Nonparametric Bayesian posterior contraction rates for discretely observed scalar diffusions

Jakob Söhl

Statistical Laboratory
University of Cambridge

joint with R. Nickl

Dynstoch, Rennes, 9 June 2015

UNIVERSITY OF CAMBRIDGE

Outline

Nonparametric Estimation for Diffusions

Diffusion process

$$\mathrm{d}X(t) = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t, \quad t \geqslant 0,$$

with drift coefficient b, diffusion coefficient σ and Brownian motion W

Discrete low-frequency observations

$$X_0, X_\Delta, \dots, X_{n\Delta}, \qquad n \to \infty,\ \Delta > 0 \text{ fixed}$$

Goal: Nonparametric estimation of b and σ

Minimax Optimal Estimation

Gobet, Hoffmann & Reiß (2004):
minimax optimal nonparametric estimation of σ and b

Reflected diffusion on $[0,1]$, smoothness class for $s > 1$:

$$\Big\{(\sigma,b)\in H^s([0,1])\times H^{s-1}([0,1])\,\Big|\,\|\sigma\|_{H^s}\leqslant C, \|b\|_{H^{s-1}}\leqslant C, \inf_x \sigma(x)\geqslant c\Big\}$$

ill-posed problem with rates

$$\begin{array}{cc} \sigma & b \\ n^{-s/(2s+3)} & n^{-(s-1)/(2s+3)} \end{array}$$

in $L^2([\alpha,\beta])$ with $0<\alpha<\beta<1$

Bayesian Estimation

Bayesian estimation methods:
Roberts & Stramer (2001), Papaspiliopoulos, Pokern, Roberts & Stuart (2012), Pokern, Stuart & van Zanten (2013), van der Meulen, Schauer & van Zanten (2014), van Waaij & van Zanten (2015), . . .

Previous analysis of Bayesian methods restricted to $\sigma \equiv 1$ and consistency in a weak topology:
van der Meulen & van Zanten (2013), Gugushvili & Spreij (2014), Koskela, Spano & Jenkins (2015)

Wavelet Series Priors I

ψ_{lk} boundary corrected Daubechies wavelets, $0 < \alpha < \beta < 1$,
$\mathcal{I} = \{(l, k) : \psi_{lk}$ supported in $[\alpha, \beta]\}$

Model diffusion coefficient σ by

$$\log(\sigma^{-2}(x)) = \sum_{(l,k)\in\mathcal{I}} \frac{2^{-l(s+1/2)}}{l^2} u_{lk}\psi_{lk}(x), \qquad u_{lk} \sim^{iid} U(-B, B).$$

Comments:

- Could replace uniform distributions $U(-B, B)$ by any distribution with bouded support and density bounded away from zero.
- Could truncate sum in l at $L_n \to \infty$ sufficiently fast.
- By connection between Hölder norms and wavelet series $\log(\sigma^{-2})$ is modelled as typical s-Hölder smooth function (with a 'convenient' log-factor) .

Wavelet Series Priors II

Model invariant density μ through

$$H(x) = \sum_{(l,k)\in\mathcal{I}} \frac{2^{-l(s+3/2)}}{l^2} \bar{u}_{lk}\psi_{lk}(x), \qquad \bar{u}_{lk} \sim^{iid} U(-B,B),$$

$$\mu = e^H / \int e^H.$$

Drift coefficient b indirectly given by

$$2b = (\sigma^2)' + \sigma^2(\log\mu)'.$$

Overall Prior is given by

$$\Pi = \mathcal{L}(\sigma^2, ((\sigma^2)' + \sigma^2 H')/2)).$$

Comments:

- Priors on b, σ^2 are not independent.
- Invariant density is modelled explicitely.

