Asymptotically efficient estimation of a drift coefficient in diffusion processes

Evgeny Pchelintsev and Svyatoslav Perelevskiy
Tomsk State University, Tomsk, Russia
e-mail: evgen-pch@yandex.ru, slavaperelevskiy@mail.ru

Abstract

The article studies the asymptotic properties of an adaptive model selection procedure for estimation an unknown drift coefficient in diffusion processes. It is shown that the procedure is asymptotically efficient, i.e. it is established that the asymptotic quadratic risk of the procedure coincides with the Pinsker constant, which provides an exact lower bound of the quadratic risk for all possible estimates.

Keywords: improved estimation, stochastic diffusion process, mean-square accuracy, oracle inequalities, Pinsker constant, asymptotic efficiency.

Introduction

Consider the problem of asymptotically efficient estimation of the unknown drift coefficient in diffusion process, described by 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 some given constant, and $S(\cdot)$ is an unknown function. Note that such models are widely used in financial markets, radio-physics, etc. [1]. The problem is to estimate the function $S(x)$, $x \in [a, b]$, from the observations $(y_t)_{0\le t\le T}$. The main goal of this paper to prove the asymptotic efficiency property of the improved model selection procedure proposed in [2] for estimating the function S in (1). The concept of asymptotic efficiency is associated with the optimal rate of convergence of the minimax risk, i.e. An important issue in the optimality results is the study of the exact asymptotic of the minimax risk. The problem of asymptotic non-parametric estimation in the model of heteroscedastic regression was studied by Efroimovich [3] and Pinsker [4]. To prove the asymptotic efficiency of the procedure, it is necessary to show that its asymptotic quadratic risk coincides with the lower bound defined by the Pinsker constant [5, 6]. In this paper, the problem is solved using an approach based on the model selection methods and sharp oracle inequalities. Recall that the model selection method appeared in the pioneering works of Akaike [7] and Mallows [8], in which proposed to introduce a penalization term in the criteria of maximum likelihood. Further, Barron, Birgé and Massart [9], Massart [10] and Kneip [11] developed this method to obtain non-asymptotic oracle inequalities in non-parametric regression models with Gaussian noise in discrete time. Unfortunately, this method cannot be applied in our case to prove an asymptotic efficiency property, since the coefficient in main term of the resulting oracle inequalities is greater then one. For this reason, in this paper we will use the method proposed in [12]. This paper deals with the estimating the unknown function $S(x)$, $a \le x \le b$, in the sense of the mean square 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 some estimate of S by 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 $\mathbf{P}_S$ of the random process $(y_t)_{0\le t\le T}$ given the drift function S. To obtain a reliable estimator of function S, it is necessary that the process (1) has the ergodicity property. For this we suppose that unknown function S belongs to the following functional class:

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

where $L > 1$, $N > |a| + |b|$, $\dot{S}(x)-$ derivative $S(x)$. For estimating the drift S in (1) Galtchouk and Pergamenshchikov [13] have proposed to apply the sequential approach. First step is a passage to a discrete time regression model by making use of the truncated sequential procedure introduced in [5]. To this end, at any point x_k of an equidistant partition of the interval $[a, b]$, we define a sequential procedure (τ_k, S^*_k) with a stopping rule τ_k and an estimators S^*_k. For $Y_k = S^*_k$ with $1 \le k \le n$, we come to the regression equation on some set $\Gamma \subseteq \Omega$ ($\sup_{S\in\Sigma_{L,N}} \mathbf{P}_S(\Gamma^c) \le \Pi_T$, where $\lim_{T\to\infty} T^m \, \Pi_T = 0$ for any $m > 0$):

$$Y_k = S(x_k) + \zeta_k \, . \tag{4}$$

Here, in contrast with the classical regression model, the noise sequence $(\zeta_k)_{1\le k\le n}$ has a complicated structure, namely,

$$\zeta_k = \sigma_k \, \xi_k + \delta_k \, , \tag{5}$$

where $(\sigma_k)_{1\le k\le n}$ is a sequence of some observed random variables, $(\delta_k)_{1\le k\le n}$ is a sequence of bounded random variables and $(\xi_k)_{1\le k\le n}$ is a sequence of i.i.d. random variables $\mathcal{N}(0, 1)$ which are independent of $(\sigma_k)_{1\le k\le n}$.

