МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
Международная лаборатория статистики случайных
процессов и количественного финансового анализа

Международная научная конференция «Робастная статистика и финансовая математика – 2017»

(03–05 июля 2017 г.)

Сборник статей

Под редакцией
д-ра физ.-мат. наук, профессора С.М. Пергаменщикова,
канд. физ.-мат. наук, доцента Е.А. Пчелинцева

Томск
Издательский Дом Томского государственного университета
2017

УДК 519
ББК 22.17
Меж43

Меж43 Международная научная конференция «Робастная статистика и финансовая математика – 2017»: сборник статей / под ред. С.М. Пергаменщикова, Е.А. Пчелинцева. – Томск : Издательский Дом Томского государственного университета, 2017. – 87 с.

ISBN 978-5-94621-639-5

В сборнике представлены статьи, посвященные актуальным проблемам математической статистики и финансовой математики, а также современным подходам и методам решения фундаментальных и прикладных задач.

Для студентов, аспирантов, специалистов в области математической статистики и ее приложений.

Сборник подготовлен к печати при поддержке Российского научного фонда, проект № 17-11-01049.

УДК 519
ББК 22.17

Организационный комитет конференции:

к.ф.-м.н., доцент Пчелинцев Е. А. (председатель оргкомитета);
Макарова И. А.;
Перелевский С. С. (ответственный секретарь);
Повзун М. А.

ISBN 978-5-94621-639-5

Содержание

Model selection for a semi - Markov continuous time regression observed in the discrete time moments *

Barbu V.S., Beltaief S., Pergamenshchikov S.

Université de Rouen Normandie, Rouen,
Tomsk State University, Tomsk
e-mail: Serge.Pergamenshchikov@univ-rouen.fr

Abstract

In this article we consider the nonparametric robust estimation problem for regression models in continuous time with semi-Markov noises observed in discrete time moments. An adaptive model selection procedure is proposed. A sharp non-asymptotic oracle inequality for the robust risks is obtained. We obtain sufficient conditions on the frequency observations under which the robust efficiency is shown. It turns out that for the semi-Markov models the robust minimax convergence rate may be faster or slower than the classical one.

Keywords: Non-asymptotic estimation; Robust risk; Model selection; Sharp oracle inequality; Asymptotic efficiency

In this paper we consider the semi-Markov regression model in continuous time introduced in [1], i.e.

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

where $S(\cdot)$ is an unknown 1-periodic function defined on $\mathbf{R}$ with values on $\mathbf{R}$, $(\xi_t)_{t\ge 0}$ is the unobserved noise process defined through a certain semi-Markov process in [1].

Our problem in the present paper is to estimate the unknown function S in the model (1) on the basis of observations

$$(y_{t_j})_{0\le j\le np}, \quad t_j = j\Delta, \quad \Delta = \frac{1}{p}, \tag{2}$$

where the integer $p \ge 1$ is the observation frequency. Firstly, this problem was considered in the framework "signal+white noise" (see, for example, [3] or [11]). Later, to introduce a dependence in the continuous time regression model in [10], [6], [4], [5] [7], the Ornstein-Uhlenbeck processes has been used to model the "color noise". Moreover, in order to introduce the dependence and the jumps in the regression model

*The work was supported by the RSF grant 17-11-01049.

(1), the papers [8] and [9] use the non Gaussian Ornstein-Uhlenbeck processes defined in [2]. The problem in all these papers is that the introduced Ornstein-Uhlenbeck type of dependence decreases with a geometric rate. So, asymptotically when the duration of observations goes to infinity, we obtain the same "signal+white noise" model very quick. To keep the dependence for sufficiently large duration of observations, in [1] it was proposed the model (1) with a semi-Markov component in the jumps of the noise process $(\xi_t)_{t\geq 0}$.

The main goal of this paper is to develop adaptive robust method from [1], that was based on continuous observations, to the estimation problem based on discrete observations given in (2). In this paper we use quadratic risk defined as

$$\mathcal{R}_Q(\tilde{S}_n, S) = \mathbf{E}_{Q,S}\, \|\tilde{S}_n - S\|^2\,, \tag{3}$$

where $\tilde{S}_n(\cdot)$ is some estimate (i.e. any periodical function measurable with respect to the observations $\sigma\{y_{t_0}, \ldots y_{t_{pn}}\}$), $\|f\|^2 = \int_0^1 f^2(s)ds$ and $\mathbf{E}_{Q,S}$ is the expectation with respect to the distribution $\mathbf{P}_{Q,S}$ of the process (1) corresponding to the unknown noise distribution Q in the Skorokhod space $\mathcal{D}[0,n]$. We assume that this distribution belongs to some distribution family $\mathcal{Q}_n$ specified in [1].

To study the properties of the estimators uniformly over the noise distribution (what is really needed in practice), we use the robust risk defined as

$$\mathcal{R}^*_n(\tilde{S}_n, S) = \sup_{Q\in\mathcal{Q}_n} \mathcal{R}_Q(\tilde{S}_n, S)\,. \tag{4}$$

Thus the goal of this paper is to develop a robust efficient model selection method based on the observations (2) for the model (1) with the semi-Markov components in the jumps of the noise $(\xi_t)_{t\geq 0}$. We use the approach proposed by Konev and Pergamenshchikov in [9] for continuous-time regression models observed in the discrete time moments. Unfortunately, we cannot use directly this method for semi-Markov regression models, since their tool essentially uses the fact that the Ornstein-Uhlenbeck dependence decreases with geometrical rate and obtain sufficiently quickly the "white noise" case. In the present paper, in order to obtain the sharp non-asymptotic oracle inequalities, we use the renewal methods from [1] developed for the model (1). As a consequence, we can obtain the constructive sufficient conditions that provide the robust efficiency for proposed model selection procedures.

In this paper we construct some special family of the weight least square estimators $(\hat{S}_\lambda)_{\lambda\in\Lambda}$, where Λ is some finite set from $[0,1]^n$. In the sequel we assume that the number of the vector of the $\nu = \mathrm{card}\Lambda$

is a function of n, i.e. $\nu = \nu_n$, such that for any $\gamma > 0$

$$\lim_{n\to\infty} \frac{\nu}{n^\gamma} = 0\,. \tag{5}$$

The model select procedure for this family is defined as

$$\hat{S}_* = \hat{S}_{\hat{\lambda}} \quad \text{and} \quad \hat{\lambda} = \text{argmin}_{\lambda\in\Lambda} J_{n,\delta}(\lambda)\,, \tag{6}$$

where $J_{n,\delta}(\lambda)$ is the cost function defined for any in [1] and $0 < \delta < 1$ is the penalty parameter.

We assume that the renewal distribution in the semi-markov component of the noise process in (1) has a density g that satisfies the following conditions.

$\mathbf{H}_1$) *Assume that, for any $x \in \mathbf{R}$, there exist the finite limits*

$$g(x-) = \lim_{z\to x-} g(z) \quad \textit{and} \quad g(x+) = \lim_{z\to x+} g(z)$$

and, for any $K > 0$, there exists $\delta = \delta(K) > 0$ for which

$$\sup_{|x|\le K} \int_0^\delta \frac{|g(x+t) + g(x-t) - g(x+) - g(x-)|}{t}\, dt \,<\, \infty. \tag{7}$$

$\mathbf{H}_2$) *For any $\gamma > 0$,*

$$\sup_{z\ge 0} z^\gamma |2g(z) - g(z-) - g(z+)| \,<\, \infty\,.$$

$\mathbf{H}_3$) *There exists $\beta > 0$ such that $\int_{\mathbf{R}} e^{\beta x}\, g(x)\, dx < \infty$.*

$\mathbf{H}_4$) *There exists $t^* > 0$ such that the function $\hat{g}(\theta - it)$ belongs to $\mathbf{L}_1(\mathbf{R})$ for any $0 \le t \le t^*$, where*

$$\hat{g}(\theta) = \frac{1}{2\pi} \int_{\mathbf{R}} e^{i\theta x}\, g(x)\, dx\,.$$

Theorem 1. *Assume that the function S is continuously differentiable and that Conditions $\mathbf{H}_1$)–$\mathbf{H}_4$) hold true. Then there exists some constant $l^* > 0$ such that for any noise distribution Q, the weight vectors set Λ, for any periodic function S for any $n \ge 1$, $p \ge 3$ and $0 < \delta \le 1/6$, the procedure (6) satisfies the following oracle inequality*

$$\mathcal{R}_Q(\hat{S}_*, S) \le \frac{1+3\delta}{1-3\delta} \min_{\lambda\in\Lambda} \mathcal{R}_Q(\hat{S}_\lambda, S) + l^* \frac{\nu}{\delta n} \tag{8}$$

and

$$\mathcal{R}^*(\hat{S}_*, S) \le \frac{1+3\delta}{1-3\delta} \min_{\lambda\in\Lambda} \mathcal{R}^*(\hat{S}_\lambda, S) + \frac{U_n^*}{\delta n}\,, \tag{9}$$

where the the rest term U_n^* is such that for any $\gamma > 0$

$$\lim_{n\to\infty} \frac{U_n^*}{n^\gamma} = 0\,.$$

Remark 1. *The oracle inequalities* (8) *and* (9) *are called sharp oracle inequalities, since the coefficients in the main terms in the right side closed to one. As we will see below this property allows to show asymptotic efficiency in the adaptive setting.*

In order to obtain the efficiency property, we specify the weight coefficients $(\lambda(j))_{1\le j\le n}$ in the procedure (6). Consider, for some fixed $0<\varepsilon<1$, a numerical grid of the form

$$\mathcal{A}=\{1,\ldots,k^*\}\times\{\varepsilon,\ldots,m\varepsilon\}, \tag{10}$$

where $m=[1/\varepsilon^2]$. We assume that both parameters $k^*\ge 1$ and ε are functions of n, i.e. $k^*=k^*(n)$ and $\varepsilon=\varepsilon(n)$, such that

$$\begin{cases} \lim_{n\to\infty} k^*(n)=+\infty\,, \quad \lim_{n\to\infty} \dfrac{k^*(n)}{\ln n}=0\,, \\ \\ \lim_{n\to\infty} \varepsilon(n)=0 \quad \text{and} \quad \lim_{n\to\infty} n^{\gamma}\varepsilon(n)\ =+\infty \end{cases} \tag{11}$$

for any $\gamma>0$. One can take, for example, for $n\ge 2$

$$\varepsilon(n)=\frac{1}{\ln n} \quad \text{and} \quad k^*(n)=k^*_0+\sqrt{\ln n}\,, \tag{12}$$

where $k^*_0\ge 0$ is some fixed constant. For each $\alpha=(\beta,\tau)\in\mathcal{A}$, we introduce the weight sequence

$$\lambda_\alpha=(\lambda_\alpha(j))_{1\le j\le p}$$

with the elements

$$\lambda_\alpha(j)=\mathbf{1}_{\{1\le j<j_*\}}+\left(1-(j/\omega_\alpha)^\beta\right)\,\mathbf{1}_{\{j_*\le j\le\omega_\alpha\}}, \tag{13}$$

where $j_*=1+[\ln\upsilon_n]$, $\omega_\alpha=(\mathbf{d}_\beta\,\tau\upsilon_n)^{1/(2\beta+1)}$,

$$\mathbf{d}_\beta=\frac{(\beta+1)(2\beta+1)}{\pi^{2\beta}\beta} \quad \text{and} \quad \upsilon_n=n/\varsigma^*\,.$$

Threshold ς^* is a function of n, i.e. $\varsigma^*=\varsigma^*(n)$ such that for any $\gamma>0$

$$\liminf_{n\to\infty} n^{\gamma}\,\varsigma^*(n)=+\infty \quad \text{and} \quad \limsup_{n\to\infty} n^{-\gamma}\,\varsigma^*(n)=0\,. \tag{14}$$

Now we define the set Λ as

$$\Lambda\ =\ \{\lambda_\alpha\,,\,\alpha\in\mathcal{A}\}\,. \tag{15}$$

These weight coefficients are used in [8, 9] for continuous time regression models to show the asymptotic efficiency. Note also that in this case the cardinal of the set Λ is $\nu=k^*m$. Therefore, the conditions in (11) yield the property (5). Moreover, to obtain the efficiency for the model selection procedure we assume the following condition on the frequency of the observations.

$\mathbf{H}_5$) *Assume that for any $n\ge 3$ the observation frequence $p\ge n^{5/6}$.*

To this end, we assume that the unknown function S in the model

(1) belongs to the Sobolev ball

$$W_r^k = \{ f \in \mathcal{C}^k_{per}[0,1] \,,\, \sum_{j=0}^{k} \|f^{(j)}\|^2 \le r \} \,, \tag{16}$$

where $r > 0$, $k \ge 1$ are some parameters, $\mathcal{C}^k_{per}[0,1]$ is the set of k times continuously differentiable functions $f : [0,1] \to \mathbf{R}$ such that $f^{(i)}(0) = f^{(i)}(1)$ for all $0 \le i \le k$. The function class W_r^k can be written as an ellipsoid in l_2, i.e.

$$W_r^k = \{ f \in \mathcal{C}^k_{per}[0,1] \,:\, \sum_{j=1}^{\infty} a_j \, \theta_j^2 \le r \} \tag{17}$$

where $a_j = \sum_{i=0}^{k} \left(2\pi[j/2]\right)^{2i}$.

Similarly to [8,9] we will show here that the asymptotic sharp lower bound for the robust risk (4) is given by

$$r^*_k = l \, \left((2k+1)r\right)^{1/(2k+1)} \left(\frac{k}{(k+1)\pi} \right)^{2k/(2k+1)} \,. \tag{18}$$

Note that this is the well-known Pinsker constant obtained for the nonadaptive filtration problem in "signal + small white noise" model (see, for example, [11]).

Let Π_n be the set of all estimators $\hat{S}_n$ measurable with respect to the sigma-algebra $\sigma\{y_t \,,\, 0 \le t \le n\}$ generated by the process (1).

Theorem 2. *Under Conditions* (14)

$$\liminf_{n\to\infty} \upsilon_n^{2k/(2k+1)} \inf_{\hat{S}_n \in \Pi_n} \sup_{S \in W_r^k} \mathcal{R}^*_n(\hat{S}_n, S) \ge r^*_k \,, \tag{19}$$

where $\upsilon_n = n/\varsigma^*$.

Note that, if the parameters r and k are known, i.e. for the nonadaptive estimation case, in order to obtain the efficient estimation for the "signal+white noise" model, Pinsker proposed in [11] to use the estimate $\hat{S}_{\lambda_0}$ with the weights (13) in which

$$\lambda_0 = \lambda_{\alpha_0} \quad \text{and} \quad \alpha_0 = (k, \tau_0) \,, \tag{20}$$

where $\tau_0 = [r/\varepsilon]\varepsilon$. For the semi-markov regression model (1) we show the same result.

Proposition 3. *The estimator* $\hat{S}_{\lambda_0}$ *satisfies the following asymptotic upper bound*

$$\lim_{n\to\infty} \upsilon_n^{2k/(2k+1)} \sup_{S \in W_r^k} \mathcal{R}^*_n(\hat{S}_{\lambda_0}, S) \le r^*_k \,.$$

For the adaptive estimation we user the model selection procedure (6) with the parameter δ defined as a function of n satisfying

$$\lim_{n\longrightarrow\infty} \delta_n = 0 \quad \text{and} \quad \lim_{n\longrightarrow\infty} n^{\gamma} \, \delta_n = 0 \tag{21}$$

for any $\gamma > 0$. For example, we can take $\delta_n = (6 + \ln n)^{-1}$. Now the oracle inequality (9) for the family (15) and Proposition 3 imply the following upper bound.

Theorem 4. *Assume that Conditions* $\mathbf{H}_1)$–$\mathbf{H}_5)$ *hold true. Then the robust risk defined in* (4) *for the procedure* (6) *with the coefficients* (13) *and the parameter* $\delta = \delta_n$ *satisfying* (21) *has the following asymptotic upper bound*

$$\limsup_{n\to\infty} \upsilon_n^{2k/(2k+1)} \sup_{S\in W_r^k} \mathcal{R}_n^*(\hat{S}_*, S) \le r_k^* \,. \tag{22}$$

Theorem 2 and Theorem 4 imply the following result.

Corollary 5. *Under the conditions of Theorem 4, the model selection procedure* $\hat{S}_*$ *is efficient, i.e.*

$$\lim_{n\to\infty} \frac{\inf_{\hat{S}_n\in\Pi_n} \sup_{S\in W_r^k} \mathcal{R}_n^*(\hat{S}_n, S)}{\sup_{S\in W_r^k} \mathcal{R}_n^*(\hat{S}_*, S)} = 1 \,. \tag{23}$$

Remark 2. *It is well known that the optimal (minimax) risk convergence rate for the Sobolev ball* W_r^k *is* $n^{2k/(2k+1)}$ *(see, for example, [11]). We see here that the efficient robust rate is* $\upsilon_n^{2k/(2k+1)}$*, i.e. if the distribution upper bound* $\varsigma^* \to 0$ *as* $n \to \infty$ *we obtain a faster rate with respect to* $n^{2k/(2k+1)}$*, and if* $\varsigma^* \to \infty$ *as* $n \to \infty$ *we obtain a slower rate. In the case when* ς^* *is constant the robuste rate is the same as the classical non robuste convergence rate.*

Conclusion. In the conclusion we would like to emphasize that in this paper :

- we construct a selection model procedure based on the weight least square estimators;
- we find conditions for which we obtained an sharp non asymptotic oracle inequalities for the simple quadratic risks and for the robust risks as well;
- using the Pinsker method we obtain a lower bound for the robust quadratic risks, then, through the obtained sharp oracle inequalities we show that the risk upper bound for the constructed procedure matters this lower bound, i.e. the procedure is efficient in the adaptative setting.

References

1. Barbu V.S. Robust adaptive efficient estimation for semi-Markov nonparametric regression models / Barbu V.S., Beltaif S., Pergamenchtchikov S.M. // Statistical inference for stochastic processes. 25 March 2017.— (Preprint / Statistical inference for stochastic processes) – URL: https://arxiv.org/pdf/1604.04516.pdf

2. Barndorff-Nielsen O. E. and Shephard N. Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial mathematics // J. Royal Stat. Soc. 2001. V. B 63. P. 167—241.

3. Ibragimov I. A. Statistical Estimation: Asymptotic Theory / Ibragimov I. A., Khasminskii R. Z.—Springer, Berlin, New York. 1981.

4. Höpfner R. and Kutoyants Yu. A. On LAN for parametrized continuous periodic signals in a time inhomogeneous diffusion // Statist. Decisions. 2009. V. 27. No. 4. P. 309—326.

5. Höpfner R., Kutoyants Yu. A. Estimating discontinuous periodic signals in a time inhomogeneous diffusion // Statistical Inference for Stochastic Processes. 2010. V.13. No. 3. P. 193—230.

6. Konev V. V., Pergamenshchikov S. M. Sequential estimation of the parameters in a trigonometric regression model with the gaussian coloured noise // Statistical Inference for Stochastic Processes. 2003. V. 6. P. 215—235.

7. Konev V. V., Pergamenshchikov S. M. General model selection estimation of a periodic regression with a Gaussian noise // Annals of the Institute of Statistical Mathematics. 2010. V. 62. P. 1083—1111.

8. Konev V. V., Pergamenshchikov S.M. Efficient robust nonparametric estimation in a semimartingale regression model // Ann. Inst. Henri Poincaré Probab. Stat. 2012. V. 48. No. 4. P. 1217—1244.

9. Konev V.V. and Pergamenshchikov S.M. Robust model selection for a semimartingale continuous time regression from discrete data //Stochastic processes and their applications. 2015. V. 294—326.

10. Kutoyants Yu. A. Estimation of signal parameter in Gaussian noise // Problems of Information Transmission. 1977. V. 13. P. 266-271.

11. Pinsker M. S. Optimal filtration of square integrable signals in gaussian white noise // Problems of Transimission information. 1981. V. 17. P. 120–133.

Adaptive prediction of continuous-time processes *

Dogadova T. V., Vasiliev V. A.

Department of Applied Mathematics and Cybernetics,
Tomsk State University, Tomsk
e-mail: aurora1900@mail.ru, vas@mail.tsu.ru

Abstract

The problem of asymptotic efficiency of adaptive one-step predictors for stable continuous-time processes with unknown parameters is considered. The proposed criteria of optimality are based on the loss function, defined as a linear combination of sample size and squared prediction error's sample mean.

Keywords: Adaptive optimal prediction, stochastic differential equations, stochastic differential equations with time delay, risk function; guaranteed parameter estimation.

The main purpose of this paper is to present applications of the truncated estimation method for constructing optimal adaptive predictors for the stochastic processes related with continuous-time dynamical systems. The proposed procedures are based on the truncated estimators which have been developed in order to estimate ratio type functionals from a wide class by dependent observations and by samples of fixed size so that they had guaranteed accuracy in the sense of the L_{2m}-norm, $m \geq 1$.

Estimators with the said quality can be applied in adaptive prediction of dynamical systems with unknown parameters. Then it is possible to optimize the prediction procedure in terms of sample mean of squared prediction errors and sample size.

In this paper we construct and investigate real-time predictors which only use past values of the multivariate diffusion-type and scalar time-delayed processes.

We consider the problem of minimization of the risk function associated with predictors of values of the process and size of a sample. The proposed procedure is shown to be asymptotically risk efficient as the cost of prediction error tends to infinity.

*Research was supported by the RNF, Project No. 17-11-01049.

Parameter estimation of a multivariate diffusion process. Assume the model

$$dx(t) = \Lambda x(t)dt + dW_t, \quad t \geq 0 \tag{1}$$

with an unknown $p \times p$ matrix parameter Λ, where $x(0)$ is zero mean random vector with variance matrix Σ_0 having all the moments, W_t is a multivariate Wiener process with independent components, $x(0)$ and W_t are mutually independent. Suppose that the process(1) is stable, i.e. all the eigenvalues of the matrix Λ have negative real parts. Consider the estimation problem of Λ with a guaranteed accuracy. Define the estimator of Λ as

$$\hat{\Lambda}'_T = \overline{G}_T^{-1}\overline{\Phi}_T, \quad T > 0,$$

where

$$\overline{G}_T = \frac{1}{T}G_T, \quad G_T = \int\limits_0^T x(t)x'(t)dt,$$

$$\overline{\Phi}_T = \frac{1}{T}\Phi_T, \quad \Phi_T = \int\limits_0^T x(t)dx'(t), \quad T > 0.$$

Define the truncated estimator

$$\tilde{\Lambda}'_T = \hat{\Lambda}'_T \cdot \chi(\overline{\Delta}_T \geq H), \quad \overline{\Delta}_T = \det(\overline{G}_T). \tag{2}$$

There exists a given number C_Λ such that for every $T > 0$,

$$\sup_{\Lambda \in \Lambda_0} E_\Lambda \|\tilde{\Lambda}_T - \Lambda\|^2 \leq \frac{C_\Lambda}{T},$$

where Λ_0 is a compact set from the stable region of the process $(x(t))$.

