ВЕСТНИК ТОМСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА

2014 Математика и механика № 5(31)

УДК 519.2

В.А. Пчелинцев, Е.А. Пчелинцев

МИНИМАКСНОЕ ОЦЕНИВАНИЕ ГАУССОВСКОЙ ПАРАМЕТРИЧЕСКОЙ РЕГРЕССИИ

Рассматривается задача минимаксного оценивания d-мерного вектора неизвестных параметров регрессии с гауссовскими шумами при квадратической функции потерь. Предлагается модификация процедуры Джеймса – Стейна, для которой найдена явная верхняя граница для среднеквадратического риска и показано, что ее риск строго меньше риска классической оценки максимального правдоподобия для размерности $d \geq 2$. Проведено численное сравнение среднеквадратических рисков рассматриваемых оценок.

Ключевые слова: *параметрическая регрессия, улучшенное оценивание, процедура Джеймса – Стейна, среднеквадратический риск, минимаксная оценка.*

Рассмотрим классическую задачу оценивания неизвестных параметров регрессионных моделей. Пусть наблюдения описываются уравнением

$$Y = \theta + \sigma\xi, \tag{1}$$

где θ – неизвестный вектор постоянных параметров из некоторого ограниченного множества $\Theta \subset \mathbb{R}^d$, ξ – гауссовский случайный вектор с нулевым средним и единичной ковариационной матрицей I_d, т.е. $Law(\xi)=\mathrm{N}_d(0, I_d)$, σ – некоторое известное положительное число. Задача состоит в том, чтобы оценить параметр θ по наблюдениям Y. В качестве меры точности оценки $\hat{\theta}$ выберем среднеквадратический риск, определяемый следующим образом:

$$R\left(\theta, \hat{\theta}\right) := \boldsymbol{E}_\theta \left|\theta - \hat{\theta}\right|^2, \ |x|^2 = \sum_{j=1}^{d} x_j^2 ,$$

$\boldsymbol{E}_\theta$ – математическое ожидание относительно меры $\boldsymbol{P}_\theta$. Напомним, что оценкой параметра θ является любая борелевская функция от наблюдений Y [1].

Известно [2], что наилучшей по точности в классе линейных несмещенных оценок является оценка по методу максимального правдоподобия

$$\hat{\theta}_{ML} = Y , \tag{2}$$

которая имеет нормальное распределение $\mathrm{N}_d\left(\theta, \sigma^2 I_d\right)$ и ее среднеквадратический риск определяется равенством

$$R\left(\theta, \hat{\theta}_{ML}\right) = d\sigma^2 .$$

В 1961 г. Джеймс и Стейн [3] предложили сжимающую оценку вида

$$\hat{\theta}_{\mathrm{JS}} = \left(1 - \frac{c}{|Y|^2}\right) Y ,$$

которая для всех $0 < c < 2(d-2)$ равномерно по θ превосходит по среднеквадратической точности оценку метода максимального правдоподобия при $d \geq 3$, т.е. для любого $\theta \in \mathbb{R}^d$ справедливо неравенство

$$R(\theta, \hat{\theta}_{\mathrm{JS}}) < R(\theta, \hat{\theta}_{ML}) .$$

Оценка Джеймса – Стейна является минимаксной оценкой [3, 4]. Полученный результат побудил многих статистиков к развитию так называемой теории улучшенного оценивания. Появилась серия работ, в которых были предложены различные минимаксные модификации оценки Джеймса – Стейна. Одной из простых модификаций является оценка

$$\hat{\theta}^{+}_{\mathrm{JS}} = \left(1 - \frac{c}{|Y|^2}\right)_{+} Y , \ a_{+} = \max(a, 0) ,$$

которая известна как положительная часть оценки Джеймса – Стейна и была предложена в 1964 году Баранчиком [5]. В этой работе было доказано, что такая оценка превосходит по среднеквадратической точности не только оценку по методу максимального правдоподобия, но и оценку Джеймса – Стейна $\hat{\theta}_{\mathrm{JS}}$ (см. рис. 1).

