 Robust signal processing in nonparametric autoregression

Evgeny Pchelintsev and Maria Povzun
Tomsk State University, Russia, Tomsk, Lenin str., 36, 634050
E-mail: evgen-pch@yandex.ru; povzunyasha@gmail.com

Abstract. In this paper we consider the problem of a robust adaptive estimation of a periodic signal modeled by the nonparametric autoregression. We develop a new sequential model selection method, using improved estimation approach and the efficient sequential kernel estimators. This procedure is based on the sequential estimators. For robust quadratic risk of proposed estimate we obtain sharp oracle inequality that allows us to establish the efficiency property of this model selection procedure. We give the Monte Carlo simulation results for numerical comparing of the risks of proposed improved procedure and ordinary least squares estimates.

1. Introduction
We consider the problem of statistical signals processing by observations \((y_k)_{k \geq 1}\) satisfies the equation of nonparametric autoregression

\[
y_k = S(t_k)y_{k-1} + \varepsilon_k, \quad t_k = \frac{k}{n}, \quad 1 \leq k \leq n, \tag{1}
\]

where \(S(\cdot)\) is an unknown signal on \([0, 1]\), \((\varepsilon_k)_{k \geq 1}\) is a sequence of unobservable identically distributed random variables with zero mean and unit variance, the initial value \(y_0\) is a fixed quantity.

Autoregressive processes are widely used for modeling of real time series, in particular in radio engineering [1, 2, 3]. Moreover, many processes in radiophysics, radiolocation, astronomy possess the characteristics repeating in time. In this case for mathematical modeling it is expedient to use the random processes with periodic structures [4]. In this paper we consider the varying nonparametric periodic coefficient autoregressive models (1). There are a lot of papers which consider such models and propose some asymptotic (as \(n \to \infty\)) methods for their statistical identification [5, 6].

In nonparametric statistical inferences, the very important problem is improving of the estimation quality for any finite numbers of observations \(n\) [7, 8]. Our goal is to construct a robust adaptive model selection procedure for estimating the unknown signal \(S\) from observations \((y_k)_{k \geq 1}\), which for any finite observations number \(n\) outperforms in mean square accuracy least squares estimates (LSE) proposed in [9]. The adaptive estimation problem arises due to the absence of information about the smoothness (regularity) of an unknown signal.

The quality of estimating the signal \(S\) will be measured by the quadratic risk

\[
R_p(\hat{S}, S) := E_{p,S}\|\hat{S} - S\|^2 \quad \text{and} \quad \|S\|^2 = \int_0^1 S^2(t)dt, \tag{2}
\]
where \( \hat{S} \) is some estimate (a measurable function of the observations), \( E_{p,S} \) is the expectation with respect to the distribution of observations given density \( p \) and the coefficient \( S \). In the case when the distribution density \( p \) is unknown we will use the robust risk

\[
R(\hat{S}, S) := \sup_{p \in \mathcal{P}} R_p(\hat{S}, S),
\]

where \( \mathcal{P} \) is a family of the distributions defined in Section 2 from [9].

2. Improved estimation method

For adaptive non-asymptotic estimation of an unknown function \( S \), we apply the approach from [10]. So, using the sequential method of truncated estimation, we pass from model (1) to its approximation by the regression model. At every point \( z_i = i/m, i = 1, m \) with \( m = \lceil \sqrt{n} \rceil \) (\([a]\) denotes integer part of \( a \)) on \([0, 1]\) we define a sequential plan \((\tau_i, \hat{S}_i)\) with stopping times

\[
\tau_i = \inf \{ l_i + 1 \leq k \leq k_i : A_{l_i,k} \geq H_i \}, \quad A_{l_i,k} = \sum_{j=l_i+1}^{k} Q_{i,j}y_{j-1}^2, \quad k_i = \min([nz_i + nh], n),
\]

and estimators

\[
\hat{S}_i = \frac{1}{H_i} \left( \sum_{j=l_i+1}^{\tau_i-1} Q_{i,j}y_{j-1}y_j + \kappa_i Q_{i,\tau_i-1}y_{\tau_i-1} \right) 1_{B_i}, \quad B_i = \{ A_{l_i,k_i-1} \geq H_i \}.
\]

