https://doi.org/10.26552/com.C.2018.1.73-77

KOMUNIKÁCIE COMMUNICATIONS

Evgeny Pchelintsev - Valeriy Pchelintsev - Serguei Pergamenshchikov*

NON ASYMPTOTIC SHARP ORACLE INEQUALITIES FOR THE IMPROVED MODEL SELECTION PROCEDURES FOR THE ADAPTIVE NONPARAMETRIC SIGNAL ESTIMATION PROBLEM

In this paper, we consider the robust adaptive non parametric estimation problem for the periodic function observed with the Levy noises in continuous time. An adaptive model selection procedure, based on the improved weighted least square estimates, is proposed. Sharp oracle inequalities for the robust risks have been obtained.

Keywords: *improved non-asymptotic estimation, weighted least squares estimates, robust quadratic risk, non-parametric regression, Levy process, model selection, sharp oracle inequality, asymptotic efficiency*

1. Introduction

In this paper, we consider a signal statistical treatment problem in the framework of a nonparametric regression model in continuous time, i.e.

$$dy_t = S(t)dt + d\xi_t, \quad 0 \le t \le n \tag{1}$$

where $S(\cdot)$ is an unknown 1-periodic signal, $(\xi_t)_{0\le t\le n}$ is an unobserved noise and n is the duration of observation. The problem is to estimate the function S on the observations $(y_t)_{0\le t\le n}$. Note that if $(\xi_t)_{0\le t\le n}$ is a Brownian motion, then we obtain the well-known "signal + white noise" model which is very popular in statistical radio-physics (see, for example, [1-3] and etc.). In this paper, we assume that in addition to the intrinsic noise in the radio-electronic system, approximated usually by the Gaussian white or color noise, the useful signal S is distorted by the impulse flow described by the Levy process, i.e. we assume that the noise process $(\xi_t)_{0\le t\le n}$ is defined as

$$\xi_t = \rho_1 w_t + \rho_2 z_t \quad \text{and} \quad z_t = x^*(\mu - \tilde{\mu})_t \tag{2}$$

where ρ_1 and ρ_2 are some unknown constants, $(w_t)_{t\ge 0}$ is a standard Brownian motion, $\mu(ds\,dx)$ is a jump measure with deterministic compensator $\tilde{\mu}(ds\,dx) = ds\Pi(dx), \Pi(\cdot)$ is a Levy measure, i.e. some positive measure on $\mathbb{R}_* = \mathbb{R}\setminus\{0\}$, such that

$$\Pi(x^2) = 1 \quad \text{and} \quad \Pi(x^6) < \infty \tag{3}$$

Here we use the notation $\Pi(|x|^m) = \int_{\mathbb{R}_*} |y|^m \Pi(dy)$. Note that the Levy measure $\Pi(\mathbb{R}_*)$ could be equal to +∞. We use * for the stochastic integrals with respect to random measures (see [4], Chs. 2 and 3), i.e.

$$x * (\mu - \tilde{\mu})_t = \int_0^t \int_{\mathbb{R}_*} y(\mu - \tilde{\mu})(ds, dy) \tag{4}$$

In the sequel we will denote by Q the distribution of the process $(\xi_t)_{0\le t\le n}$ and by Q_n^* we denote all these distributions for which the parameters $\rho_1 \ge \varsigma_*$ and $\rho_1^2 + \rho_2^2 \le \varsigma^*$ where ς_* and ς^* are some fixed positive bounds. The cause of the appearance of a pulse stream in the radio-electronic systems can be, for example, either external unintended (atmospheric) or intentional impulse noise and the errors in the demodulation and the channel decoding for the binary information symbols. Note that, for the first time the impulse noises for detection signal problems have been introduced on the basis of compound Poisson processes was

