A pre-processing technique for asymptotically correct drift estimation in multiscale diffusion processes

Assyr Abdulle [1] Giacomo Garegnani [1] Grigorios A. Pavliotis [2]
Andrew M. Stuart [3] Andrea Zanoni [1]

École polytechnique fédérale de Lausanne

Imperial College London – 7 February 2020

[1]Institute of Mathematics, EPFL
[2]Department of Mathematics, Imperial College London
[3]Department of Computing and Mathematical Sciences, Caltech

Setting - Homogenization

Multiscale SDE

$$\mathrm{d}X_t^\varepsilon = -\alpha \cdot V'(X_t^\varepsilon)\,\mathrm{d}t - \frac{1}{\varepsilon} p'\left(\frac{X_t^\varepsilon}{\varepsilon}\right)\mathrm{d}t + \sqrt{2\sigma}\,\mathrm{d}W_t.$$

Parameters:

- drift coefficient $\alpha \in \mathbb{R}^N$
- slow potential $V\colon \mathbb{R} \to \mathbb{R}^N$, $V(x) = (V_1(x), V_2(x), \dots, V_N(x))^\top$,
- fast potential $p\colon \mathbb{R} \to \mathbb{R}$, L-periodic
- diffusion coefficient $\sigma > 0$
- multiscale parameter $\varepsilon > 0$
- standard one-dimensional BM W_t

Setting - Homogenization

$$\mathrm{d}X_t^\varepsilon = -\alpha \cdot V'(X_t^\varepsilon)\,\mathrm{d}t - \frac{1}{\varepsilon}p'\left(\frac{X_t^\varepsilon}{\varepsilon}\right)\,\mathrm{d}t + \sqrt{2\sigma}\,\mathrm{d}W_t.$$

Homogenization theory: $X_t^\varepsilon \to X_t$ in law for $\varepsilon \to 0$ and

$$\mathrm{d}X_t = -A \cdot V'(X_t)\,\mathrm{d}t + \sqrt{2\Sigma}\,\mathrm{d}W_t,$$

with $A = K\alpha$, $\Sigma = K\sigma$ and

$$K = \int_0^L \left(1 + \Phi'(y)\right)^2 \mu(\mathrm{d}y), \quad \mu(\mathrm{d}y) = \frac{1}{Z}e^{-p(y)/\sigma}\,\mathrm{d}y,$$

with Φ solution of

$$-p'(y)\Phi'(y) + \sigma\Phi''(y) = p'(y), \quad 0 \le y \le L.$$

Setting - Parameter inference

$$\mathrm{d}X_t^\varepsilon = -\alpha \cdot V'(X_t^\varepsilon)\,\mathrm{d}t - \frac{1}{\varepsilon} p'\left(\frac{X_t^\varepsilon}{\varepsilon}\right)\mathrm{d}t + \sqrt{2\sigma}\,\mathrm{d}W_t \quad \rightarrow \text{data}$$

$$\mathrm{d}X_t = -A \cdot V'(X_t)\,\mathrm{d}t + \sqrt{2\Sigma}\,\mathrm{d}W_t \quad \rightarrow \text{model}$$

Goal: Estimate $A \in \mathbb{R}^N$ from observations $X^\varepsilon = (X_t^\varepsilon, 0 \le t \le T)$

Setting - Parameter inference

$$\mathrm{d}X_t^\varepsilon = -\alpha \cdot V'(X_t^\varepsilon)\,\mathrm{d}t - \frac{1}{\varepsilon}p'\left(\frac{X_t^\varepsilon}{\varepsilon}\right)\mathrm{d}t + \sqrt{2\sigma}\,\mathrm{d}W_t \quad \to \text{data}$$

$$\mathrm{d}X_t = -A \cdot V'(X_t)\,\mathrm{d}t + \sqrt{2\Sigma}\,\mathrm{d}W_t \quad \to \text{model}$$

Goal: Estimate $A \in \mathbb{R}^N$ from observations $X^\varepsilon = (X_t^\varepsilon, 0 \leq t \leq T)$

What we know:

- slow potential V (i.e., V')

What we ignore:

- the rest

Setting - Parameter inference

$$\mathrm{d}X_t^\varepsilon = -\alpha \cdot V'(X_t^\varepsilon)\,\mathrm{d}t - \frac{1}{\varepsilon} p'\left(\frac{X_t^\varepsilon}{\varepsilon}\right)\mathrm{d}t + \sqrt{2\sigma}\,\mathrm{d}W_t \quad \to \text{data}$$

$$\mathrm{d}X_t = -A \cdot V'(X_t)\,\mathrm{d}t + \sqrt{2\Sigma}\,\mathrm{d}W_t \quad \to \text{model}$$

Goal: Estimate $A \in \mathbb{R}^N$ from observations $X^\varepsilon = (X_t^\varepsilon, 0 \leq t \leq T)$

Idea: Maximize likelihood function (Girsanov)

$$L(X^\varepsilon \mid A) = \exp\left(-\frac{I(X^\varepsilon \mid A)}{2\Sigma}\right),$$

$$I(X^\varepsilon \mid A) = \int_0^T A \cdot V'(X_t^\varepsilon)\,\mathrm{d}X_t^\varepsilon + \frac{1}{2}\int_0^T \left(A \cdot V'(X_t^\varepsilon)\right)^2 \mathrm{d}t.$$

