Asymptotic equivalence for inhomogeneous jump diffusion processes and white noise.

Ester Mariucci

Laboratoire Jean Kuntzmann, Grenoble

Dynstoch, 10–12 September 2014

Table of contents

Inhomogeneous jump diffusion processes

$$X_t = \int_0^t f(s)ds + \int_0^t \sigma(s)dW_s + \sum_{i=1}^{N_t} Y_i, \quad t \geq 0,$$

- $W = \{W_t\}_{t\geq 0}$ is a standard Brownian motion;
- $N = \{N_t\}_{t\geq 0}$ is an inhomogeneous Poisson process with intensity function $\lambda(\cdot)$, independent of W;
- $(Y_i)_{i\geq 1}$ is a sequence of i.i.d. real random variables with distribution G (either concentrated on $\mathbb{Z}$ or absolutely continuous with respect to Lebesgue), independent of W and N;
- $\sigma^2(\cdot), \lambda(\cdot)$ and G are supposed to be known and $f(\cdot)$ belongs to a certain non-parametric class $\mathcal{F}$.

The problem we consider

We suppose to observe $\{X_t\}_{t\geq 0}$ at discrete times $0 = t_1 < \cdots < t_n = T_n$ such that

$$\Delta_n = \max_{1\leq i\leq n} \big\{|t_i - t_{i-1}|\big\} \downarrow 0 \text{ as } n \to \infty.$$

Problem: To estimate the drift function $f(\cdot)$ from the discrete data $(X_{t_i})_{i=1}^n$.

At least two natural questions arise:

1. How much information about the parameter $f(\cdot)$ do we lose by observing $(X_{t_i})_{i=1}^n$ instead of $\{X_t\}_{t\in[0,T_n]}$?
2. Can we construct an easier (read: mathematically more tractable), but equivalent, model from $(X_{t_i})_{i=1}^n$?

The general problem

The statistician has several experiments at his disposal to estimate a parameter θ. How to compare them?

- Experiment 1 $\mapsto \mathcal{E}_1 = \big(\mathcal{X}_1, \mathcal{T}_1, (P_\theta)_{\theta\in\Theta}\big)$,
- Experiment 2 $\mapsto \mathcal{E}_2 = \big(\mathcal{X}_2, \mathcal{T}_2, (Q_\theta)_{\theta\in\Theta}\big)$,

First idea (Bohnenblust, Sharpey, Sherman, 1949): $\mathcal{E}_1$ is more informative than $\mathcal{E}_2$ if for any bounded loss function L and any decision ρ_2 for the experiment $\mathcal{E}_2$ there exists a decision ρ_1 for the experiment $\mathcal{E}_1$ s.t.

$$R_\theta(\mathcal{E}_1, L, \rho_1) \le R_\theta(\mathcal{E}_2, L, \rho_2), \quad \forall\theta\in\Theta.$$

Problem: With this approach $\mathcal{E}_1$ and $\mathcal{E}_2$ may be non comparable.

The notion of deficiency

Le Cam idea (1964): "How much do we lose if we use the experiment $\mathcal{E}_1$ instead of the experiment $\mathcal{E}_2$?"

Definition

The **deficiency** $\delta\big(\mathcal{E}_1, \mathcal{E}_2\big)$ of $\mathcal{E}_1$ with respect to $\mathcal{E}_2$ is defined as

$$\delta(\mathcal{E}_1, \mathcal{E}_2) = \inf_K \sup_{\theta \in \Theta} \|K P_\theta - Q_\theta\|_{TV},$$

where the infimum is taken over all "randomizations".

Remark 1: Markov kernels are special cases of randomizations.

Remark 2: The deficiency is defined for any pair of statistical models indexed by the same parameter space.

The Le Cam Δ-distance

Property

Let $\varepsilon > 0$ be fixed. $\delta(\mathcal{E}_1, \mathcal{E}_2) \le \varepsilon \iff \forall$ bounded loss function L, $\forall$ decision rule ρ_2 on $\mathcal{E}_2$, $\exists$ a decision rule ρ_1 on $\mathcal{E}_1$ such that

$$R_\theta(\mathcal{E}_1, L, \rho_1) \le R_\theta(\mathcal{E}_2, L, \rho_2) + \varepsilon, \quad \forall \theta \in \Theta.$$

Definition

The so called **Δ-distance** between $\mathcal{E}_1$ and $\mathcal{E}_2$ is the pseudometric defined by:

$$\Delta(\mathcal{E}_1, \mathcal{E}_2) = \max\left(\delta(\mathcal{E}_1, \mathcal{E}_2), \delta(\mathcal{E}_2, \mathcal{E}_1)\right).$$

The experiments $\mathcal{E}_1$ and $\mathcal{E}_2$ are said to be **equivalent** if $\Delta(\mathcal{E}_1, \mathcal{E}_2) = 0$. Two sequences of statistical models $(\mathcal{E}_1^n)_{n\in\mathbb{N}}$ and $(\mathcal{E}_2^n)_{n\in\mathbb{N}}$ are called **asymptotically equivalent** if $\Delta(\mathcal{E}_1^n, \mathcal{E}_2^n) \to 0$ as $n \to \infty$.