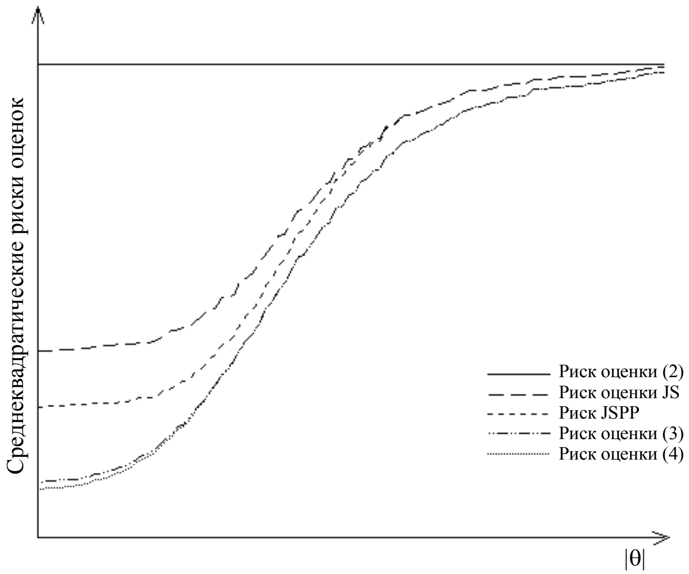

Рис. 1. Среднеквадратические риски оценок максимального правдоподобия (2), Джеймса – Стейна, её положительной части, (3) и (4) как функции от $|\theta|$

Различные оценки, обладающие аналогичным свойством, были предложены в работах [6–8]. В перечисленных работах сжимающий коэффициент не был явно определен аналитически, а лишь предложены алгоритмы его численной оптимизации. Задача Джеймса – Стейна была изучена для более общих моделей, в том числе с неизвестной ковариационной матрицей [9–11]. Значительные усилия были направлены на решение задачи улучшенного оценивания в негауссовских моделях [12–17].

В работах [14–17] для модели регрессии, в которой шум является условно-гауссовским, предложены новые минимаксные модификации оценки Джеймса – Стейна вида

$$\theta^* = \left(1 - \frac{c}{|Y|}\right) Y . \tag{3}$$

Здесь, в отличие от всех других модификаций, сжимающий коэффициент определяется множителем, содержащим $|Y|$, а не $|Y|^2$. Такая замена оправдана тем, что позволяет получить явные формулы для среднеквадратической точности и контролировать ее.

Лемма. Пусть наблюдения описываются уравнением (1). Тогда оценка θ^* с $c = (d-1)\sigma^2\delta_d$, где $\delta_d = \left(\beta + \sigma\sqrt{d}\right)^{-1}$, $\beta = \sup_{\theta\in\Theta}\{|\theta|\}$, превосходит по среднеквадратической точности оценку максимального правдоподобия для любого $d \geq 2$ и является минимаксной, причем разность рисков удовлетворяет неравенству

$$\Delta^*(\theta) := R(\theta, \theta^*) - R(\theta, \hat{\theta}) \leq -\left((d-1)\sigma^2\delta_d\right)^2 .$$

Доказательство. Рассмотрим риски оценок (2) и (3):

$$R(\hat{\theta}_{ML}, \theta) = \boldsymbol{E}_\theta \left|\hat{\theta}_{ML} - \theta\right|^2 = \sigma^2 \boldsymbol{E}_\theta |\xi|^2 = \sigma^2 d ;$$

$$R\left(\theta^*, \theta\right) = R\left(\hat{\theta}_{ML}, \theta\right) + \boldsymbol{E}_\theta \left(g(Y) - 1\right)^2 |Y|^2 + 2\sum_{j=1}^{d} \boldsymbol{E}_\theta \left(g(Y) - 1\right) Y_j (Y_j - \theta_j),$$

где $g(Y) = 1 - c/|Y|$.

Обозначив $f(Y) = \left(g(Y) - 1\right) Y_j$ и используя плотность распределения вектора Y

$$p_Y(x) = \frac{1}{(2\pi)^{d/2}\sigma} \exp\left(-\frac{|x-\theta|^2}{2\sigma^2}\right),$$

имеем

$$I_j := \boldsymbol{E}_\theta f(Y)(Y_j - \theta_j) = \int_{\mathbb{R}^d} f(x)(x_j - \theta_j) p_Y(x) dx , \quad j = \overline{1, d} .$$

Делая замену переменной $u = (x - \theta)/\sigma$ и полагая $\tilde{f}(u) = f(\sigma u + \theta)$, находим

$$I_j = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \tilde{f}(u) u_j \exp\left(-\frac{|u|^2}{2}\right) du, \quad j = \overline{1, d} .$$