Setting - Parameter inference

$$\mathrm{d}X_t^\varepsilon = -\alpha \cdot V'(X_t^\varepsilon)\,\mathrm{d}t - \frac{1}{\varepsilon} p'\left(\frac{X_t^\varepsilon}{\varepsilon}\right)\mathrm{d}t + \sqrt{2\sigma}\,\mathrm{d}W_t \qquad \to \text{data}$$

$$\mathrm{d}X_t = -A \cdot V'(X_t)\,\mathrm{d}t + \sqrt{2\Sigma}\,\mathrm{d}W_t \qquad \to \text{model}$$

Goal: Estimate $A \in \mathbb{R}^N$ from observations $X^\varepsilon = (X_t^\varepsilon, 0 \leq t \leq T)$

Idea: Maximize likelihood function (Girsanov)

$$I(X^\varepsilon \mid A) = \int_0^T A \cdot V'(X_t^\varepsilon)\,\mathrm{d}X_t^\varepsilon + \frac{1}{2}\int_0^T \left(A \cdot V'(X_t^\varepsilon)\right)^2 \mathrm{d}t,$$

Estimator:

$$\widehat{A}^\varepsilon(T) = \arg\min_{A \in \mathbb{R}^N} I(X^\varepsilon \mid A),$$

Setting - Parameter inference

$$\mathrm{d}X_t^\varepsilon = -\alpha \cdot V'(X_t^\varepsilon)\,\mathrm{d}t - \frac{1}{\varepsilon} p'\left(\frac{X_t^\varepsilon}{\varepsilon}\right)\mathrm{d}t + \sqrt{2\sigma}\,\mathrm{d}W_t \qquad \rightarrow \text{data}$$

$$\mathrm{d}X_t = -A \cdot V'(X_t)\,\mathrm{d}t + \sqrt{2\Sigma}\,\mathrm{d}W_t \qquad \rightarrow \text{model}$$

Goal: Estimate $A \in \mathbb{R}^N$ from observations $X^\varepsilon = (X_t^\varepsilon, 0 \leq t \leq T)$

Idea: Maximize likelihood function (Girsanov)

$$I(X^\varepsilon \mid A) = \int_0^T A \cdot V'(X_t^\varepsilon)\,\mathrm{d}X_t^\varepsilon + \frac{1}{2}\int_0^T \left(A \cdot V'(X_t^\varepsilon)\right)^2 \mathrm{d}t,$$

Estimator:

$$\widehat{A}^\varepsilon(T) = \left(\int_0^T V'(X_t^\varepsilon) \otimes V'(X_t^\varepsilon)\,\mathrm{d}t\right)^{-1} \int_0^T V'(X_t^\varepsilon)\,\mathrm{d}X_t^\varepsilon,$$

Setting - Parameter inference

$$\mathrm{d}X_t^\varepsilon = -\alpha \cdot V'(X_t^\varepsilon)\,\mathrm{d}t - \frac{1}{\varepsilon}p'\left(\frac{X_t^\varepsilon}{\varepsilon}\right)\mathrm{d}t + \sqrt{2\sigma}\,\mathrm{d}W_t \qquad \to \text{data}$$

$$\mathrm{d}X_t = -A \cdot V'(X_t)\,\mathrm{d}t + \sqrt{2\Sigma}\,\mathrm{d}W_t \qquad \to \text{model}$$

Goal: Estimate $A \in \mathbb{R}^N$ from observations $X^\varepsilon = (X_t^\varepsilon, 0 \leq t \leq T)$

Idea: Maximize likelihood function (Girsanov)

$$I(X^\varepsilon \mid A) = \int_0^T AV'(X_t^\varepsilon)\,\mathrm{d}X_t^\varepsilon + \frac{1}{2}\int_0^T \left(AV'(X_t^\varepsilon)\right)^2 \mathrm{d}t,$$

Estimator:

$$\widehat{A}^\varepsilon(T) = -\frac{\int_0^T V'(X_t^\varepsilon)\,\mathrm{d}X_t^\varepsilon}{\int_0^T V'(X_t^\varepsilon)^2\,\mathrm{d}t}, \qquad (N = 1).$$

Setting - Parameter inference

$$\mathrm{d}X_t^\varepsilon = -\alpha \cdot V'(X_t^\varepsilon)\,\mathrm{d}t - \frac{1}{\varepsilon}p'\left(\frac{X_t^\varepsilon}{\varepsilon}\right)\mathrm{d}t + \sqrt{2\sigma}\,\mathrm{d}W_t \quad \to \text{data}$$

$$\mathrm{d}X_t = -A \cdot V'(X_t)\,\mathrm{d}t + \sqrt{2\Sigma}\,\mathrm{d}W_t \quad \to \text{model}$$

Goal: Estimate $A \in \mathbb{R}^N$ from observations $X^\varepsilon = (X_t^\varepsilon, 0 \leq t \leq T)$

Estimator:

$$\widehat{A}^\varepsilon(T) = -\frac{\int_0^T V'(X_t^\varepsilon)\,\mathrm{d}X_t^\varepsilon}{\int_0^T V'(X_t^\varepsilon)^2\,\mathrm{d}t},$$

Homogenization: We have $X_t^\varepsilon \to X_t$ for $\varepsilon \to 0$

$$\implies \lim_{\varepsilon\to 0}\lim_{T\to\infty}\widehat{A}^\varepsilon(T) = A, \qquad \text{Wrong!}$$

