Adaptive sequential estimation for ergodic diffusion processes in quadratic metric. *

Leonid Galtchouk[1] and Sergey Pergamenshchikov[2]

[1] IRMA, Department of Mathematics, University of Strasbourg,
7 Rene Descartes str.,
67084, Strasbourg Cedex, France
galtchou@math.u-strasbg.fr

[2] Laboratoire de Mathématiques Raphael Salem UMR CNRS 6085,
UFR Sciences, Université de Rouen,
F76821 Mont St Aignan, Cedex France
Serge.Pergamenchtchikov@univ-rouen.fr

Abstract. An adaptive nonparametric estimation procedure is constructed for estimating the drift coefficient in ergodic diffusion processes. A non-asymptotic upper bound (an oracle inequality) is obtained for the quadratic risk. Asymptotic efficiency is proved, i.e. Pinsker's constant is found in the asymptotic lower bound for the minimax quadratic risk. It is shown that the asymptotic minimax quadratic risk of the constructed procedure coincides with this constant.

Keywords. Adaptive estimation, Diffusion process, Efficient estimation, Non-asymptotic bounds, Nonparametric estimation, Sequential estimation,Pinsker's constant.

1 Introduction

We consider the following stochastic differential equation

$$\mathrm{d}y_t = S(y_t)\,\mathrm{d}t + \mathrm{d}w_t\,, \quad 0 \le t \le T\,, \tag{1}$$

where $(w_t)_{t\ge 0}$ is a scalar standard Wiener process, the initial value y_0 is a given constant and $S(\cdot)$ is an unknown function. The problem is to estimate the function $S(x)$ from observations of $(y_t)_{0\le t\le T}$.

It seems that, for the first time, the problem of non-asymptotic parameter estimation for diffusion processes has been studied in [1] for wobbling analysis of the axis of the equator. There, for a special diffusion process, the exact distribution of the ML-estimators of unknown parameters has been obtained for any finite sample time T. Unfortunately, in the majority of cases, when the sample time is finite, it is difficult to study classical estimators such as LS-estimators or ML-estimators since they are non-linear functionals of observations. In particular, it is difficult to compute the mean, a minimax risk etc.

In [14] it has been shown that many difficulties in non-asymptotic parameter estimation for one-dimensional diffusion processes can be overcome by the sequential approach. It turns out that the theoretical analysis of the sequential ML-estimator is easier than the analysis of classical procedures. In particular, it is possible to calculate non-asymptotic bounds for quadratic risk in the sequential procedure. By making use of the sequential approach non-asymptotic parameter estimation problems have been studied in [11], [2] for multidimensional diffusion processes and recently in [3] for multidimensional continuous and discrete time semimartingales. In the paper [12] a truncated sequential method has been developed for parameter estimation in diffusion processes.

The sequential approach to nonparametric minimax estimation problem of the drift coefficient in ergodic diffusion processes has been developed in [4]–[7]. The papers [4],[6] and [7] deal with sequential pointwise kernel estimators of the drift coefficient. For these estimators non-asymptotic upper bounds were obtained for absolute error risks, the estimators yield also the optimal convergence rate as the sample time $T \to \infty$. In the paper [4] it is shown that this procedure is minimax and adaptive in the both cases when either the smoothness is known or unknown. The same type of the sequential

* The paper is supported by the RFFI-Grant 09-01-00172-a.

kernel estimators is used in the paper [5] for the nonparametric estimation in the $\mathbf{L}_2$- metric of the drift coefficient via model selection. A non-asymptotic upper bound for the quadratic risk is proved. The procedure is minimax and adaptive in the asymptotic setting as well. A sequential asymptotically efficient kernel estimator is constructed for pointwise drift estimation in [7].

This paper deals with the estimation of the drift coefficient $S(\cdot)$ on the interval $[a,b]$ in adaptive setting for the quadratic risk

$$\mathcal{R}(\widehat{S}_T, S) = \mathbf{E}_S \|\widehat{S}_T - S\|^2 , \quad \|S\|^2 = \int_a^b S^2(x)\mathrm{d}x , \tag{2}$$

where $\widehat{S}_T$ is an estimator of S based on observations $(y_t)_{0\le t\le T}$, $a < b$ are some real numbers. Here $\mathbf{E}_S$ is the expectation with respect to the distribution law $\mathbf{P}_S$ of the process $(y_t)_{0\le t\le T}$ given by the drift S.

