ISSN 0278-6419, Moscow University Computational Mathematics and Cybernetics, 2018, Vol. 42, No. 2, pp. 85–88. © Allerton Press, Inc., 2018.
Original Russian Text © O.V. Shestakov, 2018, published in Vestnik Moskovskogo Universiteta, Seriya 15: Vychislitel'naya Matematika i Kibernetika, 2018, No. 2, pp. 36–39.

Limit Theorems for Risk Estimate in Models with Non-Gaussian Noise

O. V. Shestakov*

Faculty of Computational Mathematics and Cybernetics, Moscow State University, Moscow, 119991 Russia;
Institute of Informatics Problems, Federal Research Center "Computer Science and Control", Russian Academy of Sciences, Moscow, 119333 Russia

Received November 15, 2017

Abstract—The problem of constructing an estimate of a signal function from noisy observations, assuming that this function is uniformly Lipschitz regular, is considered. The thresholding of empirical wavelet coefficients is used to reduce the noise. As a rule, it is assumed that the noise distribution is Gaussian and the optimal parameters of thresholding are known for various classes of signal functions. In this paper a model of additive noise whose distribution belongs to a fairly wide class, is considered. The mean-square risk estimate of thresholding is analyzed. It is shown that under certain conditions, this estimate is strongly consistent and asymptotically normal.

Keywords: thresholding, non-Gaussian noise, mean-square risk estimate.

DOI: 10.3103/S027864191802005X

1. INTRODUCTION

Many algorithms of signal and image processing are based on possibility of a sparse representation of the useful signal function in a certain basis. For a fairly wide class of functions, such sparseness is ensured by using wavelet bases [1]. This enables us to effectively separate noise from the useful signal and remove it using simple thresholding procedures [2–5]. A classical observation model assumes the presence of white Gaussian noise. The properties of estimates obtained by thresholding have been well studied, and we know the order of the mean-square risk of such procedures is found to be close to optimal [1]. Some results also testify to the asymptotic behavior of a mean-square estimate constructed from noisy observations of a signal function [6]. The strong consistency and the asymptotic normality of this estimate were demonstrated in [7, 8].

A wider class of possible noise distributions, especially ones with heavier tails than a Gaussian distribution, was considered in [9]. The parameters of so-called universal thresholding were calculated for this class, and it was shown the order of the mean-square risk was close to minimal with an accuracy up to the logarithm of the number of observations in a certain power, which depends on the distribution parameters. In this work, we prove the strong inconsistency and asymptotic normality of the mean-square risk estimate of universal thresholding in a model with a non-Gaussian noise distribution, assuming the signal function belongs to the Lipschitz class with a certain index.

2. THRESHOLDING IN A DATA MODEL WITH ADDITIVE NOISE

Let signal function f be defined on a certain interval $[a, b]$ and uniformly Lipschitz regular with an exponent $\gamma > 0$. In practice, f is given in discrete samples. We assume that the number of these samples is 2^J for a certain $J > 0$. After the discrete wavelet transform of the signal, set of wavelet coefficients

*E-mail: oshestakov@cs.msu.su.

$\{\mu_{j,k}\}_{j=0,\dots,J-1,k=0,\dots,2^j-1}$ is obtained, for which index j is called the scale, and index k is called the shift. When meeting certain conditions on the wavelet function [1], coefficients $\mu_{j,k}$ obey the inequality

$$|\mu_{j,k}| \leqslant \frac{C_f 2^{J/2}}{2^{j(\gamma+1/2)}}. \tag{1}$$

Since there is usually noise in actual observations, empirical wavelet coefficients take the form

$$Y_{j,k} = \mu_{j,k} + W_{j,k}, \quad j = 0,\dots,J-1, \quad k = 0,\dots,2^j-1,$$

where $\mu_{j,k}$ are the discrete wavelet coefficients of a pure signal, and $W_{j,k}$ represent the noise coefficients, for which it is assumed that they are independent and have a distribution with symmetric differentiable density $h(x)$. In this work, it is assumed that $\sup\limits_{x\in\mathbf{R}} |h'(x)| < A$ with a certain constant $A > 0$, and that

$$h(x) \asymp x^{\alpha} e^{-\theta x^{\beta}} \quad \text{as} \quad x \to \infty, \quad \alpha \in \mathbf{R}, \quad \theta > 0, \quad \beta > 0.$$

We denote a variance of $W_{j,k}$ by σ^2. The class of such distributions is fairly wide. Distributions from this class can have lighter and heavier tails than the Gaussian distribution.

