

УДК 519.213:519.23

ОБ ОЦЕНИВАНИИ ПАРАМЕТРОВ РЕГРЕССИОННЫХ МОДЕЛЕЙ С ИСПОЛЬЗОВАНИЕМ ВЕЙВЛЕТОВ

Тимофеев В.С., Исаева Е.В.

ФГБОУ ВО «Новосибирский государственный технический университет», Новосибирск,
e-mail: v.timofeev@corp.nstu.ru, isaeva@corp.nstu.ru

В работе рассматривается применение вейвлет-анализа при оценивании неизвестных параметров регрессионных моделей с помощью метода максимального правдоподобия. На сегодняшний день вейвлет-анализ широко используется для аппроксимации различных функций, в том числе и для восстановления функции плотности распределения. Таким образом, использование различных материнских вейвлетов дает хороший результат при оценивании параметров регрессионных зависимостей за счет возможности адаптироваться как к известным теоретическим распределениям, так и к различным распределениям, возникающим при статистической обработке данных. В качестве базисных вейвлетов в работе были рассмотрены вейвлет LITTLEWOOD & PALEY, вейвлет Морле, DOG вейвлет и вейвлет «Мексиканская шляпа». На основе данных материнских вейвлетов были построены функции правдоподобия и получены их логарифмические представления. Опираясь на теорию вейвлет-анализа, предложили универсальный алгоритм для оценивания параметров регрессионных моделей, обеспечивающих получение качественных результатов для большого числа реальных задач, в том числе при отсутствии известных предположений о свойствах случайной компоненты. Достаточная серия вычислительных экспериментов позволила сделать вывод о качестве получаемых оценок неизвестных параметров регрессионных моделей, при различных условиях формирования выборки. Отмечено влияние объема выборки на точность оценивания параметров регрессионных моделей. Установлено, что увеличение объема исходных данных дает существенное преимущество. Сравнение точности предложенного алгоритма с методом наименьших квадратов и методом полупараметрического восстановления функции плотности на основе обобщенного лямбда-распределения позволило сделать вывод о целесообразности применения вейвлет-анализа при оценивании неизвестных параметров регрессионных моделей.

Ключевые слова: уравнение регрессии, метод максимального правдоподобия, функция плотности распределения, оценка функции плотности распределения, оценивание параметров, вейвлет-анализ, вейвлет-оценка, вейвлет Морле, DOG вейвлет, вейвлет LITTLEWOOD & PALEY, вейвлет «Мексиканская шляпа»

ON ESTIMATION OF PARAMETERS OF REGRESSION MODELS WITH THE USE OF WAVELETS

Timofeev V.S., Isaeva E.V.

Novosibirsk State Technical University, Novosibirsk,
e-mail: v.timofeev@corp.nstu.ru, isaeva@corp.nstu.ru

This paper considers the use of wavelet analysis in estimating unknown parameters of regression models using the maximum likelihood method. To date, wavelet analysis is widely used to approximate various functions, including the restoration of the distribution density function. Thus, the use of different mother wavelets gives a good result in estimating the parameters of regression dependencies due to the ability to adapt both to known theoretical distributions and to various distributions that arise during statistical data processing. As basic wavelets, the LITTLEWOOD & PALEY wavelet, the Morlet wavelet, the DOG wavelet, and the «Mexican hat» wavelet were considered in the work. On the basis of these mother wavelets, the likelihood functions were constructed and their logarithmic representations were obtained. Based on the theory of wavelet analysis, a universal algorithm was proposed for estimating the parameters of regression models that provide high-quality results for a large number of real problems, including in the absence of known assumptions about the properties of a random component. A sufficient series of computational experiments made it possible to draw a conclusion about the quality of the obtained estimates of the unknown parameters of the regression models, under various conditions of sampling. The influence of the sample size on the accuracy of estimating the parameters of regression models is noted. It has been established that an increase in the volume of initial data provides a significant advantage. Comparison of the accuracy of the proposed algorithm with the least squares method and the classical adaptive method led to the conclusion that it is expedient to use wavelet analysis when estimating unknown parameters of regression models.

Keywords: regression equation, maximum likelihood method, distribution density function, estimation of distribution density function, parameter estimation, wavelet analysis, wavelet estimation, Morlet wavelet, DOG wavelet, Littlewood Paley wavelet, «Mexican Hat» wavelet

Методы теоретической и прикладной статистики данных нашли широкое применение в различных сферах, среди которых можно выделить оптимизацию сложных технологических процессов, сертификацию технических систем и изделий, геофизические исследования грунтовых оснований для получения информации о распростране-

нении талых зон на вечномерзлых грунтах с целью обеспечения устойчивости зданий и сооружений. Возникающие в реальной жизни задачи нередко заставляют исследователя сталкиваться с необходимостью поиска зависимости между входными данными, задающими условия функционирования, и выходными данными, которые

характеризуют изучаемый объект. Решение таких задач может быть выполнено путем построения регрессионных моделей, и один из этапов состоит в оценивании неизвестных параметров. Такой подход позволяет не просто восстановить исходную зависимость, но и выполнять прогнозирование поведения изучаемого объекта.