In order to estimate the function S in model (4) we make use of the model selection method based on improved weighted least squares estimates proposed [18]. Improved estimation method in nonparametric regression models has been developed in [15, 16, 17].

1 Oracle inequalities

To estimate the unknown function in model (4), we use improved weighted least squares estimates, defined in [2],

$$S^*_\lambda(x_l) = \sum_{j=1}^n \lambda(j) \, \theta^*_{j,n} \, \phi_j(x_l) \, \mathbf{1}_\Gamma , \quad 1 \le l \le n , \tag{6}$$

where $(\phi_j)_{j\leq 1}$ is an orthonormal functions system, the vector of weight coefficients $\lambda = (\lambda_1, ..., \lambda_n)$ belongs some finite set Λ from $[0,1]^n$,

$$\theta^*_{j,n} = \left(1 - \frac{c(d)}{\|\widetilde{\theta}_n\|}\mathbf{1}_{\{1\leq j\leq d\}}\right)\widehat{\theta}_{j,n}, \quad \|\widetilde{\theta}_n\|^2 = \sum_{j=1}^{d}\widehat{\theta}^2_{j,n}, \quad \widehat{\theta}_{j,n} = \frac{b-a}{n}\sum_{l=1}^{n} Y_l\phi_j(x_l)\,.$$

Here the coefficient $d \approx n^\epsilon/\ln n$, $0 < \epsilon < 1$, $c(d) \approx d/n$. Now we define the estimate for S in (1). We set for any $a \leq x \leq b$

$$S^*_\lambda(x) = S^*_\lambda(x_1)\mathbf{1}_{\{a\leq x\leq x_1\}} + \sum_{l=2}^{n} S^*_\lambda(x_l)\mathbf{1}_{\{x_{l-1}<x\leq x_l\}}\,. \tag{7}$$

In order to obtain a good estimator, we have to write a rule to choose a weight vector $\lambda \in \Lambda$ in (7). It is obvious, that the best way is to minimize the empirical squared error with respect to λ:

$$\mathrm{Err}_n(\lambda) = \|S^*_\lambda - S\|^2_n \to \min\,.$$

Making use of (7) and the Fourier transformation of S imply

$$\mathrm{Err}_n(\lambda) = \sum_{j=1}^{n}\lambda^2(j)\theta^{*2}_{j,n} - 2\sum_{j=1}^{n}\lambda(j)\theta^*_{j,n}\,\theta_{j,n} + \sum_{j=1}^{n}\theta^2_{j,n}\,.$$

Since the coefficient $\theta_{j,n}$ is unknown, we need to replace the term $\theta^*_{j,n}\,\theta_{j,n}$ by some its estimator which we choose as

$$\widetilde{\vartheta}_{j,n} = \widehat{\theta}_{j,n}\theta^*_{j,n} - \frac{b-a}{n}s_{j,n} \quad \text{with} \quad s_{j,n} = \frac{b-a}{n}\sum_{l=1}^{n}\sigma^2_l\,\phi^2_j(x_l)\,.$$

One has to pay a penalty for this substitution in the empirical squared error. Finally, we define the cost function of the form

$$J_n(\lambda) = \sum_{j=1}^{n}\lambda^2(j)\theta^{*2}_{j,n} - 2\sum_{j=1}^{n}\lambda(j)\,\widetilde{\vartheta}_{j,n} + \rho P_n(\lambda)\,,$$

where the penalty term is defined as

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

and $0 < \rho < 1$ is some positive constant which will be chosen later. We set

$$\widehat{\lambda} = \mathrm{argmin}_{\lambda\in\Lambda}\, J_n(\lambda)$$

and define an estimator of S of the form (7):

$$S^*(x) = S^*_{\widehat{\lambda}}(x) \quad \text{for} \quad a \leq x \leq b\,. \tag{8}$$

Now we obtain the non asymptotic upper bound for the quadratical risk of the estimator (8).