Setting - Parameter inference

$$\mathrm{d}X_t^\varepsilon = -\alpha \cdot V'(X_t^\varepsilon)\,\mathrm{d}t - \frac{1}{\varepsilon} p'\left(\frac{X_t^\varepsilon}{\varepsilon}\right)\mathrm{d}t + \sqrt{2\sigma}\,\mathrm{d}W_t \qquad \to \text{data}$$

$$\mathrm{d}X_t = -A \cdot V'(X_t)\,\mathrm{d}t + \sqrt{2\Sigma}\,\mathrm{d}W_t \qquad \to \text{model}$$

Goal: Estimate $A \in \mathbb{R}^N$ from observations $X^\varepsilon = (X_t^\varepsilon, 0 \leq t \leq T)$

Estimator:

$$\widehat{A}^\varepsilon(T) = -\frac{\int_0^T V'(X_t^\varepsilon)\,\mathrm{d}X_t^\varepsilon}{\int_0^T V'(X_t^\varepsilon)^2\,\mathrm{d}t},$$

Homogenization: We have $X_t^\varepsilon \to X_t$ for $\varepsilon \to 0$ [1]

$$\implies \lim_{\varepsilon\to 0}\lim_{T\to\infty} \widehat{A}^\varepsilon(T) = \alpha, \qquad \text{Problem!}$$

[1] Pavliotis and Stuart (2007)

Solutions – Literature

- Subsample the data
  - Theory: Pavliotis and Stuart (2007); Papavasiliou et al. (2009)
  - Practice: Cotter and Pavliotis (2009) (oceanography) Zhang et al. (2005); Olhede et al. (2010); Aït-Sahalia and Jacod (2014) (econometrics)
- Martingale property-based
  - Theory: Kalliadasis et al. (2015); Krumscheid et al. (2013)
  - Practice: Krumscheid et al. (2015)
- Nonparameteric / Bayesian: Pokern et al. (2009, 2013) (single scale)
- ...

Subsampling the data

Estimator:

$$\widehat{A}^{\varepsilon}(T) = -\frac{\int_0^T V'(X_t^\varepsilon)\,\mathrm{d}X_t^\varepsilon}{\int_0^T V'(X_t^\varepsilon)^2\,\mathrm{d}t}.$$

Problem: $\lim_{\varepsilon\to 0}\lim_{T\to\infty}\widehat{A}^{\varepsilon}(T) \to \alpha$.

Solution[1]: Subsample the data with step δ and compute

$$\widehat{A}_\delta^{\varepsilon}(T) = -\frac{\sum_{i=0}^{N-1} V'(X_{i\delta}^\varepsilon)\left(X_{(i+1)\delta}^\varepsilon - X_{i\delta}^\varepsilon\right)}{\delta\sum_{i=0}^{N-1} V'(X_{i\delta}^\varepsilon)^2}, \qquad N\delta = T.$$

[1] Pavliotis and Stuart (2007)

Subsampling the data

Solution[1]: Subsample the data with step δ and compute

$$\widehat{A}^{\varepsilon}_{\delta}(T) = -\frac{\sum_{i=0}^{N-1} V'(X^{\varepsilon}_{i\delta})\left(X^{\varepsilon}_{(i+1)\delta} - X^{\varepsilon}_{i\delta}\right)}{\delta \sum_{i=0}^{N-1} V'(X^{\varepsilon}_{i\delta})^2}, \qquad N\delta = T.$$

Theorem

If $\delta = \varepsilon^{\zeta}$, $0 < \zeta < 1$ and $N = \lceil \varepsilon^{-\gamma} \rceil$ with $\gamma > \zeta$, then

$$\lim_{\varepsilon \to 0} \widehat{A}^{\varepsilon}_{\delta}(T) = A, \quad \textit{in probability.}$$

Issue: How do we choose $\zeta \in (0,1)$?

[1] Pavliotis and Stuart (2007)

Subsampling the data

Solution[1]: Subsample the data with step δ and compute

$$\widehat{A}_\delta^\varepsilon(T) = -\frac{\sum_{i=0}^{N-1} V'(X_{i\delta}^\varepsilon)\left(X_{(i+1)\delta}^\varepsilon - X_{i\delta}^\varepsilon\right)}{\delta \sum_{i=0}^{N-1} V'(X_{i\delta}^\varepsilon)^2}, \qquad N\delta = T.$$

Theorem

If $\delta = \varepsilon^\zeta$, $0 < \zeta < 1$ *and* $N = \lceil \varepsilon^{-\gamma} \rceil$ *with* $\gamma > \zeta$, *then*

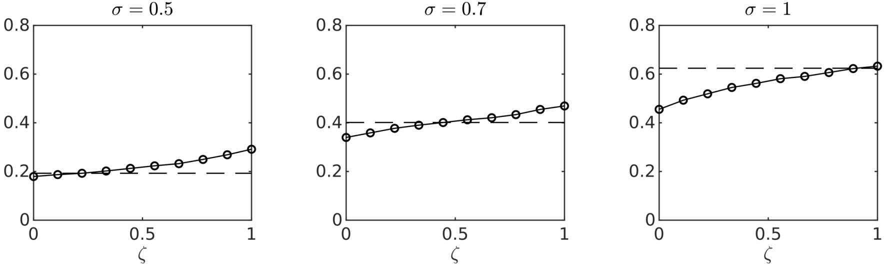
$$\lim_{\varepsilon \to 0} \widehat{A}_\delta^\varepsilon(T) = A, \quad \textit{in probability}.$$

Issue: Take $\varepsilon = 0.1$, data $\tilde{\delta} = 0.01$, $\delta = \sqrt{\varepsilon} \approx 0.3 \implies 97\%$ "garbage"

[1]Pavliotis and Stuart (2007)

Subsampling the data

Experiment: Estimate A for $V(x) = x^2/2$ (Ornstein–Uhlenbeck) with subsampling varying $\delta = \varepsilon^{\zeta}$ ($\varepsilon = 0.1$, $T = 10^3$, $p(y) = \cos(y)$)



Issue: How do we choose $\zeta \in (0, 1)$?

