estimation.

Fix a subspace $S_m \subset \mathbb{L}^2([0,1], dx)$. The estimator $\hat{b}_m$ of b on $[0,1]$ is a minimizer on S_m of the following contrast function:

$$\gamma_t(h) = \frac{1}{t}\int_0^t h^2(X_s)ds - \frac{2}{t}\int_0^t h(X_s)dX_s.$$

Denote

$$\|h\|_t^2 = \frac{1}{t}\int_0^t h^2(X_s)ds, \quad \text{and} \quad T_m(h,f) = \frac{1}{t}\int_0^t h(X_s)f(X_s)ds$$

the “empirical $\mathbb{L}^2$ norm” and the corresponding quadratic form.
Remark that by Ito

$$\gamma_t(h) = \|h-b\|_t^2 - \frac{2}{t}M_t - \|b\|_t^2,$$

hence $M_t/t \to 0$, minimization of $\gamma_t(h) \sim$ minimization of $\|h-b\|_t^2$.

To ensure the existence of $\hat{b}_m$:

Let

$$A_t = \{\forall m \in \mathcal{M}_t;\ \rho_m \geq t^{-1/2}\}$$

where ρ_m is a smallest eigenvalue of T_m.

T_m positively defined on A_t, hence minimizer exists and is unique.

$$\hat{b}_m = argmin_{h \in S_m} \gamma_t(h) \quad \text{on} \quad A_t \quad \text{and} \quad \hat{b}_m = 0 \quad \text{on} \quad A_t^c.$$

We define the risk as the expectation of the empirical norm:

$$\mathbb{E}_x \|\hat{b}_m - b\|_t^2 = \mathbb{E}_x \left(\frac{1}{t} \int_0^t (\hat{b}_m - b)^2 (X_s) ds \right).$$

Let b_m be the $\mathbb{L}^2([0,1],dx)$ projection of b onto S_m and $\|.\|$ be the $\mathbb{L}^2([0,1],dx)$ norm.

Theorem

for $t \geq t_0 = \frac{4}{p_0^2}$,

$$E_x\|\hat{b}_m - b\|_t^2 \leq 3\kappa\|b_m - b\|^2 + 16\sigma_1^2 \frac{\kappa}{p_0}\frac{D_m}{t} + C\left(b_0^2 + \sigma_1^2\right) t^{-1}$$

Here, $\kappa = \kappa(t) = \frac{2}{\sigma_0^2}(\frac{2}{t} + \frac{2\sigma_1}{\sqrt{t}} + 2b_0 + \frac{\sigma_1^2}{2})$ *and* $p_0 = const(\sigma_1, b_0)$.

in some cases we can bound $\|b_m - b\|^2$. For Besov space $B_{\alpha,2,\infty}([0,1])$, $\|b_m - b\|^2 \leq CteD_m^{-2\alpha}$. Then the best choice of D_m : $D_m = t^{\frac{1}{2\alpha+1}}$; then we obtain

$$E_x\|\hat{b}_m - b\|_t^2 \leq Ct^{-\frac{2\alpha}{2\alpha+1}} + C\left(b_0^2 + \sigma_1^2\Phi_0^2\right) t^{-1},$$

and this yields exactly the classical non-parametric rate $t^{-\frac{2\alpha}{2\alpha+1}}$.

Adaptive estimation.

The previous choice supposes the knowledge of the regularity α of the unknown drift. Adaptive estimation scheme choose automatically the best dimension D_m when α is unknown. Recall for $m \in \mathcal{M}_t$

$$\hat{b}_m = argmin_{h \in S_m} \gamma_t(h) \quad \text{on} \quad A_t \quad \text{and} \quad \hat{b}_m = 0 \quad \text{on} \quad A_t^c.$$

Put

$$\hat{m} := \arg \min_{m \in \mathcal{M}_t} \left[\gamma_t(\hat{b}_m) + pen(m) \right]. \tag{1}$$

Then the estimator that we propose is the following adaptive estimator

$$\hat{b}_{\hat{m}} := \begin{cases} \sum_n 1_{\{\hat{m}=n\}} \hat{b}_n & \text{on } A_t \\ 0 & \text{on } A_t^c \end{cases}. \tag{2}$$

Theorem

Let

$$pen(m) \geq \chi \sigma_1^2 \frac{D_m}{t},$$

where χ *is a universal constant. Then we have for all* $b \in \mathcal{M}(M_0, b_0, \gamma)$,

$$E_x ||\hat{b}_{\hat{m}} - b||_t^2 \leq 3\kappa \inf_{m \in \mathcal{M}_t} \left(||b_m - b||^2 + pen(m) \right) + \frac{K'}{t}, \qquad (3)$$

where $\kappa = \kappa(t) = \frac{2}{\sigma_0^2}(\frac{2}{t} + \frac{2\sigma_1}{\sqrt{t}} + 2b_0 + \frac{\sigma_1^2}{2})$

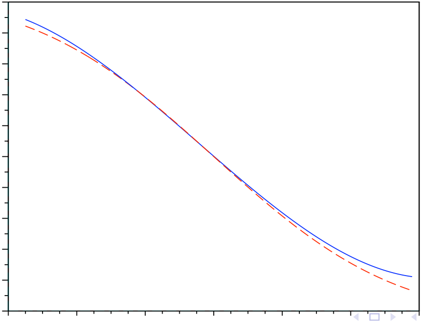
The adaptive estimator achieves automatically the bias-varance equilibrium.