* **[1]Evgeny Pchelintsev, [2]Valeriy Pchelintsev, [3]Serguei Pergamenshchikov**
[1]Department of Information Technologies and Business Analytics, Tomsk State University, Russia
[2]Department of Mathematics and Informatics, Tomsk Polytechnic University, Russia
[3]Laboratoire de Mathematiques Raphael Salem, Universite de Rouen, Saint Etienne du Rouvray, France and International Laboratory of Statistics of Stochastic Processes and Quantitative Finance of Tomsk State University, Russia
E-mail: evgen-pch@yandex.ru

by Kassam in [2]. However, the compound Poisson process can describe only the large impulses influence of fixed single frequency. There should be taking into account that in telecommunication systems, impulses are without limitations on frequencies. So, one needs to extend the framework of the observation model by making use of the Levy processes as in Equations (2). In this paper, we consider the estimation problem in the adaptive setting, i.e. when the regularity of S is unknown. Since the distribution Q of the noise process $(\xi_t)_{0\le t\le n}$ is unknown we use the robust estimation approach developed for nonparametric problems in [5]. We define the robust risk as

$$\mathcal{R}^*(\hat{S}_n, S) = \sup_{Q\in \mathcal{Q}_n^*} \mathcal{R}_Q(\hat{S}_n, S) \tag{5}$$

where $\hat{s}_n$ is an estimation of the unknown signal S, i.e. some function of $(y_t)_{0\le t\le n}$, $\mathcal{R}_Q(\cdot,\cdot)$ is the usual quadratic risk defined as

$$\mathcal{R}^*(\hat{S}_n, S) := \mathbf{E}_{Q,S}\|\hat{S}_n - S\|^2 \text{ and } \|S\|^2 = \int_0^1 S^2(t)dt \tag{6}$$

Here $\mathbf{E}_{Q,S}$ stands for the expectation with respect to the distribution $\mathbf{P}_{Q,S}$ of the process in the Equation (1) with a fixed distribution Q of the noise $(\xi_t)_{0\le t\le n}$ and a given function S.

In this paper, we consider minimax optimization criteria which aim to minimize the robust risk which is defined in the Equation (5) (see, for example, in [6]). To do this we use the model selection methods. The interest to such statistical procedures is explained by the fact that they provide adaptive solutions for a nonparametric estimation through oracle inequalities which give a non-asymptotic upper bound for a quadratic risk including a minimal risk over chosen family of estimators. It should be noted that the model selection methods for parametric models were proposed, for the first time, by Akaike [7]. Then, these methods had been developed by Barron, Birge and Massart [8] and Fourdrinier and Pergamenshchikov [9] for the nonparametric estimation and oracle inequalities for the quadratic risks. Unfortunately, the oracle inequalities obtained in these papers cannot provide the efficient estimation in the adaptive setting, since the upper bounds in these inequalities have some fixed coefficients in the main terms which are more than one. In order to obtain the efficiency property for estimation procedures, one has to obtain the sharp oracle inequalities, i.e. in which the factor at the principal term on the right-hand side of the inequality is close to unity. For this reason, one needs to use the general semimartingale approach for the robust adaptive efficient estimation of the nonparametric signals in continuous time proposed by Konev and Pergamenshchikov [5]. The goal of this paper is to develop a new sharp model selection method for estimating the unknown signal S using the improved estimation approach. Usually, the model selection procedures are based on the least square estimators. However, in this paper, we propose to use the improved least square estimators which enable us to considerably improve the non-asymptotic estimation accuracy. Such idea was proposed, for the first time, in [9]. Our goal is to develop these methods for non-Gaussian regression models in continuous time and to obtain the sharp oracle inequalities. It should be noted that to apply the improved estimation methods to the non-Gaussian regression models in continuous time one needs to modify the well-known James-Stein procedure introduced in [10] in the way proposed in [11, 12]. So, by using these estimators we construct the improved model selection procedure and we show that the constructed estimation procedure is optimal in the sense of the sharp non-asymptotic oracle inequalities for the robust risks which are defined in the Equation (5).

2. Improved estimation

Let $(\phi_j)_{j\ge 1}$ be an orthonormal basis in $\mathbf{L}_2[0,1]$. We extend these functions by the periodic way on $\mathbb{R}$, i.e. $\phi_j(t) = \phi_j(t+1)$ for any $t\in\mathbb{R}$. For estimating the unknown function S in the Equation (1) we consider it's Fourier expansion (see, for example, [13])

$$S(t) = \sum_{j=1}^{\infty} \theta_j \phi_j(t) \text{ and } \theta_j = (S,\phi_j) = \int_0^1 S(t)\phi_j(t)dt \tag{7}$$

The corresponding Fourier coefficients can be estimated as

$$\hat{\theta}_{j,n} = \frac{1}{n}\int_0^n \phi(t)dy_t \tag{8}$$

We define a class of weighted least squares estimates for $S(t)$ as

$$\hat{S}_\lambda = \sum_{j=1}^{n} \lambda(j)\hat{\theta}_{j,n}\phi_j \tag{9}$$

where the weights $\lambda\in\mathbb{R}^n$ belong to some finite set Λ from $[0, 1]^n$.