Эти величины можно переписать как

$$I_j = \sigma^2 \boldsymbol{E}_\theta \left(\frac{\partial f}{\partial u_j}(u) \Big|_{u = Y} \right), \quad j = \overline{1, d} .$$

Таким образом, квадратический риск для оценки (3) представляется в виде

$$R\left(\theta^*, \theta\right) = R\left(\hat{\theta}_{ML}, \theta\right) + \boldsymbol{E}_\theta \left(g(Y) - 1\right)^2 |Y|^2 + 2\sigma^2 \boldsymbol{E}_\theta \sum_{j=1}^{d} \left(\frac{\partial}{\partial u_j} \left[(g(u) - 1) u_j \right] \Big|_{u = Y} \right).$$

Отсюда, после несложных преобразований, получаем

$$R(\theta,\theta^*) = R(\theta,\hat{\theta}) + \boldsymbol{E}_\theta W(Y)\,,$$

где

$$W(x) = c^2 - 2(d-1)\sigma^2 \frac{c}{|x|}\,.$$

Следовательно,

$$\Delta(\theta) = c^2 - 2(d-1)\sigma^2 c\boldsymbol{E}_\theta |Y|^{-1}\,.$$

Оценим снизу величину $\boldsymbol{E}_\theta |Y|^{-1}$. Из неравенства Йенсена [18] имеем

$$\boldsymbol{E}_\theta |Y|^{-1} \geq (|\theta| + \sigma\boldsymbol{E}_\theta |\xi|)^{-1} \geq \left(\beta + \sigma\sqrt{d}\right)^{-1} = \delta_d\,.$$

Тогда для всех $\theta \in \Theta$:

$$\Delta(\theta) \leq c^2 - 2(d-1)\sigma^2\delta_d c =: \varphi(c)\,.$$

Минимизируя функцию $\varphi(c)$ по c, получим

$$\Delta^*(\theta) \leq -\left((d-1)\sigma^2\delta_d\right)^2.$$

Лемма доказана.

Цель настоящей статьи выяснить, является ли оценка

$$\theta_+^* = \left(1 - \frac{c}{|Y|}\right)_+ Y \tag{4}$$

– положительная часть оценки (3), улучшенной по сравнению с оценками метода максимального правдоподобия (2) и θ^* и, следовательно, минимаксной.

Введем следующие обозначения:

$$M = (d-1)\sigma^2\delta_d \Gamma\left(\frac{d}{2}\right) \Big/ \gamma\left(\frac{d}{2}, \frac{((d-1)\sigma\delta_d)^2}{2}\right),$$

$$\Gamma(z) = \int_0^{+\infty} t^{z-1} \mathrm{e}^{-t} dt, \quad \gamma(z,a) = \int_0^a t^{z-1} \mathrm{e}^{-t} dt$$

– полная и неполная гамма-функции соответственно,

$$\Theta_M = \{\theta \in \Theta : |\theta| < M\}.$$

Теорема. Пусть наблюдения описываются уравнением (1), причем $\theta \in \Theta_M$. Тогда оценка (4) с $c = (d-1)\sigma^2\delta_d$ превосходит по среднеквадратической точности оценку максимального правдоподобия (2) для любого $d \geq 2$, т.е.

$$\sup_{\theta\in\Theta_M} \left[R(\theta,\theta_+^*) - R(\theta,\hat{\theta})\right] < 0,$$

и является минимаксной.

Доказательство. Рассмотрим риск оценки (4):

$$R(\theta,\theta_+^*) = \boldsymbol{E}_\theta \left|\theta - \theta_+^*\right|^2 = \boldsymbol{E}_\theta \left|\theta - \theta_+^*\right|^2 \boldsymbol{I}_{(|Y|>c)} + \boldsymbol{E}_\theta \left|\theta - \theta_+^*\right|^2 \boldsymbol{I}_{(|Y|\leq c)} =$$

$$= \boldsymbol{E}_\theta \left|\theta - \theta^*_+\right|^2 \boldsymbol{1}_{(|Y|>c)} + \boldsymbol{E}_\theta |\theta|^2 \boldsymbol{1}_{(|Y|\le c)} = \boldsymbol{E}_\theta \left|\theta - \theta^*_+\right|^2 \left(1 - \boldsymbol{1}_{(|Y|\le c)}\right) + \boldsymbol{E}_\theta |\theta|^2 \boldsymbol{1}_{(|Y|\le c)} =$$

$$= R(\theta,\theta^*) - \boldsymbol{E}_\theta \left|\theta - \theta^*\right|^2 \boldsymbol{1}_{(|Y|\le c)} + |\theta|^2 \boldsymbol{P}_\theta\left(|Y| \le c\right),$$