Assumptions on σ_0 and μ_0

We define the Hölder-type space

$$\mathcal{C}^t([0,1]) := \{f \in C([0,1]) : \|f\|_{\mathcal{C}^t} < \infty\}, \quad \text{where}$$

$$\|f\|_{\mathcal{C}^t} := \sum_{k=0}^{\lfloor t \rfloor} \|D^k f\|_\infty + \sup_{h>0} \sup_{x\in[0,1]} \frac{|D^{\lfloor t \rfloor} f(x+h) - D^{\lfloor t \rfloor} f(x)|}{h^{t-\lfloor t \rfloor} \log(1/h)^{-2}}.$$

Assume diffusion coefficient $\sigma_0 \in \mathcal{C}^s$ is of form

$$\log \sigma_0^{-2}(x) = \sum_{(l,k)\in\mathcal{I}} \tau_{lk}\psi_{lk}(x), \quad x \in [0,1], \qquad \text{with } 2^{l(s+1/2)} l^2 |\tau_{lk}| \leqslant B.$$

Assume invariant density $\mu_0 \in \mathcal{C}^{s+1}$ is of form

$$\log \mu_0(x) = \sum_{(l,k)\in\mathcal{I}} \nu_{lk}\psi_{lk}(x), \quad x \in [0,1], \qquad \text{with } 2^{l(s+3/2)} l^2 |\nu_{lk}| \leqslant B.$$

Contraction Theorem

For $s \geqslant 2$ we define Θ_s by

$$\Big\{(\sigma, b) : \|\sigma\|_{\mathcal{C}^s} \leqslant D, \|b\|_{\mathcal{C}^{s-1}} \leqslant D, \inf_x \sigma(x) \geqslant d, \text{ boundary conditions}\Big\}$$

Theorem

$(X_t : t \geqslant 0)$ reflected diffusion with $(\sigma_0, b_0) \in \Theta_s$. σ_0 and μ_0 as above. Π wavelet series prior. Then for all $0 < \alpha < \beta < 1$ there exists $\gamma > 0$ such that in the $L^2([\alpha, \beta])$-norm

$$\Pi\left((\sigma, b) : \begin{array}{l} n^{-s/(2s+3)}\|\sigma^2 - \sigma_0^2\|_{L^2} > \log^\gamma n \quad \text{or} \\ n^{-(s-1)/(2s+3)}\|b - b_0\|_{L^2} > \log^\gamma n \end{array} \;\middle|\; X_0, \dots, X_{n\Delta} \right) \to 0$$

under $\mathbb{P}_{\sigma_0 b_0}$ as $n \to \infty$.

Bound on Information Theoretic Distance

Information theoretic distance

$$\mathrm{KL}((\sigma_0, b_0), (\sigma, b)) := \mathbb{E}_{\sigma_0 b_0}\left[\log\left(\frac{p_{\sigma_0 b_0}(\Delta, X_0, X_\Delta)}{p_{\sigma b}(\Delta, X_0, X_\Delta)}\right)\right],$$

$p_{\sigma b}$ transition density, expectation $\mathbb{E}_{\sigma_0 b_0}$ w.r.t. stationary distribution

Need good bound on KL:

$$\begin{aligned}
\mathrm{KL}((\sigma_0, b_0), (\sigma, b)) &\lesssim \|p_{\sigma b} - p_{\sigma_0 b_0}\|_{L^2}^2 \\
&\lesssim \|P_\Delta^{\sigma b} - P_\Delta^{\sigma_0 b_0}\|_{HS}^2 \\
&\lesssim \|e^{\Delta/L_{\sigma b}^{-1}} - e^{\Delta/L_{\sigma_0 b_0}^{-1}}\|_{HS}^2 \\
&\lesssim \|L_{\sigma b}^{-1} - L_{\sigma_0 b_0}^{-1}\|_{HS}^2,
\end{aligned}$$

where $P_\Delta^{\sigma b}$ transition operator and $L_{\sigma b}$ infinitesimal generator.

Bound on Information Theoretic Distance II

Infinitesimal generator $L_{\sigma b} f(x) = \frac{1}{2}\sigma^2(x) f''(x) + b(x) f'(x)$

Inverse of infinitesimal generator

$$L_{\sigma b}^{-1} f(x) = \int K_{\sigma b}(x,z) f(z) \mu_0(z) \,\mathrm{d}z$$

Bound distance between integral kernels

$$\begin{aligned}
\mathrm{KL}((\sigma_0, b_0), (\sigma, b)) &\lesssim \| L_{\sigma b}^{-1} - L_{\sigma_0 b_0}^{-1} \|_{HS}^2 \\
&\lesssim \int \int (K_{\sigma b} - K_{\sigma_0 b_0})^2(x,z) \mu_0(x) \mu_0(z) \,\mathrm{d}x \,\mathrm{d}z \\
&\lesssim \| \mu_{\sigma b} - \mu_{\sigma_0 b_0} \|_{L^2([0,1])}^2 + \left\| \frac{1}{\sigma^2} - \frac{1}{\sigma_0^2} \right\|_{(B_{1\infty}^1)^*}^2 + \| b - b_0 \|_{(B_{1\infty}^2)^*}^2 ,
\end{aligned}$$

with dual spaces of Besov spaces $B_{1\infty}^1$ and $B_{1\infty}^2$.