Here \( Q_{i,j} = Q((x_j - z_i)/h), Q(\cdot) = 1_{[-1,1]}(\cdot) \) is an indicator function on \([-1,1]\), \( l_i = [nz_i - nh] + 1 + q, \ h = 1/(2m), \ q = \left( (nh)^\gamma \right) \) for some \( 0 < \gamma < 1 \), and \( 0 < \kappa_i < 1 \) is a correction constant and define as \( \kappa_i^2 = (H_i - A_{l_i,k_i-1})/(Q_{i,\tau_i-1}y_{\tau_i-1}^2) \). It should be noted that \( \hat{S}_i \) is consistent estimators for \( S(z_i) \) if we define the upper thresholds \( H_i \) as in [9].

Then on the set \( B_n = \cap_{i=1}^n B_i \) (note that \( \lim_{n \to \infty} n^a P_{p,S}(\mathcal{B}_n) = 0 \) for any \( a > 0 \)), we come to the heteroscedastic regression equation

\[
Y_i = S(z_i) + \xi_i, \quad 1 \leq i \leq m,
\]

where the observations \( Y_i = \hat{S}_i(z_i)1_{B_n} \) and the noise \( \xi_i \) is sum of the stochastic term \( \eta_i \) and approximation error \( \varepsilon_i \) which are defined in [9].

Now we estimate the signal \( S \) in equation (4). For this we use its Fourier expansion in a trigonometric basis \((\phi_j)_{j \geq 1} \in L_2[0,1]\)

\[
S(x) = \sum_{j=1}^{m} \theta_j \phi_j(x), \quad \theta_j = (S, \phi_j)_m = \frac{1}{m} \sum_{i=1}^{m} S(x_i) \phi_j(x_i).
\]

For estimating the function \( S \), it is necessary to estimate the Fourier coefficients \( \theta_j \). Instead of the ordinary LSE, we define the shrinkage estimates in the form

\[
\theta_{j,n}^* = (1 - g(j))\hat{\theta}_{j,n}.
\]

Here

\[
\hat{\theta}_{j,n} = \frac{1}{m} \sum_{i=1}^{m} Y_i \phi_j(x_i), \quad g(j) = \frac{c_n}{||\theta_{n,d}||} 1_{|1 \leq j \leq d}, \quad c_n = \frac{(d - 1)\sigma_1}{n(\sqrt{d}\sigma_2/n)},
\]

Note that \( || \cdot || \) denotes the Euclidian norm in \( \mathbb{R}^d \), \( d = 2 + [\ln n], 2 \leq d \leq m \), \( r_n \) is a positive parameter such that

\[
\lim_{n \to \infty} r_n = \infty \quad \text{and} \quad \lim_{n \to \infty} \frac{r_n}{n^a} = 0.
\]
for any $a > 0$, and

$$
\sigma_1 = \frac{3 + \ln n}{2nh} \quad \text{and} \quad \sigma_2 = \frac{2 + \ln n}{(1 + \ln n)(2nh - q - 3)}.
$$

We define the class of shrinkage weighted least squares estimates of the signal $S$ for all $0 \leq t \leq 1$ in the following form:

$$
S^*_\lambda(t) = \sum_{i=1}^{m} S^*_\lambda(z_i)1\{t\leq z_i\}, \quad S^*_\lambda(z_i) = \sum_{j=1}^{m} \lambda_j \theta^*_j \phi_j(x_i)1_{g_n},
$$

(5)

where $\lambda = (1, ..., 1, \lambda_{d+1}, ..., \lambda_m)$ is the weights vector.