Filtering the data[1]

Idea: Consider

$$Z_t^\varepsilon = \int_0^t k(t,s) X_s^\varepsilon \,\mathrm{d}s,$$

where $\delta, \beta > 0$ and

$$k(t,s) = C_\beta \delta^{-1/\beta} e^{-\frac{1}{\delta}(t-s)^\beta}, \qquad C_\beta = \beta\, \Gamma(1/\beta)^{-1}.$$

Why? Subsampling data is a "smoothing" process, so why not directly smoothing the data?

[1] Abdulle et al. (2020)

Filtering the data[1]

Idea: Consider

$$Z_t^\varepsilon = \int_0^t k(t,s) X_s^\varepsilon \, \mathrm{d}s,$$

where $\delta > 0$ and $(\beta = 1)$

$$k(t,s) = \delta^{-1} e^{-\frac{1}{\delta}(t-s)}.$$

We can write

$$\mathrm{d}Z_t^\varepsilon = k(t,t) X_t^\varepsilon \, \mathrm{d}t + \int_0^t \partial_t k(t,s) X_s^\varepsilon \, \mathrm{d}s \, \mathrm{d}t = \frac{1}{\delta} \left(X_t^\varepsilon - Z_t^\varepsilon \right) \mathrm{d}t.$$

[1] Abdulle et al. (2020)

Filtering the data[1]

Idea: Consider

$$Z_t^\varepsilon = \int_0^t k(t,s) X_s^\varepsilon \, \mathrm{d}s,$$

where $\delta > 0$ and $(\beta = 1)$

$$k(t,s) = \delta^{-1} e^{-\frac{1}{\delta}(t-s)}.$$

We can write

$$\mathrm{d}X_t^\varepsilon = -\alpha \cdot V'(X_t^\varepsilon)\,\mathrm{d}t - \frac{1}{\varepsilon} p'\left(\frac{X_t^\varepsilon}{\varepsilon}\right)\mathrm{d}t + \sqrt{2\sigma}\,\mathrm{d}W_t,$$

$$\mathrm{d}Z_t^\varepsilon = \frac{1}{\delta}\left(X_t^\varepsilon - Z_t^\varepsilon\right)\mathrm{d}t. \qquad \to \text{System of coupled SDEs!}$$

[1] Abdulle et al. (2020)

Ergodic properties of the filter

Now: Something we can work on

$$\mathrm{d}X_t^\varepsilon = -\alpha \cdot V'(X_t^\varepsilon)\,\mathrm{d}t - \frac{1}{\varepsilon}p'\left(\frac{X_t^\varepsilon}{\varepsilon}\right)\mathrm{d}t + \sqrt{2\sigma}\,\mathrm{d}W_t,$$
$$\mathrm{d}Z_t^\varepsilon = \frac{1}{\delta}\left(X_t^\varepsilon - Z_t^\varepsilon\right)\mathrm{d}t.$$

We have $(X_t^\varepsilon, Z_t^\varepsilon)^\top$ geometrically ergodic with smooth invariant density.
Invariant measure $\mu^\varepsilon(\mathrm{d}x, \mathrm{d}z) = \rho^\varepsilon(x,z)\,\mathrm{d}x\,\mathrm{d}z$ satisfies stationary FP

$$\sigma\partial_{xx}^2\rho^\varepsilon(x,z) + \partial_x\left(\left(\alpha \cdot V'(x) + \frac{1}{\varepsilon}p'\left(\frac{x}{\varepsilon}\right)\right)\rho^\varepsilon(x,z)\right) + \frac{1}{\delta}\partial_z\left((z-x)\rho^\varepsilon(x,z)\right) = 0.$$

Ergodic properties of the filter

Lemma

Let us write $\rho^\varepsilon(x,z) = \varphi^\varepsilon(x)\psi^\varepsilon(z)R^\varepsilon(x,z)$, *with* φ^ε, ψ^ε *marginal densities wrt* x *and* z. *Then, it holds*

$$\varphi^\varepsilon(x) = \frac{1}{C_{\varphi^\varepsilon}} \exp\left(-\frac{1}{\sigma}\alpha \cdot V(x) - \frac{1}{\sigma} p\left(\frac{x}{\varepsilon}\right)\right),$$

Moreover, the "magic equality" holds

$$\sigma\delta \int_{\mathbb{R}} \int_{\mathbb{R}} V'(z)\varphi^\varepsilon(x)\psi^\varepsilon(z)\partial_x R^\varepsilon(x,z)\,\mathrm{d}x\,\mathrm{d}z = \mathbb{E}^{\rho^\varepsilon}[((X^\varepsilon)^2 - (Z^\varepsilon)^2)V''(Z^\varepsilon)].$$

Remark: φ^ε = invariant measure of X^ε alone.