где $\boldsymbol{1}_A(t)$ – индикатор множества A, т.е.

$$\boldsymbol{1}_A(t) = \begin{cases} 1, & t \in A, \\ 0, & t \notin A. \end{cases}$$

Поскольку второе слагаемое неотрицательно, то для риска оценки (4) получаем оценку

$$R(\theta,\theta^*_+) \le R(\theta,\theta^*) + |\theta|^2 \boldsymbol{P}_\theta\left(|Y| \le c\right). \tag{5}$$

Используя неравенство Андерсона [18], для вероятности имеем

$$\boldsymbol{P}_\theta\left(|Y| \le c\right) = \boldsymbol{P}_\theta\left(|\theta + \sigma\xi| \le c\right) \le \boldsymbol{P}_\theta\left(|\xi|^2 \le \frac{c^2}{2\sigma^2}\right).$$

Подставляя эту оценку в (5) и учитывая, что величина $|\xi|^2$ имеет хи-квадрат-распределение с d степенями свободы, получаем

$$R(\theta,\theta^*_+) \le R(\theta,\theta^*) + |\theta|^2 \frac{\gamma\left(\frac{d}{2}, \frac{c^2}{2\sigma^2}\right)}{\Gamma\left(\frac{d}{2}\right)}. \tag{6}$$

Далее, из предыдущей леммы следует, что

$$R(\theta,\theta^*_+) \le R(\theta,\hat\theta) - \left((d-1)\sigma^2\delta_d\right)^2 + |\theta|^2 \frac{\gamma\left(\frac{d}{2}, \frac{((d-1)\sigma\delta_d)^2}{2}\right)}{\Gamma\left(\frac{d}{2}\right)}.$$

Отсюда и из условий теоремы имеем следующее неравенство для рисков оценок (4) и (2):

$$R(\theta,\theta^*_+) - R(\theta,\hat\theta) < 0.$$

Теорема доказана.

Замечание. В теореме утверждается, что предложенная оценка (4) является улучшенной в смысле среднеквадратической точности по сравнению с оценкой максимального правдоподобия (2) и, следовательно, является минимаксной оценкой. Остался открытым вопрос: будет ли оценка (4) превосходить по среднеквадратической точности и оценку (3). Из неравенства (6) следует, что оценка (4) может иметь большую точность по сравнению с (3) лишь в случаях, когда неизвестный параметр лежит в некоторой малой окрестности нуля.

Для подтверждения аналитически установленных результатов, в среде Scilab проведено численное моделирование среднеквадратических рисков рассматриваемых в работе оценок. При численном моделировании наблюдений Y предполагалось, что размерность вектора параметров $d = 5$, коэффициент $\sigma = 1$ и ξ – вектор с независимыми случайными гауссовскими (0, 1) компонентами. Вектор параметров θ выбирался таким, что $|\theta|$ изменяется в пределах от 0 до 100. Оценки вычислялись по соответствующим формулам (2) – (4). Среднеквадратический риск оценки вычислялся по эмпирической формуле

$$\tilde{R}\left(\theta,\hat{\theta}\right)=\frac{1}{N}\sum_{k=1}^{N}\left|\theta-\hat{\theta}_k\right|^2,$$

где $\hat{\theta}_k$ – k-я реализация оценки $\hat{\theta}$ и $N = 10\,000$.

На представленном рисунке видно, что в заданной области наилучшей по среднеквадратической точности является предложенная минимаксная оценка (4). Отметим, что при $|\theta| \to \infty$ риски всех изучаемых в работе оценок будут стремиться к риску оценки максимального правдоподобия (2).