To obtain a good estimate of the function S, it is necessary to impose some conditions on the function S which are similar to the periodicity of the deterministic signal in the white noise model (see,e.g., [10]). One of conditions which is sufficient for this purpose is the assumption that the process $(y_t)_{0\le t\le T}$ returns to any vicinity of each point $x \in [a,b]$ infinite times. The ergodicity provides this property (see,e.g., [13]). Let $L > 1$ and $N > |a| + |b|$. We define the following functional class :

$$\begin{aligned} \varSigma_{L,N} = \{ S \in Lip_L(\mathbb{R}) \, : \, |S(N)| \le L \, ; \, \forall |x| \ge N, \, \exists \, \dot{S}(x) \in \mathbf{C}(\mathbb{R}) \\ \text{such that} - L \le \inf_{|x|\ge N} \dot{S}(x) \le \sup_{|x|\ge N} \dot{S}(x) \le -1/L \} , \end{aligned} \tag{3}$$

where

$$Lip_L(\mathbb{R}) = \left\{ f \in \mathbf{C}(\mathbb{R}) \, : \, \sup_{x,y\in\mathbb{R}} \frac{|f(x) - f(y)|}{|x-y|} \le L \right\} .$$

First of all, note that if $S \in \varSigma_{L,N}$, then the equation (1) has an unique strong solution and there exists the invariant density

$$q(x) = q_S(x) = \frac{\exp\{2\int_0^x S(z)\mathrm{d}z\}}{\int_{-\infty}^{+\infty} \exp\{2\int_0^y S(z)\mathrm{d}z\}\mathrm{d}y} . \tag{4}$$

(see,e.g., [9], Ch.4, 18, Th2).

2 Sequential procedure

We start with the partition of the interval $[a,b]$ by points $(x_k)_{1\le k\le n}$ defined as

$$x_k = a + \frac{k}{n}(b-a) , \tag{5}$$

where $n = n(T)$ is a integer-valued function of T such that

$$\lim_{T\to\infty} \frac{n(T)}{T} = 1 .$$

At any point x_k we estimate the function S by the sequential kernel estimator from [5]-[6]. We fix some $0 < t_0 < T$ and we set

$$\tau_k = \inf\{ t \ge t_0 \, : \, \int_{t_0}^t Q\left(\frac{y_s - x_k}{h}\right) \mathrm{d}s \ge H_k \}$$

and

$$S_k^* = \frac{1}{H_k} \int_{t_0}^{\tau_k} Q\left(\frac{y_s - x_k}{h}\right) \mathrm{d}y_s ,$$

where $Q(z) = \mathbf{1}_{\{|z|\le 1\}}$, $h = (b-a)/(2n)$ and H_k is a positive threshold. From (1) it is easy to obtain that

$$S^*_k \,=\, S(x_k) \,+\, \zeta_k\,.$$

The error term ζ_k is represented as the following sum of the approximation term B_k and the stochastic term:

$$\zeta_k \,=\, B_k \,+\, \frac{1}{\sqrt{H_k}}\,\xi_k\,,$$

where

$$B_k = \frac{1}{H_k}\int_{t_0}^{\tau_k} Q\left(\frac{y_s - x_k}{h}\right)\,(S(y_s)\,-\,S(x_k))\mathrm{d}s\,,$$

$$\xi_k = \frac{1}{\sqrt{H_k}}\int_{t_0}^{\tau_k} Q\left(\frac{y_s - x_k}{h}\right)\,\mathrm{d}w_s\,.$$

Taking into account that the function S is lipschitzian, we obtain the upper bound for the approximation term as

$$|B_k| \,\le\, L\,h\,.$$

It is easy to see that the random variables $(\xi_k)_{1\le k\le n}$ are i.i.d. normal $\mathcal{N}(0,1)$.