Ergodic properties of the filter

Question: What happens in the limit $\varepsilon \to 0$?

Lemma

The measure μ^ε converges weakly to $\mu^0(\mathrm{d}x, \mathrm{d}z) = \rho^0(x,z)\,\mathrm{d}x\,\mathrm{d}z$ satisfying

$$\Sigma \partial_{xx}^2 \rho^0(x,z) + \partial_x \left(A \cdot V'(x) \rho^0(x,z) \right) + \frac{1}{\delta} \partial_z \left((z-x) \rho^0(x,z) \right) = 0,$$

where A and Σ coefficients of the homogenized equation.

Back to parameter estimation

Estimator:

$$\widehat{A}^{\varepsilon}(T) = -\frac{\int_0^T V'(X_t^{\varepsilon})\,\mathrm{d}X_t^{\varepsilon}}{\int_0^T V'(X_t^{\varepsilon})^2\,\mathrm{d}t}.$$

Idea: Replace X_t^{ε} with Z_t^{ε} (but not everywhere)

$$\widehat{A}_Z^{\varepsilon}(T) = -\frac{\int_0^T V'(Z_t^{\varepsilon})\,\mathrm{d}X_t^{\varepsilon}}{\int_0^T V'(Z_t^{\varepsilon})V'(X_t^{\varepsilon})\,\mathrm{d}t}.$$

Remark: Denominator $\neq 0$ a.s. $\implies$ $\widehat{A}_Z^{\varepsilon}(T)$ well-defined.

Back to parameter estimation

Estimator:

$$\widehat{A}^{\varepsilon}(T) = -\frac{\int_0^T V'(X_t^{\varepsilon})\,\mathrm{d}X_t^{\varepsilon}}{\int_0^T V'(X_t^{\varepsilon})^2\,\mathrm{d}t}.$$

Idea: Replace X_t^{ε} with Z_t^{ε} (but not everywhere)

$$\widehat{A}_Z^{\varepsilon}(T) = -\frac{\int_0^T V'(Z_t^{\varepsilon})\,\mathrm{d}X_t^{\varepsilon}}{\int_0^T V'(Z_t^{\varepsilon})V'(X_t^{\varepsilon})\,\mathrm{d}t}.$$

Theorem

$$\lim_{\varepsilon\to 0}\lim_{T\to\infty}\widehat{A}_Z^{\varepsilon}(T) = A, \quad a.s.,$$

where A drift coefficient of homogenized equation.

Remark: True also for N parameters and d-dimensional SDE.

Back to parameter estimation

Result:

$$\lim_{\varepsilon\to 0}\lim_{T\to\infty}\widehat{A}^{\varepsilon}(T) = -\lim_{\varepsilon\to 0}\lim_{T\to\infty}\frac{\int_0^T V'(Z_t^{\varepsilon})\,\mathrm{d}X_t^{\varepsilon}}{\int_0^T V'(Z_t^{\varepsilon})V'(X_t^{\varepsilon})\,\mathrm{d}t} = A, \quad \text{a.s.}$$

Proof steps:

Replace $\mathrm{d}X_t^{\varepsilon}$ in the estimator

$$\begin{aligned}
-\frac{\int_0^T V'(X_t^{\varepsilon})\,\mathrm{d}X_t^{\varepsilon}}{\int_0^T V'(Z_t^{\varepsilon})V'(X_t^{\varepsilon})\,\mathrm{d}t} = \alpha\frac{\int_0^T V'(Z_t^{\varepsilon})V'(X_t^{\varepsilon})\,\mathrm{d}t}{\int_0^T V'(Z_t^{\varepsilon})V'(X_t^{\varepsilon})\,\mathrm{d}t} &\qquad = \alpha \\
+\frac{1}{\varepsilon}\frac{\int_0^T V'(Z_t^{\varepsilon})p'\left(\frac{X_t^{\varepsilon}}{\varepsilon}\right)\,\mathrm{d}t}{\int_0^T V'(Z_t^{\varepsilon})V'(X_t^{\varepsilon})\,\mathrm{d}t} &\qquad =: I_1^{\varepsilon}(T) \\
-\sqrt{2\sigma}\frac{\int_0^T V'(Z_t^{\varepsilon})\,\mathrm{d}W_t}{\int_0^T V'(Z_t^{\varepsilon})V'(X_t^{\varepsilon})\,\mathrm{d}t} &\qquad =: I_2^{\varepsilon}(T)
\end{aligned}$$

Back to parameter estimation

Result:

$$\lim_{\varepsilon \to 0} \lim_{T \to \infty} \widehat{A}^{\varepsilon}(T) = -\lim_{\varepsilon \to 0} \lim_{T \to \infty} \frac{\int_0^T V'(Z_t^{\varepsilon}) \, \mathrm{d}X_t^{\varepsilon}}{\int_0^T V'(Z_t^{\varepsilon}) V'(X_t^{\varepsilon}) \, \mathrm{d}t} = A, \quad \text{a.s.}$$

Proof steps:

Take care of the first remainder

$$\lim_{\varepsilon \to 0} \lim_{T \to \infty} I_2^{\varepsilon}(T) = \lim_{\varepsilon \to 0} \lim_{T \to \infty} \sqrt{2\sigma} \frac{\int_0^T V'(Z_t^{\varepsilon}) \, \mathrm{d}W_t}{\int_0^T V'(Z_t^{\varepsilon}) V'(X_t^{\varepsilon}) \, \mathrm{d}t} = 0, \quad \text{a.s.},$$

due to the strong LLN for (continuous) martingales.