To determine the set of weights we define a grid $A_n$

$$
A_n = \{1, ..., \beta^*\} \times \{\rho, ..., b\},
$$

where $b = [1/\rho^2]$, $\beta^* \geq 1$ and $\rho$ are functions of $n$, such that

$$
\beta^*(n) = \beta_0 + \sqrt{\ln(n + 1)} \quad \text{and} \quad \rho(n) = 1/\ln(n + 1)
$$

with $\beta_0 \geq 0$. For each $(\beta, r) \in A_n$, we introduce the weight coefficients $\lambda = (\lambda_j)_{j \geq 1}$

$$
\lambda_j = 1_{\{1 \leq j \leq d\}} + (1 - (j/\omega)^\beta)1_{\{d < j \leq \omega\}},
$$

(6)

where $\omega = \left(\frac{(\beta + 1)(2\beta + 1)n}{\pi^2 j^2}\right)^{1/(2\beta + 1)}$.

We denote the difference of quadratic risks of the estimates (5) and (4.9) from [9] as

$$
\Delta_p(S) = R_p(S^*_\lambda, S) - R_p(\hat{S}_\lambda, S).
$$

Theorem 2.1 Let the observations be described by the equation (1), then for any $n \geq 3$ and $\lambda = (\lambda_j)_{j \geq 1}$ uniformly over all distributions $p \in \mathcal{P}$ the signal estimate (5) outperforms in mean square accuracy the weighted LSE, i.e.

$$
\sup_{p \in \mathcal{P}} \sup_{\|S\| \leq r} \Delta_p(S) \leq -c_n^2.
$$

Note that this theorem implies that the constructed shrinkage estimates of unknown signal $S$ are improved in quadratic risks sense.

3. Model selection procedure

In this section we will construct the model selection procedure based on the family of improved estimates (5). It is required to choose the best estimate (in the sense of oracle inequality). To this end we consider the empirical squared error

$$
Err_m(\lambda) = \|S^*_\lambda - S\|^2_m = \sum_{j=1}^{m} \lambda_j^2 \theta^*_j \phi_j^2 + 2 \sum_{j=1}^{m} \lambda_j \theta^*_j \theta_{j,n} + \|S\|^2.
$$

Since $\theta^*_j, \theta_{j,n}$ are unknown, it is impossible to determine the weight coefficients $\lambda = (\lambda_j)_{j \geq 1}$ as minimum of $Err_m(\lambda)$. To do this, we replace the Fourier coefficients with estimates and add a penalty that makes sense of the payment for replacing the true value with estimates:

$$
\tilde{\theta}_{j,n} = \theta^*_j \theta_{j,n} - \frac{1}{m} s_{j,m}, \quad s_{j,m} = \frac{1}{m} \sum_{j=1}^{m} \phi_j^2(x_i)E_p S\eta_i^2 \quad \text{and} \quad P_m(\lambda) = \sum_{j=1}^{m} \lambda_j^2 s_{j,m}.
$$
Define the cost function for $0 < \delta < 1$

$$J(\lambda) = \sum_{j=1}^{m} \lambda_j^2 \theta_{j,n}^2 - 2 \sum_{j=1}^{m} \lambda_j \tilde{\theta}_{j,n} + \delta P_m(\lambda).$$

Minimizing this function by weight vectors

$$\lambda^* = \arg\min_{\lambda \in \Lambda} J_n(\lambda),$$

we get improved model selection procedures

$$S^* = S^*_{\lambda^*}.$$  \hspace{1cm} (7)

Further, a non-asymptotic upper bound was obtained for the robust risk of the procedure (7).

**Theorem 3.1** Let the observations be described by the equation (1). Then for any $n \geq 3, 0 < \delta < 1/3$ the robust risk (3) of the proposed model selection procedure (7) satisfies the following sharp oracle inequality

$$R(S^*, S) \leq \frac{1 - \delta}{1 - 2\delta} \min_{\lambda \in \Lambda} R(S^*_\lambda, S) + \frac{1}{\delta n} U_n,$$

where the term $U_n$ is such that for any $a > 0$

$$\lim_{n \to \infty} \frac{U_n}{n^a} = 0.$$

This theorem implies that the model selection procedure (7) is efficient in oracle inequalities sense.