One popular way of removing noise is thresholding of the empirical wavelet coefficients. This involves setting to zero the coefficients whose absolute values do not exceed a given threshold. Estimate $\hat{Y}_{j,k}$ is calculated using threshold function $\rho_T(Y_{j,k})$ with threshold T. Most common are the function of hard thresholding $\rho_T^{(h)}(x) = x\mathbf{1}(|x| > T)$ and soft thresholding $\rho_T^{(s)}(x) = \text{sign}(x)(|x| - T)_+$. We define the mean-square risk of the estimate obtained by thresholding as

$$R(f,T) = \sum_{j=0}^{J-1} \sum_{k=0}^{2^f-1} \mathbf{E}(\hat{Y}_{j,k} - \mu_{j,k})^2. \tag{2}$$

The way of choosing the threshold is one of the main problems when thresholding. So-called universal threshold $T_U = \left(\theta^{-1} \ln 2^J\right)^{1/\beta}$ is considered in this work. In a certain sense, this threshold is maximal among reasonable thresholds, and the mean-square risk at this threshold is close to the minimum one with an accuracy up to the logarithmic factor of the number of observations in a power that depends on the distribution parameters [9].

3. MEAN-SQUARE RISK ESTIMATE

Unknown values of pure coefficients $\mu_{j,k}$ are presented in expression (2), so we cannot calculate risk value $R(f,T)$ in practice. However, it can be estimated directly from the observed data. If $|Y_{j,k}| > T$ in the term, the contribution of this term into risk is σ^2, in the case of hard and $\sigma^2 + T^2$, in the case of soft thresholding, and if $|X_i| \leqslant T$, the contribution is $\mu_{j,k}^2$ in both cases. Since $\mathbf{E}Y_{j,k}^2 = \sigma^2 + \mu_{j,k}^2$, the value of $\mu_{j,k}^2$ must be estimated using difference $Y_{j,k}^2 - \sigma^2$. As an estimate of the mean-square risk, we can use the quantity

$$\hat{R}(f,T) = \sum_{j=0}^{J-1} \sum_{k=0}^{2^f-1} F[Y_{j,k}, T], \tag{3}$$

where $F[Y_{j,k},T] = (Y_{j,k}^2 - \sigma^2)\mathbf{1}(|Y_{j,k}| \leqslant T) + \sigma^2 \mathbf{1}(|Y_{j,k}| > T)$ in the case of hard thresholding and $F[Y_{j,k},T] = (Y_{j,k}^2 - \sigma^2)\mathbf{1}(|Y_{j,k}| \leqslant T) + (\sigma^2 + T^2)\mathbf{1}(|Y_{j,k}| > T)$ in the case of soft thresholding.

Risk estimate (3) enables us to obtain a representation of the error with which a signal function is estimated using only the observed data. If distribution $W_{j,k}$ is Gaussian, in the case of soft thresholding $\hat{R}(f,T)$ is an unbiased estimate of $R(f,T)$. Under certain additional conditions, it is also strongly consistent and asymptotically normal [7, 8]. Similar properties for the more general noise model considered in this work will be proved.

Theorem 1. *Let signal function f be defined on interval $[a,b]$ and uniformly Lipschitz regular with an exponent $\gamma > 1/2$. Then with hard and soft thresholdings*

$$\mathbf{P}\left(\frac{\hat{R}(f,T_U)-R(f,T_U)}{\sqrt{\vartheta^2 2^J}} < x\right) \to \Phi(x) \quad \text{as} \quad x \to \infty, \tag{4}$$

where ϑ^2 is the variance of $W_{j,k}^2$, and $\Phi(x)$ is the distribution function of the standard normal law.