Классические методы оценивания неизвестных параметров регрессионных зависимостей позволяют получать достаточно корректные и качественные результаты только при условии, что имеются достоверные предположения о свойствах случайной компоненты [1]. Одним из таких способов определения неизвестных оценок является метод максимального правдоподобия. Применение данного метода возможно при условии, что имеется достоверная информация о виде распределения случайных ошибок наблюдения [1]. Предположение о нормальности распределения ошибок наблюдения позволяет применить метод наименьших квадратов и тем самым упростить поиск оценок. Но на практике в большинстве случаев распределение случайной ошибки нельзя считать нормальным. В связи с этим целью данной работы является разработка универсального адаптивного алгоритма оценивания параметров регрессионных моделей, позволяющего получать корректные результаты во многих практически реализуемых задачах. Предложенный авторами подход основан на вейвлет-анализе [2, 3], который широко используется для аппроксимации различных функций [4, 5]. Применение различных материнских вейвлетов позволило создать целый ряд новых адаптивных алгоритмов оценивания параметров регрессионных зависимостей, обеспечивающих получение качественных результатов для большого числа реальных ситуаций, в том числе при отсутствии известных предположений о свойствах случайной компоненты.

*Постановка задачи
и основные предположения*

Пусть регрессионное уравнение принимает вид

$$y = X\theta + \varepsilon, \quad (1)$$

где $X = \begin{bmatrix} \varphi_1(x_{11}) & \dots & \varphi_q(x_{1q}) \\ \dots & \dots & \dots \\ \varphi_1(x_{n1}) & \dots & \varphi_q(x_{nq}) \end{bmatrix}$ – матрица

значений функций регрессионной модели, $rg(X) = q$; $\theta = (\theta_1, \theta_2, \dots, \theta_q)$ – оцениваемые неизвестные параметры модели;

$\varphi_i(x)$ – известные вещественные функции; x_{ij} – значения входных факторов в n наблюдениях; $y = (y_1, y_2, \dots, y_n)^T$ – вектор значений отклика; $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T$ – вектор независимых случайных ошибок, q – число неизвестных параметров; n – число проведенных экспериментов.

Считаем, что вектор ошибок наблюдений ε_i состоит из независимых одинаково распределенных случайных величин с функцией плотности $h(t)$, для которых выполняется

$$E(\varepsilon_i) = 0, D(\varepsilon_i) = \sigma^2 < \infty.$$

Требуется, по имеющимся значениям отклика и входных данных, выполнить наиболее точное оценивание вектора неизвестных параметров регрессионного уравнения (1).

Оценка функции плотности распределения с помощью вейвлет-анализа

Пусть на произвольном отрезке $[c, d]$ задана выборка $\{x_j \mid j = 1, n\}$, состоящая из независимых значений случайной величины

ζ , причем $\begin{cases} c = \min x_j \\ d = \max x_j \end{cases}$. Закон и функция

плотности распределения $h(t)$ случайной величины ζ считаются неизвестными. Тогда согласно [4, 5] оценка функции плотности $\hat{h}_n(t)$ случайной величины ζ может быть представлена в виде ряда

$$\hat{h}_n(t) = \sum_{i=1}^N \hat{c}_i \psi_i(t), \quad (2)$$

где N – параметр сглаживания (количество членов ряда), \hat{c}_i – коэффициенты разложения по ортонормированным базисным функциям $\psi_i(t)$, выражющиеся в следующем виде:

$$\hat{c}_i = \frac{1}{n} \sum_{j=1}^n \psi_i(x_j). \quad (3)$$

Принимая во внимание соотношение (3), запишем выражение для $\hat{h}_n(t)$:

$$\hat{h}_n(t) = \frac{1}{n} \sum_{j=1}^n W_N(t, x_j), \quad (4)$$

где $W_N(t, x_j) = \sum_{i=1}^N \psi_i(t) \psi_i(x_j)$.

Согласно [4, 5] ортонормированная система функций $\psi_i(t)$ определена на $[0, 1]$ и выражается соотношением

$$\psi_i(t) = 2^{k/2} \psi(2^k t - (j-1)), \quad (5)$$

где $k \geq 0$, $1 \leq j \leq 2^k$, $i = 2^k + j$, $\psi_i(t)$ – материнский вейвлет [4, 5]. Выражение (5) вы-

полняется, если $i > 1$. В случае $i = 1$ получаем, что $\psi_1(t) = \begin{cases} 1, & t \in [0,1] \\ 0, & t \notin [0,1] \end{cases}$.

