Efficient pointwise estimation based on discrete data in ergodic nonparametric diffusions

Serguei Pergamenchtchikov

Laboratory of Mathematics Raphaël Salem

University of Rouen, Rouen, France

joint work with Galtchouk, L.

March 17 - 20, 2015, Le Mans, France

Outline

1. Model
2. Sequential procedure
3. Functional class
4. Asymptotic efficience
5. Main tool
6. Conclusion

Model

We consider the following diffusion model:

$$\mathrm{d}y_t = S(y_t)\,\mathrm{d}t + \sigma(y_t)\mathrm{d}W_t\,, \quad 0 \le t \le T\,,$$

where $(W_t)_{t\ge 0}$ is a scalar standard Wiener process, $S(\cdot)$ and $\sigma(\cdot)$ are unknown functions.

This model appears in a number of applied problems of stochastic control, filtering, spectral analysis, identification of dynamic system, financial mathematics and others (see Liptser and Shiryaev (1978), Arato (1982), Bensoussan (1992), Karatzas and Shreve (1998), Kutoyants (2004), and others for details).

Observations

The problem is to estimate the function $S(\cdot)$ at a point x_0 basing on the discrete time observations

$$(y_{t_j})_{1\le j\le N}\,, \quad t_j = j\delta\,,$$

where $N = [T/\delta]$ and the frequency $\delta = \delta_T \in (0,1)$ is a function of T that will be specified later.

Pointwise Risk

We consider the pointwise estimation problem for the function $S(\cdot)$ at a fixed point $x_0 \in \mathbb{R}$ with unknown diffusion coefficient σ, i.e. σ is a nuisance parameter. For any estimate (i.e. any $(y_t)_{0 \le t \le T}$ measurable function) $\tilde{S}_T(x_0)$ of $S(x_0)$, we define the pointwise risk as follows

$$\mathcal{R}_\vartheta(\widetilde{S}_T) = \mathbf{E}_\vartheta \left|\widetilde{S}_T(x_0) - S(x_0)\right|,$$

where $\vartheta = (S, \sigma)$.
Our goal

$$\sup_{\theta \in \Theta} \mathcal{R}_\vartheta(\widetilde{S}_T) \to \min_{\widetilde{S}_T} .$$

Continuous data

The kernel estimator for $S(x_0)$ is

$$\widehat{S}_T(x_0) = \frac{\int_0^T Q\left(\frac{y_t - x_0}{h}\right) \mathrm{d}y_t}{\int_0^T Q\left(\frac{y_t - x_0}{h}\right) \mathrm{d}t}$$

where $h > 0$ is the bandwidth and $Q(\cdot)$ is a kernel function, such that $Q(x) = 0$ for $|x| \geq 1$. We define this estimator on the set

$$\left\{ \left| \int_0^T Q\left(\frac{y_t - x_0}{h}\right) \mathrm{d}t \right| > 0 \right\}$$

Continuous data

Using the sequential analysis approach proposed by Novikov (1971) and Liptesr and Shiriaev (1971) for parametric estimation in diffusion processes we replace the observation time T with a special stopping time τ_H, defined as

$$\tau = \tau_H = \inf\{t \geq 0 : \int_0^t Q\left(\frac{y_t - x_0}{h}\right) \mathrm{d}t \geq H\},$$

where $H > 0$ is some fixed threshold.
In this case the kernel estimator becomes as follows

$$\widehat{S}_\tau(x_0) = \frac{1}{H} \int_0^{\tau_H} Q\left(\frac{y_t - x_0}{h}\right) \mathrm{d}y_t \,.$$

Continuous data

The error in this case has the form

$$\widehat{S}_\tau(x_0) - S(x_0) = B_H^* + \frac{1}{\sqrt{H}}\eta_H^*,$$

where B_H^* is approximative term and

$$\eta_H^* = H^{-1/2}\int_0^{\tau_H} Q\left(\frac{y_t - x_0}{h}\right)\mathrm{d}W_t \quad \sim \quad \mathcal{N}(0,1).$$

