Nadaraya-Watson Estimator for I.I.D. Paths of Diffusion Processes

Nicolas MARIE

–

Université Paris Nanterre - Modal'X
nmarie@parisnanterre.fr

Introduction

The model

Consider the stochastic differential equation

$$\mathrm{X}_t = x_0 + \int_0^t b(\mathrm{X}_s)\mathrm{d}s + \int_0^t \sigma(\mathrm{X}_s)\mathrm{d}\mathrm{W}_s \ ; t \in [0, \mathrm{T}] \qquad (1)$$

where W is a Brownian motion.

How to estimate the drift function?

Two approaches:

1. Estimators based on the behavior of X when $T \to \infty$ (see Kutoyants (2004)).
2. Estimators based on

$$X^i := \mathcal{I}(x_0, W^i)\ ;\ i = 1, \dots, N,$$

where $W^1, \dots, W^N$ are i.i.d. copies of W, and $\mathcal{I}$ is the solution map for Equation (1).

Remark: Here, T is fixed but $N \to \infty$.

The Nadaraya-Watson estimator

Consider the Nadaraya-Watson (NW) estimator

$$\widehat{b}_{\mathrm{N},h,\eta}(x) := \frac{\widehat{bf}_{\mathrm{N},h}(x)}{\widehat{f}_{\mathrm{N},\eta}(x)},$$

where

$$\widehat{f}_{\mathrm{N},\eta}(x) := \frac{1}{\mathrm{N}(\mathrm{T}-t_0)} \sum_{i=1}^{\mathrm{N}} \int_{t_0}^{\mathrm{T}} \mathrm{K}_\eta(\mathrm{X}_t^i - x)\mathrm{d}t,$$

K is a kernel, and

$$\widehat{bf}_{\mathrm{N},h}(x) := \frac{1}{\mathrm{N}(\mathrm{T}-t_0)} \sum_{i=1}^{\mathrm{N}} \int_{t_0}^{\mathrm{T}} \mathrm{K}_h(\mathrm{X}_t^i - x)\mathrm{d}\mathrm{X}_t^i.$$

Objectives

To establish a risk bound on the NW estimator and to provide adaptive procedures.

Related paper

Nadaraya-Watson Estimator for I.I.D. Paths of Diffusion Proc.

N. Marie and A. Rosier. Scandinavian Journal of Statistics, accepted, 2022.

Preliminaries

Why our NW estimator seems relevant?

On the one hand,

$$\mathbb{E}(\widehat{f}_{\mathrm{N},\eta}(x)) = \int_{-\infty}^{\infty} \mathrm{K}_\eta(y-x) \underbrace{\frac{1}{\mathrm{T}-t_0}\int_{t_0}^{\mathrm{T}} p_t(y)\mathrm{d}t}_{=:f(y)} \mathrm{d}y \xrightarrow[\eta\to 0]{} f(x).$$

On the other hand,

$$\mathbb{E}(\widehat{bf}_{\mathrm{N},h}(x)) = \int_{-\infty}^{\infty} \mathrm{K}_h(y-x)b(y)f(y)\mathrm{d}y \xrightarrow[h\to 0]{} b(x)f(x).$$

Conditions on the coefficients

Condition on b. The function b is Lipschitz continuous.

Conditions on σ:

- $\sigma \in C^1(\mathbb{R})$.
- σ is bounded.
- σ satisfies the nondegeneracy condition $|\sigma(.)| > \alpha > 0$.
- σ' is Hölder continuous.

A Nikol'skii type condition

By Menozzi et al. (2021), Theorem 1.2.(iv),

$$|p_t'(x)| \leqslant \frac{\mathfrak{c}}{t_0^q} \exp\left[-\mathfrak{m}\frac{(x-x_0)^2}{t}\right] \ ; \forall t \in [t_0, \mathrm{T}].$$

As a consequence, for every $\theta \in \mathbb{R}$,

$$\int_{-\infty}^{\infty} [f(x+\theta) - f(x)]^2 \mathrm{d}x \leqslant \mathfrak{c}(t_0)(\theta^2 + |\theta|^3) \qquad (2)$$

with $\mathfrak{c}(t_0)$ of order t_0^{-2q}.