В таблице представлены примеры различных материнских вейвлетов.

Примеры материнских вейвлетов

Вейвлет-функция	Аналитическая запись
«Мексиканская шляпа»	$\psi(t) = (1-t^2)e^{-t^2/2}$
DOG-вейвлет	$\psi(t) = e^{-t^2/2} - \frac{1}{2}e^{-t^2/8}$
LITTLEWOOD & PALEY	$\psi(t) = \frac{\sin(2\pi t) - \sin(\pi t)}{\pi t}$
Морле	$\psi(t) = e^{-t^2/2} \cos 5t$

Отметим, что необходимо перейти от ортонормированной на $[0,1]$ системы базисных функций $\psi_i(t)$ к системе функций $\tilde{\psi}_i(t)$ ортонормированной на отрезке, который совпадает с областью определения случайной величины ξ . Переход от системы $\psi_i(t)$ к системе $\tilde{\psi}_i(t)$ согласно [6, 7] выполняется с помощью следующего соотношения:

$$\tilde{\psi}_i(t) = \frac{1}{\sqrt{d-c}} \psi_i\left(\frac{t-c}{d-c}\right). \quad (6)$$

Тогда вейвлет-оценка функции плотности $\hat{h}_n(t)$ случайной величины ξ определяется соотношением (4) по ортонормированным базисным функциям $\tilde{\psi}_i(t)$.

Ранее авторами было исследовано качество восстановления функции плотности распределения на основе базисных вейвлетов, приведенных в таблице. С некоторыми результатами исследования можно познакомиться в [6–9]. Выводы, сделанные в результате вычислительных экспериментов, позволили предположить возможность применения вейвлет-оценок функции плотности при оценивании неизвестных параметров регрессионных моделей.

Алгоритм оценивания параметров регрессионных моделей

Рассмотрим задачу оценивания параметров регрессионного уравнения (1), для решения которой воспользуемся методом максимального правдоподобия [10]. Будем считать, что уравнение регрессии (1) является истинным. Значения остатков $e_i = y_i - X_i \hat{\theta}$ (X_i – i -я строка матрицы X) являются независимыми случайными величинами с плотностью распределения $h(e_i, \theta)$, что следует из предположения о независимо-

сти случайных ошибок. Тогда логарифмическая функция правдоподобия принимает вид

$$\ln L(e_1, \dots, e_n, \hat{\theta}) = \ln \left(\prod_{i=1}^n h(e_i, \hat{\theta}) \right) = \sum_{i=1}^n \ln(h(e_i, \hat{\theta})). \quad (7)$$

Приведем пошаговый алгоритм, который позволит выполнять оценивание параметров регрессионных моделей на основе вейвлетов:

Шаг 1. Определить начальное значение вектора неизвестных параметров $\hat{\theta}^0 = (\hat{\theta}_1^0, \hat{\theta}_2^0, \dots, \hat{\theta}_q^0)^T$ уравнения (1), $l = 0$, где l – номер итерации. В данном алгоритме предполагается использование метода наименьших квадратов [10] для определения начального приближения.

Шаг 2. Вычислить значения остатков $e_i = y_i - X_i \hat{\theta}$ регрессионной зависимости.

Шаг 3. Получить оценку функции плотности распределения с помощью соотношения (4) по ортонормированным базисным функциям (6).

Шаг 4. Определить значение логарифмической функции правдоподобия (7).

Шаг 5. Найти значение оценки вектора неизвестных параметров

$$\hat{\theta}^{l+1} = \arg \max_{\theta} L(e_1, \dots, e_n, \hat{\theta}^l).$$

Шаг 6. Итерационный процесс завершается, если $\|\hat{\theta}^{l+1} - \hat{\theta}^l\| < \delta$, где δ – заданная погрешность вычисления. Если же $\|\hat{\theta}^{l+1} - \hat{\theta}^l\| > \delta$, то выполняется переход на шаг 2 при условии $l = l + 1$.

Оценивание параметров регрессионных моделей с использованием вейвлетов

Поиск решения для задачи оценивания параметров регрессионной зависимости (1) может быть выполнен с помощью вычислительной схемы алгоритма, предложенного выше, где логарифмическая функция правдоподобия выражается соотношением (7) с использованием различных материнских вейвлетов, представленных в таблице.

Ранее на основе вейвлета LITTLEWOOD & PALEY (таблица) авторами была получена система функций $\tilde{\psi}_i(t)$ [8], ортонормированная на произвольном отрезке $[c, d]$, которая выражается следующим соотношением:

$$\tilde{\psi}_i(t) = \frac{2^{k/2} (\sin 2\pi\tau - \sin \pi\tau)}{\pi\tau\sqrt{d-c}}, \quad (8)$$

где $\tau = \frac{2^k}{d-c} (t-c) - (j-1)$, а значения i, k, j определяются как в (5).