Moreover, in [7] it is established that the efficient kernel estimator has the stochastic term distributed as $\mathcal{N}(0, 2Thq_S(x_k))$. Therefore for the efficient estimation at each point x_k we need to estimate the ergodic density (4) from the observations $(y_t)_{0\le t\le t_0}$. We put

$$\tilde{q}_T(x_k) = \max\{\widehat{q}(x_k)\,,\,\epsilon_T\}\,,$$

where ϵ_T is positive, $0 < \epsilon_T < 1$,

$$\widehat{q}(x_k) = \frac{1}{2t_0 h}\int_0^{t_0} Q\left(\frac{y_s - x_k}{h}\right)\mathrm{d}s\,.$$

Now we choose the threshold H_k as

$$H_k = (T - t_0)(2\tilde{q}_T(x_k) - \epsilon^2_T)h\,,$$

where for $T \ge 32$,

$$t_0 = \max\{\min\{\ln^4 T\,,\,T/2\}\,,\,16\} \quad \text{and} \quad \epsilon_T = \sqrt{2}\,t_0^{-1/8}\,.$$

We set now

$$\Gamma = \{\max_{1\le k\le n}\,\tau_k\,\le\,T\}\,.$$

One can show that for any $m > 0$

$$\lim_{T\to\infty} T^m \sup_{S\in\Sigma_{L,N}} \mathbf{P}_S(\Gamma^c) = 0\,. \tag{6}$$

For $Y_k = S^*_k$ with $1 \le k \le n$, we come to the regression model on the set Γ :

$$Y_k = S(x_k) + \zeta_k\,, \quad \zeta_k = \sigma_k\,\xi_k + \delta_k\,, \tag{7}$$

where $(\xi_k)_{1\le k\le n}$ is a sequence of i.i.d. random variables $\mathcal{N}(0,1)$, $\delta_l = B_l$ and

$$\sigma^2_l \,=\, \frac{n}{(T-t_0)(\tilde{q}_T(x_l) - \epsilon^2_T/2)(b-a)} \le \frac{4}{\varepsilon_T(b_\varsigma - a_\varsigma)} := \sigma_*(T) = \sigma_*\,,$$

where

$$\lim_{T\to\infty}\frac{\sigma_*(T)}{T^m} = 0 \quad \text{for any} \quad m > 0\,.$$

Obviously that the random variables $(\xi_k)_{1\le k\le n}$ are independent of $(\sigma_k)_{1\le k\le n}$.

3 Oracle inequality

In this section we consider the estimation problem for the regression model (7). Now we fix a basis $(\phi_j)_{1\le j\le n}$ which is orthonormal with respect to the empirical inner product :

$$(\phi_i\,,\,\phi_j)_n\;=\;\frac{b-a}{n}\sum_{l=1}^{n}\phi_i(x_l)\,\phi_j(x_l)\;=\;\mathbf{Kr}_{ij}\,,$$

where $\mathbf{Kr}_{ij}$ is Kronecker's symbol.

By making use of this basis we apply the discrete Fourier transformation to (7) on the set Γ, i.e.

$$\widehat{\theta}_{j,n}=\frac{b-a}{n}\sum_{l=1}^{n}Y_l\phi_j(x_l)\,,\quad \theta_{j,n}=\frac{b-a}{n}\sum_{l=1}^{n}S(x_l)\,\phi_j(x_l)\,.$$

We estimate the function S in (7) on the sieve (5) by the weighted least squares estimator

$$\widehat{S}_\lambda(x_l)\;=\;\sum_{j=1}^{n}\lambda(j)\,\widehat{\theta}_{j,n}\,\phi_j(x_l)\,\mathbf{1}_\Gamma\,,\quad 1\le l\le n\,,$$

where the weight vector $\lambda=(\lambda(1),\ldots,\lambda(n))'$ belongs to some finite set $\Lambda\subset[0,1]^n$, the prime denotes the transposition. We set for any $x\in[a,b]$

$$\widehat{S}_\lambda(x)=\widehat{S}_\lambda(x_1)\mathbf{1}_{\{a\le x\le x_1\}}+\sum_{l=2}^{n}\widehat{S}_\lambda(x_l)\mathbf{1}_{\{x_{l-1}<x\le x_l\}}\,.$$

Now we have to write a rule to choose a weight vector $\lambda\in\Lambda$ to obtain a "good" estimator. To this end we set

$$\tilde{\theta}_{j,n}=\widehat{\theta}^2_{j,n}-\frac{b-a}{n}s_{j,n}\quad\text{with}\quad s_{j,n}=\frac{b-a}{n}\sum_{l=1}^{n}\sigma_l^2\,\phi_j^2(x_l)\,.$$

and we define the cost function as

$$J_n(\lambda)\;=\;\sum_{j=1}^{n}\lambda^2(j)\widehat{\theta}^2_{j,n}\;-2\sum_{j=1}^{n}\lambda(j)\,\tilde{\theta}_{j,n}\;+\;\rho P_n(\lambda)\,,$$

where

$$P_n(\lambda)=\frac{b-a}{n}\sum_{j=1}^{n}\lambda^2(j)s_{j,n}$$

and $0<\rho<1$ is some positive coefficient. We put

$$\widehat{S}_*=\widehat{S}_{\widehat{\lambda}}\quad\text{with}\quad\widehat{\lambda}=\text{agrmin}_{\lambda\in\Lambda}\,J_n(\lambda)\,. \tag{8}$$