Risk bounds

Risk bound on the denominator

Under usual assumptions on K,

$$\mathbb{E}(\|\widehat{f}_{\mathrm{N},\eta} - f\|_2^2) \leqslant \mathfrak{c}\eta^2 + \frac{\|\mathrm{K}\|^2}{\mathrm{N}\eta} \tag{3}$$

Remarks:

- The bias-variance tradeoff is reached by $\widehat{f}_{\mathrm{N},\eta}$ when η is of order $\mathrm{N}^{-1/3}$, leading to a rate of order $\mathrm{N}^{-2/3}$.
- This rate can be improved when b is bounded thanks to Kusuoka and Stroock (1985), Corollary 3.25.

Risk bound on the numerator

Under usual assumptions on K,

$$\mathbb{E}(\|\widehat{bf}_{\mathrm{N},h} - bf\|_2^2) \leqslant \|(bf)_h - bf\|_2^2 + \frac{\mathfrak{c}}{\mathrm{N}h} \tag{4}$$

with $(bf)_h = \mathrm{K}_h * (bf)$.

Risk bound on the (truncated) NW estimator

By Menozzi et al. (2021), Theorem 1.2.(i),

$$f(x) > m > 0 \ ; \forall x \in [\mathrm{A}, \mathrm{B}].$$

Then,

$$\int_{\mathrm{A}}^{\mathrm{B}} \mathbb{E}[|\widetilde{b}_{\mathrm{N},h,\eta}(x) - b(x)|^2] f(x) \mathrm{d}x \leqslant \frac{\mathfrak{c}}{m^2} \left(\|(bf)_h - bf\|_2^2 + \frac{1}{\mathrm{N}h} + \eta^2 + \frac{1}{\mathrm{N}\eta} \right)$$

where

$$\widetilde{b}_{\mathrm{N},h,\eta}(x) := \widehat{b}_{\mathrm{N},h,\eta}(x) \mathbf{1}_{\widehat{f}_{\mathrm{N},\eta}(x) > \frac{m}{2}}.$$

Bandwidth selection I: an extension of the PCO method

The PCO method: from density estimation to statistical inference for diffusion processes

- **Estimator Selection: a New Method with Applications to Kernel Density Estimation**
  C. Lacour, P. Massart and V. Rivoirard. Sankhya A 79, 298-335, 2017.
- **On a Nadaraya-Watson Estimator with Two Bandwidths**
  F. Comte and N. Marie. Electron. J. Statist. 15, 2566-2607, 2021.

The PCO criterion for the denominator

Let $\mathcal{H}_{\mathrm{N}}$ be a finite subset of $[\eta_0, 1]$, where $\eta_0 \geqslant \mathrm{N}^{-1}$.

Selection rule:

$$\widehat{\eta} \in \arg\min_{\eta \in \mathcal{H}_{\mathrm{N}}} \{\|\widehat{f}_{\mathrm{N},\eta} - \widehat{f}_{\mathrm{N},\eta_0}\|_2^2 + \mathrm{pen}(\eta)\},$$

where

$$\mathrm{pen}(\eta) := \frac{2}{(\mathrm{T}-t_0)^2\mathrm{N}^2}\sum_{i=1}^{\mathrm{N}}\left\langle \int_{t_0}^{\mathrm{T}} \mathrm{K}_\eta(\mathrm{X}_s^i - \cdot)\mathrm{d}s, \int_{t_0}^{\mathrm{T}} \mathrm{K}_{\eta_0}(\mathrm{X}_s^i - \cdot)\mathrm{d}s \right\rangle_2.$$

The PCO criterion for the numerator

Let $\mathfrak{H}_{\mathrm{N}}$ be a finite subset of $[h_0, 1]$, where $h_0 \geqslant \mathrm{N}^{-1/3}$.