Stopping time

$$\tau_H \quad \approx \quad \frac{H}{2q_S(x_0)h}.$$

Kernel estimator

The kernel estimator for $S(x_0)$ is

$$\widehat{S}_T(x_0) = \frac{\sum_{j=1}^{N} Q\left(\frac{y_{t_{j-1}} - x_0}{h}\right) \Delta y_{t_j}}{\sum_{j=1}^{N} Q\left(\frac{y_{t_{j-1}} - x_0}{h}\right) \Delta t_j}$$

where $h > 0$ is the bandwidth and $Q(\cdot)$ is a kernel function, such that $Q(x) = 0$ for $|x| \geq 1$. We define this estimator on the set

$$\left\{ \left| \sum_{j=1}^{N} Q\left(\frac{y_{t_{j-1}} - x_0}{h}\right) \right| > 0 \right\} .$$

Stopping time

Let us define the following stopping time

$$\nu = \inf\left\{ j \geq N_0 \, : \, \sum_{i=N_0}^{j} Q\left(\frac{y_{t_{i-1}} - x_0}{h}\right) \geq H_T \right\} \wedge N \,,$$

where $H_T > 0$ is a threshold and h is a positive bandwidth. In this case the observation time is given by stopping time

$$\tau = t_\nu = \delta \nu \,.$$

Now we have to choose the threshold H_T. Note that in order to construct an efficient estimator one should use all, i.e. N observations.

Stopping time

Therefore, the threshold H_T should provide the asymptotic relations

$$\nu \quad \approx \quad N \quad \text{as} \quad T \to \infty \,.$$

In view of the geometric ergodicity property

$$\sum_{i=N_0}^{N} Q\left(\frac{y_{t_{i-1}} - x_0}{h}\right) \quad \approx \quad 2h(N - N_0) q_\vartheta(x_0) \,.$$

Replacing here the ergodic density with its estimate yields the following definition of the threshold

$$H_T = h(N - N_0)(2\widetilde{q}_T(x_0) - \upsilon_T) \,.$$

Correction coefficient

Sequential estimator for the desecrate date

$$\widehat{S}_\tau(x_0) = \frac{\sum_{j=1}^{\nu} Q\left(\frac{y_{t_{j-1}} - x_0}{h}\right) \Delta y_{t_j}}{\sum_{j=1}^{\nu} Q\left(\frac{y_{t_{j-1}} - x_0}{h}\right) \Delta t_j}.$$

To correct here the denominator we use the same way which Borisov and Konev (1977) proposed to study the mean square accuracy of sequential parametric estimation for autoregressive processes.

Correction coefficient

Now on the set

$$\Gamma_T = \left\{ \sum_{i=N_0}^{N} Q\left(\frac{y_{t_{i-1}} - x_0}{h}\right) > H_T \right\}$$

we define the correction coefficient $\varkappa = \varkappa_T$ as

$$\sum_{j=N_0}^{\nu-1} \phi_h(y_{t_{i-1}}) + \varkappa\,\phi_h(y_{t_{\nu-1}}) = H_T\,,$$

where

$$\phi_h(y) = Q\left(\frac{y - x_0}{h}\right)\,.$$

Decomposition

Now we define the sequential estimator for $S(x_0)$ as

$$S^*_{h,T}(x_0) = \frac{\sum_{j=N_0}^{\nu-1} \phi_h(y_{t_{i-1}})\, \Delta y_{t_j} + \varkappa\, \phi_h(y_{t_{\nu-1}})\, \Delta y_{t_\nu}}{\delta H_T}\, \mathbf{1}_{\Gamma_T}\,.$$

In this case on the set Γ_T we obtain the following decomposition

$$S^*_{h,T}(x_0) = S(x_0) + B_{1,T} + B_{2,T} + \frac{\sigma(x_0)}{\sqrt{\delta H_T}}\, \eta^*_T\,,$$

and

$$\eta^*_T \quad \sim \quad \mathcal{N}(0,1)\,.$$

Non-asymptotic estimation

Theorem

For any $h > 0$ *and* $T \geq 3$ *for which* $0 < \delta \leq 1$, *one has*

$$\mathbf{E}_\vartheta \, |S^*_{h,T}(x_0) - S(x_0)| \leq L \, U^* + L \, \mathbf{P}_\vartheta \left(\Gamma^c_T\right) ,$$

where

$$U^*(\delta, h, T) = \sqrt{\delta} + h + \mathbf{E}_\vartheta \frac{1}{\sqrt{\delta \, H_T}} .$$

Functional class

It is clear that to obtain a good estimate for the function $S(\cdot)$ at the point x_0 it is necessary to impose some conditions on the function $\vartheta = (S, \sigma)$ which provide that the observed process $(y_t)_{0 \le t \le T}$ returns to any vicinity of the point x_0 infinitely many times.
We make use of the weak Hölder functions introduced in Galtchouk and Pergamenshchikov (2006), which guarantees the ergodicity property for this model.