We make use of the special weight set (see, [17], [16])

$$\Lambda\;=\;\{\lambda_\alpha\,,\,\alpha\in\mathcal{A}_\varepsilon\}\,,$$

where for $0<\varepsilon<1$ we define the set

$$\mathcal{A}_\varepsilon=\{1,\ldots,k^*\}\times\{t_1,\ldots,t_m\}\,,$$

with $t_i=i\varepsilon$, $m=[1/\varepsilon^2]$, $\varepsilon_n=1/\ln(n+2)$ and $k_n^*=\sqrt{\ln(n+2)}$.

For any $\alpha=(\beta,t)\in\mathcal{A}_\varepsilon$ we will take the weight vector $\lambda_\alpha=(\lambda_\alpha(1),\ldots,\lambda_\alpha(n))'$ of the form

$$\lambda_\alpha(j)\;=\;\begin{cases}1\,, & \text{for}\quad 1\le j\le j_0\,,\\ \left(1-(j/\omega_\alpha)^\beta\right)_+\,, & \text{for}\quad j_0<j\le n\,,\end{cases}$$

where

$$j_0 = j_0(\alpha) = [\omega_\alpha / \ln n] + 1\,, \quad \omega_\alpha = (A_\beta\, t\, n)^{1/(2\beta+1)}$$

and

$$A_\beta = \frac{(b-a)^{2\beta+1}(\beta+1)(2\beta+1)}{\pi^{2\beta}\beta}\,.$$

For the procedure (8) we show the following oracle inequality.

Theorem 1. *Assume that $S \in \Sigma_{L,N}$. Then for any $T \geq 32$ and $0 < \rho < 1/6$ the procedure $\widehat{S}_*$ satisfies, the following inequality*

$$\mathcal{R}(\widehat{S}_*, S) \,\leq\, \frac{(1+\rho)^2(1+4\rho)}{1-6\rho}\, \min_{\lambda\in\Lambda} \mathcal{R}(\widehat{S}_\lambda, S) + \frac{\mathcal{B}_T(\rho)}{T}\,, \tag{9}$$

where for any $\gamma > 0$

$$\lim_{T\to\infty} \frac{\mathcal{B}_T(\rho)}{T^\gamma} = 0\,.$$

In the following section by making use of the inequality (9) we show that the procedure (8) is asimptotically efficient.

4 Asymptotic efficiency

We define the following functional Sobolev ball

$$W^0_{k,r} = \{f \in \mathbf{C}^k_0([a,b]) \,:\, \sum_{j=0}^k \|f^{(j)}\|^2 \leq r\}\,,$$

where $r > 0$ and $k \geq 1$ are some unknown parameters, $\mathbf{C}^k_0([a,b])$ is the set of k times differentiable functions $f\,:\,\mathbb{R} \to \mathbb{R}$ such that

$$f^{(i)}(x) = 0 \quad \text{for} \quad 0 \leq i \leq k-1 \quad \text{and} \quad x \notin [a,b]\,.$$

Let S_0 be some fixed $k+1$ times continuously differentiable function from $\Sigma_{L/2,N}$. We set

$$\Theta_{k,r} = \{S = S_0 + f\,,\, f \in W^0_{k,r} \cap Lip_{L/2}(\mathbb{R})\}\,.$$

In order to formulate our asymptotic results we define the following normalizing coefficient

$$\gamma(S) \,=\, ((1+2k)r)^{1/(2k+1)} \left(\frac{J(S)k}{\pi(k+1)}\right)^{2k/(2k+1)} \tag{10}$$

with

$$J(S) = \int_a^b \frac{1}{q_S(u)} \mathrm{d}u\,.$$

It is well known that for any $S \in \Theta_{k,r}$ the optimal rate of convergence is $T^{-2k/(2k+1)}$. Now we state the following asymptotic upper bound for the quadratic risk of the estimator $\widehat{S}_*$.