Back to parameter estimation

Result:

$$\lim_{\varepsilon\to 0}\lim_{T\to\infty}\widehat{A}^{\varepsilon}(T) = -\lim_{\varepsilon\to 0}\lim_{T\to\infty}\frac{\int_0^T V'(Z_t^{\varepsilon})\,\mathrm{d}X_t^{\varepsilon}}{\int_0^T V'(Z_t^{\varepsilon})V'(X_t^{\varepsilon})\,\mathrm{d}t} = A, \quad \text{a.s.}$$

Proof steps:

Take care of the second "remainder"

$$\lim_{\varepsilon\to 0}\lim_{T\to\infty} I_1^{\varepsilon}(T) = \lim_{\varepsilon\to 0}\lim_{T\to\infty}\frac{1}{\varepsilon}\frac{\int_0^T V'(Z_t^{\varepsilon})p'\left(\frac{X_t^{\varepsilon}}{\varepsilon}\right)\,\mathrm{d}t}{\int_0^T V'(Z_t^{\varepsilon})V'(X_t^{\varepsilon})\,\mathrm{d}t} = A-\alpha, \quad \text{a.s.,}$$

applying

1. ergodic theorem for $T \to \infty$ and "magic equality"
2. convergence $\mu^{\varepsilon} \to \mu^0$ for $\varepsilon \to 0$ and "magic equality" again

Numerical experiments

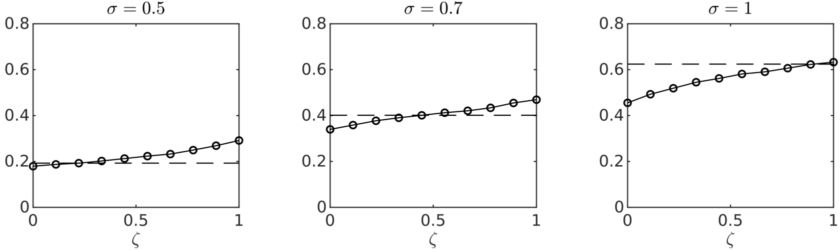
Reminder: Filter had a parameter β:

$$k(t,s) = C_\beta \delta^{-1/\beta} e^{-\frac{1}{\delta}(t-s)^\beta}, \qquad C_\beta = \beta\, \Gamma(1/\beta)^{-1}.$$

We did analysis for $\beta = 1$ but show experiments for larger values of β, too.

Numerical experiments

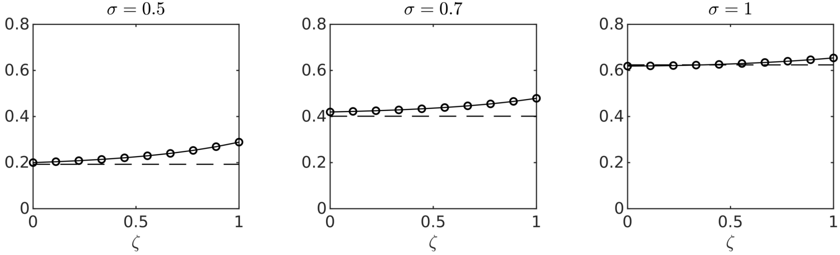
Setting: Estimate A for $V(x) = x^2/2$ (Ornstein–Uhlenbeck) with subsampling varying $\delta = \varepsilon^\zeta$ ($\varepsilon = 0.1,\ T = 10^3$, $p(y) = \cos(y)$)



Issue: How do we pick $\zeta \in (0, 1)$?

Numerical experiments

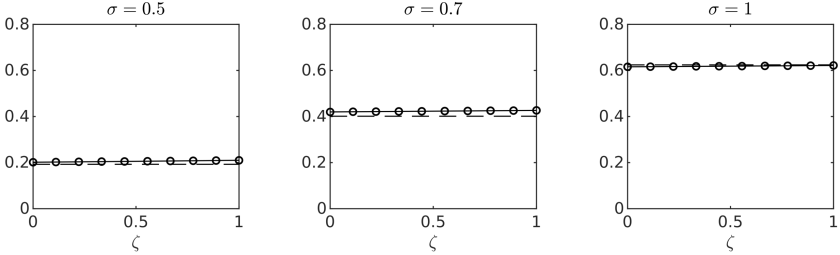
Setting: Estimate A for $V(x) = x^2/2$ (Ornstein–Uhlenbeck) with filtering $\beta = 1$ and $\delta = \varepsilon^\zeta$ ($\varepsilon = 0.1$, $T = 10^3$, $p(y) = \cos(y)$)



