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

$$X_t = x_0 + \int_0^t b(X_s)\,ds + \int_0^t \sigma(X_s)\,dW_s \ ; \ t \in [0, T]$$

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

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

where $W^1, \ldots, W^N$ 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

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

where

$$\hat{f}_{N,\eta}(x) := \frac{1}{N(T - t_0)} \sum_{i=1}^{N} \int_{t_0}^{T} K_\eta(X_t^i - x)dt,$$

$K$ is a kernel, and

$$\hat{bf}_{N,h}(x) := \frac{1}{N(T - t_0)} \sum_{i=1}^{N} \int_{t_0}^{T} K_h(X_t^i - x)dX_t^i.$$
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,

$$
\mathbb{E}(\hat{f}_N, \eta(x)) = \int_{-\infty}^{\infty} K_\eta(y - x) \frac{1}{T - t_0} \int_{t_0}^{T} p_t(y) dt dy \xrightarrow{\eta \to 0} f(x).
$$

On the other hand,

$$
\mathbb{E}(\hat{bf}_N, h(x)) = \int_{-\infty}^{\infty} K_h(y - x)b(y)f(y)dy \xrightarrow{h \to 0} b(x)f(x).
$$
Conditions on the coefficients

Condition on $b$. The function $b$ is Lipschitz continuous.

Conditions on $\sigma$:

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

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

\[ |p'_t(x)| \leq \frac{c}{t^q_0} \exp \left[ -m \frac{(x - x_0)^2}{t} \right] ; \forall t \in [t_0, T]. \]

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

\[ \int_{-\infty}^{\infty} [f(x + \theta) - f(x)]^2 dx \leq c(t_0)(\theta^2 + |\theta|^3) \quad (2) \]

with \( c(t_0) \) of order \( t_0^{-2q} \).
| Introduction | Preliminaries | Risk bounds | The PCO method | Cross-validation | Numerical experiments |
|--------------|---------------|-------------|----------------|-----------------|----------------------|
|              |               |             |                |                 |                      |

**Risk bounds**
Risk bound on the denominator

Under usual assumptions on $K$,

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

Remarks:

- The bias-variance tradeoff is reached by $\hat{f}_{N,\eta}$ when $\eta$ is of order $N^{-1/3}$, leading to a rate of order $N^{-2/3}$.
- 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 $K$,

$$
\mathbb{E}(\|\widehat{bf}_{N,h} - bf\|^2_2) \leq \|(bf)_h - bf\|^2_2 + \frac{c}{Nh} \quad (4)
$$

with $(bf)_h = K_h \ast (bf)$. 
Risk bound on the (truncated) NW estimator

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

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

Then,

\[
\int_{A}^{B} \mathbb{E}[\tilde{b}_{N,h,\eta}(x) - b(x)]^2 f(x) dx \\
\leq \frac{c}{m^2} \left( \| (bf)_h - bf \|_2^2 + \frac{1}{Nh} + \eta^2 + \frac{1}{Nh\eta} \right)
\]

where

\[ \tilde{b}_{N,h,\eta}(x) := \hat{b}_{N,h,\eta}(x) \mathbf{1}_{\hat{f}_{N,\eta}(x) > \frac{m}{2}}. \]
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 $\mathcal{H}_N$ be a finite subset of $[\eta_0, 1]$, where $\eta_0 \geq N^{-1}$.

Selection rule:

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

where

$$\text{pen}(\eta) := \frac{2}{(T - t_0)^2 N^2} \sum_{i=1}^{N} \left\langle \int_{t_0}^{T} K_\eta(X_s^i - \cdot) ds, \int_{t_0}^{T} K_{\eta_0}(X_s^i - \cdot) ds \right\rangle_2.$$
The PCO criterion for the numerator

Let $\mathcal{H}_N$ be a finite subset of $[h_0, 1]$, where $h_0 \geq N^{-1/3}$.