Concentration of Frequentist Estimators

For $\hat{\sigma}$ and $\hat{b}$ estimators by Gobet, Hoffmann & Reiß (2004) we have:

Theorem

There exists $R > 0$ such that for n large enough we have uniformly over Θ_s, $s \geqslant 2$,

$$\mathbb{P}\left(\begin{array}{c} \|\hat{\sigma}^2 - \sigma^2\|_{L^2([\alpha,\beta])} \geqslant R n^{-s/(2s+3)} \quad \text{or} \\ \|\hat{b} - b\|_{L^2([\alpha,\beta])} \geqslant R n^{-(s-1)/(2s+3)} \end{array} \right) \leqslant \exp\left(-D n^{1/(2s+3)} \right).$$

This means exponential concentration of $\hat{\sigma}$ and $\hat{b}$ at minimax rates $n^{-s/(2s+3)}$ and $n^{-(s-1)/(2s+3)}$, respectively.

Concentration Inequality

Bernstein-type inequality

There exists $\kappa > 0$ such that for all reflected diffusions $\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t$, $t \in [0, \infty)$ with $(\sigma, b) \in \Theta := \Theta_2$ and arbitrary initial distribution, $\forall f : [0,1] \to \mathbb{R}$ bounded, $\forall s > 0$ and $\forall n \in \mathbb{N}$,

$$\mathbb{P}\left(\Big| \sum_{j=0}^{n-1} (f(X_{j\Delta}) - \mathbb{E}_\mu[f]) \Big| > s \right) \leqslant \kappa \exp\left(-\frac{1}{\kappa} \min\left(\frac{s^2}{n\|f\|_{L^2(\mu)}^2}, \frac{s}{\log(n)\|f\|_\infty} \right) \right).$$

Concentration Inequality for Suprema of Empirical Processes

Class of functions $\mathcal{F} = \{f_i : i \in I\}$ with $0 \in \mathcal{F}$ and $\dim I = d$

$V^2 = \kappa n \sup_{f \in \mathcal{F}} \|f\|^2_{L^2(\mu)}$ and $U = \kappa \log n \sup_{f \in \mathcal{F}} \|f\|_\infty$

Theorem
For $\tilde{\kappa} = 18$ and for all $x \geqslant 0$ we have

$$\mathbb{P}\left(\sup_{f \in \mathcal{F}} \left| \sum_{j=0}^{n-1} (f(X_{j\Delta}) - \mathbb{E}_\mu[f]) \right| \geqslant \tilde{\kappa} \left(\sqrt{V^2(d+x)} + U(d+x) \right) \right) \leqslant 2\kappa e^{-x}.$$

Follows from chaining and previous concentration inequality.

Concentration inequality builds on results by Adamczak (2008) for Markov chains based on regeneration approach.

Regeneration Approach

$X_0, X_1, \ldots$ Markov chain on $(S, \mathcal{B})$ with transition probability $P(x, A)$

C called atom if $P(x, A) = \nu(A)$ for all $x \in C$. Markov chain regenerates when $X_n \in C$. Lower bound on transition density instead of atom.

(A1) *Minorization condition.* $\exists C \in \mathcal{B}$, $\tilde{\beta} > 0$ and probability measure ν on $(S, \mathcal{B})$ such that $\forall x \in C$ and $\forall A \in \mathcal{B}$

$$P(x, A) \geqslant \tilde{\beta}\nu(A),$$

and $\forall x \in S$ there $\exists n \in \mathbb{N}$ such that $P^n(x, C) > 0$.