Parameter estimation of one-parameter delay differential equation. Assume $w = w_t,\ t \geq 0$ is a real-valued standard Wiener process, b is a real number and $x = (x_t,\ t \geq -r)$ is a solution of the stochastic delay differential equation

$$dx_t = bx_{t-r} + dw_t,\ t \geq 0 \tag{3}$$

with some fixed initial condition $x_t = X_0(t),\ t \in [-r, 0]$, where $X_0(\cdot)$ is a cadlag stochastic process independent of $w(\cdot)$. Note that the process (3) is stable when the parameter $b \in (-\pi/2r, 0)$.

The solution x of (3) exists, it is pathwise uniquely determined and for $t \geq 0$ can be represented as

$$X_t = x_0(t)X_0(0) + b\int\limits_{-r}^{0} x_0(t-s-r)X_0(s)ds + \int\limits_0^t x_0(t-s)dw_s.$$

Here $x_0(t),\ t \geq -r$ denotes the fundamental solution of the deterministic equation

$$\dot{x}(t) = bx_0(t-r),\ x_0(0) = 1 \text{ and } x_0(t) = 0,\ t \in [-r, 0).$$

The truncated estimator of the unknown parameter b can be defined as follows

$$b_t = \frac{\int\limits_r^t x_{v-r} dx_v}{\int\limits_r^t x_{v-r}^2 dv} \chi\left(\int\limits_r^t x_{v-r}^2 dv \geq t \log^{-1} t\right),\ t > \max(u, r). \quad (4)$$

Define the numbers $\sigma_0^2 = \int\limits_0^\infty x_0^2(v) dv$ and $s_0 = \max\{r, \exp(\sigma_0^2)\}$.

For $t > u + s_0$ estimators (4) have the properties

$$E(\hat{b}_t - b)^{2p} \leq \frac{C_p}{t^p},\ p \geq 1.$$

Optimal adaptive prediction of the multivariate diffusion process. Assume the model (1). The problem is to construct a predictor for $x(t)$ defined in (1) by observations $x^{t-u} = (x(s))_{0 \leq s \leq t-u}$ which is optimal in the sense of the risk function introduced below. Here $u > 0$ is a fixed time delay.

Using the solution of (1) we obtain the following representation

$$x(t) = Bx(t-u) + \xi_{t,t-u}, \quad t \geq u, \quad (5)$$

where $\xi_{t,t-u} = \int\limits_{t-u}^t e^{\Lambda(t-s)} dW_s$, $B = e^{\Lambda u}$. Applying properties of the Ito integral it is easy to verify that

$$E\xi_{t,t-u} = 0, \qquad \sigma^2 := E\|\xi_{t,t-u}\|^2 = \int\limits_0^u \|e^{\Lambda s}\|^2 ds.$$

Optimal in the mean square sense predictor $x^0(t)$ for $x(t)$ is the conditional mathematical expectation of $x(t)$ under the condition of x^{t-u} which can be found using (5)

$$x^0(t) = Bx(t-u), \quad t \geq u.$$

Since the parameters Λ and B are unknown, we define the adaptive predictor

$$\hat{x}(t) = B_{t-u} x(t-u), \quad t \geq u, \quad (6)$$

where B_{t-u} is the estimator of B defined as follows

$$B_t = \mathrm{e}^{\tilde{\Lambda}_t u},$$

$\tilde{\Lambda}_t$ is defined in (2).

Denote the prediction errors of x_t^0 and $\hat{x}_t$ as

$$e^0(t) = x(t) - x^0(t) = \xi_{t,t-u},$$

$$e(t) = x(t) - \hat{x}(t) = (B - B_{t-u})x_(t-u) + \xi_{t,t-u}, \quad t \geq u.$$

Now we define the loss function

$$L_t = \frac{A}{t}\overline{e}^2(t) + t, \quad t \geq u,$$

where

$$\overline{e}^2(t) = \frac{1}{t}\int_u^t \|e(s)\|^2 ds$$

and the parameter $A > 0$ is the cost of prediction error.

We also define the risk function $R_t = EL_t$ as

$$R_t = \frac{A}{t}E\overline{e}^2(t) + t$$

and consider optimization problem

$$R_t \to \min_t .$$

For the optimal predictors $x^0(t)$ it is possible to optimize the corresponding risk function

$$R_t^0 = E\left(\frac{A}{t}(e^0(t))^2 + t\right) = \frac{A\sigma^2}{t} + t \to \min_t, \tag{7}$$

where $(e^0(t))^2 = \frac{1}{t}\int\limits_u^t (e_s^0)^2 ds.$

In this case the optimal duration of observations T_A^0 and the corresponding value of R_t^0 are respectively

$$T_A^0 = A^{\frac{1}{2}}\sigma, \quad R_{T_A^0}^0 = 2A^{\frac{1}{2}}\sigma, \tag{8}$$

where $\sigma := \sqrt{\sigma^2}$.

However, since Λ and consequently σ are unknown and both T_A^0 and $R_{T_A^0}^0$ depend on a, the optimal predictor can not be used in practice. We define the estimator T_A of the optimal time T_A^0 as

$$T_A = \inf\{t \geq t_A : t \geq A^{1/2}\hat{\sigma}_{t_A}\}, \tag{9}$$

where $t_A := A^{1/2} \cdot \log^{-1} A = o(A^{1/2})$. Here $\hat{\sigma}_t := \sqrt{\sigma_t^2}$ is the estimator of unknown σ, where

$$\hat{\sigma}_t^2 = \int\limits_0^u \|\mathrm{e}^{\Lambda_t s}\|^2 ds.$$

Our purpose is to prove the asymptotic equivalence of T_A and T_A^0 in the almost surely and mean senses and the optimality of the presented adaptive prediction procedure in the sense of equivalence of R_A^0 and the obviously modified risk

$$\bar{R}_A = A \cdot E\frac{1}{T_A}e^2(T_A) + ET_A. \tag{10}$$

Theorem 1. *Assume the model (1) and t_A that is defined in (9). Let the predictors $\hat{x}_t$ be defined by (6), the times T_A^0, T_A and the risk functions R_t^0, $\bar{R}_A$ defined by (8), (9) and (7), (10) respectively. Then*

$$\frac{T_A}{T_A^0} \xrightarrow[A\to\infty]{} 1 \quad \text{-}a.s., \qquad \frac{ET_A}{T_A^0} \xrightarrow[A\to\infty]{} 1, \qquad \frac{\bar{R}_A}{R_A^0} \xrightarrow[A\to\infty]{} 1.$$

One-parameter delay differential equation. Assume the model (3). In this section we construct optimal and adaptive predictors for the process (3). Optimal in the mean square sense predictor is the conditional mathematical expectation

$$z_t^{(k)}(t-u) = E(x_t|x_{t-u}),$$

that satisfies the following equation

$$z_t^{(k)}(t-u) = x_{t-u} + b \int\limits_{t-u}^{t-(u-r)\wedge t} x_{v-r}dv + b \int\limits_{t-(u-r)\wedge t}^{t} z_{v-kr}^{(0)}dv +$$

$$+b\sum_{i=1}^{k-1} \int\limits_{t}^{t-r} z_{v-(k-i)r}^{(i)}dv, \;\; kr < u \le (k+1)r, \;\; t > u. \tag{11}$$

Since the parameter b in the definition of the optimal predictors $z_t^{(k)}(t-u)$ is unknown, we define the adaptive predictor by formula (11) replacing the unknown b with $\hat{b}_{t-u}$, where $\hat{b}_{t-u}$ is the projection

$$\hat{b}_{t-u} = \text{proj}_{[-\pi/2r,0]} b_{t-u}$$

of the truncated estimator of the parameter b which is proposed in (4).

Define the numbers $\sigma_0^2 = \int\limits_r^\infty x_0^2(v)dv$ and $s_0 = \max\{r, \exp(\sigma_0^{-2})\}$.

Denote the adaptive prediction error and rewrite it in the form

$$e_t^{(k)}(t-u) := x_t - \hat{z}_t^{(k)}(t-u) = e_t^0 + \hat{e}_t^{(k)}(t-u),$$

where $e_t^0(t-u) = x_t - E(x_t|x_{t-u})$ and $\hat{e}_t^{(k)}(t-u) = = z_t^{(k)}(t-u) - \hat{z}_t^{(k)}(t-u)$. Then for every fixed $k \ge 0$

$$\overline{\lim_{t\to\infty}}\, t(E(e_t^{(k)}(t-u))^2 - \sigma_0^2) \le C.$$

If it is known that $b \in [b_0, b_1]$, $-\pi/2r < b_0 < b_1 < 0$, then for $t-u > s_1 = \max\{r, \exp(\sigma_1^{-2})\}$, where $\sigma_1^2 = \inf\limits_{b\in[b_0,b_1]} \sigma_0^2$ the non-asymptotic property is fulfilled

$$E(e_t^{(k)}(t-u))^2 - \sigma_0^2 \le \frac{C}{t}.$$

These properties can be used to prove optimality in the sense of considered above risk functions.

Non-asymptotic distribution of the sequential estimates of parameters in a first-order unstable autoregression with unknown mean *

Konev V. V., Nazarenko B. N.
Tomsk State University, Tomsk
e-mail: vvkonev@mail.tsu.ru

Abstract

The paper proposes new sequential estimates for the parameters in an autoregressive process $AR(1)$ with unknown mean. The property of uniform asymptotic normality is proved in the case of non-stable process with unspecified noise distributions. In case of Gaussian noises non-asymptotic distribution of estimates has been derived.

Keywords: sequential estimate, autoregression, fixed-accuracy estimation, uniform asymptotic normality.

Introduction. It is well-known that autoregressive models are widely used in the problems of automatic control, time series analysis, filtering theory, image processing, spectral analysis and others because they provide adequate description of different processes in applications. For estimating parameters in these models a wide variety of methods have been developed. Much efforts have been made to investigate the asymptotic properties of the least squares estimates (LSE) of an autoregressive parameter in the AR(1) model obeying the equation

$$x_k = \theta x_{k-1} + \varepsilon_k,\ k \geqslant 1, \tag{1}$$

where $\{\varepsilon_k\}$ is a sequence of i.i.d. random variables with $\mathsf{E}\varepsilon_k = 0$, $0 < \mathsf{E}\varepsilon_k^2 = \sigma^2 < \infty$. It is well-known (see [5] for details and other references) that the LSE based on $(x_0, \dots, x_n)$

$$\widehat{\theta}_n = \frac{\sum\limits_{k=1}^{n} x_{k-1}x_k}{\sum\limits_{k=1}^{n} x_{k-1}^2} \tag{2}$$

has three different limiting distributions, depending on the value of unknown parameter θ each demanding its own normalizing factor.

*This study was supported by the ministry of education and science of the Russian Federation, goszadanie no 2.3208.2017/4.6

Lai and Siegmund [4] proposed for estimating parameter θ in the model (1) to use a sequential sampling scheme based on the stopping time

$$\tau = \tau(h) = \{n \geqslant 1: \ I_n \geqslant h\}, \ h > 0,$$

where $I_n = \sum\limits_{k=1}^{n} x_{k-1}^2$ is the observed Fisher information. They have proved that the sequential estimate $\widehat{\theta}_{\tau(h)}$ obtained from (2) by replacing n with $\tau(h)$, has the important property of uniform asymptotic normality

$$\lim_{h\to\infty} \sup_{|\theta|\leqslant 1} \left[P_\theta \left(I_{\tau(h)}^{1/2} (\theta_{\tau(h)} - \theta) \leqslant z \right) - \Phi\left(\frac{z}{\sigma}\right) \right] = 0.$$

In this paper we consider the problem of estimating two unknown parameters θ_1 and θ_2 in the model of type

$$x_k = \theta_1 x_{k-1} + \theta_2 + \varepsilon_k, \ k \geqslant 1, \tag{3}$$

where $(\varepsilon_k)_{k\geqslant 1}$ is a sequence of i.i.d. random variables with $\mathsf{E}\varepsilon_k = 0$, $0 < \mathsf{E}\varepsilon_k^2 = \sigma^2 < \infty$.

As is pointed out in [5], the multiparametric case is more complicated and should be treated differently to obtain the estimator with the property of uniform asymptotic normality. Galtchouk and Konev [2] treated problem of estimating parameters θ_1 and θ_2 in the model (3) from the standpoint of sequential analysis. To construct the sequential estimate of the vector $\theta = (\theta_1, \theta_2)'$ they use the LS-estimate based on the sample $(x_0, \dots, x_n)$

$$\widehat{\theta}(n) = M_n^{-1} \sum_{k=1}^{n} Y_{k-1} x_k,$$

where $Y_k = (x_k, 1)'$, the prime denotes the transposition; M_n is the sample Fisher information matrix, that is

$$M_n = \sum_{k=1}^{n} Y_{k-1} Y_{k-1}^{'}. \tag{4}$$

The sequential estimate is defined as

$$\widehat{\theta}_{\tau(h)} = M_{\tau(h)}^{-1} \sum_{k=1}^{\tau(h)} Y_{k-1} x_k \tag{5}$$

where $h > 0$, $\tau(h)$ is the stopping time of the form

$$\tau(h) = \inf \left\{ n \geqslant 1: \sum_{k=1}^{n} \|Y_{k-1}\| \geqslant h \right\}, \ \inf\{\emptyset\} = \infty,$$

$$\|Y_{k-1}\|^2 = 1 + x_{k-1}^2.$$

The paper [2] has established the uniform asymptotic normality property of estimate (5) which is given in the following theorem.

Theorem 1. *Let the process (3) be stable, $|\theta_1| < 1$, and (ε_n) be i.i.d. with mean 0 and unit variance and be independent of x_0. Then for any $0 \leqslant l < \infty$, $0 < s < 1$,*

$$\lim_{h\to\infty} \sup_{\theta\in\Theta_{l,s}} \sup_{t\in R^2} \left[P_\theta \left(M^{1/2}_{\tau(h)} (\widehat{\theta}_{\tau(h)} - \theta) \leqslant t \right) - \Phi_2(t) \right] = 0,$$

where $\Phi_2(t) = \Phi(t_1)\Phi(t_2)$, Φ is the standard normal distribution function,

$$\Theta_{l,s} = \{(\theta_1, \theta_2)' : \ -1 + s \leqslant \theta_1 \leqslant 1 - s, \ |\theta_2| \leqslant l\}.$$

Remark 1. *In order to estimate parameters θ_1 and θ_2 in (3) with prescribed mean square precision one can apply the theory of guaranteed estimation developed in the paper of Konev and Pergamenshchikov [3].*

We aim here at constructing sequential estimates which possess the following properties.

First, they are asymptotically uniformly normal for an unstable process (3).

Second, in the case of Gaussian disturbances $\{\varepsilon_n\}_{n\geqslant 1}$ and appropriate choice of normalizing factor, the estimates have standard two-dimensional normal distribution.

Construction of sequential estimates We will need the following sample Fisher information matrices

$$M_{1,n} = \sum_{k=1}^{n} Y_{k-1} Y_{k-1}^{'}, \quad M_{2,m} = \sum_{k=n+1}^{m} Y_{k-1} Y_{k-1}^{'}, \ m > n, \tag{6}$$

calculated by the sets of observations $(x_0, \ldots, x_{n-1})$ and $(x_n, x_{n+1}, \ldots, x_{m-1})$ respectively.

Further we introduce two stopping rules $\tau_1 = \tau_1(h)$ and $\tau_2 = \tau_2(h)$, for each $h > 0$ as

$$\tau_1(h) = \inf \left\{ n \geqslant 1 : \sum_{k=1}^{n} x_{k-1}^2 \geqslant h \right\},$$

$$\tau_2(h) = \inf \left\{ n > \tau_1(h) : \sum_{j=\tau_1(h)+1}^{n} 1 \geqslant h \right\}, \ \inf\{\varnothing\} = \infty,$$

and modify the matrices (6) as follows

$$\widehat{M}_{1,\tau_1(h)} = \sum_{k=1}^{\tau_1(h)} \sqrt{\beta_{1,k}(h)} Y_{k-1} Y_{k-1}^{'}, \ \widehat{M}_{2,\tau_2(h)} = \sum_{j=\tau_1(h)+1}^{\tau_2(h)} \sqrt{\beta_{2,j}(h)} Y_{j-1} Y_{j-1}^{'}, \tag{7}$$

where

$$\beta_{1,k}(h) = \begin{cases} 1 & if \quad k < \tau_1(h), \\ \alpha_1(h) & if \quad k = \tau_1(h); \end{cases} \quad \beta_{2,j}(h) = \begin{cases} 1 & if \quad j < \tau_2(h), \\ \alpha_2(h) & if \quad j = \tau_2(h). \end{cases}$$

Here $\alpha_1(h)$ and $\alpha_2(h)$, $0 < \alpha_i(h) \leqslant 1$, $i = 1, 2$, are the correction factors defined by the equations

$$\sum_{k=1}^{\tau_1(h)-1} x_{k-1}^2 + \alpha_1(h) x_{\tau_1(h)-1}^2 = h, \qquad \sum_{j=\tau_1(h)+1}^{\tau_2(h)-1} 1 + \alpha_2(h) = h. \tag{8}$$

Let $\upsilon(h) = (\upsilon_1(h), \upsilon_2(h))'$ be the vector with the components

$$\upsilon_1(h) = \sum_{k=1}^{\tau_1(h)} \sqrt{\beta_{1,k}(h)} x_{k-1} x_k; \quad \upsilon_2(h) = \sum_{j=\tau_1(h)+1}^{\tau_2(h)} \sqrt{\beta_{2,j}(h)} x_j. \tag{9}$$

By making use of matrices (7) we define a sequential version of the sample Fisher information matrix (4) as

$$m(h) = ||m_{i,j}(h)|| = \begin{bmatrix} \langle \widehat{M}_{1,\tau_1(h)} \rangle_{11} & \langle \widehat{M}_{1,\tau_1(h)} \rangle_{12} \\ \langle \widehat{M}_{2,\tau_2(h)} \rangle_{21} & \langle \widehat{M}_{2,\tau_2(h)} \rangle_{22} \end{bmatrix}, \tag{10}$$

where $\langle A \rangle_{ij}$ denotes (i,j)-th element of a matrix A.

Finally, we construct the sequential estimate for vector $\theta = (\theta_1, \theta_2)'$ in (3) as

$$\theta^*(h) = \begin{pmatrix} \theta_1^*(h) \\ \theta_2^*(h) \end{pmatrix} = m^{-1}(h) \upsilon(h). \tag{11}$$

Uniform asymptotic normality of $\theta^*(h)$. First we will study the asymptotic distribution of the estimate (11). We assume that the process (3) is unstable, that is $|\theta_1| \leqslant 1$ and that the distribution of disturbances ε_k is not specified. In this case we arrive at the following result.

Theorem 2. *Let sequential estimates $\theta^*(h) = (\theta_1^*(h), \theta_2^*(h))'$ for $\theta = (\theta_1, \theta_2)'$ in (3) be defined, for each $h > 0$, by (11). If the process (3) is unstable, that is $|\theta| \leqslant 1$, and $\{\varepsilon_n\}$ are i.i.d. random variables with $\mathsf{E}\varepsilon_n = 0$, $0 < \mathsf{E}\varepsilon_n^2 = \sigma^2 < \infty$ and are independent of x_0, then for any $0 < L < \infty$*

$$\lim_{h \to \infty} \sup_{\theta \in \Theta_L} \sup_{t \in R^2} \left| P_\theta \left(\frac{m(h)}{\sqrt{h}} (\theta^*(h) - \theta) \leqslant t \right) - \Phi_2 \left(\frac{t}{\sigma} \right) \right| = 0$$

where $t = (t_1, t_2)'$, $\Phi_2(t) = \Phi(t_1)\Phi(t_2)$, Φ is the standard normal distribution function,

$$\Theta_L = \{\theta = (\theta_1, \theta_2)' : \ |\theta_1| \leqslant 1, |\theta_2| \leqslant L\}.$$

Proof. Substituting x_k from (3) in (11) yields

$$m(h)\theta^*(h) = v(h) = m(h)\theta + \eta(h),\ \eta(h) = (\eta_1(h), \eta_2(h))'$$

where

$$\eta_1(h) = \sum_{k=1}^{\tau_1(h)} \sqrt{\beta_{1,k}(h)} x_{k-1}\varepsilon_k, \quad \eta_2(h) = \sum_{j=\tau_1(h)+1}^{\tau_2(h)} \sqrt{\beta_{2,j}(h)}\varepsilon_j.$$

This implies that

$$\frac{m(h)}{\sqrt{h}}(\theta^*(h) - \theta) = \frac{1}{\sqrt{h}}\eta(h). \tag{12}$$

Therefore we have to establish that $\eta(h)/\sqrt{h}$ converges uniformly in $\theta \in \Theta_L$ to two-dimensional normal distribution, that is

$$\frac{1}{\sqrt{h}}\eta(h) \Rightarrow \mathcal{N}(0, \sigma^2 I), \quad I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

as $h \to \infty$. To this end it suffices to show that for each constant vector $\lambda = (\lambda_1, \lambda_2)' \in R^2$ with $||\lambda|| = 1$, $\lambda_1 \neq 0$, $\lambda_2 \neq 0$,

$$\lim_{h\to\infty} \sup_{\theta\in\Theta_L} \sup_{-\infty<z<\infty} \left| P_\theta\left(\lambda' \frac{\eta(h)}{\sqrt{h}} \leqslant z\right) - \Phi\left(\frac{z}{\sigma}\right) \right| = 0. \tag{13}$$

The linear combination

$$\zeta(h) = \lambda_1\eta_1(h) + \lambda_2\eta_2(h)$$

can be rewritten as

$$\zeta(h) = \sum_{j=1}^{\tau_2(h)} y_{j-1}\varepsilon_j + \zeta_1(h)$$

where

$$y_j = \lambda_1 x_j \chi_{(\tau_1(h)>j+1)} + \lambda_1\sqrt{\alpha_1(h)} x_{\tau_1(h)-1}\chi_{(\tau_1(h)=j+1)} + \lambda_2\chi_{(\tau_1(h)<j+1)},$$

$$\zeta_1(h) = (\sqrt{\alpha_2(h)} - 1)\varepsilon_{\tau_2(h)}.$$

Further we prove that

$$\lim_{h\to\infty} \sup_{\theta\in\Theta_L} \sup_{-\infty<z<\infty} \left| P_\theta\left(\lambda' \frac{1}{\sqrt{h}} \sum_{j=1}^{\tau_2(h)} y_{j-1}\varepsilon_j \leqslant z\right) - \Phi\left(\frac{z}{\sigma}\right) \right| = 0. \tag{14}$$

and that for any $\Delta > 0$

$$\lim_{h\to\infty} \sup_{\theta\in\Theta_L} P_\theta\left(\frac{1}{\sqrt{h}}|\zeta_1(h) > \Delta|\right) = 0. \tag{15}$$

The argument in the proof of (14) is similar to that of Proposition 2.1 of the paper [4]. Combining (14) and (15) one comes to (13). Hence Theorem 2. □

Non-asymptotic normality of $\theta^*(h)$. Now we assume that the noises $\{\varepsilon_k\}$ in (3) is a sequence of i.i.d. random variables with Gaussian distributions. In this case we can derive non-asymptotic distribution of the sequential estimate (11) under appropriate choice of the normalizing factor.