Simulations

$$dX_t = -\frac{\gamma X_t}{1+X_t^2}\,dt + dW_t, \quad X_0 = x, \quad \gamma > \frac{1}{2}.$$

This diffusion is positive recurrent with stationary distribution

$$\mu(dx) \sim \frac{dx}{(1+x^2)^{\gamma}}.$$



Polynomial deviation inequality.

Denote by μ the invariant measure of X. The main tool in the proof of the statistical result is the following deviation inequality:

$$\mathbb{P}_\nu \left(\left| \frac{1}{t} \int_0^t f(X_s) ds - \mu(f) \right| \geq \epsilon \right) \leq K(p) \frac{1}{t^{p/2}} \frac{1}{\epsilon^p} A(f)^p$$

where ν is an initial distribution, f bounded or bounded and compactly supported and $A(f) = \|f\|_\infty$ when f is bounded and $A(f) = \mu(|f|)$ when f is bounded and compactly supported.

What is the speed p ? To formulate conditions on p we start by introducing the notion of polynomial degree of recurrence.

Polynomial degree of recurrence.

Assume: $\sigma^2(x) > 0$ for all x, β, σ continuous, non exploding solution exist.

$$dX_t = \beta(X_t)dt + \sigma(X_t)dW_t, \quad X_0 = x,$$

Denote $T_y = \inf\{t \geq 0,\ X_t = y\}$.

X is recurrent if and only if $\mathbb{P}_x(T_y < \infty) = 1$ for all x, y.
X is positive recurrent if and only if $E_x T_y < \infty$ for all x, y.

Definition

The degree of recurrence of X is polynomial of order $p \geq 1$, if for all x, y we have $\mathbb{E}_x T_y^p < \infty$ and $\mathbb{E}_x T_y^{p+1} = \infty$.

It is known (Maryama, Tanaka) $\mathbb{E}_x T_y^p < \infty$ or $= \infty$ simultaneously for all couples x, y : We show the following theorem:

Theorem

1. *Let $x < b$ and $n \in \mathbb{N}^*$.*
   - (1i) $\mathbb{E}_x T_b^n < \infty$ *if and only if* $\int_{-\infty}^{x} \mathbb{E}_\xi T_b^{n-1} m(\xi) d\xi < \infty$.
   - (1ii) *If for one couple $x < b$, $\mathbb{E}_x T_b^n < \infty$, then for all couples $x' < b'$, $\mathbb{E}_{x'} T_{b'}^n < \infty$.*
     Moreover, $x' \to \mathbb{E}_{x'} T_{b'}^n$ is continuous.
2. *Let $a < x$ and $n \in \mathbb{N}^*$.*
   - (2i) $\mathbb{E}_x T_a^n < \infty$ *if and only if* $\int_{x}^{+\infty} \mathbb{E}_\xi T_a^{n-1} m(\xi) d\xi < \infty$.
   - (2ii) *If for one couple $a < x$, $\mathbb{E}_x T_a^n < \infty$, then for all couples $a' < x'$, $\mathbb{E}_{x'} T_{a'}^n < \infty$.*
     Moreover, $x' \to \mathbb{E}_{x'} T_{a'}^n$ is continuous.

here $m(x)$ is the density of the invariant distribution.

If X has the degree of recurrence at least p, the deviation inequality is true with this p.

$$\mathbb{P}_\nu\left(|\frac{1}{t}\int_0^t f(X_s)ds - \mu(f)| \geq \epsilon\right) \leq K(p)\frac{1}{t^{p/2}}\frac{1}{\epsilon^p}A(f)^p$$

Fix two points $a < b$. Define: $S_0 = 0,\ R_0 = 0,$

$$S_1 = \ \{t \geq 0 : X_t = b\}, \quad R_1 := \inf\{t \geq S_1 : X_t = a\},$$

and for $n \geq 1$,

$$S_{n+1} := R_n + S_1 \circ \partial_{R_n}, \quad R_{n+1} := R_n + R_1 \circ \partial_{R_n}.$$

For any measurable and bounded function $f : \mathbb{R} \to \mathbb{R}$, we put

$$\xi_n = \int_{R_n}^{R_{n+1}} f(X_s)ds, \quad n \geq 1. \tag{4}$$

Lemma

For any initial distribution ν, the sequence $(\xi_n)_{n\geq 1}$ is an i.i.d. sequence under $\mathbb{P}_\nu$. For all $n \geq 1$, the law of ξ_n under $\mathbb{P}_\nu$ is equal to the law of $\int_0^{R_1} f(X_s)ds$ under $\mathbb{P}_a$.