Examples

Parametric case: estimation of a real parameter $\theta \in \Theta \subset \mathbb{R}$.

1. $\mathcal{E}_1^n = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), (P_{n,\theta} : \theta \in \Theta)),\ P_{n,\theta} = \otimes_{i=1}^n \mathcal{L}\big(\mathcal{N}(\theta, 1)\big)$.
2. $\mathcal{E}_2^n = (\mathbb{R}, \mathcal{B}(\mathbb{R}), (Q_{n,\theta} : \theta \in \Theta)),\ Q_{n,\theta} = \mathcal{L}\big(\mathcal{N}(\theta, n^{-1})\big)$.

Non parametric case: estimation of a function $h : [0,1] \to \mathbb{R}$

1. $Y_i = h(i/n) + \sigma(i/n)\xi_i$, $\xi_i \sim \mathcal{N}(0,1)$, $i = 1, \dots, n$, i.i.d; h is an unknown function in $\mathcal{H}$, σ is supposed to be known.
   $\mathcal{E}_1^n = \big(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), (P_{n,h} : h \in \mathcal{H})\big),\ P_{n,h} = \mathcal{L}\big((Y_1, \dots, Y_n)\big)$
2. $dY_t = h(t)dt + \frac{\sigma(t)}{\sqrt{n}} dW_t,\ t \in [0,1], (W_t)$ SBM.
   $\mathcal{E}_2^n = \big(C[0,1], \mathcal{C}, (Q_{n,h} : h \in \mathcal{H})\big),\ Q_{n,h} = \mathcal{L}Y$.

Asymptotic equivalence in a non parametric framework

If, for the estimation of a function f, the sequences of experiments $(\mathcal{E}_1^n)_{n\in\mathbb{N}}$ and $(\mathcal{E}_2^n)_{n\in\mathbb{N}}$ are asymptotically equivalent in the Le Cam's sense:

$$\Delta(\mathcal{E}_1^n, \mathcal{E}_2^n) \to 0,$$

then asymptotic properties of any inference problem are the same for these experiments (rates of convergence, minimax exact constants) $\Longrightarrow$ it is enough to choose the simplest one when studying these properties.

- Brown and Low (1996): regression and white noise
- Nussbaum (1996): density estimation and white noise
- + numerous papers showing the global asymptotic equivalence between non-parametric experiences (generalized linear models, time series, diffusion models without jumps, GARCH model, functional linear regression, spectral density estimation).

Notations

- $(D,\mathcal{D})$: Skorokhod space (càdlàg functions);
- $P^{(f,\sigma^2,\lambda G)}$: law of the process $\{X_t\}_{t\in[0,T_n]}$ on $(D,\mathcal{D})$;
- $Q_n^{(f,\sigma^2,\lambda G)}$: law of the vector $(X_{t_1},\dots,X_{t_n})$ on $(\mathbb{R}^n,\mathcal{B}(\mathbb{R}^n))$;
- $\mathscr{P}^{(f,\sigma^2,\lambda G)}=\Big(D,\mathcal{D},\big(P^{(f,\sigma^2,\lambda G)}\big)_{f\in\mathcal{F}}\Big)$;
- $\mathscr{Q}_n^{(f,\sigma^2,\lambda G)}=\Big(\mathbb{R}^n,\mathcal{B}(\mathbb{R}^n),\big(Q_n^{(f,\sigma^2,\lambda G)}\big)_{f\in\mathcal{F}}\Big)$.

Remark

$\mathscr{P}^{(f,\sigma^2,0)}$ is the experiment associated with the observation of a trajectory of the process:

$$X_t^c=\int_0^t f(s)ds+\int_0^t\sigma(s)dW_s,\quad t\in[0,T_n].$$

$\mathcal{F}$: a class of α-Hölder, uniformly bounded functions on $\mathbb{R}$; diffusion coefficient $\sigma_0 < \sigma(\cdot) < \sigma_1$ such that $\sigma'(\cdot) \in L_\infty(\mathbb{R})$; intensity $\lambda(\cdot) \in L_\infty(\mathbb{R})$.