Remark: Results still depend on δ - less than subsampling

Numerical experiments

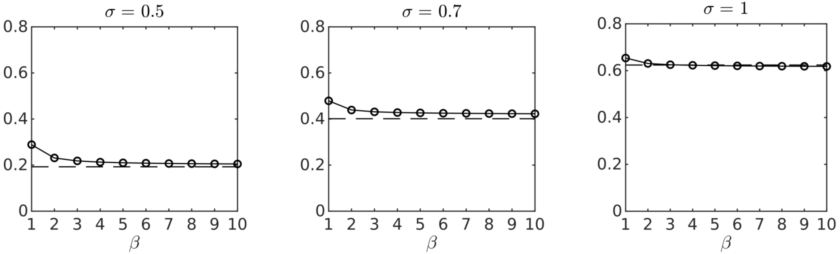
Setting: Estimate A for $V(x) = x^2/2$ (Ornstein–Uhlenbeck) with filtering $\beta = 5$ and $\delta = \varepsilon^\zeta$ ($\varepsilon = 0.1,\ T = 10^3$, $p(y) = \cos(y)$)



Remark: Dependence on δ disappeared

Numerical experiments

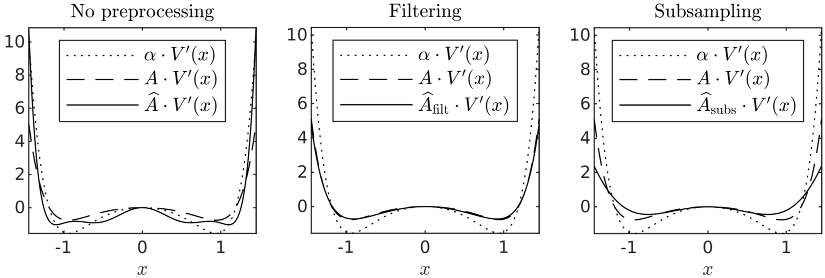
Setting: Estimate A for $V(x) = x^2/2$ (Ornstein–Uhlenbeck) with filtering variable β and δ fixed ($\varepsilon = 0.1$, $T = 10^3$, $p(y) = \cos(y)$)



Remark: Results stabilize fast wrt β

Numerical experiments

Setting: Estimate $A \in \mathbb{R}^4$ for $V_i(x) = x^{2i}/(2i)$, $i = 1, \dots, 4$ with, no pre-processing, subsampling and filtering $\beta = 1$ ($\varepsilon = 0.05$, $T = 10^3$, $p(y) = \cos(y)$)



Remark: Estimate with filter can be done in multi-dimensional case, too.

The Bayesian paradigm

A step back: Likelihood function

$$L(X^\varepsilon \mid A) = \exp\left(-\frac{I(X^\varepsilon \mid A)}{2\Sigma}\right),$$

$$I(X^\varepsilon \mid A) = \int_0^T A \cdot V'(X_t^\varepsilon)\,\mathrm{d}X_t^\varepsilon + \frac{1}{2}\int_0^T \left(A \cdot V'(X_t^\varepsilon)\right)^2 \mathrm{d}t.$$

$\implies$ $\log L(X^\varepsilon \mid A)$ quadratic function of A.

Prior: Fix $\mu_0 = \mathcal{N}(A_0, C_0)$ on A, density p_0

Posterior: Bayes' rule gives (densities)

$$p(A \mid X^\varepsilon) = \frac{1}{Z} L(X^\varepsilon \mid A)\, p_0(A),$$

The Bayesian paradigm

A step back: Likelihood function

$$L(X^\varepsilon \mid A) = \exp\left(-\frac{I(X^\varepsilon \mid A)}{2\Sigma}\right),$$
$$I(X^\varepsilon \mid A) = \int_0^T A \cdot V'(X_t^\varepsilon)\,\mathrm{d}X_t^\varepsilon + \frac{1}{2}\int_0^T \left(A \cdot V'(X_t^\varepsilon)\right)^2 \mathrm{d}t.$$

$\implies \log L(X^\varepsilon \mid A)$ quadratic function of A.

Prior: Fix $\mu_0 = \mathcal{N}(A_0, C_0)$ on A, density p_0

Posterior: Bayes' rule gives (densities)

$$p(A \mid X^\varepsilon) = \frac{1}{Z} L(X^\varepsilon \mid A)\, p_0(A), \implies \text{Gaussian!}$$

The Bayesian paradigm

Posterior: $\mu = \mathcal{N}(m_T, C_T)$ with (complete the squares)

$$C_T^{-1} = C_0^{-1} + TM,$$
$$C_T^{-1} m_T = C_0^{-1} A_0 - Th,$$

where

$$M = \frac{1}{2\Sigma T}\int_0^T V'(X_t^\varepsilon) \otimes V'(X_t^\varepsilon)\,\mathrm{d}t, \quad h = \frac{1}{2\Sigma T}\int_0^T V'(X_t^\varepsilon)\,\mathrm{d}X_t^\varepsilon.$$

Issue 1: Σ unknown – estimate with e.g. subsampling

The Bayesian paradigm

Posterior: $\mu = \mathcal{N}(m_T, C_T)$ with (complete the squares)

$$C_T^{-1} = C_0^{-1} + TM,$$
$$C_T^{-1} m_T = C_0^{-1} A_0 - Th,$$

where

$$M = \frac{1}{2\Sigma T} \int_0^T V'(X_t^\varepsilon) \otimes V'(X_t^\varepsilon)\,\mathrm{d}t, \quad h = \frac{1}{2\Sigma T} \int_0^T V'(X_t^\varepsilon)\,\mathrm{d}X_t^\varepsilon.$$

Issue 2: μ collapses to MLE for $T \to \infty$ ...But MLE is wrong (α)

The Bayesian paradigm – filtering solution

Issue 2: μ collapses to MLE for $T \to \infty$ ...But MLE is wrong (α)