Theorem 3. *Let sequential estimates $\theta^*(h) = (\theta_1^*(h), \theta_2^*(h))'$ for $\theta = (\theta_1, \theta_2)'$ in (3) be defined by (11). If $\{\varepsilon_n\}$ are i.i.d. standard normal random variables, that is $\varepsilon_n \sim \mathcal{N}(0,1)$, then for any $\theta \in R^2$ and any $h > 0$*

$$P_\theta\left(\frac{m(h)}{\sqrt{h}}(\theta^*(h) - \theta) \leqslant t\right) = \Phi_2(t)$$

where $m(h)$ is defined by (10).

Proof. Taking into account (12) we establish that the characteristic function of the vector $\eta(h)/\sqrt{h}$ has the form

$$\varphi_{\frac{\eta(h)}{\sqrt{h}}}(u) = \mathsf{E}exp(i(\eta, u)) = exp\left(-\frac{u_1^2}{2} - \frac{u_2^2}{2}\right).$$

This completes the proof of Theorem 3. □

Monte-Carlo simulation results. In this section we report some numerical results to verify the asymptotic uniform normality property of sequential estimate proved in Theorem 2. The basic experiment consisted of 20000 replications of the sequential procedure (11) with $h = 20$ for each value of the parameter vector $\theta = (\theta_1, \theta_2)'$ indicated in Table 1.

After each simulation of the procedure the normalized deviation vector

$$z(h) = (z_1(h), z_2(h))' = \frac{m(h)}{\sqrt{h}}(\theta^*(h) - \theta)$$

was calculated. Table 1 gives frequency estimates for the probability

$$P(z_1(h) \leqslant a,\ z_2(h) \leqslant b)$$

for two points (a, b) and different values of unknown parameters θ_1, θ_2.

The results of simulations show good performance of the procedure (11) and confirm the results of Theorem 2.

Concluding remarks. In this paper we propose new sequential estimates for estimating parameters in autoregressive process $AR(1)$ with unknown mean.

We prove the property of uniform asymptotic normality when the process is unstable and the noise distributions are not specified.

Table 1 — Test on the uniform normality

$a = -0.3,\ b = 1.7,\ h = 20,\ p = 0.365$							
$\theta_1 \| \theta_2$	-2.1	-1.4	-0.7	0	0.7	1.4	2.1
-1	0.367	0.364	0.365	0.365	0.367	0.369	0.365
-0.5	0.370	0.355	0.365	0.363	0.369	0.365	0.357
0	0.366	0.364	0.368	0.361	0.360	0.363	0.365
0.5	0.365	0.368	0.368	0.367	0.362	0.361	0.366
1	0.369	0.367	0.364	0.369	0.363	0.364	0.368
$a = 0.2,\ b = -0.5,\ h = 20,\ p = 0.179$							
$\theta_1 \| \theta_2$	-2.1	-1.4	-0.7	0	0.7	1.4	2.1
-1	0.175	0.179	0.182	0.180	0.181	0.180	0.182
-0.5	0.181	0.182	0.181	0.181	0.180	0.183	0.183
0	0.181	0.177	0.178	0.174	0.176	0.177	0.182
0.5	0.179	0.183	0.182	0.182	0.178	0.175	0.178
1	0.175	0.176	0.179	0.182	0.181	0.177	0.180

References

1. Borisov V. Z. On sequential parameters estimation in discrete time processes /V. Z. Borisov, V. V. Konev // Automat. Remote Control. – 1977. – No 6. – pp. 58–64.

2. Galtchouk L. On Uniform Asymptotic Normality of Sequential Estimators for the Parameters in a Stable AR(1) / L. Galtchouk, V. V. Konev // Sequential Analysis. – 2003. – vol. 22. – Nos 1 & 2. – pp.31-54.

3. Konev V. V. Sequential Parameter Estimation with Guaranteed Mean-Square Accuracy for Unstable Linear Stochastic Systems / V. V. Konev, S. M. Pergamenshchikov // Problemy Peredachi Informatsii. – 1992. – Vol. 28. – No. 4. – pp. 35–48.

4. Lai T. L. Fixed-accuracy estimation of an autoregressive parameter / T. L. Lai and D. Siegmund // Ann. Statist. – 1983. – Vol. 11. – pp. 478–485.

5. Shiryaev A. N. Statistical experiments and decisions: Asymptotic theory /Shiryaev A. N. and Spokoiny V. G. – World Scientific. – 2000

Improved estimation of a function in continuous regression with semimartingale noise *

Pchelintsev E. A., Pchelintsev V. A., Pergamenshchikov S. M.
Tomsk State University, Tomsk
e-mail: evgen-pch@yandex.ru

Abstract

In this paper, we consider the robust adaptive non parametric estimation problem for the periodic function observed with the semimartingale 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.

Key words: Improved non-asymptotic estimation, Weighted least squares estimates, Robust quadratic risk, Non-parametric regression, Model selection procedure, Oracle inequality.

Introduction. Consider a regression model in continuous time

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

where S is an unknown 1-periodic $\mathbb{R} \to \mathbb{R}$ function, $S \in \mathbf{L}_2[0,1]$; $(\xi_t)_{t\ge 0}$ is an unobservable conditionally gaussian semimartingale with the values in the Skorokhod space $\mathbf{D}[0,n]$ such that, for any cadlag $[0,n] \to \mathbb{R}$ function f from $\mathbf{L}_2[0,n]$, the stochastic integral

$$I_n(f) = \int_0^n f(s)\mathrm{d}\xi_s \tag{2}$$

is well defined and has the following properties

$$\mathbf{E}_Q I_n(f) = 0 \quad \text{and} \quad \mathbf{E}_Q I_n^2(f) \le \varkappa_Q \int_0^n f^2(s)\mathrm{d}s\,. \tag{3}$$

Here $\mathbf{E}_Q$ denotes the expectation with respect to the distribution Q in $\mathcal{D}[0,n]$ of the process $(\xi_t)_{0\le t\le n}$, which is assumed to belong to some probability family $\mathcal{Q}_n^*$ specified below; $\varkappa_Q > 0$ is some positive constant depending on the distribution Q.

The problem is to estimate the unknown function S in the model (1) on the basis of observations $(y_t)_{0\le t\le n}$.

The class of the disturbances ξ satisfying conditions (3) is rather wide and comprises, in particular, the Lévy processes which are used in different applied problems. The models (1) with the Lévy's type

* This study was supported by RSF grant, project no 17-11-01049

noise naturally arise in the nonparametric functional statistics problems. Moreover, non-Gaussian Ornstein–Uhlenbeck-based models, enter this class.

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]. To this end we define the robust risk as

$$\mathcal{R}^*(\widehat{S}_n, S) = \sup_{Q\in\mathcal{Q}_n^*} \mathcal{R}_Q(\widehat{S}_n, S) \tag{4}$$

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

$$\mathcal{R}_Q(\widehat{S}_n, S) := \mathbf{E}_{Q,S}\,\|\widehat{S}_n - S\|^2 \quad \text{and} \quad \|S\|^2 = \int_0^1 S^2(t)\mathrm{d}t\,.$$

In this paper, we consider a minimax optimisation criteria which aims to minimize the robust risk (4). 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 [1]. Then, these methods had been developed by Barron, Birgé and Massart [2] and Fourdrinier and Pergamenshchikov [3] for the nonparametric estimation and oracle inequalities for the quadratic risks. Unfortunately, the oracle inequalities obtained in these papers can not 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 semi - martingale approach for the robust adaptive efficient estimation of the nonparametric signals in continuous time proposed by Konev and Pergamenshchikov in [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 [3]. 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 [4] in the way proposed in [6, 7]. 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 (4).

Example of noise processes. Let the useful signal S is distorted by the impulse flow described by the Lévy process, i.e. we assume that the noise process $(\xi_t)_{0\le t\le n}$ is defined as

$$\xi_t = \varrho_1 w_t + \varrho_2 z_t \quad \text{and} \quad z_t = x * (\mu - \widetilde{\mu})_t \,, \tag{5}$$

where, ϱ_1 and ϱ_2 are some unknown constants, $(w_t)_{t\ge 0}$ is a standard brownian motion, $\mu(\mathrm{d}s\,\mathrm{d}x)$ is a jump measure with deterministic compensator $\widetilde{\mu}(\mathrm{d}s\,\mathrm{d}x) = \mathrm{d}s\Pi(\mathrm{d}x)$, $\Pi(\cdot)$ is a Lévy measure, i.e. some positive measure on $\mathbb{R}_* = \mathbb{R}\setminus\{0\}$, such that

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

Here we use the notation $\Pi(|x|^m) = \int_{\mathbb{R}_*} |z|^m \,\Pi(\mathrm{d}z)$. Note that the Lévy measure $\Pi(\mathbb{R}_*)$ could be equal to $+\infty$. Then the class $\mathcal{Q}^*_n$ of distributions of the process $(\xi_t)_{0\le t\le n}$ includes all distributions for which the parameters $\varkappa_Q = \varrho_1 \ge \varsigma_*$ and $\varrho_1^2 + \varrho_2^2 \le \varsigma^*$, where ς_* and ς^* are some fixed positive bounds.

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 (1) we consider it's Fourier expansion

$$S(t) = \sum_{j=1}^{\infty} \theta_j \phi_j(t) \quad \text{and} \quad \theta_j = (S,\phi_j) = \int_0^1 S(t)\,\phi_j(t)\,\mathrm{d}t\,.$$

The corresponding Fourier coefficients can be estimated as

$$\widehat{\theta}_{j,n} = \frac{1}{n}\int_0^n \phi_j(t)\,\mathrm{d}y_t\,.$$

In view of (1), one obtains

$$\widehat{\theta}_{j,n} = \theta_j + \frac{1}{\sqrt{n}}\xi_{j,n}\,, \quad \xi_{j,n} = \frac{1}{\sqrt{n}} I_n(\phi_j)\,, \tag{6}$$

where $I_n(\phi_j)$ is given in (2).

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

$$\widehat{S}_\lambda = \sum_{j=1}^{n} \lambda(j) \widehat{\theta}_{j,n} \phi_j \,, \tag{7}$$

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 (7) we use the improved estimation method proposed for parametric models in [7]. To this end we set $\widetilde{\theta}_n = (\widehat{\theta}_{j,n})_{1\le j\le d}$. 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 $\mathbb{R}^d$. Now we define the shrinkage estimators as

$$\theta^*_{j,n} = (1 - g(j))\, \widehat{\theta}_{j,n} \,, \tag{8}$$

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

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

We denote the difference of quadratic risks of the estimates (7) and (9) as $\Delta_Q(S) := \mathcal{R}_Q(S^*_\lambda, S) - \mathcal{R}_Q(\widehat{S}_\lambda, S)$. 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 $n \ge 1$*

$$\sup_{Q\in\mathcal{Q}^*_n} \sup_{\|S\|\le \mathbf{r}} \Delta_Q(S) < -\mathbf{c}_n^2 \,. \tag{10}$$

Remark. The inequality (10) means that non asymptotically, i.e. for any $n \ge 1$ the estimate (9) outperforms in mean square accuracy the estimate (7). Moreover, as we will see below, $n\mathbf{c}_n \to \infty$ as $d \to \infty$. This means that improvement is considerable may better than for the parametric regression (see, [7]).

Model selection procedure. This Section gives the construction of a model selection procedure for estimating a function S in (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 (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

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

In order to obtain a good estimate, we have to write a rule to choose a weight vector $\lambda \in \Lambda$ in (9). It is obvious, that the best way is to minimise the empirical squared error with respect to λ. Making use the estimate definition (9) and the Fourier transformation of S implies

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

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

$$\widetilde{\theta}_{j,n} = \theta^*_{j,n}\,\widehat{\theta}_{j,n} - \frac{\widehat{\sigma}_n}{n}\,,$$

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

$$\widehat{\sigma}_n = \sum_{j=[\sqrt{n}]+1}^{n} \widehat{t}^{\,2}_{j,n} \quad \text{and} \quad \widehat{t}_{j,n} = \frac{1}{n}\int_0^n \mathrm{Tr}_j(t)\mathrm{d}y_t\,.$$

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

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

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

$$\widehat{P}_n(\lambda) = \frac{\widehat{\sigma}_n\,|\lambda|^2_n}{n}\,.$$

Substituting the weight coefficients, minimizing the cost function

$$\lambda^* = \mathrm{argmin}_{\lambda\in\Lambda}\, J_n(\lambda)\,, \tag{11}$$

in (7) leads to the improved model selection procedure

$$S^* = S^*_{\lambda^*}\,. \tag{12}$$

It will be noted that λ^* exists because Λ is a finite set. If the minimizing sequence in (11) λ^* is not unique, one can take any minimizer. In the case, when the value of σ_Q is known, one can take $\widehat{\sigma}_n = \sigma_Q$ and $P_n(\lambda) = \sigma_Q\,|\lambda|^2_n/n$.

Theorem 2. *For any $n \ge 2$ and $0 < \delta < 1/3$, the robust risks* (4) *of estimate* (12) *for continuously differentiable function S satisfies the* *oracle inequality*

$$\mathcal{R}^*(S^*_{\lambda^*}, S) \le \frac{1+3\delta}{1-3\delta} \min_{\lambda\in\Lambda} \mathcal{R}^*(S^*_\lambda, S) + \frac{B^*_n}{n\delta}\,, \qquad (13)$$

where the rest term is such that $B^*_n/n^\epsilon \to 0$ as $n \to \infty$ for any $\epsilon > 0$.

Remark. The inequality (13) means that the procedure (12) is optimal in the oracle inequalities sense. This property enables to provide asymptotic efficiency in the adaptive setting, i.e. when information about the function regularity is unknown.

References

1. Akaike H. A new look at the statistical model identification // IEEE Trans. on Automatic Control. 1974. V. 19. P. 716–723.

2. Barron A., Birgé L. and Massart P. Risk bounds for model selection via penalization. // Probab. Theory Relat. Fields. 1999. V. 113. P. 301–415.

3. Fourdrinier D., Pergamenshchikov S. Improved selection model method for the regression with dependent noise // Annals of the Institute of Statistical Mathematics. 2007. V.59. No. 9. P. 435–464.

4. James W., Stein C. (1961). Estimation with quadratic loss // In Proceedings of the Fourth Berkeley Symposium Mathematics, Statistics and Probability. University of California Press, Berkeley. 1961. V. 1. P. 361–380.

5. Konev V. V., Pergamenshchikov S. M. Robust model selection for a semimartingale continuous time regression from discrete data // Stochastic processes and their applications. 2015. V.125. P. 294 – 326.

6. Konev V. (2014) Estimation of a regression with the pulse type noise from discrete data / Konev V., Pergamenshchikov S., Pchelintsev E. // Theory Probab. Appl. 2014. V. 58. No. 3. P. 442–457.

7. Pchelintsev E. Improved estimation in a non-Gaussian parametric regression // Stat. Inference Stoch. Process. 2013. V. 16. No. 1. P. 15 – 28.

Adaptive estimation of heavy-tailed distributions *

Politis D. N.*, **Vasiliev V. A.**+, **Vorobeychikov S. E.**+

*University of California, San Diego

+Tomsk State University, Tomsk

e-mail: dimipoli@hotmail.com, vas@mail.tsu.ru, sev@mail.tsu.ru

Abstract

The problem of tail indexes estimation of heavy tail distributions is considered. We propose the estimators of tail indexes using the truncated estimation method developed for ratio type functionals. It is shown that the truncated estimators constructed on the sample of fixed size have a guaranteed accuracy in the sense of the L_{2m}-norm, $m \geq 1$. Moreover they have optimal or suboptimal rates of convergence. These estimators are used to construct the adaptive estimators of distributions. The rates of decreasing of the χ^2 divergence in the almost surely sense between distributions and their adaptive estimators are found.

Keywords: Heavy-tailed distributions, estimation of distribution, Pareto's distribution, Hall's distribution, tail indexes estimation, fixed sample size.

Introduction. The models with heavy tail distributions are of interest in many applications connected with financial mathematics, insurance theory, telecommunication and physics. Usually it is assumed that the distribution function contains as an unknown multiplier a slowly varying function. The problem of tail index estimation was studied by Hill [4] who proposed the estimators based on the order statistics. The estimator is optimal in mean square sense on the class of distribution functions with heavy tails in presence of unknown slowly varying function. It should be noted that the Hill's estimators are unstable and can diverge essentially from the estimated parameter for large sample sizes.

Later another approaches to estimation problem were proposed (see, e.g., [2] and references therein).

*Research was supported by the RNF, Project No. 17-11-01049.

In this paper, the truncated estimation method of ratio type functionals, proposed by Vasiliev [5], is used to obtain estimators with guaranteed accuracy in the sense of the L_{2m}-norm, $m \geq 1$. The estimators are constructed on the basis of empirical functionals without usage of non-parametric approach in an effort to obtain (or get close to) the parametric optimal rate of convergence. These estimators are used to construct the adaptive estimators of distribution functions. It allows one to find the rates of decreasing $\varphi_\varepsilon^{-1}(n),\ \varepsilon > 0$ of the χ^2 divergence in the almost surely sense between distributions and their adaptive estimators.

These rates of decreasing are found for Pareto distribution perturbed by logarithm and Hall distributions.

Adaptive distribution estimation. Let $\mathcal{P} = \{P_\Delta(x),\ x \in G \subseteq \mathcal{R}^1,\ \Delta \in \mathcal{D} \subseteq \mathcal{R}^q\}$ be the parametric family of heavy tail distributions.

The problem is to construct estimators P_{Δ_n} of concrete well-known distributions P_Δ on the basis of a special type parameter estimators Δ_n with known rates of decreasing $\varphi_\varepsilon^{-1}(n),\ \varepsilon > 0$ of the χ^2 divergence in the following sense

$$\lim_{n\to\infty} \varphi_\varepsilon(n)\chi^2(P_\Delta, P_{\Delta_n}) = 0 \ \text{ a.s.} \tag{1}$$

Suppose that for every $\Delta \in \mathcal{D}$ the density $f_\Delta(x) = dP\Delta(x)/dx$ exists.

Suppose the following

Assumption (A). Assume there exists the number $\delta_0 > 0$ such that the set $\mathcal{D}_0 = \{\delta :\ \Delta + \delta \subseteq \mathcal{D},\ ||\delta|| \leq \delta_0\}$ is not empty and

$$\sup_{\delta\in\mathcal{D}_0} \int\limits_G ||\nabla_\Delta f_{\Delta+\delta}(x)||^2 f_\Delta^{-1}(x)dx < \infty.$$

Then, using the Taylor expansion for the function $f_{\Delta_n}(x)$ on the set $\{\omega :\ ||\Delta_n - \Delta|| \leq \delta_0\}$ we have

$$\chi^2(P_\Delta, P_{\Delta_n}) \leq C||\Delta_n - \Delta||^2.$$

Thus to prove (1) it is enough to find the functions $\varphi_\varepsilon(n)$ and estimators Δ_n such that

$$\lim_{n\to\infty} \varphi_\varepsilon(n)||\Delta_n - \Delta||^2 = 0 \ \text{ a.s.} \tag{2}$$

In the following sections we consider well known distributions and propose estimators of tail indexes with the properties

$$E||\Delta_n - \Delta||^{2p} \leq r^{-1}(n,p), \quad n \geq 1, \tag{3}$$

which are fulfilled for every $p \geq 1$ and some functions $r(n,p) \to \infty$ as $n \to \infty$ and/or $p \to \infty$.

Estimators with these properties will be constructed by making use the general truncated estimation procedure, presented in [5].

Define $\varphi(n,p) = (n^{-2}r(n,p))^{1/p}$. Then, using the Borel-Cantelli lemma for every $p \geq 1$ in particular it follows

$$\lim_{n\to\infty} \varphi(n,p)||\Delta_n - \Delta||^2 = 0 \quad \text{a.s.} \tag{4}$$

We will investigate the adaptive estimation problem of the following distributions:

- Pareto perturbed by logarithm: tail distribution function $P(x) = c_\gamma x^{-\gamma} \log^\nu x$, $G = \{x : \ x \geq c > 0\}$. The number c is supposed to be unknown. Here $\mathcal{D} = \{\Delta = (\gamma, c_\gamma), \ \gamma > 0, \ c_\gamma > 0\}$;

- Hall model: $P(x) = C_1 x^{-\gamma}(1 + C_2 x^{-\gamma(\rho+1)})$. In this case we put $G = \{x : \ x > 0\}$ and $\mathcal{D} = \{\Delta = \gamma, \ \gamma > 0\}$.

Theorem 1. *For every $\varepsilon > 0$ there exist numbers p_ε such that the property (1) is fulfilled with $\varphi_\varepsilon(n) = \varphi(n, p_\varepsilon)$, where*

$\varphi_\varepsilon(n) = n^{1-\varepsilon}$ for the adaptive estimator in Pareto model and

$\varphi_\varepsilon(n) = n^{\frac{\rho+1}{\rho+2}-\varepsilon}$ for the Hall model.

Estimation of heavy tail indexes. Pareto distribution perturbed by logarithm. The problem is to estimate the parameters γ and c_γ based on i.i.d. observations $X_1, \ldots, X_n$ from the perturbed Pareto distribution function considered, e.g., in Grama and Spokoiny [2]

$$F(x) = 1 - c_\gamma x^{-\gamma} \log^\nu x, \quad x \geq c.$$

In this context, Grama and Spokoiny [2] proposed a general estimation algorithm that ensures the rate of convergence of the estimator of $\theta = 1/\gamma$ in L_2-norm as $1/\log n$; they also mention that according to Drees [1] this rate is optimal for the estimator of distribution parameters from a wide class of functions which the perturbed Pareto distribution function belongs to.

To solve the estimation problem we apply Theorem 1 from [5].

First note that the density function exists

$$f(x) = c_\gamma x^{-(\gamma+1)} \log^{\nu-1} x[\gamma \log x - \nu]$$

and Assumption (A) is fulfilled with $\mathcal{D}_0 = \{\delta = (\delta_{01}, \delta_{02}) : \ |\delta_{01}| \leq \gamma/2, \ 0 < \delta_{02} \leq c_2\}$ with $\delta_0 = \sqrt{(\gamma/2)^2 + c_2^2}$.

Define the function $P(x) = 1 - F(x) = c_\gamma x^{-\gamma} \log^\nu x$.