Theorem (M., 2014)

$$\left.\begin{array}{r}\Delta\Big(\mathscr{Q}_n^{(f,\sigma^2,\lambda G)}, \mathscr{P}^{(f,\sigma^2,0)}\Big) \to 0\\ \Delta\Big(\mathscr{P}^{(f,\sigma^2,\lambda G)}, \mathscr{Q}_n^{(f,\sigma^2,\lambda G)}\Big) \to 0\end{array}\right. \quad \textit{as } n \to \infty,$$

under either of the following two sets of conditions:

1. Y_1 *is discrete with support on* $\mathbb{Z}$, $\alpha \geq \frac{1}{2}$ *and* $T_n\Delta_n \to 0$ *as* $n \to \infty$*; in this case the rate of convergence is* $O(\sqrt{T_n\Delta_n})$.
2. Y_1 *admits a density with respect to the Lebesgue measure on* $\mathbb{R}$, $\alpha \geq \frac{1}{4}$ *and* $T_n\sqrt{\Delta_n} \to 0$ *as* $n \to \infty$*; in this case the rate of convergence is* $O\Big(T_n^{\frac{1}{2}}\Delta_n^{\frac{1}{4}}\Big)$.

Gaussian white model

Recall: The main interest in the Le Cam asymptotic theory lies in the approximation of general statistical models by simpler ones $\implies$ we can reduce the more complicated model $\mathscr{Q}_n^{(f,\sigma^2,\lambda G)}$ to a simpler one since $\mathscr{P}^{(f,\sigma^2,0)}$ is essentially a Gaussian white noise model.
Indeed, when $\sigma^2(\cdot) = \sigma^2$ is constant, the experiment associated with

$$dX_t^c = f(t)dt + \sigma dW_t, \quad t \in [0, T_n]$$

is equivalent to that associated with

$$dY_u = F(u)du + \varepsilon dW_u, \quad u \in [0, 1],$$

where $F(u) := \frac{f(uT_n)}{T_n}$ and $\varepsilon := \sigma T_n^{-\frac{3}{2}}$.

Example (Diffusion + inhomogeneous Poisson process)

$$X_t = \int_0^t f(s)ds + \int_0^t \sigma(s)dW_s + N_t.$$

This is a special case of (1), with $Y_1 \equiv 1$.

Example (Merton model, inhomogeneous in time)

$$X_t = \int_0^t f(s)ds + \int_0^t \sigma(s)dW_s + \sum_{i=1}^{N_t} Y_i, \quad t \geq 0,$$

where Y_i are Gaussian r.v. $\mathcal{N}(m, \Gamma^2)$, $\Gamma > 0$. This is a special case of (2).

Main ideas

Goal: $(X_{t_i})_{i=1}^n \overset{\Delta}{\Longleftrightarrow} (X_t^c)_{t\in[0,T_n]}$.

- $(X_{t_i})_{i=1}^n \overset{\Delta}{\Longleftrightarrow} (X_{t_i}-X_{t_{i-1}})_{i=1}^n$;
- $X_{t_i}-X_{t_{i-1}} \sim N_i * \sum_{j=1}^{\mathscr{P}(\lambda_i)} Y_j$ with $N_i \sim \mathcal{N}(m_i,\sigma_i^2)$, $m_i = \int_{t_{i-1}}^{t_i} f(s)ds$, $\sigma_i^2 = \int_{t_{i-1}}^{t_i} \sigma^2(s)ds$, $\lambda_i = \int_{t_{i-1}}^{t_i} \lambda(s)ds$.

Step 1: Reduce to having in each interval at most one jump (Bernoulli approximation);

Step 2: Filter it out with an explicit Markov kernel $\Longrightarrow$ reducing ourselves to $(N_i)_{i=1}^n$;

Step 3: Apply an argument similar to that in Brown and Low.

Bernoulli approximation

Let $(\varepsilon_i)_{i=1}^n$ be a sequence of Bernoulli independent r.v. with parameter $\alpha_i = \lambda_i e^{-\lambda_i}$, then:

Lemma

$$\Big\| \bigotimes_{i=1}^n N_i * \sum_{j=1}^{\mathscr{P}(\lambda_i)} Y_j - \bigotimes_{i=1}^n N_i * \varepsilon_i Y_1 \Big\|_{TV} \leq 2\sqrt{\sum_{i=1}^n \lambda_i^2}.$$

Conclusion: $\Delta\Big((X_{t_i})_{i=1}^n, \otimes_{i=1}^n N_i * \varepsilon_i Y_1\Big) = O\Big(\sqrt{T_n \Delta_n}\Big).$