Idea: Use the filter as before

$$\widetilde{L}^{\varepsilon}(X \mid A) = \exp\left(-\frac{\widetilde{I}^{\varepsilon}(X \mid A)}{2\Sigma}\right),$$

$$\widetilde{I}^{\varepsilon}(X \mid A) = \int_0^T A \cdot V'(Z_t^{\varepsilon})\,\mathrm{d}X_t^{\varepsilon} + \frac{1}{2}\int_0^T \left(A \cdot V'(Z_t^{\varepsilon})\right)\left(A \cdot V'(X_t^{\varepsilon})\right)\,\mathrm{d}t.$$

The Bayesian paradigm – filtering solution

Issue 2: μ collapses to MLE for $T \to \infty$ ...But MLE is wrong (α)

Posterior: $\mu = \mathcal{N}(\widetilde{m}_T, \widetilde{C}_T)$ with (complete the squares)

$$\widetilde{C}_T^{-1} = C_0^{-1} + T\widetilde{M}_S,$$
$$\widetilde{C}_T^{-1}\widetilde{m}_T = C_0^{-1}A_0 - T\widetilde{h},$$

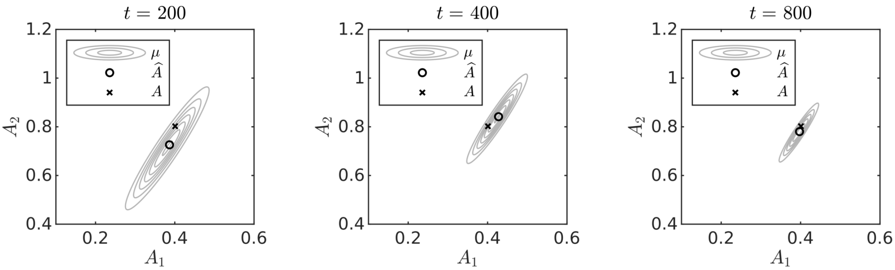
where

$$\widetilde{M} = \frac{1}{2\Sigma T}\int_0^T V'(Z_t^\varepsilon) \otimes V'(X_t^\varepsilon)\,\mathrm{d}t, \quad \widetilde{h} = \frac{1}{2\Sigma T}\int_0^T V'(X_t^\varepsilon)\,\mathrm{d}X_t^\varepsilon,$$

and $\widetilde{M}_S = \left(\widetilde{M} + \widetilde{M}^\top\right)/2 \implies \widetilde{\mu}$ collapses to A wrt T.

Numerical Experiment

Setting: Bistable potential $V_1(x) = x^4/4$, $V_2(x) = -x^2/2$, $\alpha_1 = 1$, $\alpha_2 = 2$. Filtering $\beta = 1$, $\delta = 0.2$ ($\sigma = 0.8$, $\varepsilon = 0.05$, $T = 800$, $p(y) = \cos(y)$)



Remark: As expected, posterior shrinks to MLE (and to A) wrt t

References I

Abdulle, A., Garegnani, G., Pavliotis, G. A., Stuart, A. M., and Zanoni, A. (2020). Drift estimation of multiscale diffusions via filtering. In preparation.

Aït-Sahalia, Y. and Jacod, J. (2014). *High-frequency financial econometrics*. Princeton University Press.

Cotter, C. J. and Pavliotis, G. A. (2009). Estimating eddy diffusivities from noisy Lagrangian observations. *Commun. Math. Sci.*, 7(4):805–838.

Kalliadasis, S., Krumscheid, S., and Pavliotis, G. A. (2015). A new framework for extracting coarse-grained models from time series with multiscale structure. *J. Comput. Phys.*, 296:314–328.

Krumscheid, S., Pavliotis, G. A., and Kalliadasis, S. (2013). Semiparametric drift and diffusion estimation for multiscale diffusions. *Multiscale Model. Simul.*, 11(2):442–473.

Krumscheid, S., Pradas, M., Pavliotis, G. A., and Kalliadasis, S. (2015). Data-driven coarse graining in action: Modeling and prediction of complex systems. *Physical Review E*, 92(4):042139.

Olhede, S. C., Sykulski, A. M., and Pavliotis, G. A. (2010). Frequency domain estimation of integrated volatility for Itô processes in the presence of market-microstructure noise. *Multiscale Model. Simul.*, 8(2):393–427.

Papavasiliou, A., Pavliotis, G. A., and Stuart, A. M. (2009). Maximum likelihood drift estimation for multiscale diffusions. *Stochastic Process. Appl.*, 119(10):3173–3210.

References II

Pavliotis, G. A. and Stuart, A. M. (2007). Parameter estimation for multiscale diffusions. *J. Stat. Phys.*, 127(4):741–781.

Pokern, Y., Stuart, A. M., and van Zanten, J. H. (2013). Posterior consistency via precision operators for Bayesian nonparametric drift estimation in SDEs. *Stochastic Process. Appl.*, 123(2):603–628.

Pokern, Y., Stuart, A. M., and Vanden-Eijnden, E. (2009). Remarks on drift estimation for diffusion processes. *Multiscale Model. Simul.*, 8(1):69–95.

Zhang, L., Mykland, P. A., and Aït-Sahalia, Y. (2005). A tale of two time scales: determining integrated volatility with noisy high-frequency data. *J. Amer. Statist. Assoc.*, 100(472):1394–1411.