ЛИТЕРАТУРА

1. *Fourdrinier D.* Statistique Inférentielle. Paris: Dunod, 2002.
2. *Lehmann E.L.*, *Casella G.* Theory of Point Estimation. 2nd edition. N.Y.: Springer, 1998.
3. *James W.*, *Stein C.* Estimation with quadratic loss // Proceedings of the Fourth Berkeley Symposium on Mathematics Statistics and Probability. V. 1. Berkeley: University of California Press, 1961. P. 361–380.
4. *Stein C.* Estimation of the mean of a multivariate normal distribution // The Annals of Statistics. 1981. V. 9(6). P. 1135–1151.
5. *Baranchik A.J.* Multiple regression and estimation of the mean of a multivariate normal distribution // Technical Report / Department of Statistics, Stanford University. 1964. V. 51.
6. *Strawderman W.E.* Proper Bayes minimax estimators of the multivariate normal distribution // Annals of Mathematical Statistics. 1971. V. 42. P. 385–388.
7. *Guo Y.Y.*, *Pal N.* A sequence of improvements over the James – Stein estimator // J. Multivariate Analysis. 1992. V. 42. P. 302–317.
8. *Shao P.Y.-S.*, *Strawderman W.E.* Improving on the James – Stein positive-part estimator // The Annals of Statistics. 1994. V. 22. P. 1517–1538.
9. *Efron B.*, *Morris C.* Families of minimax estimators of the mean of a multivariate normal distribution // The Annals of Statistics. 1976. No. 4. P. 11–21.
10. *Berger J.O.*, *Haff L.R.* A class of minimax estimators of a normal mean vector for arbitrary quadratic loss and unknown covariance matrix // Statist. Decisions. 1983. No. 1. P. 105–129.
11. *Gleser L.J.* Minimax estimators of a normal mean vector for arbitrary quadratic loss and unknown covariance matrix // The Annals of Statistics. 1986. V. 14. P. 1625–1633.
12. *Fourdrinier D.*, *Pergamenshchikov S.* Improved selection model method for the regression with dependent noise // Ann. of the Inst. of Statist. Math. 2007. V. 59 (3). P. 435–464.
13. *Fourdrinier D.*, *Strawderman W.E.*, *William E.* A unified and generalized set of shrinkage bounds on minimax Stein estimates // J. Multivariate Anal. 2008. V. 99. P. 2221–2233.
14. *Пчелинцев Е.А.* Процедура Джеймса – Стейна для условно-гауссовской регрессии // Вестник Томского государственного университета. Математика и механика. 2011. № 4(16). С. 6–17.
15. *Конев В.В.*, *Пчелинцев Е.А.* Оценивание параметрической регрессии с импульсными шумами по дискретным наблюдениям // Вестник Томского государственного университета. Математика и механика. 2012. № 1(17). С. 20–35.
16. *Pchelintsev E.* Improved estimation in a non-Gaussian parametric regression // Statistical Inference for Stochastic Processes. 2013. V. 16 (1). P. 15–28.
17. *Конев В.В.*, *Пергаменщиков С.М.*, *Пчелинцев Е.А.* Оценивание регрессии с шумами импульсного типа по дискретным наблюдениям // ТВП. 2013. V. 58(3). С. 454–471.
18. *Ибрагимов И.А.*, *Хасьминский Р.З.* Асимптотическая теория оценивания. М.: Наука, 1979.

Статья поступила 15.07.2014 г.

Pchelintsev V.A., Pchelintsev E.A. MINIMAX ESTIMATION OF THE GAUSSIAN PARAMETRIC REGRESSION

The paper considers the problem of estimating a $d\geq 2$ dimensional mean vector of a multivariate normal distribution under quadratic loss. Let the observations be described by the equation

$$Y = \theta + \sigma\xi, \tag{1}$$

where θ is a d-dimension vector of unknown parameters from some bounded set $\Theta \subset \mathbb{R}^d$, ξ is a Gaussian random vector with zero mean and identity covariance matrix I_d, i.e. $Law(\xi)=N_d(0, I_d)$ and σ is a known positive number. The problem is to construct a minimax estimator of the vector θ from observations Y. As a measure of the accuracy of estimator $\hat{\theta}$ we select the quadratic risk defined as

$$R(\theta,\hat{\theta}) := \boldsymbol{E}_\theta |\theta - \hat{\theta}|^2, \quad |x|^2 = \sum_{j=1}^{d} x_j^2,$$

where $\boldsymbol{E}_\theta$ is the expectation with respect to measure $\boldsymbol{P}_\theta$.

We propose a modification of the James – Stein procedure of the form

$$\theta_+^* = \left(1 - \frac{c}{|Y|}\right)_+ Y$$

where $c > 0$ is a special constant and $a_+ = \max(a,0)$ is a positive part of a. This estimate allows one to derive an explicit upper bound for the quadratic risk and has a significantly smaller risk than the usual maximum likelihood estimator and the estimator

$$\theta^* = \left(1 - \frac{c}{|Y|}\right) Y$$

for the dimensions $d \geq 2$. We establish that the proposed procedure θ_+^* is minimax estimator for the vector θ.