To estimate the unknown parameters we evaluate this function in two different points $x_1 > x_2 > \max(c, e)$:

$$\log P(x_1) = \log c_\gamma - \gamma \log x_1 + \nu \log\log x_1,$$
$$\log P(x_2) = \log c_\gamma - \gamma \log x_2 + \nu \log\log x_2,$$

and consider the difference

$$\gamma[\log x_1 - \log x_2] = \log P(x_2) - \log P(x_1) + \nu \log\log x_2 - \nu \log\log x_1.$$

The parameter γ is the ratio type functional

$$\gamma = \frac{\log(P(x_2)/P(x_1)) + b}{a},$$

$$a = \log(x_1/x_2), \quad b = \nu \log[(\log x_1)/(\log x_2)].$$

Define the estimator

$$\gamma_n = \frac{\log(P_n(x_2)/P_n(x_1)) + b}{a} \cdot \chi(P_n(x_1) \geq \log^{-1} n), \tag{5}$$

where $P_n(x)$ is the empirical tail distribution function

$$P_n(x) = \frac{1}{n}\sum_{k=1}^{n} \chi(X_k \geq x)$$

and the estimator c_n of c_γ

$$c_n = x^{\gamma_n} \log^{-\nu} x P_n(x).$$

In view of the number c is unknown, to ensure the inequalities $x_1 > x_2 > c$ we choose the diverging sequences $x_1 = x_1(n)$ and $x_2 = x_2(n)$. As a concrete example we may let

$$x_1(n) = \log\log n \quad \text{and} \quad x_2(n) = \log^{1-\beta} \log n, \tag{6}$$

where $\beta \in (0,1)$ is a given number and the parameter x in the estimator c_n of c_γ one should choose as a slowly increasing to the infinity sequence $x = x(n)$ as well e.g., as follows

$$x(n) = (\log\log\log n)^{\frac{2p-1}{2p}}. \tag{7}$$

Estimators γ_n of γ and c_n of c_γ satisfy the following properties (for an integer $p \geq 1$), namely

$$E(\gamma_n - \gamma)^{2p} \leq C\frac{\log^A \log n}{n^p}, \tag{8}$$

$$E(c_n - c_\gamma)^{2p} \leq C\frac{\log^A \log n}{n^p}, \tag{9}$$

where $A > 0$ is some number.

Then

$$r(n,p) = \frac{Cn^p}{\log^A \log n}, \quad C > 0, \quad A > 0$$

and for $\varepsilon > 0$ the number p_ε and the function $\varphi_\varepsilon(n)$ can be chosen as $p_\varepsilon > 2\varepsilon^{-1}$ and $\varphi_\varepsilon(n) = n^{1-\varepsilon}$.

Hall model. The problem is to estimate by i.i.d. observations $X_1, \ldots, X_n$ the parameter $\gamma = 1/\beta$ of the Hall distribution function

$$F(x) = 1 - C_1 x^{-1/\beta} - C_2 x^{-1/\delta}, \quad x \geq x_0,$$

where $\beta > 0$, $\delta > 0$; $\beta = \delta + \theta \geq \beta_0 > 0$.

Then the tail distribution function

$$P(x) = C_1 x^{-\gamma}(1 + C_3 x^{-\gamma(\rho+1)}),$$

where $C_3 = C_2/C_1$, $\gamma = 1/\beta$, $\rho = \theta/\delta - 1$.

The density function has the form

$$f(x) = C_1 \gamma x^{-(\gamma+1)} + C_2/\delta x^{-(1/\delta+1)}$$

and Assumption (A) is fulfilled for $\mathcal{D}_0 = \{\delta : |\delta| \le \gamma/2\}$ and $\delta_0 = \gamma/2$.

For some $X = (x_1, x_2)$, $x_1 > x_2 > x_0$ by the definition of $P(x)$ we have

$$\log P(x_1) = \log C_1 - \gamma \log x_1 + \log(1 + C_3 x_1^{-\gamma(\rho+1)}),$$
$$\log P(x_2) = \log C_1 - \gamma \log x_2 + \log(1 + C_3 x_2^{-\gamma(\rho+1)})$$

and it is natural to define the estimators γ_n of γ as follows

$$\gamma_n(X) = \frac{\log(P_n(x_2)/P_n(x_1))}{\log(x_1/x_2)}.$$

To get the estimator γ_n with the optimal rate of convergence (in the sense of L_2-norm see [1–3]), we put for $p \ge 1$ the sequence $X(n) = (x_1(n), x_2(n)$, where

$$x_1(n) = \mathrm{e} \cdot x_2(n), \qquad x_2(n) = n^{\frac{p}{\gamma[2p(\rho+2)-1]}} \tag{10}$$

Then, using Theorem 1 in [5] we have

$$E(\gamma_n(X(n)) - \gamma)^{2p} \le C \cdot n^{-\frac{2p(\rho+1)}{2p(\rho+1)+2p-1}p}$$

and we can put

$$p_\varepsilon \ge 2\varepsilon^{-1}, \qquad \varphi_\varepsilon(n) = n^{\frac{\rho+1}{\rho+2}-\varepsilon}.$$

Note that proposed parameter estimation procedure gives estimator γ_n with optimal convergence rate. At the same time the sequences (10) depend on the unknown model parameters. Then the adaptive estimation procedure can be constructed on the presented scheme using some estimators of γ and ρ.

Consider, for instance, the case of known ρ. Define the known deterministic sequence $(m_n)_{n\ge1}$, $m_n = n^\kappa$, $\kappa \in (0,1)$ and pilot estimator $\tilde\gamma_n = \tilde\gamma_n(\tilde X(m_n))$ of γ as follows

$$\tilde\gamma_n(\tilde X(m_n)) = \frac{\log(P_{m_n}(\tilde x_2(m_n))/P_{m_n}(\tilde x_1(m_n)))}{\log(\tilde x_1(m_n)/\tilde x_2(m_n))},$$

where $\tilde X(n) = (\tilde x_1(n), \tilde x_2(n))$,

$$\tilde x_1(n) = \mathrm{e} \cdot \tilde x_2(n), \qquad \tilde x_2(n) = n^{\frac{p}{\gamma_0[2p(\rho+2)-1]}}, \quad \gamma_0 = \beta_0^{-1}.$$

This estimator has the property

$$E(\tilde\gamma_n - \gamma)^{2p} \le C \cdot r_0^{-1}(n,p), \quad p \ge 1,$$
$$r_0(n,p) = n^{\frac{2p\gamma\kappa(\rho+1)}{\gamma_0[2p(\rho+1)+2p-1]}p}$$

and is strongly consistent with the following rate

$$n^\nu(\tilde\gamma_n - \gamma) \to 0 \quad \text{a.s.} \tag{11}$$

for every

$$0 < \nu < \frac{\gamma\kappa(\rho+1)}{\gamma_0(2\rho+3)}.$$

Define the adaptive estimator of γ as follows

$$\hat{\gamma}_n = \frac{\log(\tilde{P}_n(\hat{x}_2(n))/\tilde{P}_n(\hat{x}_1(n)))}{\log(\hat{x}_1(n)/\hat{x}_2(n))},$$

where $\tilde{P}_n$ is the empirical tail distribution function

$$\tilde{P}_n(x) = \frac{1}{n-m_n} \sum_{k=m_n+1}^{n} \chi(X_k \geq x)$$

and $\hat{X}(n) = (\hat{x}_1(n), \hat{x}_2(n))$,

$$\hat{x}_1(n) = \mathrm{e} \cdot \hat{x}_2(n), \qquad \hat{x}_2(n) = n^{\frac{p}{\hat{\gamma}_n[2p(\rho+2)-1]}}.$$

The estimator $\hat{\gamma}_n$ has the property

$$E[(\hat{\gamma}_n - \gamma)^{2p} | \mathcal{F}_{m_n}] \leq C \cdot n^{-\frac{2p\gamma(\rho+1)}{\hat{\gamma}_n[2p(\rho+1)+2p-1]}p}, \quad p \geq 1,$$

where σ-algebra $\mathcal{F}_{m_n} = \sigma\{X_1, \ldots, X_{m_n}\}$.

Thus, using the Borel-Cantelli lemma and strong consistency of the pilot estimator $\hat{\gamma}_n$ it is easy to prove the last property of Theorem 1 for the adaptive estimator $P_{\hat{\gamma}}$ in the Hall model.

It can be shown $\varphi(n,p)(\hat{\gamma}_n - \gamma)^2 \to 0$ a.s. as $n \to \infty$ and Theorem 1 is fulfilled with

$$p_\varepsilon > (\rho+1)\max[\frac{\gamma_0}{\kappa\gamma(\rho+1)}, \frac{2}{\rho+1-\varepsilon(\rho+2)}].$$

References

1. Drees H. Optimal rates of convergence for estimates of the extreme value index // Ann. Statist. 1998. V. 26. No. 1. P. 434-448.

2. Grama I., Spokoiny V. Statistics of extremes by oracle estimation // Ann. Statist. 2008. V. 36. No. 4. P. 1619-1648.

3. Hall P., Welsh A.H. Best attainable rates of convergence for estimates of parameters of regular variation // Ann. Statist. 1984. V. 12. No. 3. P. 1079-1084. MR0751294.

4. Hill B.M. A simple general approach to inference about the tail of a distribution // Ann. Statist. 1975. V. 3. No. 3. P. 1163-1174.

5. Vasiliev V. A truncated estimation method with guaranteed accuracy // Annals of the Institute of Statistical Mathematics. 2014. V. 66. No. 1. P. 141-163.

Вычисление среднего времени работы системы с использованием свойств оптимальных стратегий *

Губин В. Н.

Томский политехнический университет, Томский государственный университет, Томский государственный университет систем управления и радиоэлектроники, г. Томск
e-mail: gubinvn@tpu.ru

Аннотация

В статье изложены основные свойства оптимальных стратегий для модели динамического резервирования Райкина-Герцбаха, когда в качестве критерия оптимизации выбрано среднее время безотказной работы на бесконечном промежутке времени. На основе изложенных свойств построен алгоритм для вычисления среднего времени безотказной работы системы и оптимальной стратегии. Полученный алгоритм значительно сокращает время вычисления оптимальной стратегии по сравнению с простым перебором, что весьма важно в прикладных задачах надежности.

Ключевые слова: оптимальная стратегия, среднее время безотказной работы, динамического резервирование, отказ системы.

Введение. Проблемой оптимального резервирования в надежности стали заниматься в 50-х годах двадцатого века. Эта проблема остается актуальной и по сей день в различных отраслях промышленности, поскольку отказ сложных систем приводит иногда к катастрофическим последствиям, а также к огромным затратам для их восстановления.

Первые работы в области динамического резервирования принадлежат А.Л. Райкину [1]и И.Б. Герцбаху [2]. Также задачами динамического резервирования занимались А.С. Мандель, Г.Г. Пестов, В.В. Конев, Л.В. Ушакова, В.А. Томиленко, И.Б. Голдовский и другие.

*Работа выполнена при финансовой поддержке РНФ, проект No 17-11-01049.

Постановка задачи и основные результаты. Пусть система состоит из конечного числа параллельно включенных (в смысле надежности) идентичных элементов. В моменты времени $t_n = n\Delta, n \in \mathbb{N}$ проводится проверка исправности включенных в работу элементов. Вероятность безотказной работы одного элемента между моментами проверок равна p, а соответствующая вероятность отказа – q $(p+q=1)$. Отказавшие элементы не восстанавливаются. Система исправна, если в работу включено по крайней мере m исправных элементов. Будем обозначать такую систему через S_m.

Определение 1. *Под стратегией резервирования будем понимать такую функцию $K(r)$, принимающую значения из отрезка $[m, r]$, которая задает количество элементов, которое включается в работу при наличии r исправных.*

В различных оптимизационных задачах рассматриваются различные *критерии качества* резервирования. Будем в данной работе в качестве критерия качества рассматривать среднее время безотказной работы системы.

Ясно, что слишком малое количество включаемых в работу элементов приведет к быстрому отказу системы, а включение всех элементов сразу нерационально, так как они начнут сразу расходовать свой ресурс и долго система не проработает. В связи с этим введем следующее определение.

Определение 2. *Стратегию $K_0(r)$, которая обращает в максимум среднее время безотказной работы системы, будем называть оптимальной по данному критерию.*

Введем следующие обозначения:

$T(k,r)$ математическое ожидание времени безотказной работы системы, если на первом шаге в работу включено k элементов из r имеющихся в наличии, а в дальнейшем используется стратегия, оптимальная по критерию среднему времени работы системы на промежутке $[0, +\infty)$;

$T(r)$ математическое ожидание времени работы системы при стратегии, оптимальной по критерию среднего времени работы системы, если в начальный момент имеется ровно r исправных элементов.

По формуле полного математического ожидания [3] имеем

$$T(k,r) = \sum_{i=0}^{k-m} C_k^i p^{k-i} q^i T(r-i) + 1 \tag{1}$$

Для нахождения оптимальной стратегии исследуем сначала некоторые свойства функций $T(k,r), T(r)$ и $K_0(r)$. Очевидно, что справедлива

Лемма 1. $T(r) > 0$, $T(r)$ возрастает с ростом r;

Рассмотрим линейный оператор σ, который действует следующим образом: $\sigma T(r) = T(r-1)$.

Тогда можно обобщить этот оператор для функции $T(k,r)$ и, очевидно, $\sigma T(k,r) = T(k, r-1)$.

Рассмотрим разность $T(k+1,r) - T(k,r)$. Используя известное свойство биномиальных коэффициентов, получим:

$$T(k+1,r) - T(k,r) = -q \sum_{i=0}^{k-m} C_k^i \ p^{k-i} q^i \ T(r-i) +$$

$$+ \sum_{j=0}^{k-m} C_k^j \ p^{k-j} q^{j+1} \ T(r-j-1) + C_k^{k+1-m} \ p^m q^{k+1-m} \ T(r-k-1+m) =$$

$$= q(\sigma - 1) \sum_{i=0}^{k-m} C_k^i \ p^{k-i} (q\sigma)^i T(r) + C_k^{k+1-m} p^m q^{k+1-m} T(r-k-1+m) =$$

$$= q(\sigma-1) T(k,r) + C_k^{k+1-m} \ p^m q^{k+1-m} T(r-k-1+m). \qquad (2)$$

Из выражения (2) можно сделать вывод, что разность вида $T(k+1,r) - T(k,r)$ может быть как положительной, так и отрицательной. В [4] доказана следующая теорема.

Теорема 1. *При* $p \geqslant \dfrac{m+1}{m+3}$ *функция* $T(k,r)$ *для системы* S_m *имеет не более двух максимумов при фиксированном* r*, причем она выпукла вверх по* k *в области* $m \leqslant k \leqslant K_0(r) + 1$ *и не возрастает при* $K_0(r) < k \leqslant r$.

Из теоремы 1 получаем как следствие области возрастания и убывания функции $T(k,r)$ по k.

Следствие. *Для любого* $r > 0$ *если* $T(k,r) > T(k-1,r)$*, то* $K_0(r) \geqslant k$*; если же* $T(k,r) < T(k-1,r)$*, то* $K_0(r) \leqslant k$.

В [4] было показано, что оптимальная стратегия является нестрого возрастающей функцией резерва. В [5] было установлено, что если с увеличением резерва на единицу оптимальная стратегия возрастает, то не более, чем на единицу. Таким образом, имеет место

Теорема 2. *Для любого* $r \geqslant 1$ *выполнены следующие неравенства:*

$$K_0(r) \leqslant K_0(r+1) \leqslant K_0(r) + 1. \qquad (3)$$

Графически неравенство (3) означает, что функция $K_0(r)$ ведет себя как кусочно-постоянная функция, причем «скачки» у функции всегда высотой 1. Таким образом, весь промежуток по резерву $[1, r]$ разбивается на непересекающиеся промежутки K_0-постоянства.

Кроме того, в работе [4] доказаны следующие асимптотические свойства оптимальных стратегий:

1. $\lim\limits_{r\to\infty} \dfrac{T(r+1)}{T(r)} = 1;$
2. Существует $\lim\limits_{r\to\infty} K_0(r)$, причем, если система допускает включение в работу неограниченного числа элементов, то этот предел равен бесконечности.

Первое свойство можно интерпретировать следующим образом: если производить увеличение числа резервных элементов, то вклад в увеличение среднего времени работы системы каждого последующего элемента уменьшается. Второе соотношение довольно естественно. Если имеется ограниченного число «гнезд» для элементов, то оптимальной стратегией будет полное их заполнение при наличии достаточного количества резервных элементов.

Упрощенный алгоритм вычисления оптимальной стратегии. Будем вычислять оптимальную стратегию рекуррентно. Пусть $r = m$, тогда, очевидно, $K_0(r) = m$, иначе наступит отказ системы. В этом случае $T(m)$ легко вычислить. Из уравнения (1)

$$T(k, r) = p^k T(r) + \sum_{i=1}^{k-m} C_k^i p^{k-i} q^i T(r-i) + 1 \qquad (4)$$

Далее возможны два случая.

1) Если $k = K_0(r) = k_0$, то

$$T(r) = p^{k_0} T(r) + \sum_{i=1}^{k_0-m} C_{k_0}^i p^{k_0-i} q^i T(r-i) + 1$$

Отсюда получаем рекуррентное соотношение для вычисления $T(r)$

$$T(r) = \frac{1}{1-p^{k_0}} \left(\sum_{i=1}^{k_0-m} C_{k_0}^i p^{k_0-i} q^i T(r-i) + 1 \right). \qquad (5)$$

2)Если в уравнение (4) подставить $k = k_1 \neq k_0$, то

$$T(r) \geqslant \frac{1}{1-p^{k_1}} \left(\sum_{i=1}^{k_1-m} C_{k_1}^i p^{k_1-i} q^i T(r-i) + 1 \right). \qquad (6)$$

Таким образом, $T(r) = \max\limits_{m\leqslant k\leqslant r} \frac{1}{1-p^k} \left(\sum\limits_{i=1}^{k-m} C_k^i p^{k-i} q^i T(r-i) + 1 \right).$

3) Пусть уже вычислены $K_{m+1}, K_0(m+2), \ldots, K_0(r)$ и, соответственно, $T(m+1), T(m+2), \ldots, T(r)$. Тогда для вычисления $K_0(r+1)$ достаточно взять два значения параметра k: $k = K_0(r)$ и $k = K_0(r)+1$ и вычислить $f(K_0(r), r+1)$ и $f(K_0(r)+1, r+1)$, где $f(k,r) = \frac{1}{1-p^k}\left(\sum\limits_{i=1}^{k-m} C_k^i p^{k-i} q^i T(r-i)+1\right)$. Выбирая больше значение, получаем значение $T(r+1)$, а соответствующее ему значение параметра k и будет $K_0(r+1)$.

Таким образом, с помощью полученных автором свойств оптимальных стратегий резервирования построен эффективный алгоритм для вычисления оптимальной стратегии и среднего времени безотказной работы системы.

Литература

1. Райкин А.Л. Маневрирование аппаратурной избыточностью в реальных системах // Труды Ш Всесоюзного совещания по автоматическому управлению. Одесса, 1965. М.: Наука, 1967. Т.5: Технические средства автоматики. С. 94.

2. *Герцбах И.Б.* Об оптимальном управлении включением резервных элементов // Изв. АН СССР. Сер. Техническая кибернетика. 1966. №5. С. 75-80.

3. Чистяков В.П. Курс теории вероятностей. 5-е изд. М.: Агар; - 2000, 256 с.

4. Губин В.Н. Об одном классе резервируемых устройств / В.Н. Губин, Г.Г. Пестов // Вестник Томского госуниверситета. Математика и механика. 2014. №4(29). С. 14–23.

5. Губин В.Н. Две задачи динамического резервирования / В.Н. Губин, В.В. Травкина // Вестник Томского государственного университета. Математика и механика. 2013. № 5(25). С. 5-12.

Gubin V. N.(Tomsk State University, Tomsk, 2017) **Calculating of the mean time of life system using the properties of optimal strategies.**

Abstract. The main properties of optimal strategies for the Raykin-Gerzbach dynamic reservation model are described in the article, when the average time of failure-free operation for an infinite time interval was chosen as an optimization criterion. Based on the above properties, an algorithm is constructed for calculating the average time of trouble-free operation of the system and the optimal strategy. The obtained algorithm significantly shortens the calculation time of the optimal strategy in comparison with a simple search, which is very important in applied reliability problems.

Key words: optimal strategy, average uptime, dynamic redundancy, system failure.

О последовательном оценивании периодического сигнала на фоне аддитивных зависимых шумов авторегрессионного типа

Емельянова Т. В., Конищева А. А.

Томский государственный университет, Томск
e-mail: tv_em@mail.ru

Аннотация

Рассматривается задача оценивания периодического сигнала, наблюдаемого в дискретные моменты времени на фоне аддитивного шума авторегрессионного типа с неизвестной спектральной плотностью и распределением. Предлагается оценка для параметров сигнала по методу наименьших квадратов(МНК) со специальным правилом прекращения наблюдений. В отличие от классической оценки МНК такая оценка позволяет контролировать среднеквадратическую точность оценок. Получены формулы для среднеквадратической погрешности последовательной оценки, а также средней длительности предложенной процедуры. Приводятся результаты численного моделирования по методу Монте-Карло, дается сравнение с оценками МНК.

Ключевые слова: последовательное оценивание, заданная среднеквадратическая точность, тригонометрическая регрессия, момент остановки, авторегрессионный шум.

Введение. В задачах, связанных с анализом сигналов в информационно-измерительных комплексах, с обработкой изображений, с автоматическим управлением и адаптивным регулированием широко используются регрессионные модели с дискретным и непрерывным временем. В настоящее время разработаны различные методы оценивания параметров сигналов таких моделей на фоне аддитивных помех при различных уровнях заранее известной информации о типе сигнала и виде помех [1] - [6]. Для случая дискретного времени задача выделения сигналов наиболее полно исследована в случае, когда помехи являются последовательностью независимых случайных величин.

Менее изучена проблема оценивания параметров сигнала при коррелированных шумах с неизвестными спектральными свойствами. Наличие дополнительных неизвестных (мешающих) параметров спектра шума существенно усложняет задачу вычисления точности оценок параметров сигнала (см., например, [6]). В работе [4] была предложена последовательная процедура оценивания периодического сигнала на фоне авторегрессионных помех с неизвестными параметрами, обладающая хорошими асимптотическими свойствами и гарантирующая оценивание параметров сигнала с любой заданной среднеквадратической точностью. Эта процедура, однако, оказывается достаточно сложной для практической реализации в случае нескольких неизвестных параметров, поскольку является двухэтапной и включает построение серии последовательных оценок МНК.

В данной работе предлагается одноэтапная последовательная процедура оценивания периодического сигнала при авторегрессионном шуме с неизвестными параметрами, которая обеспечивает оценивание параметров сигнала с заданной среднеквадратической точностью в отличие от классической оценки МНК.

Постановка задачи. Построение последовательной процедуры. Рассмотрим задачу оценивания параметров $\mu_1, \mu_2, \beta_{j_1}, \beta_{j_2}$, $j = 1, \ldots, r$ тригонометрического сигнала

$$S_n = \mu_1 + (-1)^n \mu_2 + \sum_{j=1}^{r} \beta_{j_1} cos\omega_j n + \beta j_2 sin\omega_j n \tag{1}$$

по наблюдениям

$$x_n = S_n + \xi_n, \tag{2}$$

где ξ_n - шум, являющийся стационарным процессом авторегрессии p-го порядка:

$$\xi_n = \lambda_1 \xi_{n-1} + \ldots + \lambda_p \xi_{n-p} + \varepsilon_n. \tag{3}$$

Здесь $\{\varepsilon_n\}$ - последовательность независимых одинаково распределенных случайных величин, $E\varepsilon_n = 0, E\varepsilon_n^2 = \sigma^2$;
$\lambda_1, \ldots, \lambda_p$ - неизвестные параметры такие, что все корни характеристического полинома $P(q) = q^p - \lambda q^{p-1} - \ldots - \lambda_p$ лежат внутри единичного круга комплексной плоскости.