Selection rule:

$$\hat{h} \in \arg \min_{h \in \mathcal{H}_N} \{ \| \hat{b}f_{N,h} - \hat{b}f_{N,h_0} \|^2_{2,\delta} + \text{pen}_\delta(h) \},$$

where

$$\text{pen}_\delta(h) := \frac{2}{(T - t_0)^2 N^2} \times \sum_{i=1}^N \left\langle \int_{t_0}^T K_h(X^i_s - \cdot) dX^i_s, \int_{t_0}^T K_{h_0}(X^i_s - \cdot) dX^i_s \right\rangle_{2,\delta},$$

$\delta$ is a kernel belonging to $C^1_b(\mathbb{R}; (0, \infty))$, 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 $1 - c|\mathcal{H}_N|e^{-\lambda}$,

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

On the other hand, with probability larger than $1 - c|\mathcal{H}_N|e^{-\lambda}$,

$$
\|\hat{b}f_{N,h} - bf\|_{2,\delta}^2 \leq (1 + \theta) \min_{h \in \mathcal{H}_N} \|\hat{b}f_{N,h} - bf\|_{2,\delta}^2 \\
+ \frac{m}{\theta} \left[ \|(bf)_{h_0} - bf\|_{2,\delta}^2 + \frac{(1 + \lambda)^3}{N} \right].
$$
Risk bound on the PCO adaptive NW estimator

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

\[
\int_A^B \mathbb{E}[|\tilde{b}_{N,h,\hat{\eta}}(x) - b(x)|^2] f(x)dx \\
\leq \frac{c}{m^3} \left( (1 + \theta) \min_{(\eta, h) \in \mathcal{H}_N \times \mathcal{S}_N} \{ \mathbb{E}(\|\hat{b}_N - b\|_2^2) + \mathbb{E}(\|\hat{f}_N - f\|_2^2) \} \\
+ \frac{1}{\theta} \left( \|b_N - b\|_2^2 + \|f_N - f\|_2^2 + \frac{1}{N} \right) \right).
\]
Bandwidth selection II: an extension of the leave-one-out cross-validation (LOO-CV) method
A discrete-time NW estimator

For $h = \eta$, consider

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

with

$$\omega^i_j(x) := \frac{K_h(X^i_{t_j} - x)}{\sum_{k=1}^{N} \sum_{\ell=0}^{n-1} K_h(X^k_{t_\ell} - x)(t_{\ell+1} - t_\ell)}.$$ 

Clearly,

$$\sum_{i=1}^{N} \sum_{j=0}^{n-1} \omega^i_j(x)(t_{j+1} - t_j) = 1.$$
The LOO-CV criterion

Selection rule:

\[ \hat{h} \in \arg \min_{h \in \mathcal{H}_N} CV(h) \]

with

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

\[
-2 \sum_{j=0}^{n-1} \hat{b}_{n,N,h}^{-i}(X_{t_j}^i)(X_{t_{j+1}}^i - X_{t_j}^i) \]

and

\[
\hat{b}_{n,N,h}^{-i}(x) := \sum_{k \in \{1,...,N\} \setminus \{i\}} \sum_{j=0}^{n-1} \omega_{j}^{k}(x)(X_{t_{j+1}}^{k} - X_{t_{j}}^{k}).
\]
Why is the LOO-CV criterion relevant?

It’s wrong... but assume that $\text{d}X_t = Y_t \text{d}t$.

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

$$
CV^*(h) := \sum_{i=1}^{N} \sum_{j=0}^{n-1} (Y_{t,j}^i - \hat{b}_{n,N,h}^{-i}(X_{t,j}^i))^2 (t_{j+1} - t_j)
$$

$$
\approx CV(h) + \sum_{i=1}^{N} \sum_{j=0}^{n-1} (Y_{t,j}^i)^2 (t_{j+1} - t_j) \text{ for } n \text{ large enough.}
$$
Numerical experiments
Model 1 (Langevin): \( dX_t = -X_t dt + 0.1 dW_t \)
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

| Model  | 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!