Now, for the first d Fourier coefficients in Equations (7) we use the improved estimation method proposed for parametric models in [12]. To this end we set $\tilde{\theta}_n = (\hat{\theta}_{j,n})_{1\le j\le d}$. In the sequel we will use the norm $|x|_d^2 = \sum_{j=1}^{d} x_j^2$ for any vector $x = (x_j)_{1\le j\le d}$ from $\mathbb{R}^n$. Now we define the shrinkage estimators as

$$\theta^*_{j,n} = (1 - g(j))\hat{\theta}_{j,n} \tag{10}$$

where $g(j) = \frac{\mathbf{c}_n}{|\tilde{\theta}_n|_d}\mathbf{1}_{\{1\le j\le d\}}$, $\mathbf{1}_A$ is the indicator of the set A and $\mathbf{c}_n$ is some known parameter such that $\mathbf{c}_n \approx \frac{d}{n}$ as $n\to\infty$. Now we introduce a class of shrinkage weighted least squares estimates for S as

$$S^*_\lambda = \sum_{j=1}^{n} \lambda(j)\theta^*_{j,n}\phi_j \tag{11}$$

We denote the difference of quadratic risks of the estimates in Equation (11) and Equation (9) as $\Delta_Q(S) := \mathcal{R}_Q(S^*_\lambda, S) - \mathcal{R}_Q(\hat{S}_\lambda, S)$. Now for this deviation we obtain the following result.

Theorem 1. *Assume that for any vector $\lambda \in \Lambda$ there exists some fixed integer $d = d(\lambda)$ such that their first d components equal to one, i.e. $\lambda(j) = 1$ for $1 \le j \le d$ for any $\lambda \in \Lambda$. Then for any $n \ge 1$ and $r>0$*

$$\sup_{Q \in \mathcal{Q}_n} \sup_{\|S\| \le r} \Delta_Q(S) < -\mathbf{c}_n^2 \tag{12}$$

The inequality (12) means that non-asymptotically, i.e. for any $n \ge 1$ the estimate in the Equation (11) outperforms in mean square accuracy the estimate in the Equation (9). Moreover, as we will see below, $n\mathbf{c}_n \to \infty$ as $d \to \infty$. This means that improvement is considerable may be better than for the parametric regression [11].

3. Model selection

This Section gives the construction of a model selection procedure for estimating a function S in the Equation (1) on the basis of improved weighted least square estimates and states the sharp oracle inequality for the robust risk of proposed procedure.

The model selection procedure for the unknown function S in the Equation (1) will be constructed on the basis of a family of estimates $(S^*_\lambda)_{\lambda \in \Lambda}$.

The performance of any estimate S^*_λ will be measured by the empirical squared error

$$\mathrm{Err}_n(\lambda) = \| S^*_\lambda - S \|^2 \tag{13}$$

In order to obtain a good estimate, we have to write a rule to choose a weight vector $\lambda \in \Lambda$ in the Equation (6). It is obvious, that the best way is to minimize the empirical squared error with respect to λ. Making use the estimate definition in the Equation (6) and the Fourier transformation of S implies

$$\mathrm{Err}_n(\lambda) = \sum_{j=1}^{n} \lambda^2(j)(\theta^*_{j,n})^2 - 2\sum_{j=1}^{n} \lambda(j)\theta^*_{j,n}\theta_j + \sum_{j=1}^{n} \theta_j^2 \tag{14}$$

Since the Fourier coefficients $(\theta_j)_{j \ge 1}$ are unknown, the weight coefficients $(\lambda_j)_{j \ge 1}$ cannot be found by minimizing this quantity. To circumvent this difficulty one needs to replace the terms $\theta^*_{j,n}\theta_j$ by their estimators $\tilde{\theta}_{j,n}$. We set

$$\tilde{\theta}_{j,n} = \theta^*_{j,n}\hat{\theta}_{j,n} - \frac{\hat{\sigma}_n}{n} \tag{15}$$

where $\hat{\sigma}_n$ is the estimate for the noise variance of $\sigma_Q = \mathbf{E}_Q \xi^2_{j,n} = \rho_1^2 + \rho_2^2$ which we choose in the following form

$$\hat{\sigma}_n = \sum_{j=[\sqrt{n}]+1}^{n} \hat{t}^2_{j,n} \quad \text{and} \quad \hat{t}_{j,n} = \frac{1}{n}\int_0^n \mathrm{Tr}_j(t)\,dy_t \tag{16}$$