С учетом (1) и (3), наблюдаемый процесс (2) удовлетворяет уравнению

$$x_n = m_1 + (-1)^n m_2 + \sum_{j=1}^{r} (\gamma_{j_1} cos\omega_j n + \gamma_{j_2} sin\omega_j n) +$$

$$+\sum_{k=1}^{p}\lambda_k x_{n-k}+\varepsilon_n, n\geq p+1 \qquad (4)$$

Используя обозначения

$$Y_n=\begin{pmatrix}\Phi_n\\ X_{n-1}\end{pmatrix}, X_n=\begin{pmatrix}x_n\\ .\\ .\\ .\\ x_{n-p+1}\end{pmatrix}, \Phi_n=\begin{pmatrix}\phi_1(n)\\ .\\ .\\ .\\ \phi_{2r+2}(n)\end{pmatrix},$$

$\phi_1(n)=1, \phi_2(n)=(-1)^n, \phi_k(n)=cos\omega_{k-2}n$ при $3\leq k\leq r+2$,

$\Lambda\phi_k(n)=sin\omega_{k-r-2}n$ при $r+3\leq k\leq 2r+2$,

уравнение (4) записывается в векторной форме:

$$X_n=\alpha' Y_n+\varepsilon_n, n\geq p+1, Y_n=\begin{pmatrix}\Phi_n\\ X_{n-1}\end{pmatrix}, \alpha\in\Lambda, \qquad (5)$$

где $\alpha=(m_1,m_2,\gamma_{11},\gamma_{12},\ldots,\gamma_{r1},\gamma_{r2})$ - вектор оцениваемых параметров; Λ - множество всех допустимых значений вектора, учитывающее требования на параметры $\lambda_1 ... \lambda_p$ в (3). Таким образом, исходная задача сводится к оцениванию вектора параметров по наблюдениям процессов $(X_n, Y_n), n=1,\ldots,N$.

Оценка МНК вектора параметров α по наблюдениям процесса (5) имеет вид

$$\alpha(n)=M_n^{-1}\sum_{k=1}^{n}Y_k x_k, \qquad (6)$$

где $\sum_{k=1}^{n}Y_k Y_k'$, - выборочная информационная матрица Фишера. Будем предполагать, что минимальное собственное значение $\lambda_1(M_n)$ матрицы M удовлетворяет с вероятностью единица условию

$\lambda_1(M_n)\to\infty$ при $n\to\infty$.

Как следует из (5), обратная матрица M_n^{-1} в (6) является случайной. Это создает трудности при анализе среднеквадратической точности оценки вектора параметров α. Чтобы обойти эти трудности, будем использовать последовательную оценку МНК со специальным правилом остановки наблюдений. Выбор такого правила можно осуществить, используя оценку уклонения оценки МНК в модели типа (5), полученную в [5]. Для процесса (5) эта оценка будет иметь вид

$$\|\alpha(n)-\alpha\|^2\leq\|M_n^{-2}\|\|m_n\|^2, \qquad (7)$$

где $m_n=\sum_{k=1}^{n}Y_k\varepsilon_k$.

Поскольку в силу условия на матрицу множитель $\|M_n^{-2}\|$ в правой части (7) монотонно стремится к нулю с ростом объема выборки n, это можно использовать для выбора момента остановки $\tau(h)$

последовательной процедуры.

Для любого положительного h
$$\tau=\tau(h)=inf\{n\geq 1:\|M_n^{-2}\|^{\frac{1}{2}}\leq\frac{1}{h}\}. \qquad (8)$$
Последовательную оценку МНК $\alpha^*(h)$ параметра α определим равенством
$$\alpha^*(h)=\tilde{M}^{-1}_{\tau(h)}\sum_{k=p+1}^{\tau(h)}\beta_k Y_k x_k,\tilde{M}^{-1}_{\tau(h)}=\sum_{k=p+1}^{\tau(h)}\beta_k Y_k Y'_k, \qquad (9)$$
$$\beta_k=\begin{cases}1, & \text{если } k<\tau(h),\\ \nu(h), & \text{если } k=\tau(h),\end{cases}$$
где корректирующий множитель $\nu(h),0<\nu(h)\leq 1$ находится из равенства
$$\left\|\left(\sum_{k=1}^{\tau(h)-1}Y_kY_k^{'}+\nu(h)Y_{\tau(h)}Y_{\tau(h)}^{'}\right)^{-2}\right\|^{\frac{1}{2}}=\frac{1}{h}. \qquad (10)$$

Теоретические свойства последовательной процедуры. При изучении свойств последовательного плана (8), (9) в зависимости от выбора порога h будем предполагать, что вектор неизвестных параметров α в уравнении (7), принадлежит некоторому известному компакту K из параметрического множества Λ. Результаты анализа последовательного плана (8), (9), относящиеся к средней длительности последовательной процедуры и качеству последовательных оценок неизвестных параметров, сформулируем в виде теорем.

Теорема 1. *Пусть $(\varepsilon_n)_{n\geq 1}$ - последовательность н.о.р. случайных величин, $E\varepsilon_n=0, E\varepsilon_n^2=\sigma^2, E\varepsilon_n^8<\infty, \Lambda$ - множество допустимых значений вектора параметров α. Тогда для любого компакта $K\subset\Lambda$*
$$\lim_{h\to\infty}\sup_{\alpha\in K}\left|E_\alpha\frac{\tau(h)}{h}-\left\|F^{-2}\right\|^{\frac{1}{2}}\right|=0,$$
где F – положительно определенная матрица, имеющая вид: $\begin{pmatrix} M_0 & M_1 \\ M_1^{'} & F_0+DM_0D \\ \dots & a_{2n}\end{pmatrix}$, *в которой* $M_0=diag\left(1,1;\frac{1}{2},\dots,\frac{1}{2}\right)$.
$$V_k=\begin{pmatrix}1 & 0 & 0 & 0\\ 0 & (-1)^k & 0 & 0\\ 0 & 0 & V_1(k) & V_2(k)\\ 0 & 0 & -V_2(k) & V_1(k)\end{pmatrix},$$

причем
$V_1(k) = diag(cos\omega_1 k, \dots, cos\omega_r k),\ V_2(k) = diag(sin\omega_1 k, \dots, sin\omega_r k),$
$$A = \begin{pmatrix} \lambda_1 & \dots & \lambda_p \\ I_{p-1} & \dots & 0 \end{pmatrix}, \Gamma = \begin{pmatrix} m_1 & m_2 & \gamma_{11} & \gamma_{12} & \dots & \gamma_{r1} & \gamma_{r2} \\ 0 & \dots & \dots & \dots & \dots & \dots & 0 \\ 0 & \dots & \dots & \dots & \dots & \dots & 0 \end{pmatrix},$$
$$D = \sum_{k\geq 0} A^k \Gamma V(k), M_1 = M_0 V'(1) D', F_0 = \lim_{n\to\infty} \frac{1}{N} \sum_{n=p+1}^{N} \zeta_n \zeta_n' \text{п.н.},$$
$$\zeta_n = A\zeta_{n-1} + \eta_n, \zeta_p = 0, \eta_n = (\varepsilon_n, 0, ..., 0)'.$$

Заметим, что матрица M_n обладает требуемым свойством: $\lambda_{min}(M_n) \to \infty$ при $n \to \infty$.

Теорема 2. *Для любого компактного множества $K \subset \Lambda$ среднеквадратическая точность последовательной оценки удовлетворяет неравенству*
$$\sup_{\alpha\in K} E_\alpha(\|\alpha^*(h) - \alpha\|^2) \leq \frac{b_k}{h}(1 + o(1)),$$
где $b_k = \sup_{\alpha\in K} \phi(\alpha), \phi(\alpha) = Q(\alpha)\|F^{-2}\|^{\frac{1}{2}}, o(1) \to 0$ при $h \to \infty$,
$$Q(\alpha) = trF_0 + 2r + 2 + (2r+2)^2 \left(\|D\|^2 + \frac{c}{1-q}\right).$$

Замечание 1. *Теорема 2 позволяет контролировать среднеквадратическую точность последовательной оценки с помощью выбора порога h, учитывая, что величина b_k может быть вычислена априори. При этом средняя длительность процедуры, согласно теореме 1, растет линейно с ростом h.*

Экспериментальное исследование процедуры. В данном разделе представлены результаты численного моделирования предлагаемой последовательной процедуры (8)-(9) оценивания параметров регрессии с дискретным временем по наблюдениям с аддитивным шумом Орнштейна-Уленбека (3)
$$X_n = S_n + \xi_n \tag{11}$$
на примере функции
$$S_n = a_1 \cos\frac{\pi n}{6} + a_2 \sin\frac{\pi n}{3}. \tag{12}$$

Для проверки согласия выборочных свойств оценок (8), (9) с теоретическими результатами теорем 1, 2 проводилось моделирование по методу Монте-Карло, включавшее 100 повторений процедуры при различных значениях порога H определяющего момент остановки $\tau = \tau(H)$.

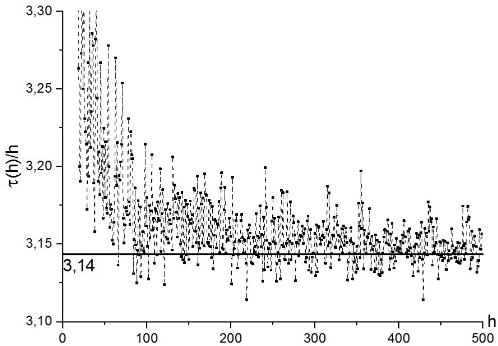

Рис. 1. Сходимость выборочной средней длительности процедуры к теоретическому значению при $\lambda = -0.6$

Выборочные характеристики, касающиеся точности последовательных оценок, полученных по 100 реализациям процедуры с порогом $h = 50$, представлены в таблице 1.

Таблица 1 — Эмпирические среднеквадратические точности оценок.

a_1	0.5	0.5	0.5	2.5	2.5
a_2	0.7	0.7	0.7	3.5	3.5
λ	-0.6	-0.6	-0.9	-0.6	-0.9
H	50	100	100	100	100
$SD(a_1^*)$	0.0062	0.001209	0.002835	0.000093	0.000212
$SD(a_2^*)$	0.0030	0.004286	0.004606	0.000216	0.000471
$SD(\tilde{a}_1^*)$	0.0006	0.003383	0.000584	0.000407	0.002096
$SD(\tilde{a}_2^*)$	0.0376	0.002441	0.005686	0.000546	0.008286

Здесь, наряду со значениями параметров модели a_1, a_2, в следующих за ними строках, приводятся среднеквадратические точности полученных последовательных оценок. В строках с заголовками $SD(\tilde{a}_1^*), SD(\tilde{a}_2^*)$ даются среднеквадратические уклонения обычных оценок МНК (7), вычисленных по реализации длины N. Сравнение последовательных и обычных оценок МНК показывает, что они близки по точности. При этом, однако, следует учитывать, что последовательные оценки позволяют контролировать среднеквадратическую точность путем выбора порога h.

Заключение. В работе построена последовательная процедура оценивания параметров тригонометрического сигнала, наблюдаемого на фоне авторегрессионного шума. Используется специальное правило прекращения наблюдений, определяемое по выборочной информационной матрице Фишера, которое гарантирует заданную среднеквадратическую точность оценок неизвестных параметров. Результаты работы могут быть использованы в задачах автоматического управления и идентификации.

Литература

1. Андерсон Т. Статистический анализ временных рядов. М.: Мир, 1976. — 756 с.

2. Ибрагимов И.А. Асимптотическая теория оценивания. / Ибрагимов И.А., Хасьминский Р.З. — М.: Наука, 1979. – 527 с.

3. Липцер Р.Ш. Статистика случайных процессов. / Липцер Р.Ш., Ширяев А.Н. — М.: Наука, 1974.– 696 с.

4. Конев В.В. Гарантированное оценивание периодического сигнала на фоне авторегрессионных помех с неизвестными параметрами / Конев В.В., Пергаменщиков С.М. // Проблемы передачи информации. 1997. Т.33. Вып. 4. С.26-44

5. Galtchouk L., Konev V. On Sequential Least Squares Estimates of Autoregressive Parameters // Sequential Analysis. 2005. V. 24. No. 4. P. 335-364.

6. Konev V., Pergamenshchikov S. On guaranteed estimation of the mean of an autoregressive process // Ann. Statist. 1997. V. 25, No. 5. P. 2127-2163.

Emelyanova T. V., Konishcheva A. A. (Tomsk State University, Tomsk, 2017) **On sequential estimation of a periodic signal distorted by an additive autoregressive noise.**

Abstract. The problem of estimating coefficients of a periodic signal in a discrete time from observations with an additive noise described by a stationary autoregressive process with unknown parameters and unknown distribution is considered. One-step sequential procedure to estimate signal coefficients is proposed, which provides a given mean-square accuracy of estimates for any values of the nuisance parameters. An asymptotic formula for the mean duration of the procedure is constructed. The results of Monte-Carlo simulation of the sequential procedure and its comparison with the LS-estimates are given.

Key words: sequential estimation, given mean-square accuracy, trigonometric regression, stopping time, autoregressive noise

Об асимптотической нормальности одноэтапного последовательного плана оценивания параметров непрерывной авторегрессии

Емельянова Т. В., Шерстобитова А. О.

Томский государственный университет, г. Томск
e-mail: annaivashchenko06@gmail.com

Аннотация

В настоящей работе предлагается одноэтапная последовательная процедура оценивания неизвестных параметров процесса авторегрессии с непрерывным временем, использующая специальное правило остановки наблюдений, которая позволяет оценить среднеквадратическую точность оценок. Процедура строится на основе классических оценок по методу наименьших квадратов (МНК), но в отличие от них обеспечивает контроль за среднеквадратической точностью оценок. Формулируется теорема об асимптотической нормальности последовательных оценок.

Ключевые слова: модель авторегрессии с непрерывным временем, одноэтапная последовательная процедура оценивания, среднеквадратическая точность, асимптотическая нормальность.

В теоретических и прикладных исследованиях, связанных с задачами обработки временных рядов и их спектральным анализом, задачами автоматического управления и регулирования, идентификации и фильтрации, в физике и финансовой инженерии широко используются динамические системы, описываемые стохастическими дифференциальными и стохастическими разностными уравнениями. Широкое применение эти модели в последнее время находят в стохастической финансовой математике [7], [1] для описания динамики цен, построения инвестиционных стратегий, расчета опционов различных типов и других финансовых расчетов. В прикладных задачах эти уравнения задаются с точностью до параметров, поэтому решению основных задач фильтрации, прогнозирования, управления обычно предшествует этап идентификации, заключающийся в оценивании неизвестных параметров.

В настоящее время существует достаточно много исследований, посвященных задачам асимптотического оценивания [5]- [3]. Однако, для практических задач типична неасимптотическая проблема оценивания, когда требуется определить длину реализации, при которой оценки достигают заданной точности [4]. Для решения задач в неасимптотической постановке требуются методы, позволяющие контролировать точность оценок при малых и средних объемах данных. В практических задачах объем доступных данных всегда конечен и желательно иметь представление о качестве оценок, вычисленных по наблюдениям на ограниченном временном интервале. Одним из подходов к задачам оценивания в неасимптотической постановке является подход с позиции последовательного анализа, который характеризуется тем, что длительность наблюдений не фиксируется заранее и определяется специальными правилами остановки.

Целью работы является исследование асимптотического распределения оценок параметра, полученных с помощью одноэтапной последовательной процедуры, предложенной в работе [2].

Пусть наблюдаемый p-мерный процесс $X_t = (X_1(t), \ldots, X_p(t))'$ описывается системой линейных дифференциальных уравнений

$$dX_t = AX_t dt + BdW_t \tag{1}$$

в которой A и B – квадратные матрицы постоянных коэффициентов размера $p \times p$, W_t – стандартный p-мерный процесс броуновского движения.

$$X_t = \begin{pmatrix} x_t^1 \\ x_t^2 \\ \ldots \\ x_t^p \end{pmatrix}; A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ . & . & . & 1 & 0 \\ . & . & . & . & 0 \\ \theta_p & \theta_{p-1} & \ldots & \ldots & \theta_1 \end{pmatrix};$$

$$B = \begin{pmatrix} 0 & \ldots & 0 \\ 0 & \ldots & 0 \\ 0 & \ldots & \sigma \end{pmatrix}; \sigma > 0. \tag{2}$$

Задача состоит в том, чтобы оценить неизвестные коэффициенты матрицы $A = \|a_{ij}\|$ по наблюдениям процесса X_t. К этой задаче сводится задача оценивания параметров стационарного гауссовского процесса авторегрессии p-го порядка ($AR(p)$), описываемого уравнением

$$dx_t^{p-1} = \left(\theta_1 x_t^{p-1} + \ldots + \theta_p x_t\right) dt + \sigma dw_t \tag{3}$$

с рациональной спектральной плотностью, имеющей вид

$$f(\lambda) = \frac{\sigma^2}{|Q(i\lambda)^2|}.$$

Предположим, что процесс авторегрессии (1) устойчив, т. е. все корни характеристического полинома $Q(z) = z^p - \theta_1 z^(p-1) - \ldots - \theta_p$ лежат в единичном круге. Пусть $H > 0$.

При построении последовательного плана будет использоваться лемма [2], дающая оценку нормы уклонения оценки (4)

$$\hat{\theta}_T = M_T^{-1} \int\limits_0^T X_s d\langle X_t \rangle_p \tag{4}$$

от ее истинного значения.

Лемма 1. *Пусть матрица M_T в (5) невырождена. Тогда квадрат нормы уклонения оценки (4) удовлетворяет неравенству*

$$\|\hat{\theta}_T - \theta\|^2 \leq \|M_T^{-2}\| \cdot \|m_T\|^2, \tag{5}$$

где $m_T = \int\limits_0^T X_s dW_s$.

Заметим, что в силу леммы 1 [2] $\|M_T^{-2}\|^{\frac{1}{2}}$ монотонно убывает, поэтому определим длительность наблюдений процесса

$$\tau = \tau(H) = inf\{t > 0 : \|M_T^{-2}\|^{\frac{1}{2}} \leq \frac{1}{H}\} \tag{6}$$

и последовательную оценку МНК неизвестных параметров

$$\theta^*(H) = M_{\tau(H)}^{-1} \int\limits_0^\tau (H) X_s d\langle X_t \rangle_p. \tag{7}$$

Последовательный план (6, 7) позволяет контролировать средне-квадратическую точность получаемых оценок за счет выбора порога процедуры H. Длительность процедуры при этом пропорциональна порогу процедуры.

Асимптотическое распределение последовательных оценок устанавливает следующая теорема.

Теорема 1. *Пусть задан процесс вида (1)*

$$dX_t = AX_t dt + BdW_t \tag{1}$$

где A и B – квадратные матрицы размера $p \times p$, W_t – стандартный p-мерный процесс броуновского движения.

$$X_t = \begin{pmatrix} x_t^1 \\ x_t^2 \\ \ldots \\ x_t^p \end{pmatrix}; A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ . & . & . & 1 & 0 \\ . & . & . & . & 0 \\ \theta_p & \theta_{p-1} & \ldots & \ldots & \theta_1 \end{pmatrix};$$

$$B = \begin{pmatrix} 0 & \dots & 0 \\ 0 & \dots & 0 \\ 0 & \dots & \sigma \end{pmatrix}; \sigma > 0.$$

Пусть неизвестные параметры θ_i, таковы, что все корни характеристического полинома $Q(z) = z^p - \theta_1 z^{p-1} - \ldots - \theta_p$ лежат в единичном круге. Последовательный план $(\tau_H, \theta^*(H)$ задается формулами

$$\tau = \tau(H) = inf\{t > 0 : \|M_T^{-2}\|^{\frac{1}{2}} \leq \frac{1}{H}\}, \tag{6}$$

$$\theta^*(H) = M_{\tau(H)}^{-1} \int_0^{\tau} (H) X_s d\langle X_t \rangle_p, \tag{7}$$

где $H > 0$ – пороговое значение.

Тогда вектор $\frac{1}{\sqrt{H}} (\theta^*(H) - \theta)$ имеет асимптотически нормальное распределение с параметрами $\left(0, F^{-1}\right)$.

Результат теоремы может быть использован для построения доверительных интервалов для параметров модели авторегрессии, а также для исследования оптимальности одноэтапной последовательной процедуры оценивания.

Литература

1. Karatzas, I. Methods of Mathematical Finance/ I. Karatzas, S. Shreeve. – New York : Columbia Univ. Press, 1995. – 321 p.

2. Емельянова Т. В. О последовательном оценивании параметров непрерывной авторегрессии / Емельянова Т. В., Конев В. В. // Вестник Томского гос. у-та. Математика и механика. 2013. №5(25). С. 12-25.

3. Ибрагимов И.А. Асимптотическая теория оценивания. / Ибрагимов И.А., Хасьминский Р.З. — М.: Наука, 1979. – 527 с.

4. Конев В. В. О последовательном оценивании параметров случайных процессов диффузионного типа / Конев В. В., Пергаменщиков С. М. // Пробл. Передачи информ. —1985. — Т. 21, вып. 1. — С.48-61.

5. Кутоянц Ю. А. Оценивание параметров случайных процессов. / Кутоянц Ю. А. – Ереван. : АН АрмССР, 1980.

6. Холево А. С. Оценка параметра снова диффузионного процесса методом стохастической аппроксимации // Исследование по теории самонастраивающихся систем. –М.: Наука, 1967. – С. 179-200.

7. Ширяев, А. Н. Основы стохастической финансовой математики / А. Н. Ширяев. – М. : Фазис, 1998. – Т. 1. – 512 с. – Т. 2. – 544 с.

Emelyanova T. V., Sherstobitova A. O. (Tomsk State University, Tomsk, 2017) **Asymptotic Normality of a One-Stage Sequential Plan for Estimating the Continuous Autoregessive Model's Parameters.**

Abstract. We propose a one-step sequential procedure for estimating unknown parameters of an autoregressive process with continuous time using a special observation stopping rule that allows to estimate the root-mean-square accuracy of estimates. The procedure is based on classical least-squares (LS) estimation method and provides control over the root-mean-square accuracy of estimates. There is formulated a theorem about asymptotic normality of sequential estimates.

Key words: autoregressive model with continuous time, a one-step procedure of sequential estimation, mean-square accuracy, asymptotic normality.

О последовательном оценивании параметра авторегрессии по наблюдениям с аддитивными шумами. *

Конев В. В., Назаренко Б. Н.

Томский государственный университет, Томск
e-mail: vvkonev@mail.tsu.ru

Аннотация

Рассматривается задача оценивания параметра процесса авторегрессии первого порядка по наблюдениям с аддитивным шумом. Предполагается, что распределения аддитивных шумов, а также шумов, задающих динамику процесса $AR(1)$, являются гауссовскими. На основе оценки Юла-Уокера построена последовательная оценка неизвестного параметра. Установлена гауссовость этой оценки при специальной нормировке. Рассмотрена задача построения доверительного интервала фиксированной ширины с заданным коэффициентом доверия. Приводятся результаты численного моделирования.

Ключевые слова: последовательная оценка, авторегрессия, аддитивные шумы, гарантированное оценивание, доверительное оценивание, оценка Юла-Уокера.