Theorem 1. *Let $\Lambda \subset [0,1]^n$ be any finite set such that the first $d \le n$ components of the weight vector λ are equal to 1. Then, for any $n \ge 3$ and $0 < \rho < 1/6$, the estimator* (8) *satisfies the following oracle inequality*

$$\mathbf{E}_S \|S^* - S\|_n^2 \le \frac{1+6\rho}{1-6\rho} \min_{\lambda \in \Lambda} \mathbf{E}_S \|\widehat{S}_\lambda - S\|_n^2 + \frac{\Psi_n(\rho)}{n}\,,$$

where $\lim_{n\to\infty} \Psi_n(\rho)/n = 0$.

Now we consider the estimation problem (1) via model (4). We apply the estimating procedure (8) with special weight set introduced in [5] to the regression scheme (4). Denoting $S^*_\alpha = S^*_{\lambda_\alpha}$ we set

$$S^* = S^*_{\widehat{\alpha}} \quad \text{with} \quad \widehat{\alpha} = \text{argmin}_{\alpha \in \mathcal{A}_\varepsilon} J_n(\lambda_\alpha)\,.$$

Theorem 2. *Assume that $S \in \Sigma_{L,N}$ and the number of the points $n = n(T)$ in the model* (4) *. Then the procedure S^* satisfies, for any $T \ge 32$, the following inequality*

$$\mathcal{R}(S^*, S) \;\le\; \frac{(1+\rho)^2(1+6\rho)}{1-6\rho} \min_{\alpha \in \mathcal{A}_\varepsilon} \mathcal{R}(S^*_\alpha, S) + \frac{\mathcal{B}_T(\rho)}{n}\,,$$

where $\lim_{T\to\infty} \mathcal{B}_T(\rho)/n(T) = 0$.

2 Asymptotic efficiency

In order to study the asymptotic efficiency we define the following functional Sobolev ball

$$W^0_{k,r} = \{f \in \mathbf{C}^k_0([a,b]) \,:\, \sum_{j=0}^{k} \|f^{(j)}\|^2 \le r\}\,, \tag{9}$$

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

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

We will call such functions *periodic on the interval* $[a,b]$. Let S_0 be a fixed $k+1$ times continuously differentiable function from $\Sigma_{L,N}$. We set

$$\Theta_{k,r} = \{S = S_0 + f\,,\, f \in W^0_{k,r}\}\,. \tag{10}$$

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

$$\gamma(S) \;=\; ((1+2k)r)^{1/(2k+1)} \left(\frac{J(S)k}{\pi(k+1)}\right)^{2k/(2k+1)} \tag{11}$$

with

$$J(S) = \int_a^b \frac{1}{q_S(x)} \mathrm{d}x\,, \quad q_S(x) \;=\; \frac{\exp\{2\int_0^x S(z)\mathrm{d}z\}}{\int_{-\infty}^{+\infty} \exp\{2\int_0^y S(z)\mathrm{d}z\}\mathrm{d}y}.$$

It is well known that for any $S \in \Theta_{k,r}$ the optimal rate of convergence is $T^{-2k/(2k+1)}$ (see, for example, [18]). On the basis of the model selection procedure (8) in the next section we construct the adaptive procedure S^* for which we obtain the following asymptotic upper bound for the quadratic risk.

Theorem 3. *The quadratic risk* (2) *for the estimating procedure* S^* *has the following asymptotic upper bound*

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

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

Theorem 4. *For any estimator* $\widehat{S}$ *of* S *measurable with respect to* $\mathcal{F}^y_T$,

$$\liminf_{T\to\infty} \inf_{\widehat{S}} T^{2k/(2k+1)} \sup_{S\in\Theta_{k,r}} \frac{\mathcal{R}(\widehat{S},S)}{\gamma(S)} \geq 1\,, \tag{13}$$

where $\mathcal{F}^y_T$ *is a* $\sigma-$*field generated by observations* $(y_t)_{0\leq t\leq T}$.