Here we denoted by $(\mathrm{Tr}_j)_{j \ge 1}$ the trigonometric basis in $\mathbf{L}_2[0,1]$. For this change in the empirical squared error, one has to pay some penalty. Thus, one comes to the cost function of the form

$$J_n(\lambda) = \sum_{j=1}^{n} \lambda^2(j)(\theta^*_{j,n})^2 - 2\sum_{j=1}^{n} \lambda(j)\tilde{\theta}_{j,n} + \delta \hat{P}_n(\lambda) \tag{17}$$

where δ is some positive constant, $\hat{P}_n(\lambda)$ is the penalty term defined as

$$\hat{P}_n(\lambda) = \frac{\hat{\sigma}_n |\lambda|_n^2}{n} \tag{18}$$

Substituting the weight coefficients, minimizing the cost function

$$\lambda^* = \mathrm{agr\,min}_{\lambda \in \Lambda} J_n(\lambda) \tag{19}$$

in the Equation (11) leads to the improved model selection procedure

$$S^* = S^*_{\lambda^*} \tag{20}$$

It will be noted that λ^* exists because Λ is a finite set. If the minimizing sequence in the Equation (19) λ^* is not unique, one can take any minimizer. In the case, when the value of σ_Q is known, one can take $\hat{\sigma}_n = \sigma_Q$ and $P_n(\lambda) = \sigma_Q |\lambda|_n^2 n^{-1}$.

Theorem 2. *For any $n \ge 2$ and $0 < \delta < 1/3$, the robust risks defined in the Equation (5) of estimate in the Equation (20) for continuously differentiable function S satisfies the oracle inequality*

$$\mathcal{R}^*(S^*_{\lambda^*}, S) \le \frac{1+3\delta}{1-3\delta} \min_{\lambda \in \Lambda} \mathcal{R}^*(S^*_\lambda, S) + \frac{B^*_n}{n\delta} \tag{21}$$

where the rest term is such that $B^*_n n^{-\varepsilon} \to 0$ as $n \to \infty$ for any $\varepsilon > 0$.

The inequality (21) means that the procedure in the Equation (20) is optimal in the oracle inequalities sense. This property enables to provide asymptotic efficiency in the adaptive setting, i.e. when information about the signal regularity is unknown.

4. Monte Carlo simulations

In this section we report the results of a Monte Carlo experiment to assess the performance of the proposed model selection procedure in the Equation (20). In the Equation (1) we choose 1-periodic function S which, for $0 \le t \le 1$, is defined as $S(t) = \left| t - \frac{1}{2} \right|$, if $\frac{1}{4} \le t \le \frac{3}{4}$ and $S(t) = \frac{1}{4}$ elsewhere. We simulate the Equation (1) with the noise process defined as

$$\xi_t = 0.5 w_t + 0.5 z_t \tag{22}$$

Table 1 Empirical risks

n	$\overline{R}(\hat{S},S)$	$\overline{R}(S^*,S)$	$\overline{R}(\hat{S},S)/\overline{R}(S^*,S)$
20	0.0462	0.0331	1.40
100	0.0262	0.0133	1.97
200	0.0126	0.00824	1.53
1000	0.00129	0.00094	1.37

where $z_t = \sum_{j=1}^{N_t} Y_j$, N_t is a Poisson process with the intensity λ=1 and $(Y_j)_{j\geq 1}$ is i.i.d. Gaussian (0,1). We use the model selection procedure defined in the Equation (20) with the weights proposed in [5]: $k^* = 100 + \sqrt{\ln n}$, $\varepsilon = 1/\ln n$ and $m = [1/\varepsilon^2]$. We used the cost function with $\delta = (3 + \ln n)^{-2}$. We define the empirical risk as $\overline{R}(\tilde{S},S) = \frac{1}{p}\sum_{j=1}^{p} \hat{\mathbf{E}}(\tilde{S}_n(t_j) - S(t_j))^2$ and $\hat{\mathbf{E}}(\tilde{S}_n(\cdot) - S(\cdot))^2 = \frac{1}{N}\sum_{l=1}^{N}(\tilde{S}_n^l(\cdot) - S(\cdot))^2$ with the frequency of observations *p*=100001 and numbers of replications *N*=10000.