Постановка задачи. Пусть $\{x_k\}_{k\geqslant 0}$ – процесс авторегрессии первого порядка, удовлетворяющий уравнению

$$x_k = \theta x_{k-1} + \varepsilon_k, \quad k \geqslant 1, \tag{1}$$

где $(\varepsilon_k)_{k\geqslant 1}$ – ненаблюдаемый процесс (шум), θ – неизвестный параметр. Требуется оценить параметр θ по наблюдениям процесса

$$y_k = x_k + \eta_k, \quad k \geqslant 1. \tag{2}$$

Предположим, что шумы $\{\varepsilon_k\}$, $\{\eta_k\}$ являются независимыми последовательностями одинаково распределенных гауссовских случайных величин с нулевыми средними и дисперсиями 1 и Δ соответственно, т.е. $\varepsilon_k \sim \mathcal{N}(0,1)$ и $\eta_k \sim \mathcal{N}(0,\Delta)$. Начальное значение x_0 является случайной величиной, не зависящей от процессов $\{\varepsilon_k\}$ и $\{\eta_k\}$. Параметр θ предполагается ограниченным известной постоянной L, т.е. $|\theta| \leqslant L < \infty$.

*Работа выполнена при финансовой поддержке Министерства образования и науки РФ, госзадание № 2.3208.2017/4.6

Замечание 1. *Для оценивания параметра θ в модели (1), (2) с гарантированной среднеквадратической точностью можно применять метод, предложенный в работе [3].*

Построим последовательную оценку параметра θ на основе классической оценки Юла-Уокера (см, например, Андерсон [1])

$$\widehat{\theta}_n = \frac{1}{\sum\limits_{k=1}^{n} y_{k-2}y_{k-1}} \sum_{k=1}^{n} y_{k-2}y_k. \tag{3}$$

Построение последовательной оценки Юла-Уокера. Основной результат. Для каждого числа $h > 0$ введем момент остановки

$$\tau = \tau(h) = \inf\left\{ n \geqslant 1 : \sum_{k=1}^{n} y_{k-2}^2 \geqslant \frac{h}{\Delta + 1} \right\}. \tag{4}$$

Построение последовательной оценки осуществляется по схеме, предложенной в работе [2].

Вычислим корректирующий множитель $\alpha(h)$, $0 < \alpha(h) \leqslant 1$, исходя из уравнения

$$\sum_{k=1}^{\tau(h)-1} y_{k-2}^2 + \alpha(h) y_{\tau(h)-2}^2 = \frac{h}{1+\Delta}. \tag{5}$$

Определим последовательную оценку Юла-Уокера по формуле

$$\theta^*_{\tau(h)} = \frac{1}{\sum\limits_{k=1}^{\tau(h)} \sqrt{\beta_k} y_{k-2}y_{k-1}} \sum_{k=1}^{\tau(h)} \sqrt{\beta_k} y_{k-2}y_k, \tag{6}$$

где

$$\beta_k = \begin{cases} 1, & \text{если } 1 \leqslant k < \tau(h); \\ \alpha(h), & \text{если } k = \tau(h). \end{cases}$$

Для получения свойств этой оценки нам потребуется следующий результат.

Теорема 1. *Пусть $(M_k, \mathcal{F}_k)_{k\geqslant 0}$ – квадратично-интегрируемый мартингал с квадратической характеристикой $(\langle M \rangle_n)_{n\geqslant 1}$ такой, что*

1. $P(\langle M \rangle_\infty = +\infty) = 1$;
2. *Приращения $\Delta M_k = M_k - M_{k-1}$, $k \geqslant 1$, являются условно-гауссовскими, т.е.*

$$Law(\Delta M_k | \mathcal{F}_{k-1}) = \mathcal{N}(0, \sigma_{k-1}^2).$$

Для каждого $h > 0$ положим

$$\tau = \tau(h) = \inf\left\{ n \geqslant 1 : \sum_{k=1}^{n} \sigma_{k-1}^2 \geqslant h \right\},$$

$$m(h) = \frac{1}{\sqrt{h}} \sum_{k=1}^{\tau(h)} \sqrt{\beta_k(h)} \Delta M_k,$$

где $\beta_k(h)$ – последовательность весов

$$\beta_k(h) = \begin{cases} 1, & \text{если } 1 \leqslant k < \tau(h); \\ \alpha(h), & \text{если } k = \tau(h); \end{cases}$$

где $\alpha(h)$ – корректирующий множитель, определяемый из уравнения

$$\sum_{k=1}^{\tau(h)-1} \sigma_{k-1}^2 + \alpha(h)\sigma_{\tau(h)-1}^2 = h.$$

Тогда $m(h)$ является стандартной нормально распределенной случайной величиной, т.е. $m(h) \sim \mathcal{N}(0,1)$.

Замечание 2. *Доказательство этого результата, а также различные его применения для построения точечных и доверительных последовательных оценок на основе метода наименьших квадратов приводятся в работах (Конев [4], Konev and Vorobeychikov [5]).*

Сформулируем основной результат работы.

Теорема 2. *Для каждого $-L \leqslant \theta \leqslant L$ и $h > 0$*

$$P_\theta\left(z(h) \frac{\theta_{\tau(h)}^* - \theta}{\sqrt{\Delta(1+\theta^2)+1}} \leqslant t \right) = \Phi(t), \ -\infty < t < \infty,$$

где

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{u^2}{2}}\, du,$$

$$z(h) = \sqrt{\frac{\Delta+1}{h}} \sum_{k=1}^{\tau(h)} \sqrt{\beta_k(h)} y_{k-2} y_{k-1}.$$

Доказательство. Согласно (1) и (2) процесс y_k удовлетворяет уравнению

$$y_k = \theta y_{k-1} + \xi_k,$$

где $\xi_k = \eta_k - \theta\eta_{k-1} + \varepsilon_k$.

Подставляя (2) в (6) и учитывая (4), получаем

$$\theta^*_{\tau(h)} = \theta + \frac{1}{\sum\limits_{k=1}^{\tau(h)} \sqrt{\beta_k} y_{k-2} y_{k-1}} \sum_{k=1}^{\tau(h)} \sqrt{\beta_k} y_{k-2} \varepsilon_k.$$

Перепишем это уравнение в виде

$$\theta^*_{\tau(h)} = \theta + \frac{\sqrt{d}\sqrt{\frac{h}{\Delta+1}}}{\sum\limits_{k=1}^{\tau(h)} \sqrt{\beta_k} y_{k-2} y_{k-1}} m(h), \tag{7}$$

где $d = \Delta(1+\theta^2)+1$,

$$m(h) = \frac{1}{\sqrt{\frac{h}{\Delta+1}}} \sum_{k=1}^{\tau(h)} \sqrt{\beta_k} y_{k-2} \widetilde{\xi}_k, \tag{8}$$

$$\widetilde{\xi}_k = \frac{\xi_k}{\sqrt{d}} \sim \mathcal{N}(0,1).$$

Далее нам потребуется естественная фильтрация $\{\mathcal{F}_k\}_{k\geqslant 0}$, порожденная процессами ε_k и η_k:

$$\mathcal{F}_0 = \sigma(x_0),\ \mathcal{F}_k = \sigma(x_0, \varepsilon_1, \ldots, \varepsilon_k, \eta_1, \ldots, \eta_k),\ k \geqslant 1.$$

Введем мартингал относительно этой фильтрации

$$M_0 = 0,\ M_n = \sum_{k=1}^{n} y_{k-2} \widetilde{\xi}_k\ n \geqslant 1.$$

Этот мартингал удовлетворяет условиям Теоремы 1, причем

$$\langle M \rangle_n = \sum_{k=1}^{n} y_{k-2}^2,$$

$$Law(\Delta M_k | \mathcal{F}_{k-1}) = \mathcal{N}(0, y_{k-2}^2).$$

Применяя Теорему 1 к случайной величине (8) и учитывая равенство (7), приходим к утверждению Теоремы 2. □

Замечание 3. *Согласно Теореме 2 последовательная оценка (6) при выборе соответствующей нормировки является гауссовской при любых значениях параметра θ. Это позволяет строить доверительный интервал для неизвестного параметра θ заданной ширины с заданным коэффициентом доверия.*

Пусть $0 < \gamma < 1$ – доверительная вероятность и a_γ – число, определяемое равенством

$$\sqrt{\frac{2}{\pi}} \int_0^{a_\gamma} e^{-\frac{u^2}{2}}\, du = \gamma.$$

Положим

$$\delta_\gamma = a_\gamma \sqrt{\Delta(1+L^2)+1}.$$

Из Теоремы 2 получаем, что для любого $h > 0$ справедливо неравенство

$$\sup_{|\theta|\leqslant L} P_\theta \left(\left| \sum_{k=1}^{\tau(h)} \sqrt{\beta_k} y_{k-2} y_{k-1} \right| \cdot |\widehat{\theta}_{\tau(h)} - \theta| \geqslant \delta_\gamma \right) \leqslant 1-\gamma.$$

Таблица 1 — Результаты проверки гипотезы на нормальность распределения оценки (6) по 500 реализациям процедуры при $h = 30$

θ	p-value для теста Колмогорова	p-value для теста Шапиро-Уилка
-2.1	0.1583	0.0247
-1.8	0.4452	0.7473
-1.5	0.2810	0.1943
-1.2	0.6209	0.9613
-0.9	0.1173	0.6081
-0.6	0.6586	0.4939
-0.3	0.8822	0.7408
0	0.8012	0.4002
0.3	0.0434	0.4833
0.6	0.9028	0.4706
0.9	0.1008	0.2610
1.2	0.1151	0.0659
1.5	0.4448	0.0971
1.8	0.7654	0.6111
2.1	0.6446	0.0841

Результаты численного моделирования. Для проверки результата Теоремы 2 было проведено численное моделирование схемы наблюдений (1)-(2).

В пакете прикладных программ MatLab моделировались уравнения (1)-(2) при различных значениях параметра θ, указанных в первом столбце Таблицы 1 и дисперсии аддитивного шума $\Delta = 0.01$. По полученным значениям процесса $\{y_k\}$ строилась последовательная оценка (6) при пороге $h = 30$. Для каждого значения θ проце-

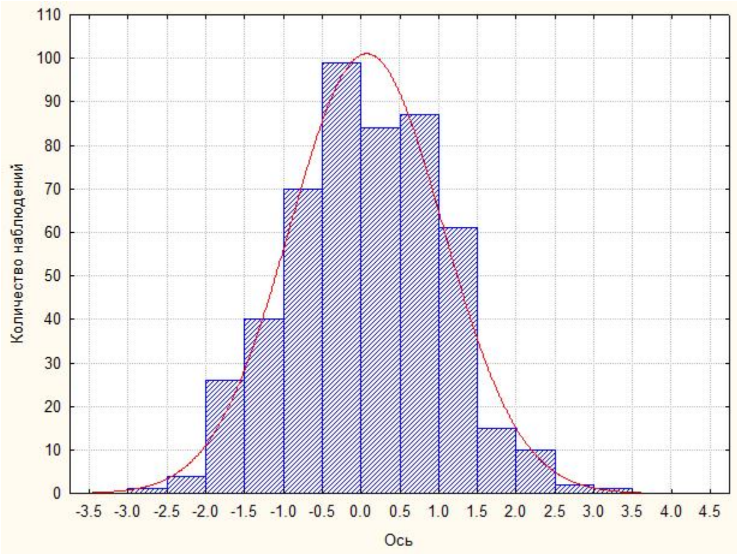

Рис. 1. Гистограмма выборки оценок (6) при $\theta = 0.3$

дура оценивания повторялась 500 раз. Полученные оценки, с соответствующим нормирующим множителем, проверялись в пакете R на нормальность с использованием известных тестов Колмогорова и Шапиро-Уилка. Для каждой выборки в Таблице 1 во втором и третьем столбце приведены значения p-value для указанных тестов. При превышении значения p-value выбранного уровня значимости 0.05 гипотеза о нормальности выборки принималась.

На основе данных Таблицы 1 можно сделать вывод, что численные результаты согласуются с утверждением Теоремы 2.

На рисунке 1 приводится гистограмма для $\theta = 0.3$ в масштабах наблюдений.

Литература

1. Anderson T. W. The statistical analysis of time series. – John Wiley & Sons, Inc. – 1971. – 704 p.

2. Борисов В. З. О последовательном оценивании параметров дискретных процессов /В. З. Борисов, В. В. Конев // Автомат. и телемех. – 1977. – № 10. – С. 58–64.

3. Васильев В. А. Последовательное оценивание параметров динамических систем при наличии мультипликативной и аддитивной

помех в наблюдениях /В. А. Васильев, В. В. Конев // Автомат. и телемех. – 1985. – № 6. С. 33–43.

4. Конев В. В. Об одном свойстве мартингалов с условно-гауссовскими приращениями и его применении в теории неасимптотических выводов // ДАН. – 2016. – Т. 471. – № 5. – С. 523-527.

5. Konev V. V. Non-asymptotic confidence estimation of the parameters in stochastic regression models with Gaussian noises /V. V. Konev, S. E. Vorobeychikov // Sequential Analysis. – 2017. – Vol. 36. – № 1. – P. 55-75.

Konev V. V., Nazarenko B. N.(Tomsk State University, Tomsk, 2017) **On sequential estimation of autoregressive parameter by observations with superimposed errors.**

Abstract. The paper considers the problem of estimating the parameter in an autoregressive process of first order by observations with superimposed errors. It is assumed that both the distributions of the superimposed errors and those of the noises determining the dynamics of the $AR(1)$ process are gaussian. This estimate is established to have non-asymptotic normal distribution if one uses special normalizing factor. The problem of constructing the confidence interval of fixed width with a given coverage probabilities is considered. The results of numerical simulations are presented.

Key words: sequential estimate, autoregression, superimposed noise, fixed-accuracy estimation, confidence estimation, Yle-Walker estimate.

Об оценивании функции сноса в диффузионных процессах *

Макарова И. А., Пчелинцев Е. А.

Томский государственный университет, Томск
e-mail: star_ irish@bk.ru

Аннотация

В настоящей работе рассматривается адаптивная задача непараметрического оценивания коэффициента сноса в диффузионных процессах. Предложена процедура выбора модели на основе улучшенных взвешенных оценках наименьших квадратов.

Ключевые слова: улучшенная оценка, стохастический диффузионный процесс , среднеквадратическая точность, выбор модели.

Введение. Пусть на стохастическом базисе $(\Omega, \mathcal{F}, (\mathcal{F})_{t\geq 0}, \mathbf{P})$ определено стохастическое дифференциальное уравнение

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

где $(w_t)_{t\geq 0}$ – стандартный винеровский процесс, начальное значение y_0 – некоторая известная постоянная, а $S(\cdot)$ – неизвестная функция. Задача заключается в том, чтобы оценить функцию $S(x)$, $x \in [a, b]$, по наблюдениям процесса $(y_t)_{0\leq t\leq T}$. Задача калибровки модели (1) является важной во многих приложениях. В частности, она возникает при построении оптимальных стратегий поведения инвесторов на диффузионных финансовых рынках. Известно, что оптимальная стратегия зависит от неизвестных параметров рынка, в том числе и от неизвестной функции сноса S. Поэтому в практических финансовых расчетах необходимо использовать статистические оценки для S, которые являются надежными на некотором фиксированном временном интервале $[0, T]$ [3]. Ранее, проблема неасимптотического оценивания параметров диффузионных процессов изучалась многими авторами (см., например, монографию [5] и ссылки в ней). Установлено также, что многие трудности неасимптотического оценивания параметров для одномерных

*Работа выполнена при финансовой поддержке РНФ, проект No 17-11-01049.

диффузионных процессов можно преодолеть, используя последовательный подход. Оказывается, что теоретический анализ последовательных оценок проще, чем анализ классических процедур. В частности, можно рассчитать неасимптотические оценки для квадратического риска.

В настоящей работе рассматривается задача оценивания неизвестной функции $S(x)$, $a \le x \le b$, в смысле среднеквадратического риска

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

где $\widehat{S}_T$ – оценка для S по наблюдениям $(y_t)_{0 \le t \le T}$, $a < b$ – некоторые вещественные числа. Здесь $\mathbf{E}_S$ обозначает математическое ожидание относительно распределения $\mathbf{P}_S$ случайного процесса $(y_t)_{0 \le t \le T}$ при заданной функции S.

Цель этой статьи – построить адаптивную оценку S^* коэффициента сноса S в (1) и показать, что квадратический риск этой оценки меньше, чем риск оценки, предложенной в [1], т.е. предлагается построить улучшенную оценку в смысле среднеквадратической точности. Для этого мы используем подход к улучшенному оцениванию, предложенный в [6] и [4] для параметрических моделей регрессии и развитый недавно в [7] для задач непараметрического оценивания. Более того, в статье рассматривается задача оценивания в адаптивной постановке, т.е. когда регулярность функции S неизвестна. Для этого используется метод выбора модели, предложенный в [7].

Далее, в разделе 2 статьи исходная задача сводится к задаче оценивания в непараметрической регрессионной модели с дискретным временем. В разделе 3 предлагаются улучшенные взвешенные оценки наименьших квадратов. В разделе 4 для оценивания функции S строится процедура выбора модели на основе улучшенных оценках МНК.

Неоднородная регрессионная модель с дискретным временем. Чтобы получить надежную оценку функции S, необходимо наложить на нее определенные условия, аналогичные периодичности детерминированного сигнала в модели белого шума [2]. Одним из условий, достаточного для этого, является предположение, что процесс $(y_t)_{t \ge 0}$ в (1) возвращается в любую окрестность каждой точки $x \in [a, b]$. Чтобы получить эргодичность процесса (1), предположим, что функция S принадлежит функциональному классу

$\Sigma_{L,N}$ ($L > 1$, $N > |a| + |b|$), определенному в [1]. Отметим, что если функция $S \in \Sigma_{L,N}$, то существует инвариантная плотность

$$q(x) \,=q_S(x) \,=\, \frac{\exp\{2\int_0^x S(z)\mathrm{d}z\}}{\int_{-\infty}^{+\infty} \exp\{2\int_0^y S(z)\mathrm{d}z\}\mathrm{d}y}\,. \tag{3}$$

Также функции из класса $\Sigma_{L,N}$ являются равномерно ограниченными на $[a, b]$, т.е.

$$s^* = \sup_{a\le x\le b} \sup_{S\in\Sigma_{L,N}} S^2(x) \,<\infty\,.$$

Рассмотрим разбиение интервала $[a, b]$ точками $(x_k)_{1\le k\le n}$, определенными как

$$x_k = a + \frac{k}{n}(b-a)\,, \tag{4}$$

где $n = n(T)$ – целочисленная функция от T такая, что

$$n(T) \le T \quad \text{and} \quad \lim_{T\to\infty} \frac{n(T)}{T} = 1\,. \tag{5}$$

Теперь в каждой точке x_k оценим функцию S последовательными ядерными оценками. Пусть $0 < t_0 < T$ – некоторый фиксированный момент, тогда положим

$$\begin{cases} \tau_k &= \inf\{t\ge t_0\,:\, \int_{t_0}^t Q\left(\frac{y_s-x_k}{h}\right)\mathrm{d}s \ge\, H_k\}\,; \\ \\ \widetilde{S}_k &= \frac{1}{H_k}\,\int_{t_0}^{\tau_k} Q\left(\frac{y_s-x_k}{h}\right)\mathrm{d}y_s\,, \end{cases} \tag{6}$$

где $Q(z) = \mathbf{1}_{\{|z|\le 1\}}$, $\mathbf{1}_A$ – индикатор множества A, $h = (b-a)/(2n)$ и H_k – положительные пороги, которые определены ниже. Из (1) нетрудно видеть, что

$$\widetilde{S}_k = S(x_k) + \zeta_k\,.$$

Здесь шум ζ_k является суммой аппроксимирующей и стохастической частей, т.е.

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

где $\Delta S(y, x) = S(y) - S(x)$ и

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

являются независимыми $\mathcal{N}(0, 1)$.

Далее необходимо оценить плотность (3) по наблюдениям $(y_t)_{0\le t\le t_0}$. Пусть

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

где $0 < \epsilon_T < 1$,

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

Тогда определим порог H_k в (6) следующим образом:

$$H_k = (T - t_0)(2\widetilde{q}_T(x_k) - \epsilon_T^2)h\,.$$

Предположим, что параметры $t_0 = t_0(T)$ и ϵ_T удовлетворяют следующим условиям:

$\mathbf{H}_1$) *Для всех $T \geq 32$,*

$$16 \leq t_0 \leq T/2 \quad \text{и} \quad \sqrt{2}/t_0^{1/8} \leq \epsilon_T \leq 1\,.$$

$\mathbf{H}_2$)

$$\lim_{T\to\infty} t_0(T) = \infty\,, \quad \lim_{T\to\infty} \epsilon_T = 0\,, \quad \lim_{T\to\infty} T\epsilon_T/t_0(T) = \infty\,.$$

$\mathbf{H}_3$) *Для всех $\nu > 0$ и $m > 0$,*

$$\lim_{T\to\infty} T\epsilon_T^m = \infty \quad \text{и} \quad \lim_{T\to\infty} T^m\, e^{-\nu\sqrt{t_0}} = 0\,.$$

Например, при $T \geq 32$,

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

Пусть

$$\Gamma = \{\max_{1\leq l\leq n} \tau_l \,\leq\, T\} \quad \text{и} \quad Y_k = \widetilde{S}_k\, \mathbf{1}_\Gamma\,. \tag{7}$$

Тогда на множестве Γ существует временная гетероскедастичная регрессионная модель

$$Y_k = S(x_k) + \zeta_k\,, \quad \zeta_k = \sigma_k\, \xi_k + \delta_k \tag{8}$$

с $\delta_k = B_k$ и

$$\sigma_k^2 = \frac{n}{(T - t_0)(\widetilde{q}_T(x_k) - \epsilon_T^2/2)(b - a)}\,.$$

Заметим, что из (5) и $\mathbf{H}_1$), находим верхнюю границу

$$\max_{1\leq k\leq n} \sigma_k^2 \leq \frac{4}{(b-a)\epsilon_T} = \sigma_* \tag{9}$$

для которой из условия $\mathbf{H}_3$) имеем

$$\lim_{T\to\infty} \frac{\sigma_*}{T^m} = 0 \quad \text{для всех} \quad m > 0\,.$$

Чтобы оценить S по наблюдениям (8), необходимо изучить некоторые свойства множества Γ в (7).

Предложение 1. *Пусть параметры t_0 и ϵ_T удовлетворяют условиям $\mathbf{H}_1$) – $\mathbf{H}_3$). Тогда*

$$\sup_{S\in\Sigma_{L,N}} \mathbf{P}_S(\Gamma^c) \leq \Pi_T\,,$$

где $\lim_{T\to\infty} T^m\, \Pi_T = 0$ для всех $m > 0$.

Основной результат. В этом параграфе мы рассмотрим задачу оценивания для модели (8). В ней $S(\cdot)$ – неизвестная функция, которую требуется оценить по наблюдениям $Y_1, \ldots, Y_n$.