Remark 1. *It should be noted that the choice of the functional class* $\Theta_{k,r}$ *in the form of* (10) *is related to the ergodicity of the process* (1). *This property is provided when the drift derivative is negative on the outside of a finite interval. The last excludes the choice of periodic functions as a class of admissible drifts. For this reason, we use the Sobolev ball of periodic functions with a non periodic center* S_0 *as a class of admissible drift functions.*

Remark 2. *Note that the inequalities* (14) *and* (13) *imply that the function* (11) *is the Pinsker constant in this case (cf. [4]).*

Corollary 1. *From Theorems 2 and 3 it follows that the procedure for choosing a model* S^*, *defined in* (8), *is asymptotically efficient, i.e.*

$$\lim_{T\to\infty} T^{2k/(2k+1)} \sup_{S\in\Theta_{k,r}} \frac{\mathcal{R}(S^*,S)}{\gamma(S)} = 1\,. \tag{14}$$

3 Numerical simulations

We suppose that in the model (1)

$$S(x) = x^2\sin(2\pi x) + x^2(1-x)\cos(4\pi x).$$

For weight coefficients we choose $n = T$,

$$k^* = 100 + \sqrt{\ln n}, \quad \varepsilon = \frac{1}{\ln n}, \quad m = \ln^2 n, \quad \omega_\alpha = 100 + (A_\beta t n)^{\frac{1}{2\beta+1}}\,.$$

The empirical risk:

$$\mathcal{R}(S^*,S) = \frac{1}{1000}\sum_{m=1}^{1000} \|S^*_m - S\|^2_n.$$

Table 1 shows the results of the behavior of empirical mean-square risks for the proposed estimation procedure (8).

Table 1: Empirical quadratic asymptotic risks

n	501	1001	2001	10001
$T^{2k/(2k+1)}\frac{\mathcal{R}(S^*,S)}{\gamma(S)})$	4.7257	2.0856	1.0072	0.9012

From Table 1 it is clear that with an increase in the number of observations n, the normalized empirical mean-square risks tend to unity, which confirm numerically the Corollary 1.

The figures show the behavior of observation processes $(y_t)_{0\le t\le 1}$, function S (red line), and improved estimate S^* (green line):

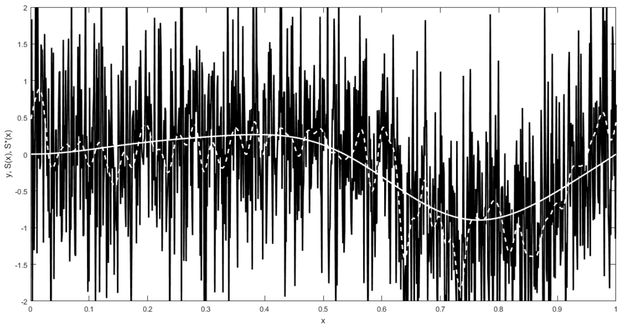

Figure 1: n=501

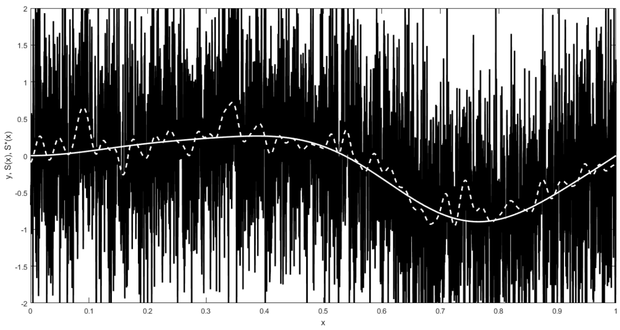

Figure 2: n=1001

Acknowledgements

This work was supported by the RSF grant no 17-11-01049.