(A2) *Drift condition.* $\exists \lambda < 1$, $K < \infty$ constants and $V : S \to [1, \infty)$ s.t.

$$PV(x) \leqslant \begin{cases} \lambda V(x), & \text{if } x \notin C, \\ K, & \text{if } x \in C. \end{cases}$$

(A3) *Strong aperiodicity condition.* $\exists \beta > 0$ such that $\tilde{\beta}\nu(C) \geqslant \beta$

Small Ball Probability Condition

$\mathcal{B} \subseteq \Theta$ with a σ-field $\mathcal{S}$, Π prior distribution on $\mathcal{S}$, $(\sigma_0, b_0) \in \Theta$, $\varepsilon_n \to 0$, $\sqrt{n}\varepsilon_n \to \infty$, and C, r fixed constants
Suppose Π satisfies

$$\Pi\left(B_{\varepsilon_n, r}\right) \geqslant e^{-Cn\varepsilon_n^2},$$

where

$$B_{\varepsilon,r} = \Bigg\{(\sigma, b) \in \mathcal{B} : \mathrm{KL}((\sigma_0, b_0), (\sigma, b)) \leqslant \varepsilon^2, \\ \mathrm{Var}_{\sigma_0 b_0}\left(\log \frac{p_{\sigma b}(\Delta, X_0, X_\Delta)}{p_{\sigma_0 b_0}(\Delta, X_0, X_\Delta)}\right) \leqslant 2\varepsilon^2, \\ \mathrm{K}(\mu_{\sigma_0 b_0}, \mu_{\sigma b}) \leqslant r, \mathrm{Var}_{\sigma_0 b_0}\left(\log \frac{\mu_{\sigma b}(X_0)}{\mu_{\sigma_0 b_0}(X_0)}\right) \leqslant 2r\Bigg\}.$$

with transition density $p_{\sigma b}$ and invariant density $\mu_{\sigma b}$.

General Contraction Theorem

Ghosal, Ghosh & van der Vaart (2000)-type theorem

X reflected diffusion started in stationary distribution
Small ball probability condition: C, L, r constants so that

$$\Pi\left(B_{\varepsilon_n, r}\right) \geqslant e^{-Cn\varepsilon_n^2},$$

and $\Pi(\mathcal{B} \backslash \mathcal{B}_n) \leqslant Le^{-(C+4)n\varepsilon_n^2}$ for some sequence $\mathcal{B}_n \subseteq \mathcal{B}$

Tests: Sequence of tests $\Psi_n \equiv \Psi(X_0, \ldots, X_{n\Delta})$ and of metrics d_n such that for $M > 0$ large enough,

$$\mathbb{E}_{\sigma_0 b_0}[\Psi_n] \to_{n \to \infty} 0, \qquad \sup_{(\sigma,b) \in \mathcal{B}_n : d_n((\sigma,b),(\sigma_0,b_0)) \geqslant M\varepsilon_n} \mathbb{E}_{\sigma b}[1 - \Psi_n] \leqslant Le^{-(C+4)n\varepsilon_n^2}.$$

Give posterior contraction: Then the posterior $\Pi(\cdot | X_0, \ldots, X_{n\Delta})$ satisfies

$$\Pi((\sigma, b) : d_n((\sigma, b), (\sigma_0, b_0)) > M\varepsilon_n | X_0, \ldots, X_{n\Delta}) \to 0$$

under $\mathbb{P}_{\sigma_0 b_0}$, as $n \to \infty$.

References

Adamczak, R. (2008): A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electron. J. Probab.* 13(34), 1000-1034.

Ghosal S., Ghosh, J.K. and van der Vaart, A.W. (2000): Convergence rates of posterior distributions. *Ann. Statist.* 28, 500–531.

Gobet, E., Hoffmann, M. and Reiß, M. (2004): Nonparametric estimation of scalar diffusions based on low frequency data. *Ann. Statist.* 32(5), 2223-2253.

Nickl, R. and Söhl, J. (2015): Nonparametric Bayesian posterior contraction rates for discretely observed scalar diffusions. *Arxiv:1510.05526*, submitted.

References

Adamczak, R. (2008): A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electron. J. Probab.* 13(34), 1000-1034.

Ghosal S., Ghosh, J.K. and van der Vaart, A.W. (2000): Convergence rates of posterior distributions. *Ann. Statist.* 28, 500–531.

Gobet, E., Hoffmann, M. and Reiß, M. (2004): Nonparametric estimation of scalar diffusions based on low frequency data. *Ann. Statist.* 32(5), 2223-2253.

Nickl, R. and Söhl, J. (2015): Nonparametric Bayesian posterior contraction rates for discretely observed scalar diffusions. *Arxiv:1510.05526*, submitted.

Thank you for your attention!