Selection rule:

$$\widehat{h} \in \arg\min_{h \in \mathfrak{H}_{\mathrm{N}}} \{\|\widehat{bf}_{\mathrm{N},h} - \widehat{bf}_{\mathrm{N},h_0}\|_{2,\delta}^2 + \mathrm{pen}_\delta(h)\},$$

where

$$\mathrm{pen}_\delta(h) := \frac{2}{(\mathrm{T} - t_0)^2 \mathrm{N}^2} \times \sum_{i=1}^{\mathrm{N}} \left\langle \int_{t_0}^{\mathrm{T}} \mathrm{K}_h(\mathrm{X}_s^i - \cdot)\mathrm{d}\mathrm{X}_s^i, \int_{t_0}^{\mathrm{T}} \mathrm{K}_{h_0}(\mathrm{X}_s^i - \cdot)\mathrm{d}\mathrm{X}_s^i \right\rangle_{2,\delta},$$

δ is a kernel belonging to $\mathrm{C}_{\mathrm{b}}^1(\mathbb{R}; (0, \infty))$, and $\langle \varphi, \psi \rangle_{2,\delta} := \langle \varphi, \psi\delta \rangle_2$.

Risk bounds on the PCO adaptive estimators

On the one hand, with probability larger than $1 - \mathfrak{c}|\mathcal{H}_{\mathrm{N}}|e^{-\lambda}$,

$$\|\widehat{f}_{\mathrm{N},\widehat{\eta}} - f\|_2^2 \leqslant (1+\theta)\min_{\eta\in\mathcal{H}_{\mathrm{N}}}\|\widehat{f}_{\mathrm{N},\eta} - f\|_2^2 + \frac{\mathfrak{m}}{\theta}\left[\|f_{\eta_0} - f\|_2^2 + \frac{(1+\lambda)^3}{\mathrm{N}}\right].$$

On the other hand, with probability larger than $1 - \mathfrak{c}|\mathfrak{H}_{\mathrm{N}}|e^{-\lambda}$,

$$\|\widehat{bf}_{\mathrm{N},\widehat{h}} - bf\|_{2,\delta}^2 \leqslant (1+\theta)\min_{h\in\mathfrak{H}_{\mathrm{N}}}\|\widehat{bf}_{\mathrm{N},h} - bf\|_{2,\delta}^2 + \frac{\mathfrak{m}}{\theta}\left[\|(bf)_{h_0} - bf\|_{2,\delta}^2 + \frac{(1+\lambda)^3}{\mathrm{N}}\right].$$

Risk bound on the PCO adaptive NW estimator

Since $f(x) > m > 0$ for every $x \in [\mathrm{A}, \mathrm{B}]$,

$$\int_{\mathrm{A}}^{\mathrm{B}} \mathbb{E}[|\widetilde{b}_{\mathrm{N},\widehat{h},\widehat{\eta}}(x) - b(x)|^2] f(x)\mathrm{d}x$$

$$\leqslant \frac{\mathfrak{c}}{m^3}\left[(1+\theta)\min_{(\eta,h)\in\mathcal{H}_{\mathrm{N}}\times\mathfrak{H}_{\mathrm{N}}}\{\mathbb{E}(\|\widehat{bf}_{\mathrm{N},h} - bf\|_2^2) + \mathbb{E}(\|\widehat{f}_{\mathrm{N},\eta} - f\|_2^2)\}\right.$$

$$\left. + \frac{1}{\theta}\left(\|(bf)_{h_0} - bf\|_2^2 + \|f_{\eta_0} - f\|_2^2 + \frac{1}{\mathrm{N}}\right)\right].$$

Bandwidth selection II: an extension of the leave-one-out cross-validation (LOO-CV) method

A discrete-time NW estimator

For $h = \eta$, consider

$$\widehat{b}_{n,\mathrm{N},h}(x) := \sum_{i=1}^{\mathrm{N}} \sum_{j=0}^{n-1} \omega_j^i(x)(\mathrm{X}^i_{t_{j+1}} - \mathrm{X}^i_{t_j})$$

with

$$\omega_j^i(x) := \frac{\mathrm{K}_h(\mathrm{X}^i_{t_j} - x)}{\sum_{k=1}^{\mathrm{N}} \sum_{\ell=0}^{n-1} \mathrm{K}_h(\mathrm{X}^k_{t_\ell} - x)(t_{\ell+1} - t_\ell)}.$$