4. **Monte Carlo simulation**

Now, we consider the results of a numerical comparison of the empirical quadratic risks of the proposed improved procedure and the LSE (4.17) from [9]. Suppose that in the model (1) the signal $S$ is defined on $[0, 1]$ and has the following form

$$S(t) = t \cos(2\pi t) + t^2 (1 - t) \cos(2\pi t),$$

where the noise $(\varepsilon_k)_{k \geq 1}$ is the sequence of independent identically distributed random variables which are mixed from the independent Gaussian $(0, 1)$ and exponential (of intensity 1) random variables. To define the vector $\lambda$ we set

$$\delta = (3 + \ln 2)^{-2},$$

Empirical quadratic risks are calculated using the approximate formula

$$\tilde{R}(S^*, S) \simeq \frac{1}{N} \sum_{l=1}^{N} \|S^*_l - S\|_n^2,$$  \hspace{1cm} (8)

where $S^*_l$ is the estimate calculated from the $l$-th replication of the sample, wherein $N = 1000$. The results are shown in the table.

It can be seen that the risks of the proposed procedure are less than for the procedure based on the LSE. We note that risks tend to zero as $n \to \infty$. Then the improved estimation approach is useful in practice when number of observations is limited.

Figures show the behavior of the observations $(y_k)_{1 \leq k \leq n}$ (black line), the real signal (white continuous line) the improved model selection procedure (white dashed line) and the procedure based on the LSE (white dotted line) depending on the values of observations $n$. From the Figures 1-3, for the proposed shrinkage procedure, we can conclude that the benefit is considerable for non large $n$. 

4
Table 1. The sample quadratic risks for different optimal $\lambda$.

| $n$   | 100       | 500       | 1000      | 10000     |
|-------|-----------|-----------|-----------|-----------|
| $\tilde{R}(S^*, S)$ | 2.4391    | 0.5299    | 0.0789    | 0.0039    |
| $\tilde{R}(\hat{S}, S)$ | 6.9832    | 2.2997    | 0.7296    | 0.0619    |
| $\tilde{R}(S^*, S)/\tilde{R}(\hat{S}, S)$ | 2.9       | 4.3       | 9.2       | 15.9      |

Figure 1. Observations, the real signal and its estimates for $n = 100$.

Figure 2. Observations, the real signal and its estimates for $n = 500$.

Figure 3. Observations, the real signal and its estimates for $n = 1000$.

Figure 4. Observations, the real signal and its estimates for $n = 10000$.

5. Conclusion
The proposed model selection procedure allows us to improve the non-asymptotic quality for the signal processing. The construction of the procedure is based on a special shrinkage algorithm. In this case, additional a priori information and an increase in the observation volume in comparison with other methods are not required. The presented theoretical results are corroborated by the corresponding experimental data found during the simulation. As a result we can see that the gain in mean square accuracy is significant. Moreover, the procedure is robust, i.e. it has stable accuracy characteristics with respect to changes over a wide range of noise distributions in the channels transmission of signals.
Acknowledgments
This work was supported by the Grant of the President of the Russian Federation, project No. MK-834.2020.9

References
[1] Koshelev V.I., Andreev V.G. 1993 Radioelectron. Commun. Syst. 36(7) 6
[2] Marple S.L. Jr. 1987, Digital Spectral Analysis with Applications, Prentice Hall.
[3] Quinn B.G. 2009 Digital Signal Processing 19 942
[4] Zvarich V.N., Marchenko B.G. 2011 Radioelectron. Commun. Syst. 54(7) 367.
[5] Belitser E. 2000 Annals of Statistics. 26 860
[6] Fan J., Zhang W. 2008 Statistics and Its Interface. 1 179
[7] Pchelintsev E.A., Pchelintsev V.A., Pergamenshchikov S.M. 2019 Journal of Nonparametric Statistics. 31 612
[8] Pchelintsev E.A., Pergamenshchikov S. M., Marcokova M. 2019 Adv. Electr. Electron. Eng. 17 (3) 270
[9] Arkoun O., Brua J.-Y., Pergamenshchikov S. 2019 Sequential Analysis. 18(4) 437
[10] Galtchouk L., Pergamenshchikov S. 2011 Journal of Nonparametric Statistics. 23 255