The successive visits (R_n) cut trajectory on i.i.d. cycles:

$$\int_0^t f(X_s)ds = \int_0^{R_1} f(X_s)ds + \sum_{n=1}^{N_t} \xi_n + \int_{N_t}^t f(X_s)ds$$

where $N_t = sup\{n : \quad R_n \leq t\}$ is a number of visits R_n before t.

$$\text{Main term:} \quad \sum_{n=1}^{N_t} \xi_n$$

To control the main term $\sum_{n=1}^{N_t} \xi_n$ we need to control N_t and ξ_n

- The processes $(N_t)_{t\geq 0}$ and $(R_n)_{n\in\mathbb{N}}$ are mutually inverse: $\{N_t \geq n\} = \{R_n \leq t\}$ and $\{N_t \leq n\} = \{R_n \geq t\}$.
- $R_n = R_0 + (R_1 - R_0) + \ldots + (R_n - R_{n-1})$, sum i.i.d.
- $\xi_n = \int_{R_n}^{R_{n+1}} f(X_s)ds$

Conditions in ergodic theorem involve integrability of cycles $(R_k - R_{k-1})$, but

$$E_\nu(R_k - R_{k-1})^p \leq 2^{p-1}\left(E_a T_b^p + E_b T_a^p\right).$$

Theorem

Let f be a measurable bounded function, or measurable bounded with compact support, such that $\mu(f) \neq 0$. Suppose that (X) is recurrent of degree at least $p \geq 1$. For all x

(a) *For a bounded function f such that $\|f\|_\infty \leq M$:*

$$\mathbb{P}_x\left(\left|\frac{1}{t}\int_0^t f(X_s)ds - \mu(f)\right| > \epsilon\right) \leq K(p)\frac{1}{t^{p/2}}\frac{1}{\epsilon^p}M^p, \quad (5)$$

where $K(p)$ does not depend on f, t, ϵ.

(b) *For a bounded function f with compact support:*

$$\mathbb{P}_x\left(\left|\frac{1}{t}\int_0^t f(X_s)ds - \mu(f)\right| > \epsilon\right) \leq K(p)\frac{1}{t^{p/2}}\frac{1}{\epsilon^p}(\mu(|f|))^p, \quad (6)$$

where $K(p)$ does not depend on f, t, ϵ.

To have this theorem under $\mathbb{P}_\nu$ we need to ensure $\mathbb{E}_\nu T_a^p < \infty$.

Existence of hitting times moments

Assume:

- There exist positive constants σ_0 and σ_1 such that

$$\sigma_0 \leq \sigma(x) \leq \sigma_1, \text{ for all } x \in \mathbb{R}. \quad (7)$$

- There exist two constant $R \geq r > 0$ such that

$$r \leq -x\beta(x) \leq R \quad \text{for } |x| > M_0. \quad (8)$$

It is known (Veretennikov, Balaji and Ramasubramanian) that under these assumptions for $x > a > M_0$ $\mathbb{E}_x T_a^p$ is finite for $p < r/\sigma_1^2 + 1/2$ and infinite for $p > R/\sigma_0^2 + 1/2$,

but we need a finer control on $\mathbb{E}_x T_a^p$ to be able to estimate $\mathbb{E}_\nu T_a^p$.

Theorem

If $x > a > M_0$ (or $x < a < -M_0$) and $n < r/\sigma_1^2 + 1/2$ then

$$\frac{(x-a)^{2n}}{R_n} \le \mathbb{E}_x T_a^n \le \frac{x^{2n}}{r_n} + \ldots = P_{2n}(x)$$

where R_n and r_n depend explicitly on σ_0, σ_1 and n.

Bibliography

1. Balaji, S., Ramasubramanian, S., Passage time Moments for Multidimensional Diffusions, Journal Of Applied Probability, Vol. 37, No. 1 (2000), pp. 246-251
2. Genon-Catalot, V., Laredo, C., Nussbaum, M., Asymptotic equivalence of estimating a poisson intensity and a positive diffusion drift, The Annals of Statistics, 2002, Vol.30., No 3, 731-753.
3. Maruyama, G., Tanaka, H. Some properties of one-dimensional diffusion processes. Memoirs of the Faculty of Science, Kyusyu University, Ser.A. Vol.11, No.2, 1957.
4. Veretennikov, A.Yu., On polynomial mixing bounds for stochastic differential equations, Stochastic processes and their applications, 70, (1997) 115-127.