Convergence rates of posterior distributions for Brownian semimartingale models

Frank van der Meulen
Harry van Zanten, Aad van der Vaart

Vrije Universiteit Amsterdam

European Meeting of Statisticians, Oslo
July 2005

vrije Universiteit amsterdam

Outline of talk

Convergence rates of posterior distributions for Brownian semimartingale models

- Brownian semimartingale Model (BSM)
  examples

- Bayesian estimation
  prior, likelihood, posterior - consistency - rate of convergence

- Main result: gives lower bounds on the Bayesian rate of convergence for the BSM

- Example: ergodic diffusion

Brownian semimartingale model

- Θ: parameter set. θ_0 true value
  B: standard Brownian motion

- $\beta^{\theta,n}$ and σ^n (suitably measurable) processes
  X^n: weak solution of

$$dX_t^n = \beta^{\theta_0,n}(t, X^n)dt + \sigma^n(t, X^n)dB_t, \quad t \in [0, T_n].$$

- Suppose we observe X^n continuously on $[0, T_n]$. Estimate θ_0.

- Behavior of Bayesian estimators in the large-sample limit $n \to \infty$.

Examples

- Gaussian white-noise model

$$dX_t^n = \theta(t)dt + \sigma_n(t)dB_t, \quad t \in [0,T],$$

where (σ_n) is a sequence of positive numbers tending to zero, as $n \to \infty$.

- Perturbed dynamical system

$$dX_t^n = \theta(X_t^n)dt + \sigma_n(t)dB_t, \quad t \in [0,T],$$

where (σ_n) is a sequence of positive numbers tending to zero, as $n \to \infty$.

- Ergodic diffusion model

$$dX_t = \beta^\theta(X_t)dt + \sigma(X_t)dB_t, \quad t \in [0,T_n],$$

where $T_n \to \infty$.

Motivation

In practice one never observes such a process continuously, so why study estimators for it?

- Limit model of “high-frequency” data
- Results provide an upper bound on what can be achieved.
- Analysis of the same problem for discrete time observations is harder.
- Bayesian computations for the BSM based on discrete time observations are promising, and feasible by MCMC-algorithms. However, no theoretical justification.
- Certain forms of the Gaussian white-noise model are “statistically equivalent” to regression problems.

Bayesian estimation (1)

Ingredients:

- Prior distribution Π on Θ.
- Likelihood function L: given observation X, $L(X;\theta)$ is conditional density of X given $\Theta = \theta$.

$$L(\theta; X^n) = \exp\left(\int_0^{T_n} \frac{\beta^{\theta,n}}{(\sigma^n)^2} dX^n - \frac{1}{2} \int_0^{T_n} \left(\frac{\beta^{\theta,n}}{\sigma^n} \right)^2 dt \right)$$

- Posterior distribution $\Pi(\cdot|X^n)$ obtained via Bayes' Theorem: conditional distribution of Θ, given X^n

Bayesian estimation (2)

- Consistency: "The posterior distribution concentrates arbitrarily close to the true one, as we obtain more data".

$$\forall \varepsilon > 0 \quad \Pi_n(d(\theta, \theta_0) \geq \varepsilon | X^n) \to 1, \quad P_{\theta_0} - a.s.$$

- Bayesian convergence rate: "Maximal rate at which we can shrink balls around θ_0, while still capturing almost all posterior mass"

  Let $\varepsilon_n \downarrow 0$ $(n \to \infty)$. The sequence of posteriors converges to θ_0 (at least) at rate ε_n if for all sequences $M_n \to \infty$,

$$\Pi_n(\theta \in \Theta \; : \; d(\theta, \theta_0) \geq M_n \varepsilon_n | X^n) \to 0$$

  in $P_{\theta_0}^{(n)}$-probability.

- Derive conditions from which we can infer the convergence rate.

Type of conditions

To obtain consistency and rates of convergence:

- Condition on the prior (sufficient mass near the true value)

  If the prior Π excludes θ_0 consistency cannot even occur!

- Entropy conditions

  Birgé and Le Cam:

  - d_n, e_n: semimetrics on Θ^n.
  - ε_n: sequence of positive numbers $\downarrow 0$.
  - There exist estimators $\hat{\theta}_n = \hat{\theta}_n(X^n)$ such that $d_n(\hat{\theta}_n, \theta_0) = O_P(\varepsilon_n)$ if

$$\sup_{\varepsilon > \varepsilon_n} \log N(\varepsilon\xi, \{\theta \ : \ d_n(\theta, \theta_0) \leq \varepsilon\}, e_n) \leq n\varepsilon_n^2.$$

Working assumption

- Measure distance on Θ by the Hellinger (semi)metric

$$h_n^2(\theta,\psi) = \int_0^T \left(\frac{\beta^{\theta,n} - \beta^{\psi,n}}{\sigma^n}\right)^2 dt, \quad \theta,\psi \in \Theta.$$

- There exist semimetrics d_n and $\bar{d}_n$ on Θ such that, as n gets large, with large probability

$$c_n d(\theta,\psi) \lesssim h_n(\theta,\psi) \lesssim c_n \bar{d}(\theta,\psi),$$

for c_n a sequence of positive numbers.

- $B(\theta_0,\varepsilon)$: ball of d-radius ε around θ_0. $\bar{B}(\theta_0,\varepsilon)$: for ball of $\bar{d}$-radius ε around θ_0

Main result

- ε_n: sequence of positive numbers such that $c_n\varepsilon_n$ is bounded away from zero.

- Assume for every $a > 0$ there exists a constant $g(a) < \infty$ such that

$$\sup_{\varepsilon>\varepsilon_n} \log N(a\varepsilon, B(\theta_0,\varepsilon), \bar{d}) \leq (c_n\varepsilon_n)^2 g(a).$$

If for every $\xi > 0$ there is a constant J such that for $j \geq J$

$$\frac{\Pi_n(B(\theta_0, j\varepsilon_n))}{\Pi_n(\bar{B}(\theta_0,\varepsilon_n))} \leq e^{\xi c_n^2 \varepsilon_n^2 j^2}.$$

- Then for every $M_n \to \infty$, we have

$$P_{\theta_0}^{(n)}[\Pi_n(\theta \in \Theta_n : h_n(\theta,\theta_0) \geq M_n c_n \varepsilon_n | X^n)] \to 0.$$

Example: Ergodic diffusion

- Under "conditions"

$$dX_t = \beta^\theta(X_t)dt + \sigma(X_t)dB_t, \quad t \in [0, T_n],$$

defines a regular diffusion in an interval I

- semimetrics: measure

$$\bar{d}(\theta,\psi) = \Big\|\frac{\theta-\psi}{\sigma}\Big\|_{L^2(\mu_0)}, \qquad d(\theta,\psi) = \Big\|\frac{\theta-\psi}{\sigma}1_J\Big\|_{L^2(\mu_0)}.$$

(μ_0 is the invariant probability measure, $J \subseteq I$ compact)

- Lipschitz condition on drift. Assume

$$\underline{b}(x)\|\theta-\psi\| \leq |b_\theta(x) - b_\psi(x)| \leq \bar{b}(x)\|\theta-\psi\|, \, \forall\, x \in I.$$

- If $\Theta \subseteq \mathbb{R}^k$ bounded and the prior density is bounded away from zero near θ_0, then as $T_n \to \infty$

$$P^{\theta_0,n}\Pi^n\Big(\theta \in \Theta \;:\; \|\theta-\theta_0\| \geq M_n/\sqrt{T_n}\,|X^n\Big) \to 0.$$