Точность оценивания будем измерять эмпирическим среднеквадратическим риском вида

$$\|\widehat{S} - S\|_n^2 = (\widehat{S} - S, \widehat{S} - S)_n = \frac{b-a}{n} \sum_{l=1}^{n} (\widehat{S}(x_l) - S(x_l))^2 \,.$$

Пусть $(\phi_j)_{1 \le j \le n}$ – ортонормированный базис в смысле эмпирического скалярного произведения:

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

где $\mathbf{Kr}_{ij}$ – символ Кронекера. Воспользовавшись этим, применим дискретное преобразование Фурье к (8) и получим коэффициенты Фурье

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

Из (8) следует, что

$$\widehat{\theta}_{j,n} \, = \, \theta_{j,n} \, + \, \zeta_{j,n} \quad \text{с} \quad \zeta_{j,n} = \sqrt{\frac{b-a}{n}} \xi_{j,n} + \delta_{j,n} \,,$$

где

$$\xi_{j,n} = \sqrt{\frac{b-a}{n}} \sum_{l=1}^{n} \sigma_l \xi_l \phi_j(x_l) \quad \text{и} \quad \delta_{j,n} = \frac{b-a}{n} \sum_{l=1}^{n} \delta_l \, \phi_j(x_l) \,.$$

Введем класс взвешенных оценок МНК для S в (8) на сетке (4):

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

где весовой вектор $\lambda = (\lambda(1), \ldots, \lambda(n))$ принадлежит некоторому конечному множеству $\Lambda \subset [0,1]^n$. Тогда для каждого $a \le x \le b$

$$\widehat{S}_\lambda(x) = \widehat{S}_\lambda(x_1) \mathbf{1}_{\{a \le x \le x_1\}} + \sum_{l=2}^{n} \widehat{S}_\lambda(x_l) \mathbf{1}_{\{x_{l-1} < x \le x_l\}} \,. \tag{10}$$

Далее предположим, что первые $d \le n$ координат вектора λ равны 1, т.е. $\lambda(j) = 1$ для всех $1 \le j \le d$.

Определим новый класс оценок для S в (8):

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

где

$$\theta^*_{j,n} = \left(1 - \frac{c(d)}{\|\widetilde{\theta}_n\|} \mathbf{1}_{\{1 \le j \le d\}} \right) \widehat{\theta}_{j,n},$$

и

$$c(d) = \frac{(d-1)\sigma_*^2 L(b-a)^{1/2}}{n(s^* + \sqrt{d\sigma_*/n})}\,, \quad \|\widetilde{\theta}_n\|^2 = \sum_{j=1}^{d} \widehat{\theta}^2_{j,n}.$$

Тогда для каждого $a \le x \le b$ положим

$$S^*_\lambda(x) = S^*_\lambda(x_1)\mathbf{1}_{\{a\le x\le x_1\}} + \sum_{l=2}^{n} S^*_\lambda(x_l)\mathbf{1}_{\{x_{l-1}<x\le x_l\}}\,. \qquad (11)$$

Обозначим разность рисков оценок (11) и (10), как

$$\Delta_n(S) := \mathbf{E}_S\|S^*_\lambda - S\|^2_n - \mathbf{E}_S\|\widehat{S}_\lambda - S\|^2_n.$$

Оценки (11) позволяют контролировать точность.

Теорема 1. *Оценка* (11) *превосходит по среднеквадратической точности оценку* (10), *т.е.*

$$\sup_{S\in\Sigma_{L,N}} \Delta_n(S) < -c^2(d).$$

Улучшенная процедура выбора модели. Чтобы получить достаточно хорошую оценку, необходимо описать правило выбора весового вектора $\lambda \in \Lambda$ в (11). Ясно, что наилучшим способом является минимизация эмпирической среднеквадратической ошибки относительно λ:

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

Используя (11) и преобразование Фурье для S, имеем

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

Поскольку коэффициент $\theta_{j,n}$ неизвестен, необходимо заменить величины $\theta^*_{j,n}\,\theta_{j,n}$ их некоторыми оценками. Положим

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

Нужно заплатить штраф за эту замену в эмпирической среднеквадратической ошибке. Поэтому определим платежную функцию

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

где "штрафное" слагаемое

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

и $0 < \rho < 1$ – некоторое положительное число. Пусть

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

тогда выбирается следующая улучшенная оценка для S из класса (10):

$$S^*(x) = S^*_{\hat{\lambda}}(x) \quad \text{для} \quad a \le x \le b\,.$$

Литература

1. Galtchouk L.I., Pergamenshchikov S.M. Asymptotically efficient sequential kernel estimates of the drift coefficient in ergodic diffusion processes // Statistical Inference for Stochastic Processes. 2006. No 9. P. 1-16.

2. Ibragimov I.A., Hasminskii R.Z. Statistical Estimation: Asymptotic Theory. Springer, New York. 1979.

3. Karatzas I., Shreve S.E. Methods of Mathematical Finance. Springer, New York. 1998.

4. Konev V., Pergamenshchikov S., Pchelintsev E. Estimation of a regression with the pulse type noise from discrete data // Theory Probab. Appl. 2014. V. 58. No 3. P. 442–457.

5. Kutoyants Yu. Statistical inference for ergodic diffusion processes. Springer-Verlag, London. 2004.

6. Pchelintsev E. Improved estimation in a non-Gaussian parametric regression // Stat. Inference Stoch. Process. V. 16. No 1. P. 15 – 28.

7. Pchelintsev E., Pchelintsev V. and Pergamenshchikov S. Improved robust model selection methods for the Lévy nonparametric regression in continuous time // Preprint http://arxiv.org/abs/1710.03111. P. 1 – 32.

Makarova I.A., Pchelintsev E.A. (Tomsk State University, Tomsk, 2017) **On the estimation of a drift function in diffusion processes.**

Abstract. In this paper, we consider the robust adaptive nonparametric estimation problem for the drift coefficient in diffusion processes. An adaptive model selection procedure, based on the improved weighted least square estimates, is proposed.

Key words: Improved estimation, stochastic diffusion process, mean-square accuracy, model selection.

Асимптотически эффективное оценивание функции неоднородной регрессии *

Перелевский С. С., Пчелинцев Е. А.

Томский Государственный Университет, Томск
e-mail: slavaperelevskiy@email.ru

Аннотация

В статье рассматриваются некоторые асимптотические свойства адаптивной процедуры, предложенные в статье авторов [10] для оценки неизвестной функции неоднородной регрессии. Устанавливается, что процедура оценивания является асимптотически эффективной в смысле среднеквадратического риска, т.е. доказывается, что асимптотический среднеквадратический риск процедуры совпадает с соответствующей константой Пинскера, обеспечивающей точную нижнюю границу риска по всевозможным оценкам.

Ключевые слова: неоднородная регрессия, адаптивная процедура выбора модели, асимптотический среднеквадратический риск, константа Пинскера.

1. Введение. Предположим, что наблюдения $(y_j)_{1\le j\le n}$ описываются уравнением неоднородной регрессии

$$y_j = S(x_j) + \sigma_j \xi_j, \tag{1}$$

где $x_j = j/n$, $S(\cdot) \in \mathcal{L}_2[0,\,1]$ – неизвестная функция, которую требуется оценить, $(\xi_j)_{1\le j\le n}$ – последовательность независимых одинаково распределенных случайных величин таких, что $E(\xi_j) = 0$, $E(\xi_j^2) = 1$, $(\sigma_j)_{1\le j\le n}$ – неизвестные коэффициенты волатильности, которые, зависят от x_j и функции S, n – число наблюдений.

Модели типа (1) являются обобщением непараметрической модели ANCOVA и встречаются в эконометрических исследованиях [1] при анализе инвестиционного поведения фирм. Отметим, что модель (1) является важной в задачах аппроксимации диффузионных процессов с непрерывным временем, путем использования последовательных ядерных оценок, имеющих асимптотически минимальные отклонения (см., например, [2, 3]).

*Работа выполнена при финансовой поддержке РФФИ (грант № 16-01-00121 А) и Минобрнауки (госзадание № 2.3208.2017/4.6).

Для оценивания функции S в модели (1) была построена в работах [10, 11] адаптивная процедура выбора модели, которая для конечного объема наблюдений имеет более высокую среднеквадратическую точность по сравнению с оценками МНК, а также получено оракульное неравенство для среднеквадратического риска построенной процедуры выбора модели. Повышение неасимптотической точности оценивания в регрессионных моделях стало возможно благодаря работам [6, 7].

Цель данной работы – установить, что предложенная адаптивная процедура оценивания является асимптотически эффективной в смысле среднеквадратического риска. Для этого доказывается, что асимптотический среднеквадратический риск процедуры совпадает с соответствующей константой Пинскера, обеспечивающей точную нижнюю границу риска по всевозможным оценкам.

Как правило, понятие асимптотической оптимальности связано с оптимальной скоростью сходимости минимаксного риска (см., например, [8]). Важным вопросом в результатах оптимальности является изучение точной асимптотики минимаксного риска. Подобные асимптотические результаты были получены только в ряде случаев. Проблемы непараметрической оценки для моделей регрессии были изучены в [9]. Для изучения асимптотических свойств процедуры оценивания в данной работе применяется подход на основе оракульных неравенств, развитый в работах [4, 5]. Хорошо известно, что для доказательства асимптотической эффективности процедуры оценивания нужно показать, что ее асимптотический риск совпадает с нижней границей, определяемой константой Пинскера. В работе решаются две задачи: во-первых, получена верхняя граница для риска путем использования неасимптотического оракульного неравенства, во-вторых, доказано, что эта верхняя граница совпадает с постоянной Пинскера.

Статья состоит из следующих разделов. В разделе 2 предлагаются улучшенные взвешенные оценки наименьших квадратов. В разделе 3 строится адаптивная процедура выбора моделей для оценивания функции, основанная на взвешенных улучшенных оценках наименьших квадратов. Получено точное неасимптотическое оракульное неравенство для риска процедуры. В разделе 4 представлена теорема об асимптотической эффективности процедуры выбора модели в минимаксном смысле. Раздел 5 содержит результаты численного сравнения данной адаптивной улучшенной процедуры с процедурой, предложенной в [4].

2. Улучшенные взвешенные оценки МНК. Для оценивания неизвестной функции S в модели (1) воспользуемся ее разложением в ряд Фурье по тригонометрическому базису $(\phi_j)_{j\geq 1}$ в пространстве $\mathcal{L}_2[0,\,1]$

$$S(x) = \sum_{j\geq 1} \theta_j \phi_j(x)\,, \tag{2}$$

где коэффициенты Фурье

$$\theta_j = (S, \phi_j)_n = \frac{1}{n}\sum_{l=1}^{n} S(x_l)\phi_j(x_l)$$

– эмпирическое скалярное произведение функций S и ϕ_j на сетке $(x_j)_{1\leq j\leq n}$. Поскольку они зависят от неизвестной функции S, то также являются неизвестными и подлежат оцениванию. Заметим, что, если n нечетно, то $(\phi_j)_{j\geq 1}$ есть ортонормированный базис относительно введенного эмпирического скалярного произведения, т.е. для $1 \leq j \leq n$

$$(\phi_i, \phi_j)_n = \frac{1}{n}\sum_{l=1}^{n} \phi_l(x_l)\phi_j(x_l) = \delta_{ij},$$

где δ_{ij} – символы Кронекера. Используя этот факт, получаем, что оценки МНК коэффициентов Фурье равны

$$\widehat{\theta}_{j,n} = (Y, \phi_j)_n,$$

где $Y = (y_1, ..., y_n)'$ (штрих обозначает транспонирование). Из (1) следует, что оценки МНК удовлетворяют следующему равенству

$$\widehat{\theta}_{j,n} = \theta_j + \frac{1}{n}\xi_{j,n}, \quad \xi_{j,n} = \frac{1}{\sqrt{n}}\sum_{l=1}^{n} \sigma_l \xi_l \phi_j(x_l). \tag{3}$$

Теперь, как и в [4], определим класс взвешенных оценок МНК функции S следующим образом

$$\widehat{S}_\lambda(x) = \sum_{j=1}^{n} \lambda(j)\widehat{\theta}_{j,n}\phi_j(x), \tag{4}$$

где $x \in [0, 1]$, вектор весовых коэффициентов $\lambda = (\lambda_1, ..., \lambda_n)$ принадлежит некоторому конечному множеству $\Lambda \subset [0,\,1]^n$, где $n \geq 3$. Пусть $\nu = |\Lambda|$. Предположим, что существует целое число $j_* = j_*(n) \leq n$ такое, что $\lambda_j = 1$ для $1 \leq j \leq j_*$. Учитывая такой вид весовых коэффициентов, вместо взвешенных оценок МНК предлагается использовать оценки вида

$$S^*_\lambda(x) = \sum_{j=1}^{j_*} \theta^*_{j,n}\phi_j(x) + \sum_{j=j_*+1}^{n} \lambda(j)\widehat{\theta}_{j,n}\phi_j(x), \tag{5}$$

где

$$\theta^*_{j,n} = \left(1 - \frac{c_n}{\| \tilde{\theta_n} \|}\right) \widehat{\theta}_{j,n}, \quad \|\tilde{\theta_n}\|^2 = \sum_{j=1}^{j_*} \widehat{\theta}^2_{j,n}.$$

Чтобы определить коэффициент c_n, предположим, что в модели (1) коэффициенты σ_j такие, что

$$\max_{1\le j\le n} \sigma_j^2 \le \sigma^*, \ \min_{1\le j\le n} \sigma_j^2 \ge \sigma_* \tag{6}$$

и $\|S\| \le r$, где σ_*, σ^* и r – некоторые известные величины. Тогда

$$c_n = \frac{(j_* - 1)\sigma_*^2}{n(r + \sqrt{j_* \sigma^{*2}/n})}.$$

Далее обозначим разность эмпирических среднеквадратических рисков предложенной оценки (5) и оценки МНК (4) как

$$\Delta_n(S) = \mathbf{E}_S \|S^*_\lambda - S\|_n^2 - \mathbf{E}_S \|\widehat{S}_\lambda - S\|_n^2, \tag{7}$$

где $\mathbf{E}_S$ – математическое ожидание относительно распределения наблюдений $Y = (y_1, ..., y_n)'$ при фиксированной функции S и

$$\|S\|_n^2 = \frac{1}{n} \sum_{l=1}^{n} S^2(x_l).$$

Предложенная оценка (5) обладает следующим свойством.

Теорема 1. *Пусть в модели* (1) *коэффициенты волатильности* $(\sigma_j)_{1\le j\le n}$ *удовлетворяют неравенствам* (6). *Тогда оценка* (5) *превосходит по среднеквадратической точности взвешенную оценку МНК* (4). *Кроме того, разность среднеквадратических рисков* (7) *удовлетворяет следующему неравенству:*

$$\Delta_n(S) \le -c_n^2\,. \tag{8}$$

3. Выбор модели. В [11] предложена процедура выбора модели S^* для оценивания функции S в модели (1) на основе улучшенных оценок (5):

$$S^* = S^*_{\lambda^*}\,, \tag{9}$$

где вектор весовых коэффициентов λ^* выбирается из условия минимизации ошибки оценивания и имеет вид, как в [4].

Далее получим неасимптотическую верхнюю границу для среднеквадратического риска процедуры выбора модели (9).

Теорема 2. *Пусть наблюдения описываются уравнением* (1). *Тогда для любого* $0 < \rho < 1/4$ *среднеквадратический риск, предложенной процедуры выбора модели* (9) *для функции* S *удовлетворяет следующему оракульному неравенству*

$$R(S^*, S) \le \frac{1+4\rho}{1-4\rho} \min_{\lambda\in\Lambda} R(S^*_\lambda, S) + \frac{2nc_n + nc_n^2 + 4\nu\sigma^*}{\rho n} \Psi_n(\rho). \tag{10}$$

4. Асимптотическая эффективность. Предположим, что неизвестная функция S в модели (1) принадлежит соболевскому шару

$$W_r^k = \{f \in \mathcal{C}_{per}^k[0,1]\,,\ \sum_{j=0}^{k} \|f^{(j)}\|^2 \le r\}\,,$$

где $r > 0\,,\ k \ge 1$ – некоторые параметры, $\mathcal{C}_{per}^k[0,1]$ множество k раз непрерывно-дифференцируемых функций $f\,:\,[0,1] \to \mathbf{R}$ таких, что $f^{(i)}(0) = f^{(i)}(1)$ для всех $0 \le i \le k$.

Справедлива следующая теорема

Теорема 3. *Для модели* (1) *робастный среднеквадратический риск процедуры выбора модели* S^* *удовлетворяет следующему асимптотическому равенству*

$$\lim_{n\to\infty} n^{\frac{2k}{2k+1}} \sup_{S\in W_r^k} \frac{R(S^*,S)}{\gamma_k(S)} = 1,$$

где $\gamma_k(S)$ *– константа Пинскера, определенная в [5].*

5. Численное моделирование. В этом разделе проиллюстрируем теоретически установленные результаты с помощью численного моделирования в среде Matlab. В качестве функции S выберем

$$S(x) = x^2\sin(4\pi x) + x^3\cos(4\pi x) + \cos(2\pi x),\ \ \sigma_j^2 = 2 + S(x_j).$$

Для вычисления весовых коэффициентов положим

$$k^* = 100 + \sqrt{\ln n},\quad \varepsilon = \frac{1}{\ln n},\quad m = \ln^2 n,$$

$$\rho = \frac{1}{3+\ln^2 n},\quad \omega_\alpha = 100 + (A_\beta t n)^{\dfrac{1}{2\beta+1}}\,.$$

В таблице 1 приведены результаты моделирования среднеквадратических рисков по $N = 1000$ реализациям процедур выбора модели (9), построенным на основе предложенных улучшенных оценок (5) и на основе оценок МНК (4):

$$\tilde{R}(S^*,S) = \frac{1}{N}\sum_{m=1}^{N} \|S_m^* - S\|_n^2,$$

где S_m^*– оценка полученная по m–ой реализации выборки.

Из таблицы видим, что эмпирический риск процедуры (9) меньше, чем для поцедуры выбора модели $\widehat{S}_{\lambda^*}$, построенной на основе оценок МНК (4).

Из таблицы 2 видно, что с ростом n нормированный риск стремится к 1, что численно подтверждает теорему 3.

Далее на рисунке представлены графики истинной функции S и ее оценок S^* и $\widehat{S}_{\lambda^*}$ при $n = 401$.

Таблица 1 — Эмпирические квадратические риски

n	5	10	101	401	1001
$R(S^*, S)$	8.5911	3.3576	0.2779	0.0596	0.0196
$R(\widehat{S}_{\lambda^*}, S)$	47.3385	10.0099	0.9276	0.1297	0.1011

Таблица 2 — Нормированные эмпирические квадратические риски

n	5	10	101	401	1001
$n^{\frac{2k}{2k+1}} \dfrac{R(S^*, S)}{\gamma_k(S)}$	58.9561	27.0694	8.1458	2.2210	1.1256

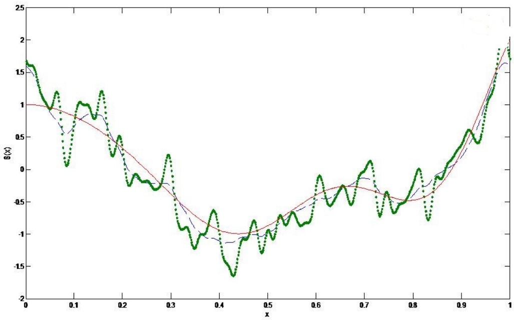

Рис. 1. Графики функции S, ее оценки МНК $\widehat{S}_\lambda$ и улучшенной оценки $\widehat{S}^*$ при n=401.

Здесь сплошная линия – график функции S, пунктирная линия – график оценки МНК $\widehat{S}^*_\lambda$ и штрих линия – график улучшенной оценки S^*. Из рисунка видно, что функция S^* приближает функцию S лучше, чем $\widehat{S}^*_\lambda$, что также подтверждает теорему 1.

Литература

1. Cai T., Wang L. Adaptive variance function estimation in heteroscedactic nonparametric regression // Annals of Statistics. 2008. V. 36. No. 5. P. 2025-2054.

2. Galtchouk L., Pergamenshchikov S. Nonparametric sequential estimation of the drift in diffusion processes // Mathematical Methods of Statistics. 2004. V. 13. No. 1. P. 25-49.

3. Galtchouk L., Pergamenshchikov S. Adaptive sequential estimation for ergodic diffusion processes in quadratic metric // Journal of Nonparametric Statistics. 2011. V. 23. No. 2. P. 255-285.

4. Galtchouk L., Pergamenshchikov S. Sharp non-asymptotic oracle inequalities for nonparametric heteroscedastic regression models // Journal of Nonparametric Statistics. 2009. V. 21. No. 1. P. 1-16.

5. Galtchouk L., Pergamenshchikov S. Adaptive asymptotically efficient estimation in heteroscedastic nonparametric regression // Journal of the Korean Statistical Society. 2009. V. 38. No. 4. P. 305-322.

6. Pchelintsev E. Improved estimation in a non-Gaussian parametric regression // Statistical Inference for Stochastic Processes. 2013. V. 16. No. 1. P. 15-28.

7. Konev V., Pergamenshchikov S. and Pchelintsev E. Estimation of a regression with the pulse type noise from discrete data // Theory Probab. Appl. 2014. V. 58. No. 3. P. 442–457.

8. Ibragimov I.A., Khasminskii R.Z. Statistical Estimation: Asymptotic Theory. Springer, Berlin, New York, 1981.

9. Pinsker M.S. Optimal filtration of square integrable signals in gaussian white noise // Problems of Transimission information. 1981. V.17. P. 120–133.

10. Перелевский С.С. Адаптивное улучшенное оценивание функции гетероскедастичной регрессии в непрерывном времени / Перелевский С.С., Пчелинцев Е.А. // Обозрение прикладной и промышленной математики. – 2015 – Т.23. – № 3. – С. 491-492.

11. Перелевский С.С. Улучшенная процедура выбора модели для оценивания функции регрессии по дискретным данным / Перелевский С.С., Пчелинцев Е.А. // Обозрение прикладной и промышленной математики. – 2016 – Т.23. – № 4. – С. 367-377.

Perelevskiy S.S., Pchelintsev E.A. (Tomsk State University, Tomsk, 2017) **Asymptotically efficient estimation of the heteroscedastic regression function.**

Abstract. In the paper, we consider some asymptotic properties of the adaptive procedure proposed in the article by the authors [10] for estimating the unknown function of heteroscedastic regression. It is established that the estimation procedure is asymptotically efficient in the sense of mean-square risk, i.e. it is proved that the asymptotic

mean square risk of the procedure coincides with the corresponding Pinsker constant, which provides the exact lower bound of the risk by all possible estimates.

Key words: heteroscedastic regression, adaptive model selection procedure, asymptotic quadratic risk, Pinsker constant.

Оценивание параметров в модели регрессии с нелинейными шумами [*]

Повзун М. А., Пчелинцев Е. А.

Томский государственный университет, Томск
e-mail: povzunyasha@gmail.com

Аннотация

В работе рассматривается задача минимаксного оценивания d–мерного вектора неизвестных параметров регрессии с нелинейными условно–гауссовскими шумами типа AR/GARCH. Предлагается новая оценка, превосходящая по среднеквадратической точности оценку по методу наименьших квадратов. Приводятся результаты численного сравнения эмпирических рисков предлагаемой улучшенной оценки и оценки МНК.