Proof. Let us prove the theorem statement for hard thresholding. (The proof is similar for soft thresholding.) We have

$$\hat{R}(f,T_U)-R(f,T_U)=\hat{R}(f,T_U)-\mathbf{E}\hat{R}(f,T_U)+\mathbf{E}\hat{R}(f,T_U)-R(f,T_U).$$

We estimate quantity $\mathbf{E}\hat{R}(f,T_U)-R(f,T_U)$. We consider the part of this sum (denoted by S_1) at which terms the inequality $|\mu_{j,k}| < cT_U$ may not be fulfilled when c is a positive constant that will be chosen later on. Summation over j in S_1 is performed up to $j_1 \approx J/(2\gamma+1)-\log_2(cT_U)/(2\gamma+1)+\log_2 C_f$ in virtue of the inequality (1). In addition, $|\mathbf{E}F[Y_{j,k},T]-\mathbf{E}(\hat{Y}_{j,k}-\mu_{j,k})^2| \leqslant C_s T_U^2$ is fulfilled for a certain constant $C_s > 0$ (See [9]). There is thus a constant $C_1 > 0$ that

$$|S_1| \leqslant C_1 T_U^2 2^{\frac{J}{2\gamma+1}}.$$

We now consider the remainder of the sum where $|\mu_{j,k}| < cT_U$. We denote it by S_2. For terms of this sum, we find

$$\mathbf{E}F[Y_{j,k},T]-\mathbf{E}(\hat{Y}_{j,k}-\mu_{j,k})^2 = 2\mathbf{E}\mathbf{1}(|Y_{j,k}|>T)[\sigma^2+\mu_{j,k}Y_{j,k}-Y_{j,k}^2]. \tag{5}$$

Taking into account the form of (5) and repeating the arguments of [9], we obtain the estimate

$$|\mathbf{E}F[Y_{j,k},T]-\mathbf{E}(\hat{Y}_{j,k}-\mu_{j,k})^2| \leqslant C_2 T_U^{3+\alpha-\beta} 2^{-J(1-c)^\beta},$$

where C_2 is a certain positive constant. Therefore,

$$|S_2| \leqslant C_2 T_U^{3+\alpha-\beta} 2^{J(1-(1-c)^\beta)}.$$

Choosing $0 < c < 1-(\frac{2\gamma}{2\gamma+1})^{1/\beta}$, we find here exists such a constant $C > 0$ that

$$|\mathbf{E}\hat{R}(f,T_U)-R(f,T_U)| \leqslant C T_U^2 2^{\frac{J}{2\gamma+1}}.$$

Consequently,

$$\frac{\mathbf{E}\hat{R}(f,T_U)-R(f,T_U)}{\sqrt{\vartheta^2 2^J}} \to 0 \quad \text{as} \quad J \to \infty. \tag{6}$$

We now consider difference $\hat{R}(f,T_U)-\mathbf{E}\hat{R}(f,T_U)$. Let p be such that $(2\gamma+1)^{-1} < p < 1/2$. We write

$$\begin{aligned}\hat{R}(f,T_U)-\mathbf{E}\hat{R}(f,T_U) &= \sum_{j=0}^{[pJ]}\sum_{k=0}^{2^f-1}(F[Y_{j,k},T_U]-\mathbf{E}F[Y_{j,k},T_U]) \\ &+ \sum_{j=[pJ]+1}^{J-1}\sum_{k=0}^{2^f-1}(F[Y_{j,k},T_U]-\mathbf{E}F[Y_{j,k},T_U]).\end{aligned} \tag{7}$$

Since

$$|F[Y_{j,k},T_U]-\mathbf{E}F[Y_{j,k},T_U]| \leqslant C_F T_U^2 \ \text{ a. e.,} \tag{8}$$

where C_F is a certain positive constant, then

$$\frac{\sum_{j=0}^{[pJ]}\sum_{k=0}^{2^j-1}(F[Y_{j,k},T_U]-\mathbf{E}F[Y_{j,k},T_U])}{\sqrt{\vartheta^2 2^J}} \to 0 \ \text{ a. e. as } J \to \infty.$$

By virtue of (1), $\mu_{jk} \to 0$ as $J \to \infty$ for all terms in the second sum of (7).