Functional class

We say that a function S satisfies the weak Hölder condition at the point $x_0 \in \mathbb{R}$ with the parameters h,$\epsilon > 0$ and exponent $\beta = 1 + \alpha$, $\alpha \in (0, 1)$, if the function $S \in \mathbf{C}^1(\mathbb{R})$ and its derivative satisfies the following inequality

$$\left|\Omega_{x_0,h}(S)\right| \leq \epsilon\, h^{\alpha}\,,$$

where

$$\Omega_{x_0,h}(S) = \int_{-1}^{1} z \int_{0}^{1} \left(\dot{S}(x_0 + uzh) - \dot{S}(x_0)\right) \mathrm{d}u\, \mathrm{d}z\,.$$

Functional class

We denote by $\mathcal{H}^{w}_{x_0,M}(\epsilon,\beta,h)$ the set of all functions D satisfying the weak Hölder condition at the point $-x_* < x_0 < x_*$ (for some fixed $x_* > 0$) such that

$$\sup_{x\in\mathbb{R}}\left(|D(x)|+|\dot{D}(x)|\right)\le M$$

and

$$D(x)=0 \quad \text{for} \quad |x|\ge x_*\,.$$

Functional class

Let S_0 be a function with $\sup_{|x|\geq x_*} \dot{S}(x) < 0$ such that

$$\lim_{h\to 0} h^{-\beta}\,\Omega_{x_0,h}(S_0) = 0\,.$$

For example, $S_0(x) = -x$ for $|x| > x_* + 1$ and $S_0(x) = 0$ for $|x| \leq x_*$. Finally we set

$$\Theta = \mathcal{U}_M(x_0,\beta) \times \mathcal{V} \quad \text{and} \quad \mathcal{U}_M(x_0,\beta) = S_0 + \mathcal{H}^{w}_{x_0,M}(\epsilon,\beta,h)\,,$$

where $h = T^{-1/(2\beta+1)}$ and $\epsilon = \epsilon_T = (\ln T)^{-3/2}$. Here $\mathcal{V}$ is a set of diffusion coefficients σ bounded with below by some constant and with bounded by the second derivative.

Upper bound

The frequency condition

$$\delta = \mathrm{O}\left(\frac{1}{T(\ln T)^{1+d}}\right) \quad \text{for some} \quad 0 < d < 1\,.$$

Proposition

For any $a > 0$ *and* $h \geq T^{-1/2}$

$$\lim_{T\to\infty} T^{a} \sup_{\vartheta\in\Theta} \mathbf{P}_{\vartheta}\left(\Gamma^{c}_{T}\right) = 0\,.$$

Upper bound

We set

$$\varsigma^*_\vartheta(x_0) = \frac{2q_\vartheta(x_0)}{\sigma^2(x_0)}\,.$$

We study the pointwise risk normalized by the rate $\varphi_T = T^{\beta/(2\beta+1)}$.

Theorem

The estimator $S^*_{h,T}$ with $h = T^{-1/(2\beta+1)}$ has the following wooer bound

$$\overline{\lim}_{T\to\infty}\,\varphi_T \sup_{\vartheta\in\Theta} \sqrt{\varsigma^*_\vartheta(x_0)}\,\mathcal{R}_\vartheta(S^*_{h,T}) \le \mathbf{E}|\xi|\,,$$

where ξ is a $(0,1)$ gaussian random variable.

Lower bound

Theorem

The point wise risk admits the following lower bound

$$\underline{\lim}_{T\to\infty}\varphi_T \inf_{\widetilde{S}_T} \sup_{\vartheta\in\Theta} \sqrt{\varsigma^*_\vartheta(x_0)}\,\mathcal{R}_\vartheta(\widetilde{S}_T) \geq \mathbf{E}|\xi|\,,$$

where infimum is taken over all possible estimators $\widetilde{S}_T$, ξ *is a* $(0,1)$ *gaussian random variable.*

Uniform Geometric Ergodicity

Galtchouk and Pergamenshchikov (2014)