Theorem 2. *The quadratic risk* (2) *for sequential procedure $\widehat{S}_*$ has the following asymptotic upper bound*

$$\limsup_{T\to\infty}\, T^{2k/(2k+1)} \sup_{S\in\Theta_{k,r}} \frac{\mathcal{R}(\widehat{S}_*, S)}{\gamma(S)} \leq 1\,. \tag{11}$$

Moreover, we show that this upper bound is sharp in the following sense.

Theorem 3.

$$\liminf_{T\to\infty} \inf_{\widehat{S}}\, T^{2k/(2k+1)} \sup_{S\in\Theta_{k,r}} \frac{\mathcal{R}(\widehat{S}, S)}{\gamma(S)} \geq 1\,. \tag{12}$$

Note that the inequalities (11) and (12) imply that the function (10) is the Pinsker constant in this case (see Pinsker (1981)).

References

[1] Arato, M., Kolmogorov, A.N. and Sinai, Ya.G. (1962) On parameter estimation of a complex stationary gaussian process. *Doklady Acad. Sciences SSSR*, **146**, 747-750.

[2] Galtchouk, L. I. and Konev, V.V. (1997) On Sequential Estimation of Parameters in Continuous-Time Stochastic Regression. *in: Yu.M. Kabanov, B.L. Rozovskii, A.N. Shiryaev (Eds), Statistics and Control of Stochastic Processes. The Liptser Festschrift, World Scientific*, 123-138.

[3] Galtchouk, L.I. and Konev, V.V. (2001) On Sequential Estimation of Parameters in Semimartingale Regression Models with Continuous Time Parameter. *The Annals of Statistics*, **29**, 1508-2035.

[4] Galtchouk, L.I. and Pergamenshchikov, S.M. (2001). Sequential nonparametric adaptive estimation of the drift coefficient in diffusion processes. *Mathematical Methods of Statistics*, **10**, 316-330.

[5] Galtchouk, L.I. and Pergamenshchikov, S.M. (2004). Nonparametric sequential estimation of the drift in diffusion via model selection. *Mathematical Methods of Statistics*, **13**, 25-49.

[6] Galtchouk, L.I. and Pergamenshchikov, S.M. (2005). Nonparametric sequential minimax estimation of the drift in diffusion processes. *Sequential Analysis*, **24**, 303-330.

[7] Galtchouk, L.I. and Pergamenshchikov, S.M. (2006). Asymptotic efficient sequential kernel estimates of the drift coefficient in ergodic diffusion processes. *Statistical Inference for Stochastic Processes*, **9**, 1-16.

[8] Galtchouk, L.I. and Pergamenshchikov, S.M. (2007). Uniform concentration inequality for ergodic diffusion processes. *Stochastic Process. Appl.*, **7**, 830-839.

[9] Gihkman, I.I. and Skorohod, A.V. *Stochastic differential equations.* Naukova Dumka, Kiev, 1968.

[10] Ibragimov, I.A. and Hasminskii, R.Z. *Statistical Estimation: Asymptotic Theory*. Springer, New York, 1979.

[11] Konev, V.V. and Pergamenshchikov, S.M. (1981) Sequential Estimation of the Parameters of Random Processes with Continuous Time. *in: Math. Statistics and Appl., Tomsk State University, Russia*, 93-101.

[12] Konev, V.V. and Pergamenshchikov, S.M. (1992) On Truncated Sequential Estimation of the Parameters of Diffusion Processes. *in: Methods of Economical Analysis, Central Economical and Mathematical Institute of Russian Academy of Science, Moscow*, 3-31.

[13] Kutoyants, Yu.A. (2003) *Statistical Inferences for Ergodic Diffusion Processes.* Springer, Berlin.

[14] Liptser, R.Sh. and Shiryaev, A.N. *Statistics of a random process.* Springer, New York, v.I, II, 1978.

[15] Novikov, A.A. (1971) Sequential estimation of the parameters of the diffusion processes. *Theory Probability and Appl.*, **16**, 394-396.

[16] Nussbaum, M. (1992) Spline smoothing in regression models and asymptotic efficiency in L_2. *Ann. Statist.* **13**, 984-997.

[17] Pinsker, M.S. (1981) Optimal filtration of square integrable signals in gaussian noise. *Problems Inform. Transmission*, **17**, 120-133.