Repeating arguments from [7], it can be shown that

$$\lim_{J\to\infty} \frac{\mathsf{D}\sum_{j=[pJ]+1}^{J-1}\sum_{k=0}^{2^j-1}(F[Y_{j,k},T_U]-\mathbf{E}F[Y_{j,k},T_U])}{\sqrt{\vartheta^2 2^J}} = 1.$$

Finally, the Lindeberg condition is fulfilled: for any $\epsilon > 0$ as $J \to \infty$

$$\frac{1}{\vartheta^2 2^J}\sum_{j=[pJ]+1}^{J-1}\sum_{k=0}^{2^j-1}\mathbf{E}\big[\big(F[Y_{j,k},T_U]-\mathbf{E}F[Y_{j,k},T_U]\big)^2\mathbf{1}\big(|F[Y_{j,k},T_U] - \mathbf{E}F[Y_{j,k},T_U]| > \epsilon\vartheta^2 2^J\big)\big] \to 0, \tag{9}$$

since starting from a certain J, all indicators in (9) vanish. Hence, statement (4) holds true. The theorem is proved.

We now prove the strong consistency of estimate (3).

Теорема 2. *Let $f \in L^2(\mathbf{R})$. Then with hard and soft thresholdings for any $\lambda > 1/2$*

$$\frac{\hat{R}(f,T_U)-R(f,T_U)}{2^{\lambda J}} \to 0 \quad \text{a. e. as} \quad J\to\infty. \tag{10}$$

Proof. Using the Hoeffding inequality in light of (8) and the form of T_U, we find that for any $\delta > 0$ there is such a constant $C_\delta > 0$ that

$$p_J = \mathbf{P}\left(\frac{|\hat{R}(f,T_U)-\mathbf{E}\hat{R}(f,T_U)|}{2^{\lambda J}} > \delta\right) \leqslant \exp\left\{-C_\delta\frac{2^{2\lambda J-J}}{J^{4/\beta}}\right\}.$$

Hence, $\sum_{J=1}^{\infty} p_J < \infty$. By virtue of the Borel–Cantelli lemma and (6), (10) is fulfilled. The theorem is proved.

ACKNOWLEDGMENTS

This work was supported in part by the Russian Foundation for Basic Research, project no. 16-07-00736.

REFERENCES

1. S. Mallat, *A Wavelet Tour of Signal Processing* (Academic, New York, 1999).
2. D. Donoho and I. M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," Biometrika **81**, 425–455 (1994).
3. J. S. Marron, S. Adak, I. M. Johnstone, M. H. Neumann, and P. Patil, "Exact risk analysis of wavelet regression," J. Comput. Graph. Stat. **7**, 278–309 (1998).
4. A. Antoniadis and J. Fan, "Regularization of wavelet approximations," J. Am. Statist. Assoc. **96** (455), 939–967 (2001).
5. M. Jansen, *Noise Reduction by Wavelet Thresholding,* Vol. 161 of *Lecture Notes in Statistics* (Springer, New York, 2001).
6. D. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," J. Am. Stat. Assoc. **90**, 1200–1224 (1995).
7. A. V. Markin, "The limiting distribution of the risk assessment for threshold processing of wavelet coefficients," Inform. Primen. **3** (4), 57–63 (2009).
8. O. V. Shestakov, "On the strong consistency of the adaptive risk estimator for wavelet thresholding," J. Math. Sci. **214** (1), 115–118 (2016).
9. O. V. Shestakov, "Universal thresholding in the models with non-Gaussian noise," Inform. Primen. **11** (2), 122–125 (2017).

Translated by D. Churochkin