A numerical comparison of the quadratic risks of the considered procedures is given. In conclusion it is shown that the proposed minimax estimator θ_+^* is the best estimator in the mean square sense.

Keywords: parametric regression; improved estimation; James – Stein procedure; mean squared risk, minimax estimator.

Pchelintsev Valery Anatolyevich (Candidate of Physics and Mathematics, Assoc. Prof.,
Tomsk Polytechnic University, Tomsk, Russian Federation)
E-mail: vpchelintsev@vtomske.ru

Pchelintsev Evgeny Anatolyevich (Candidate of Physics and Mathematics, Assoc. Prof.,
Tomsk State University, Tomsk, Russian Federation)
E-mail: evgen-pch@yandex.ru

REFERENCES

1. Fourdrinier D. *Statistique Inférentielle*. Paris, Dunod Publ., 2002.
2. Lehmann E.L., Casella G. *Theory of Point Estimation*. 2nd ed. N.Y., Springer, 1998.
3. James W., Stein C. Estimation with quadratic loss. *Proceedings of the Fourth Berkeley Symposium on Mathematics Statistics and Probability*. Berkeley, University of California Press., 1961, vol. 1, pp. 361–380.
4. Stein C. Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics*, 1981, vol. 9(6), pp. 1135–1151.
5. Baranchik A.J. Multiple regression and estimation of the mean of a multivariate normal distribution. *Technical Report. Department of Statistics*, Stanford University, 1964, vol. 51.

6. Strawderman W.E. Proper Bayes minimax estimators of the multivariate normal distribution. *Annals of Mathematical Statistics*, 1971, vol. 42, pp. 385–388.
7. Guo Y.Y., Pal N. A sequence of improvements over the James – Stein estimator. *J. Multivariate Analysis*, 1992, vol. 42, pp. 302–317.
8. Shao P.Y.-S., Strawderman W.E. Improving on the James – Stein positive-part estimator. *The Annals of Statistics*, 1994, vol. 22, pp. 1517–1538.
9. Efron B., Morris C. Families of minimax estimators of the mean of a multivariate normal distribution. *The Annals of Statistics*, 1976, no. 4, pp. 11–21.
10. Berger J.O., Haff L.R. A class of minimax estimators of a normal mean vector for arbitrary quadratic loss and unknown covariance matrix. *Statist. Decisions*, 1983, no. 1, pp. 105–129.
11. Gleser L.J. Minimax estimators of a normal mean vector for arbitrary quadratic loss and unknown covariance matrix. *The Annals of Statistics*, 1986, vol. 14, pp. 1625–1633.
12. Fourdrinier D., Pergamenshchikov S. Improved selection model method for the regression with dependent noise. *Ann. of the Inst. of Statist. Math.*, 2007, vol. 59 (3), pp. 435–464.
13. Fourdrinier D., Strawderman W.E., William E. A unified and generalized set of shrinkage bounds on minimax Stein estimates. *J. Multivariate Anal.*, 2008, vol. 99, pp. 2221–2233.
14. Pchelintsev E.A. Protsedura Dzheymsa – Steyna dlya uslovno-gaussovskoy regressii. *Vestnik Tomskogo gosudarstvennogo universiteta. Matematika i mekhanika*, 2011, no. 4(16), pp. 6–17. (in Russian)
15. Konev V.V., Pchelintsev E.A. Otsenivanie parametricheskoy regressii s impul'snymi shumami po diskretnym nablyudeniyam. *Vestnik Tomskogo gosudarstvennogo universiteta. Matematika i mekhanika*, 2012, no. 1(17), pp. 20–35. (in Russian)
16. Pchelintsev E. Improved estimation in a non-Gaussian parametric regression. *Statistical Inference for Stochastic Processes*, 2013, vol. 16 (1), pp. 15–28.
17. Konev V.V., Pergamenshchikov S.M., Pchelintsev E.A. Estimation of a regression with the noise of pulse type from discrete data. *Theory of Probability and its Applications*, 2014, vol. 58, no. 3, pp. 454–471.
18. Ibragimov I.A., Khas'minskii R.Z. *Statistical Estimation. Asymptotic Theory*. New York, Springer-Verlag, 1981.