Theorem

For any $\epsilon > 0$*, there exist constants* $R = R(\epsilon) > 0$ *and* $\kappa = \kappa(\epsilon) > 0$ *such that*

$$\sup_{u \geq 0} e^{\kappa u} \sup_{\|g\|_* \leq 1} \sup_{x \in \mathbb{R}} \sup_{\vartheta \in \Theta} \frac{|\mathbf{E}_{\vartheta,x}\, g(y_u) - q_\vartheta(g)|}{1 + |x|^\epsilon} \leq R\,,$$

where $\mathbf{E}_{\vartheta,x}\,(\cdot) = \mathbf{E}_\vartheta\,(\cdot|y_0 = x)$, $\|g\|_* = \sup_x |g(x)|$ *and*

$$q_\vartheta(g) = \int_{\mathbb{R}} g(x)\, q_\vartheta(x) \mathrm{d}x\,.$$

Concentration inequality

For any $\mathbb{R} \to \mathbb{R}$ function f belonging to $\mathbf{L}_1(\mathbb{R})$, we set

$$\mathbf{D}_n(f) = \sum_{k=1}^{n} \left(f(y_{t_k}) - q_\vartheta(f) \right) .$$

Galtchouk and Pergamenshchikov (2013)

Theorem

For any $a > 0$,

$$\lim_{T\to\infty} T^a \sup_{h \ge T^{-1/2}} \sup_{\vartheta\in\Theta} \mathbf{P}_\vartheta \left(|\mathbf{D}_N(\phi_h)| \ge \epsilon_T \, N \, h \right) = 0 \, .$$

Density estimation

We choose

$$N_0 = N^{\gamma_0} \quad \text{and} \quad 2/3 < \gamma_0 < 1\,.$$

We will make use of the following kernel estimator

$$\widehat{q}_T(x_0) = \frac{1}{2(N_0 - 1)\varsigma} \sum_{j=1}^{N_0 - 1} Q\left(\frac{y_{t_j} - x_0}{\varsigma}\right),$$

where $Q(y) = \mathbf{1}_{(|y| \le 1)}$ and $\varsigma = \varsigma_T$ is a function of T.

Density estimation

We set

$$\widetilde{q}_T(x_0) = L_{v_T}(\widehat{q}_T(x_0))$$

where $v_T = \ln^{-\gamma} T$. For any $0 < \varepsilon < 1$ and $x > 0$

$$L_\varepsilon(x) = \varepsilon \mathbf{1}_{\{x \le \varepsilon\}} + x \mathbf{1}_{\{\varepsilon \le x \le \epsilon^{-1}\}} + \varepsilon^{-1} \mathbf{1}_{\{x \ge \varepsilon^{-1}\}} \,.$$

Proposition

For any $a > 0$,

$$\lim_{T\to\infty} T^a \sup_{\vartheta\in\Theta} \mathbf{P}_\vartheta(|\widetilde{q}_T(x_0) - q_\vartheta(x_0)| > \upsilon_T) = 0 \,.$$

Adaptive pointwise estimation

Here, we studied the estimation problem of the function S when its smoothness is known. In the case of unknown smoothness, similarly to Galtchouk and Pergamenshchikov (2001) for continuous time observations we construct the adaptive estimate based on Lepskii's procedure. Indeed,

$$S^*_{h,T}(x_0) = S(x_0) + B_T(x_0) + \frac{\sigma(x_0)}{\sqrt{\delta H_T}} \eta^*_T(x_0) \,,$$

Note that Lepskii's procedure works here just thanks to sequential estimating since, for the sequential estimate of the function S, the stochastic term in the deviation is a gaussian random variable.

Oracle inequality

To estimate in the quadratic risk, i.e.

$$\mathbf{E}_{\vartheta} \int_a^b (\widehat{S}(t) - S(t))^2 \mathrm{d}t \,,$$

we apply the selection model estimation.
To this end we divide the interval (a, b) by the points $(x_k)_{1 \le k \le m}$, i.e.

$$a = x_0 < \ldots < x_m = b \,,$$

and we estimate the function S in each point x_k on the set Γ_k.

Oracle inequality

Through the sequential kernel estimator we pass to the following heteroscedastic regression model on the set $\cap_{k=1}^{m} \Gamma_k$

$$y_k = S^*_{h,T}(x_k) = S(x_k) + \zeta_k \,, \quad \zeta_k = B_k + \sigma_k \eta_k \,,$$

where the random variable $(\eta_k)_{1\le k\le m}$ are i.i.d. $(0,1)$ gaussian.
We can apply now the selection model procedure from Galtchouk and Pergamenshchikov (2009) to find the Pinsker constant for this model.

Thanks

THANK YOU VERY MUCH

FOR YOUR ATTENTION !!!