Explicit construction of Markov kernels: discrete case

Lemma

Define the Markov kernel:

$$K(x,A)=\mathbb{I}_A(x-[x]),\quad \forall A\in\mathcal{B}(\mathbb{R}).$$

For n big enough s.t. $|m_i|\leq\frac{1}{3}$, one has

$$\Big\|\bigotimes_{i=1}^n K(N_i*\varepsilon_i Y_1)-\bigotimes_{i=1}^n N_i\Big\|_{TV}\leq \sqrt{2\sum_{i=1}^n\Big(\frac{6}{\sigma_i}\varphi\big(\frac{1}{6\sigma_i}\big)+4\Phi\big(\frac{-1}{6\sigma_i}\big)\Big)}.$$

Here Φ stands for the cumulative distribution of a r.v. $\mathcal{N}(0,1)$ and φ for its derivative.

Continuous case

Lemma

Let $0 < \varepsilon < 1$ *be fixed and* $\beta_i := 1 + \sigma_i^{1-\varepsilon}$*. Define*

$$K_i(x, A) = \begin{cases} \mathbb{I}_A(x) & \text{if } x \in [-\beta_i, \beta_i], \\ \frac{1}{\sqrt{2\pi\sigma_i^2}} \int_A e^{-\frac{y^2}{2\sigma_i^2}} dy, & \text{otherwise.} \end{cases}$$

For n *big enough s.t.* $|m_i| \leq 1$*, one has*

$$\Big\| \bigotimes_{i=1}^n K_i(N_i * \varepsilon_i Y_1) - \bigotimes_{i=1}^n N_i \Big\|_{TV} \leq$$

$$\sqrt{2 \sum_{i=1}^n \bigg(8\Phi(-\sigma_i^{-\varepsilon}) + \frac{\alpha_i |m_i|}{\sqrt{2}\sigma_i} + 2\alpha_i \int_{-2\beta_i}^{2\beta_i} G'(y) dy \bigg)}$$

Conclusion: $\Delta\big((X_{t_i})_{i=1}^n, \otimes_{i=1}^n N_i\big) = O\Big(T_n^{\frac{1}{2}} \Delta_n^{\frac{1}{4}}\Big)$.

The parametric space

(F1) $\sup_{t\in\mathbb{R}}\{|f(t)| : f \in \mathcal{F}\} = B < \infty$.

(F2) Defining:

$$\bar{f}_n(t) = \begin{cases} f(t_i) & \text{if} \quad t_{i-1} \leq t < t_i, \quad i = 1, \dots, n; \\ f(T_n) & \text{if} \quad t = T_n; \end{cases}$$

we have

$$\lim_{n\to\infty} \sup_{f\in\mathcal{F}} \int_0^{T_n} \frac{(f(t) - \bar{f}_n(t))^2}{\sigma^2(t)} dt = 0$$

(F3) $\forall i = 1, \dots, n$, let γ_i and η_i be in $[t_{i-1}, t_i]$ and s.t.

$$\int_{t_{i-1}}^{t_i} \sigma^2(s)ds = \sigma^2(\eta_i)(t_i - t_{i-1}), \int_{t_{i-1}}^{t_i} f(s)ds = f(\gamma_i)(t_i - t_{i-1}).$$

Then we ask:

$$\lim_{n\to\infty} \sup_{f\in\mathcal{F}} \sum_{i=1}^{n} \frac{(f(t_i) - f(\gamma_i))^2}{\sigma^2(\eta_i)}(t_i - t_{i-1}) = 0.$$

Conclusion and extensions

Asymptotic framework: $n \to \infty$, $\Delta_n \to 0$, T_n can be fixed or go to infinity.

Extension to the case of unknown $\lambda(\cdot)$ and $G(\cdot)$: Work in progress.

Extension to the case of unknown $\sigma(\cdot)$: The statistical procedures to estimate f generally do not use the knowledge of $\sigma(\cdot)$ which is considered as a nuisance parameter. How to extend our result to the case where $\sigma(\cdot)$ is unknown is still an open problem.

Remark: Carter(2007), *Asymptotic approximation of nonparametric regression experiments with unknown variances.*

References

- Lawrence D. Brown and Mark G. Low, *Asymptotic equivalence of nonparametric regression and white noise*, Ann. Statist. **24** (1996).
- Andrew V. Carter, *Asymptotic approximation of nonparametric regression experiments with unknown variances*, Ann. Statist. **35** (2007).
- Ester Mariucci, *Asymptotic equivalence for inhomogeneous jump diffusion processes and white noise*, arXiv preprint arXiv:1405.0480 (2014).
- Michael Nussbaum, *Asymptotic equivalence of density estimation and Gaussian white noise*, Ann. Statist. **24** (1996).