Table 1 give the values for the sample risks for different numbers of observation period *n*.

5. Conclusion

In conclusion, we would like to emphasize that in this paper we developed new model selection procedures based on the improved versions of the least square estimators. It turns out that the improvement effect in the nonparametric estimation is more important than for the parameter estimation problems since the accuracy improvement is proportional to the parameter dimension. We remember that for the nonparametric estimation this dimension tends to infinity, but in the parametric case it is always fixed. Therefore, the gain in the non-asymptotic quadratic accuracy from the application of the improved estimation methods is much more significant in statistical treatment problems of nonparametric signals. Moreover, as shown by the results of numerical simulation, the improved estimate for non-asymptotic accuracy exceeds the usual least squares estimators approximately in 1.5-2 times for the chosen experimental model.

Acknowledgement

The results of section 2 of this work are supported by the Ministry of Education and Science of the Russian Federation in the framework of the research project no. 2.3208.2017/4.6 and the RFBR Grant 16-01-00121 A. The results of Section 3 are supported by the RSF grant number 14-49-00079 (National Research University "MPEI", Moscow, Russia). The last author is partially supported by the Russian Federal Professor program (project no. 1.472.2016/1.4, Ministry of Education and Science) and by the project XterM-Feder, University of Rouen.

Ce projet est cofinancé par l'Union européenne.
L'Europe s'engage en Normandie
avec le Fonds européen de Développement Régional (FEDER)

RÉGION NORMANDIE

References

[1] CHERNOYAROV, O. V., VACULIK, M., SHIRIKYAN, A., SALNIKOVA, A. V.: Statistical Analysis of Fast Fluctuating Random Signals with Arbitrary - Function Envelope and Unknown Parameters. Communications - Scientific Letters of the University of Zilina, 17(1a), 35-43, 2015.

[2] KASSAM, S. A.: Signal Detection in Non-Gaussian Noise. Springer, New York, 1988.

[3] IBRAGIMOV, I. A., KHASMINSKII, R. Z.: Statistical Estimation: Asymptotic Theory. Springer, New York, 1981.

[4] CONT, R., TANKOV, P.: Financial Modelling with Jump Processes. Chapman & Hall, London, 2004.

[5] KONEV, V. V., PERGAMENSHCHIKOV, S.: Robust Model Selection for a Semimartingale Continuous Time Regression from Discrete Data. Stochastic Processes and their Applications, 125, 294-326, 2015.

[6] JANACEK, J., KVET, M.: Min-Max Optimization of Emergency Service System by Exposing Constraints. Communications - Scientific Letters of the University of Zilina, 17(2), 15-22, 2015.

[7] AKAIKE, H.: *A* New Look at the Statistical Model Identification. IEEE Transactions on Automatic Control, 19(7), 716 – 723, 1974.

[8] BARRON, A., BIRGE, L., MASSART, P.: Risk Bounds for Model Selection via Penalization. Probability Theory and Related Fields, 113, 301-415, 1999.

[9] FOURDRINIER, D., PERGAMENSHCHIKOV, S.: Improved Selection Model Method for the Regression with Dependent Noise. Annals of the Institute of Statistical Mathematics, 59(3), 435-464. 2007.

[10] JAMES, W., STEIN, C.: Estimation with Quadratic Loss. Proceedings of the Fourth Berkeley Symposium Mathematics, Statistics and Probability, California, USA, 1, 361-380, 1966.
[11] KONEV, V., PERGAMENSHCHIKOV, S., PCHELINTSEV, E.: Estimation of a Regression with the Pulse Type Noise from Discrete Data. Theory of Probability and Its Applications, 58(3), 442-457, 2014.
[12] PCHELINTSEV, E.: Improved Estimation in a Non-Gaussian Parametric Regression. Statistical Inference for Stochastic Processes, 16(1), 15-28, 2013.
[13] MARCOKOVA, M.: Equiconvergence of Two Fourier series. Journal of Approximation Theory, 80(2), 151-163, 1995.