Clearly,

$$\sum_{i=1}^{\mathrm{N}} \sum_{j=0}^{n-1} \omega_j^i(x)(t_{j+1} - t_j) = 1.$$

The LOO-CV criterion

Selection rule:

$$\widehat{h} \in \arg\min_{h \in \mathcal{H}_{\mathrm{N}}} \mathrm{CV}(h)$$

with

$$\mathrm{CV}(h) := \sum_{i=1}^{\mathrm{N}} \left[\sum_{j=0}^{n-1} \widehat{b}_{n,\mathrm{N},h}^{-i}(\mathrm{X}_{t_j}^i)^2 (t_{j+1} - t_j) \right. \\ \left. -2 \sum_{j=0}^{n-1} \widehat{b}_{n,\mathrm{N},h}^{-i}(\mathrm{X}_{t_j}^i)(\mathrm{X}_{t_{j+1}}^i - \mathrm{X}_{t_j}^i) \right]$$

and

$$\widehat{b}_{n,\mathrm{N},h}^{-i}(x) := \sum_{k \in \{1,\dots,\mathrm{N}\} \setminus \{i\}} \sum_{j=0}^{n-1} \omega_j^k(x)(\mathrm{X}_{t_{j+1}}^k - \mathrm{X}_{t_j}^k).$$

Why is the LOO-CV criterion relevant?

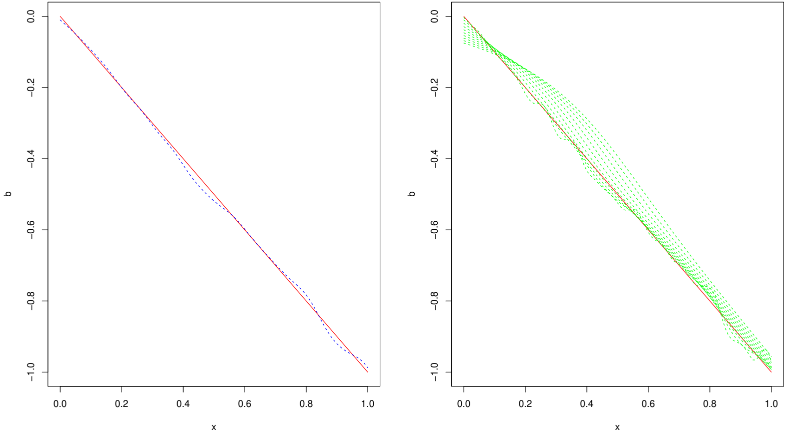
It's wrong... but assume that $\mathrm{dX}_t = \mathrm{Y}_t \mathrm{d}t$.

Then, a natural extension of the *usual* LOO-CV criterion is given by

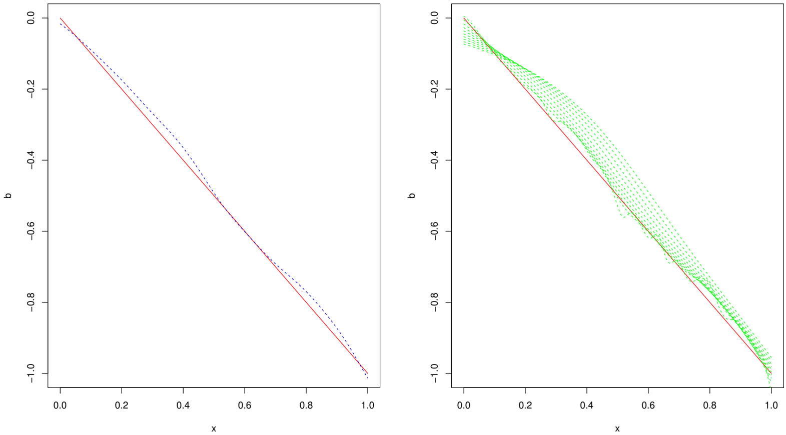
$$\begin{aligned}\mathrm{CV}^*(h) &:= \sum_{i=1}^{\mathrm{N}} \sum_{j=0}^{n-1} (\mathrm{Y}_{t_j}^i - \widehat{b}_{n,\mathrm{N},h}^{-i}(\mathrm{X}_{t_j}^i))^2 (t_{j+1} - t_j) \\ &\approx \mathrm{CV}(h) + \sum_{i=1}^{\mathrm{N}} \sum_{j=0}^{n-1} (\mathrm{Y}_{t_j}^i)^2 (t_{j+1} - t_j) \text{ for } n \text{ large enough.}\end{aligned}$$