Ключевые слова: регрессия, улучшенное оценивание, среднеквадратический риск, условно–гауссовский шум, процесс типа AR/GARCH .

Постановка задачи. Рассмотрим задачу оценивания d-мерного вектора регрессии с нелинейным условно-гауссовским шумом. Пусть на вероятностном пространстве $(\Omega, \mathcal{F}, \mathbf{P})$ наблюдения описываются уравнением

$$Y = \theta + \nu\xi, \tag{1}$$

где $\theta \in \Theta \subset \mathbb{R}^d$– вектор неизвестных параметров, ν – известное положительное число, ξ – первые d значений процесса типа $AR(p)/GARCH(q_1, q_2)$:

$$\xi_t = \gamma_0 + \sum_{i=1}^{p} \gamma_i \xi_{t-j} + \sqrt{\alpha_0 + \sum_{j=1}^{q_1} \alpha_j \xi_{t-j}^2 + \sum_{k=1}^{q_2} \beta_k \sigma_{t-k}^2}\varepsilon_t, \tag{2}$$

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{q_1} \alpha_j \xi_{t-j}^2 + \sum_{k=1}^{q_2} \beta_k \sigma_{t-k}^2,$$

где $(\alpha_j)_{0 \leqslant j \leqslant q_1}$, $(\beta_k)_{1 \leqslant k \leqslant q_1}$ – положительные коэффициенты, $(\varepsilon_t)_{t \geqslant 0}$ – белый шум. Предположим, что ξ имеет условно-гауссовское распределение относительно некоторой σ-алгебры $\mathcal{G}$ с нулевым средним и условной ковариационной матрицей $D(\mathcal{G})$, у которой

$$trD(\mathcal{G}) - \lambda_{max}(D(\mathcal{G})) \geqslant \kappa(d) > 0,$$

[*]Работа выполнена при финансовой поддержке Министерства образования и науки РФ, Госзадание No 2.3208.2017/4.6.

а математическое ожидание максимального собственного значения
$$\mathbf{E}\lambda_{max}(D(\mathcal{G})) \leq \lambda^*.$$
Требуется оценить вектор неизвестных параметров θ по наблюдениям Y.

Основной результат. Известно, что в классе линейных несмещенных оценок лучшей является оценка по методу максимального правдоподобия
$$\widehat{\theta}_{ML} = Y.$$
На практике часто такая оценка не может справится с шумами, которые импульсно влияют на данные. В работах [1–4,7] предлагается модификация этой оценки для дискретных и непрерывных моделей. Согласно этому подходу, в данной работе предлагается для оценивания вектора параметров θ использовать следующую сжимающую процедуру
$$\theta^* = \left(1 - \frac{c}{|Y|}\right) Y, \tag{3}$$
где $c = \nu^2 \lambda_*(d-1)\delta_d$, $\delta_d = \left(\rho + (2\lambda^*)^{-1/2}\nu \frac{\Gamma\left(\frac{d+1}{2}\right)}{\Gamma\left(\frac{d}{2}\right)}\right)^{-1}$, $\rho = \sup_{\theta\in\Theta}\{|\theta|\}$.

В настоящей работе в качестве меры точности оценки выберем среднеквадратический риск
$$R(\theta, \widehat{\theta}) = \mathbf{E}_\theta |\theta - \widehat{\theta}|^2, \quad |x|^2 = \sum_{j=1}^{d} x_j^2.$$
Предлагаемая оценка (3) обладает следующим свойством.

Теорема 1. *Существует $d_0 \geqslant 2$ такое, что для всех $d \geqslant d_0$ оценка* (3) *превосходит по среднеквадратической точности оценку МНК $\widehat{\theta} = Y$. Более того, разность рисков*
$$\Delta_\theta = R(\theta, \theta^*) - R(\theta - \widehat{\theta}) \leqslant -c^2.$$

Замечание 1. *Теорема 1 утверждает, что предложенная оценка* (3) *является улучшенной в понимании среднеквадратической точности в сравнении с оценкой максимального правдоподобия. При этом минимальный выигрыш в точности равен $-c^2$.*

Численный анализ эмпирических рисков. Для численного подтверждения теоремы 1 проведено имитационное моделирование в среде MATLAB. Предполагаем, что θ – единичный вектор размерности d, ξ_t – значения процесса $AR(1)/ARCH(1)$ $(AR(1)/GARCH(1,0))$, с коэффициентами $\beta_0 = \alpha_0 = 0, \nu = 1$, а

$$\mathbf{E}_\theta(\xi_0) = 0, \mathbf{D}_\theta(\xi_0) = s^2,$$
$$\mathbf{E}_\theta(\varepsilon_t) = 0, \mathbf{D}_\theta(\varepsilon_t) = \sigma^2.$$

При этом шум будет описываться следующим уравнением:

$$\xi_t = \beta_1 \xi_{t-1} + \sqrt{\alpha_1 \xi_{t-1}^2} \varepsilon_t. \quad (4)$$

Остальные параметры меняются. Далее в таблицах приводятся результаты моделирования при изменении эмпирических рисков оценок параметров шума

Таблица 1 — Поведение эмпирических среднеквадратических рисков оценок при изменении параметра α_1 с $\beta_1 = 0,1; \sigma^2 = 1; s^2 = 1; d = 5$

α_1	$R(\widehat{\theta}, \theta)$	$R(\theta^*, \theta)$	Δ_θ	$-c^2$
0.1	0.6514	0.6383	-0.0131	-0.0013
0.2	2.3788	2.2696	-0.1092	-0.0149
0.4	36.5023	34.4281	-2.0741	-0.3936
0.8	502.7098	478.9157	-23.7941	-2.9729

Таблица 2 — Поведение эмпирических среднеквадратических рисков оценок при изменении параметра β_1 с $\alpha_1 = 0,1; \sigma^2 = 1; s^2 = 1; d = 5$

β_1	$R(\widehat{\theta}, \theta)$	$R(\theta^*, \theta)$	Δ_θ	$-c^2$
0.1	0.6900	0.6800	-0.0100	-0.0013
0.2	0.6521	0.6316	-0.0206	-0.0042
0.4	1.0896	1.0357	-0.0540	-0.0195
0.8	5.4522	5.1229	-0.3293	-0.0470

Таблица 3 — Поведение эмпирических среднеквадратических рисков оценок при изменении параметра s^2 с $\alpha_1 = 0,1; \beta_1 = 0,1; \sigma^2 = 1; d = 5$

s^2	$R(\widehat{\theta}, \theta)$	$R(\theta^*, \theta)$	Δ_θ	$-c^2$
1	0.6582	0.6459	-0.0124	-0.0013
2	2.4482	2.3790	-0.0691	-0.0040
4	10.8724	10.5511	-0.3213	-0.0116
8	40.6176	39.4273	-1.1902	-0.0315
16	143.0420	139.0146	-4.0275	-0.0810

Таблица 4 — Поведение эмпирических среднеквадратических рисков оценок при изменении параметра σ^2 с $\alpha_1 = 0,1; \beta_1 = 0,1; s^2 = 1; d = 5$

σ^2	$R(\widehat{\theta}, \theta)$	$R(\theta^*, \theta)$	Δ_θ	$-c^2$
2	0.7752	0.7611	-0.0140	-0.0013
4	27.7204	27.2507	-0.4697	-0.0149
8	5.3665e+04	5.3611e+04	-53.6798	-0.3936
16	2.3232e+07	2.3228e+07	-3.86e+03	-2.9729

Как видно из таблиц 1-4, при изменении параметров риск улучшенной оценки меньше, чем риск оценки МНК, что численно подтверждает основной результат.

Таблица 5 — Поведение эмпирических среднеквадратических рисков оценок при изменении параметра d с $\alpha_1 = 0,2; \beta_1 = 0,7; s^2 = 1; \sigma^2 = 1,5$

d	$R(\widehat{\theta}, \theta)$	$R(\theta^*, \theta)$	Δ_θ	$-c^2$
2	7.3268	6.6627	-0.6641	-0.0657
3	10.7353	9.4675	-1.2678	-0.2025
4	15.7660	13.8821	-1.8839	-0.3863
5	20.2577	17.7924	-2.4653	-0.6079
6	24.7658	21.8361	-2.9297	-0.8609
10	25.5038	21.6807	-3.8231	-2.0996

При увеличении размерности d в таблице 5 видно, что улучшенная оценка на порядок точнее, а разность рисков Δ_θ возрастает.

Замечание 2. *При малых значениях дисперсий σ^2, s^2 шум ξ быстро затухает с ростом размерности d, а когда дисперсии увеличиваются, процесс выходит из области стационарности.*

В заключении отметим, что в случаях, когда шумы оказывают сильное влияние на данные, предлагаемая оценка (4) лучше с ними справляется, чем оценка МНК. Полученные результаты рекомендуется применять при статистической идентификации моделей, шум которых описывается резко меняющимися (скачкообразными) временными рядами.

Литература

1. Конев В.В. Оценивание параметрической регрессии с импульсными шумами по дискретным наблюдениям / В.В.Конев, Е.А. Пчелинцев // Вестн Том. гос. ун-та. Математика и механика. 2012. № 1(17). С. 20-35.

2. Конев В.В. Оценивание параметрической регрессии с импульсными шумами по дискретным наблюдениям / В.В.Конев, С.М. Пергаменщиков, Е.А. Пчелинцев// Теория вероятностей и ее применения.- М. :Математический институт им. В.А. Стеклова Российской Академии Наук, 2013. № 3. С. 454-471.

3. Пчелинцев Е. А. Процедура Джеймса – Стейна для условно-гауссовской регрессии // Вестн Том. гос. ун-та. Математика и механика. 2011. № 4(16). С. 6–17.

4. Пчелинцев Е. А. Минимаксное оценивание гауссовской параметрической регрессии / Е.А. Пчелинцев, В.А. Пчелинцев // Вестн Том. гос. ун-та. Математика и механика. 2014. № 5(31). С. 40–47.

5. Ширяев А.Н. Основы стахостической финансовой математики. Том 1. Факты. Модели. – М. : ФАЗИС, 1998. – 512 с.

6. James W., Stein C. Estimation with quadratic loss // in: Proceedings of the Fourth Berkeley Symposium on Mathematics Statistics and Probability, University of California Press, Berkeley. 1961. V. 1. P. 361-380.

7. Pchelintsev E. Improved estimation in a non-Gaussian parametric regression // Statistical Inference for Stochastic Processes. 2013. V. 1. P. 16-28.

Povzun M. A., Pchelintsev E. A. (Tomsk State University, Tomsk, 2017) **Estimation of the parameters in regression with nonlinear noises.**

Abstract. This paper considers the problem of estimating the d-dimensional vector of unknown parameters in regression with nonlinear conditionally Gaussian noise of the AR/GARCH type. We propose the new estimate which outperforms the LSE in mean-square accuracy. The results of numerical modeling of the empirical risks of the proposed improved estimate and LSE are given.

Key words: regression, improved estimation, mean square risk, conditionally Gaussian noise, AR/GARCH process.

Исследование качества моделей временных рядов с применением бутстрап-методов *

Пчелинцев Е. А., Филимонова Ю. О.

НИ ТГУ, Томск
e-mail: ylya2911@mail.ru

Аннотация

В работе предлагается алгоритм бутстрапирования данных для статистического анализа качества нелинейных моделей временных рядов типа AR/GARCH. Устанавливается, что предлагаемый бутстрап метод позволяет повысить достоверность статистических выводов для AR/GARCH моделей.

Ключевые слова: временные ряды, авторегрессионные модели, AR/GARCH процесс, бутстрап-метод.

Введение. Часто на практике объем данных весьма ограничен, чтобы можно было делать достоверные выводы о них. Поэтому в последнее время получили развитие новые статистические методы под общим названием ресемплинг. Методы ресемплинга включают три основных подхода, отличающихся по технике, но близкие по сути: метод "складного ножа", метод "рандомизации" и наиболее популярный - "бутстрап-метод" [1]. Стремительные перемены современной жизни, связанные с большими достижениями информационных технологий и вычислительной техники, дают нам возможность быстрого и эффективного анализа огромных массивов данных. В [2] дано объяснение развитию новой группы альтернативных компьютерно-интенсивных (computer-intensive) методов, включающих в себя рандомизацию и бутстрап. Эти технологии выполняют многократный анализ разных фрагментов исходного массива эмпирических данных, как бы рассматривая их под различными углами зрения и сравнивая полученные таким образом результаты, им не нужна никакая априорной информации о законе распределения исследуемой случайной величины, в отличие от подходов в [3].

Цель данной работы заключается в построении модели временного ряда типа AR/GARCH и разработке алгоритма бутстрапирования для анализа ее качества.

---

*Работа выполнена при финансовой поддержке РНФ, проект No 17-11-01049.

Постановка задачи и основной результат. Моделирование и прогнозирование изменчивости различных показателей, например акций, курсов валют на финансовых рынках в наше время является объектом последних исследований и теоретических работ. Все в мире меняется, цены не исключение, и предсказать их поведение становится все сложнее.Такие традиционные модели временных рядов, как ARMA, не всегда могут справедливо учитывать характеристики, которыми обладают финансовые временные ряды. Соответственно, требуется расширение таких моделей. Одна из особенностей финансовых рынков состоит в том, что неопределенность присущая рынку, изменяется во времени. Из этого наблюдается «кластеризация волатильности». Термин «волатильность» используется, для неформального обозначения разброса переменной. Формальной мерой волатильности служит дисперсия. Эффект кластеризации волатильности отмечен для таких рядов, как изменение цен акций, валютных курсов и др. Изменения в дисперсии имеют весьма важное значение для понимания финансовых рынков, так как инвесторы требуют более высокую ожидаемую доходность в качестве компенсации рисковых активов. Официальным названием для изменчивости дисперсии на различных интервалах наблюдения является гетероскедастичность. Для более адекватного описания таких ситуаций были предложены авторегрессионные модели с условной неоднородностью и их обобщения [4, 5].

Остановимся на рассмотрении AR/GARCH моделей. Пусть на вероятностном пространстве (Ω, F, P) наблюдения описываются AR/GARCH процессом, т.е. уравнениями

$$X_t = \theta + \sum\nolimits_{i=1}^{p} \alpha_i X_{t-i} + \varepsilon_t, \tag{1}$$

$$\varepsilon_t = \sigma_t v_t, \tag{2}$$

$$\sigma_t^2 = \beta_0 + \sum\nolimits_{j=1}^{p} \beta_j \varepsilon_{t-j}^2 + \sum\nolimits_{k=1}^{q} \gamma_k \sigma_{t-k}^2. \tag{3}$$

Первой основной задачей является оценивание неизвестных параметров модели $\theta, \alpha_i, \beta_j, \gamma_k$. Качество оценивания напрямую зависит от того, как долго мы наблюдаем процесс $(X_t)_{0<t\leq T}$. Некоторые методы статистической идентификации моделей, обладающие робастностью и достаточно высокой точностью предлагаются в работах [6–10]. Затем после качественной идентификации модели возникает необходимость исследования качества уравнения в целом и качества прогноза. И поскольку для этого требуются достаточно большие объемы данных, которые часто не доступны в практиче-

ских задачах, то актуальной является задача разработки бутстрап-методов для таких моделей [1].

Далее рассмотрим задачу построения математической модели динамики цен на нефть марки WTI (данные взяты с сайта finam.ru) и исследования ее качества. Подбирается AR(p)/GARCH(q_1,q_2) - модель. Используя автокорреляционные и частные автокорреляционные функции, с помощью визуального анализа полученных графиков были оценены параметры p, q_1, q_2. Выбранная модель динамики цен является модель AR(1)/GARCH(1,1). Оценивание неизвестных коэффициентов проводилось по базовому методу регрессионного анализа - методу наименьших квадратов (МНК). Таким образом, запишем уравнения процесса в явном виде

$$\sigma_t^2 = -1.9403 + 0.5589\varepsilon_{t-1}^2 + 4.6887\sigma_{t-1}^2;$$

$$\hat{X}_t = 29.7071 + 0.2218X_{t-i} + \varepsilon_t\,, \quad \varepsilon_t = \sigma_t v_t.$$

Проведем статистический анализ полученной модели [5]. Возникает задача проверки гипотез относительно значимости коэффициентов данного процесса и его адекватности в целом. Рассмотрим t-критерий, основанный на распределении Стьюдента, с его помощью можно проверить, не отличается ли оценка данного коэффициента регрессии b_i от нуля только за счет случайных возмущений.

Сформулируем гипотезу $H_0 : b_i = 0$ и альтернативу $H_1 : b_i \neq 0$.

Зададим уровень значимости $\alpha = 0.95$, на котором в дальнейшем и будет сделан вывод о справедливости гипотезы.

Вычисляется t- статистика и строятся доверительные интервалы для исследуемого параметра:

$$b_i - \frac{b_i * t}{\sqrt{n}} < b_i < b_i + \frac{b_i * t}{\sqrt{n}}.$$

Таким образом, для параметра θ и α доверительная область:

$$26.652 < \theta < 32.286$$

$$0.1652 < \alpha < 0.2824$$

следовательно, коэффициенты являются значимыми. С помощью критерия Фишера проверяется адекватность данной модели в целом: H_0: уравнение регрессии статистически незначимо, а соответствующая ей альтернатива H_1: уравнение регрессии статистически значимо. Зададим уровень значимости $\alpha = 0.05$, на котором в дальнейшем и будет сделан вывод о справедливости гипотезы. Вычисляется F-статистика, а также квантиль распределения Фишера при заданном уровне значимости. Эти величины сравниваются и если $F > F_T$, в нашем случае $F = 8.3902e + 04$ и $F_T = 0.0039$, следовательно, модель статистически не значима. Далее необходимо убедиться в том, насколько достоверны статистические выводы, полученные для исходной модели. Для решения этой задачи в работе предлагается новая процедура бутстрапирования данных.

Бутстрап-метод и результаты его численной реализации. На основе изученных ранее методов бутстрапирования предлагается следующий модифицированный пошаговый алгоритм для AR/GARCH модели:

- Оцениваем параметры части GARCH (3) по методу наименьших квадратов;
- Организуем процедуру бутстрапирования, набирая остатки v_t^{*2} из стандартного нормального распределения;
- Пересчитываем по формуле (3) σ_t^{*2};
- Пересчитываем остатки по формуле (2): $\varepsilon_t^{*2} = \sigma_t^{*2} * v_t^{*2}$;
- Оцениваем параметры части AR(p) по методу наименьших квадратов:
$$\hat{X}_t = \hat{\theta} + \sum\nolimits_{i=1}^{p} \hat{\alpha}_i X_{t-i} + \varepsilon_t;$$
- Строим бутстрапированный ряд
$$X_t^* = \hat{\theta} + \sum\nolimits_{i=1}^{p} \hat{\alpha}_i X_{t-i}^* + \varepsilon_t^*.$$

Применим этот алгоритм для исследования достоверности статитстических выводов. Исходный объем выборки $n = 262$. Таким образом, сначала получили новую выборку объемом равный исходным данным, во втором эксперименте она была увеличена практически в 2 раза до $n = 500$, затем объем модифицированной выборки получили $n = 1000$ и 5000. Ключевым вопросом при анализе данных является объем наблюдений. Очевидно, что чем больше данных имеет исследователь, тем точнее выводы он может сделать. Для того, чтобы проиллюстрировать эффективность новой процедуры бутстрапирования, описанной выше, покажем некоторые результаты экспериментов моделирования:

Таким образом, независимо от объема данных общая картина статистического анализа исходной модели сохраняется. Следовательно, полученные выводы о качестве модели можно считать достаточно обоснованными. Полученные результаты могут применяться при анализе данных в эконометрике, социологии, нефтедобывающей отрасли и других областях, где объемы данных имеют ограниченные размеры.

Таблица 1 — Доверительная область для параметра θ и F-статистика.

n	t-критерий	F-критерий
262	$24.6467<\theta<34.7569$	F=8.5341e+10
500	$26.0517<\theta<33.3581$	F=29.7490
1000	$27.1276<\theta<32.2840$	F=27.4738
5000	$28.5459<\theta<30.8556$	F=4.3870e+03

Таблица 2 — Доверительная область для параметра θ и F-статистика.

n	t-критерий	F-критерий
262	$0.1841<\alpha<02595$	F_T=0.0039
500	$0.1948<\alpha<0.2488$	F_T=0.0039
1000	$0.2025<\alpha<0.2415$	F_T=0.0039
5000	$0.2132<\alpha<0.2304$	F_T=0.0039

Литература

1. Jochen V. A., Holditch S. A., Spivey J. P. Probabilistic Reserves Estimation Using Decline Curve Analysis with the Bootstrap Method // Society of Petroleum Engineers. 1996. P. 1-8.

2. Efron B. Bootstrap confidence intervals for a class of parametric problems // Biometrica. 1985.

3. Lahiri S.N. Theoretical comparisons of block bootstrap methods // The Annals of Statistics. 1999. V. 27. P. 386-404.

4. Mousazadeh S., Cohen I. AR-GARCH in presence of noise: parameter estimation and its application to voice activity detection. 2011. P. 1-11.

5. Айвазян С.А. Основы эконометрики / Айвазян С.А. – М.: «Юнити», 2001. – Т.2.

6. Конев В.В. Оценивание параметрической регрессии с импульсными шумами по дискретным наблюдениям / В.В.Конев, Е.А. Пчелинцев // Вестн Том. гос. ун-та. Математика и механика. 2012. № 1(17). С. 20-35.

7. Конев В.В. Оценивание параметрической регрессии с импульсными шумами по дискретным наблюдениям / В.В.Конев,

С.М. Пергаменщиков, Е.А. Пчелинцев// Теория вероятностей и ее применения. 2013. № 3. С. 454-471.

8. Пчелинцев Е. А. Процедура Джеймса – Стейна для условно-гауссовской регрессии // Вестн Том. гос. ун-та. Математика и механика. 2011. № 4(16). С. 6–17.

9. Пчелинцев Е. А. Минимаксное оценивание гауссовской параметрической регрессии / Е.А. Пчелинцев, В.А. Пчелинцев // Вестн Том. гос. ун-та. Математика и механика. 2014. № 5(31). С. 40–47.

10. Pchelintsev E. Improved estimation in a non-Gaussian parametric regression // Statistical Inference for Stochastic Processes. 2013. V. 1. P. 16-28.

Pchelintsev E. A., Filimonova Yu. O.(Tomsk State University, Tomsk, 2017) **The study of the quality of time series models using bootstrap methods.**

Abstract. In the paper, a data bootstrapping algorithm is proposed for statistical analysis of the quality of nonlinear time series models of the AR/GARCH type. It established that the proposed bootstrap method can improve the accuracy of statistical inference for AR/GARCH models.

Key words: time series, autoregressive model, AR/GARCH process, bootstrap method.

Научное издание

Международная научная конференция «Робастная статистика и финансовая математика – 2017» Томск, 03–05 июля 2017 г.

Сборник статей

Под редакцией
д-ра физ.-мат. наук, профессора С. М. Пергаменщикова,
канд. физ.-мат. наук, доцента Е.А. Пчелинцева

ISBN 978-5-94621-639-5