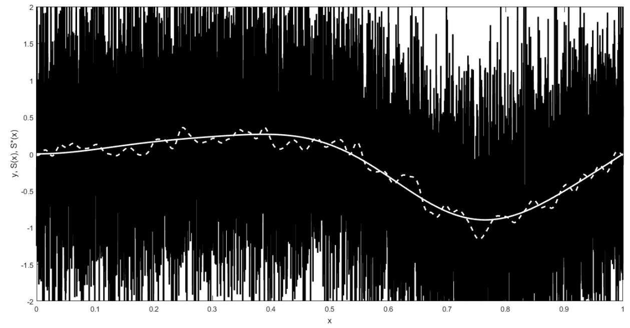

Figure 3: n=10001

References

[1] Kutoyants Yu. (2004). *Statistical inference for ergodic diffusion processes.* Springer-Verlag, London.

[2] Pchelintsev E.A., Perelevskiy S.S., Makarova I.A. (2018). Improved nonparametric estimation of the drift in diffusion processes. *Uchenye Zapiski Kazanskogo Universiteta. Seriya Fiziko-Matematicheskie Nauki.* Vol. **160**, No **2**, pp. 364–372.

[3] Efroimovich S. (2007). Sequential design and estimation in heteroscedastic nonparametric regression. *Sequential Analysis.* No **26**, pp. 3–25. Springer, New York, 1998.

[4] Pinsker M.S. (1981) Optimal filtration of square integrable signals in gaussian white noise. *Problems Transimis. information.* No **17**, pp. 120–133.

[5] Galtchouk L., Pergamenshchikov S. (2006) Asymptotically efficient sequential kernel estimates of the drift coefficient in ergodic diffusion processes. *Statistical Inference for Stochastic Process.* Vol. **9**, No. **1**, pp. 1–16.

[6] Ibragimov I.A., Khasminskii R.Z. (1981). *Statistical Estimation: Asymptotic Theory.* Springer, New York.

[7] Akaike H. (1974). A new look at the statistical model identification. *IEEE Trans. on Automatic Control.* Vol. **19**, No. **7**, pp. 716–723.

[8] Mallows C. (1973). Some comments on C_p. *Technometrics.* Vol. **15**, pp. 661–675.

[9] Barron A., Birgé L., Massart P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Relat. Fields.* Vol. **113**, pp. 301–415.

[10] Massart P. (2005). *A non-asymptotic theory for model selection.* European Congress of Mathematics. Eur. Math. Soc., Zurich.

[11] Kneip A. (1994). Ordered linear smoothers. *Annals of Statistics*. Vol. **22**, pp. 835–866.

[12] Galtchouk L.I., Pergamenshchikov S.M. (2009). Adaptive asymptotically efficient estimation in heteroscedastic nonparametric regression. *J. Korean Stat. Soc.* Vol. **38**, No **4**, pp. 305–322.

[13] Galtchouk L.I., Pergamenshchikov S.M. (2011). Adaptive sequential estimation for ergodic diffusion processes in quadratic metric. *Journal of Nonparametric Statistics*. Vol. **23**, No **2**, pp. 255–285.

[14] Pchelintsev E., Perelevskiy S. (2019). Adaptive estimation in heteroscedastic nonparametric regression. *Tomsk State University Journal of Mathematics and Mechanics*. No. **57**, pp. 38–52.

[15] Pchelintsev E.A., Pergamenshchikov S.M. (2018). Oracle inequalities for the stochastic differential equations. *Statistical Inference for Stochastic Processes*. Vol. **21**, No **2**, pp. 469–483.

[16] Pchelintsev E., Pchelintsev V., Pergamenshchikov S. (2018). Non asymptotic sharp oracle inequalities for the improved model selection procedures for the adaptive nonparametric signal estimation problem. *Communications - Scientific Letters of the University of Zilina*. Vol. **20**, No **1**, pp. 73–77.

[17] Pchelintsev E. A., Pergamenshchikov S. M. (2019). Improved model selection Method for an adaptive estimation in semimartingale regression models. *Tomsk State University Journal of Mathematics and Mechanics*. Vol. **58**, pp. 14–31.

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