Numerical experiments

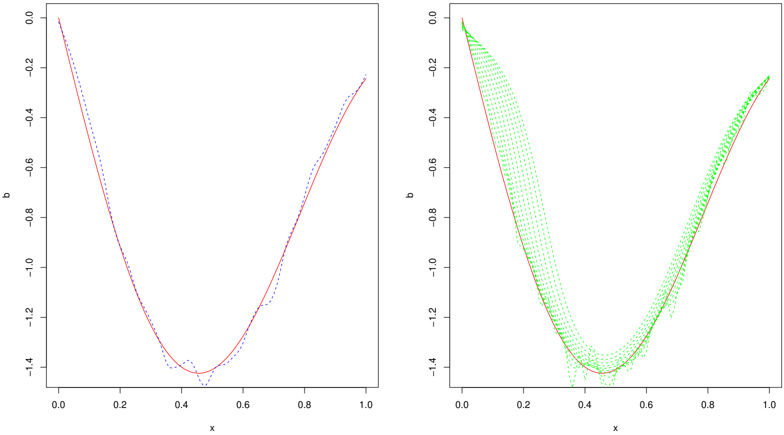
Model 1 (Langevin): $\mathrm{d}X_t = -X_t\mathrm{d}t + 0.1\mathrm{d}W_t$



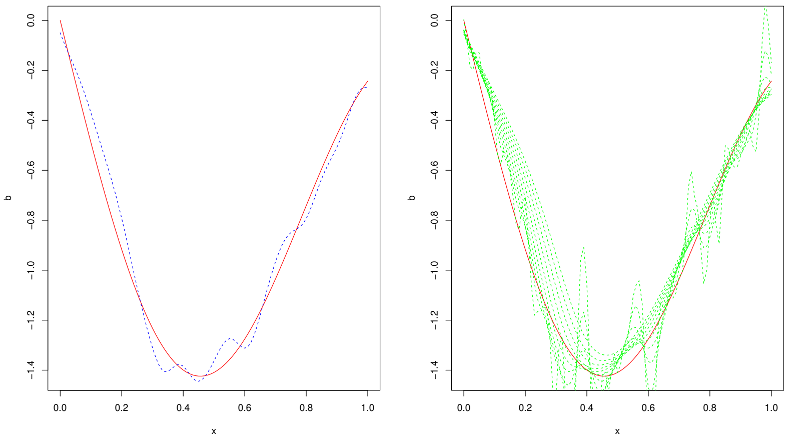
Model 2 (hyperbolic): $dX_t = -X_t dt + 0.1\sqrt{1 + X_t^2} dW_t$



Model 3: $dX_t = -(X_t + \sin(4X_t))dt + 0.1dW_t$



Model 4: $dX_t = -(X_t + \sin(4X_t))dt + 0.1(2 + \cos(X_t))dW_t$



Mean MISEs of 100 LOO-CV adaptive NW estimations compared to the oracle estimations

	LOO-CV	Oracle
Model 1	$3.03 \cdot 10^{-4}$	$2.67 \cdot 10^{-4}$
Model 2	$6.52 \cdot 10^{-4}$	$4.96 \cdot 10^{-4}$
Model 3	$2.45 \cdot 10^{-3}$	$1.99 \cdot 10^{-3}$
Model 4	$9.15 \cdot 10^{-3}$	$6.02 \cdot 10^{-3}$

Thank you for your attention!