Вейвлет-оценка функции плотности $\hat{h}_n(t)$ с использованием вейвлета LITTLEWOOD & PALEY определяется разложением (4) по ортонормированым базисным функциям (8). Принимая во внимание (8) и тот факт, что ошибки наблюдений независимы, функцию правдоподобия (7) можно записать в следующем виде:

$$L(e_1, \dots, e_n, \hat{\theta}) = \frac{1}{n^n} \prod_{s=1}^n \sum_{j=1}^n \sum_{i=1}^N \frac{2^k}{d-c} \cdot \frac{\sin(2\pi\tau_s) - \sin(\pi\tau_s)}{\pi\tau_s} \frac{\sin(2\pi\tau_j) - \sin(\pi_j\tau)}{\pi\tau_j}, \quad (9)$$

где $\tau_s = \frac{2^k}{d-c}(y_s - X_s \hat{\theta} - c) - (j-1)$, $\tau_j = \frac{2^k}{d-c}(y_j - X_j \hat{\theta} - c) - (j-1)$, а i, k, j такие же, как в (5).

Логарифмическое представление (9) выражается соотношением

$$\begin{aligned} \ln(L(e_1, \dots, e_n, \hat{\theta})) = -n \ln n + \sum_{s=1}^n \ln \left(\sum_{j=1}^n \sum_{i=1}^N \frac{2^k}{d-c} \cdot \right. \\ \left. \cdot \frac{\sin(2\pi\tau_s) - \sin(\pi\tau_s)}{\pi\tau_s} \frac{\sin(2\pi\tau_j) - \sin(\pi_j\tau)}{\pi\tau_j} \right). \end{aligned} \quad (10)$$

В работе [9] с помощью вейвлета Морле (таблица) авторами была построена система функций $\tilde{\psi}_i(t)$, ортонормированная на произвольном отрезке $[c, d]$, которая представима в виде

$$\tilde{\psi}_i(\tau) = \frac{z}{\sqrt{d-c}} 2^{\frac{k}{2}} e^{-\frac{\tau}{2}} \cos(5\tau), \quad (11)$$

где $\tau = \frac{2^k}{d-c}(t-c) - (j-1)$, а значения i, k, j такие же, как в (5).

Параметр $z = \frac{\sqrt{2}}{\sqrt[4]{\pi}} (1 + e^{-25})^{-\frac{1}{2}}$ и получен авторами в работе [8].

Соотношение (4), где базисные функции определяются выражением (11), является оценкой функции плотности $\hat{h}_n(t)$ случайной величины на основе вейвлета Морле. Тогда используя ортонормированную систему (11) и учитывая, что ошибки наблюдений независимы, функцию правдоподобия (7) можно записать в следующем виде:

$$L(e_1, \dots, e_n, \hat{\theta}) = \frac{1}{n^n} \prod_{s=1}^n \sum_{j=1}^n \sum_{i=1}^N \frac{z^2 2^k}{d-c} e^{\frac{-\tau_s^2}{2}} \cos(5\tau_s) e^{\frac{-\tau_j^2}{2}} \cos(5\tau_j), \quad (12)$$

где $\tau_s = \frac{2^k}{d-c}(y_s - X_s \hat{\theta} - c) - (j-1)$, $\tau_j = \frac{2^k}{d-c}(y_j - X_j \hat{\theta} - c) - (j-1)$, а i, k, j такие же,

как в (5), а z из (11).

Логарифмическое представление (12) выражается соотношением

$$\ln(L(e_1, \dots, e_n, \hat{\theta})) = -n \ln n + \sum_{s=1}^n \ln \left(\sum_{j=1}^n \sum_{i=1}^N \frac{z^2 2^k}{d-c} e^{\frac{-\tau_s^2}{2}} \cos(5\tau_s) e^{\frac{-\tau_j^2}{2}} \cos(5\tau_j) \right). \quad (13)$$

Ортонормированная система функций $\tilde{\psi}_i(t)$ на произвольном отрезке $[c, d]$ с использованием DOG вейвлета (таблица) получена авторами ранее в работе [7] и выражается соотношением

$$\tilde{\psi}_i(t) = \frac{z}{\sqrt{d-c}} 2^{k/2} \left(e^{-\tau^2/2} - \frac{1}{2} e^{-\tau^2/8} \right), \quad (14)$$

где $\tau = \frac{2^k}{d-c}(t-c) - (j-1)$, а значения i, k, j такие же, как в (5).

Значение параметра получено авторами в работе [6] и определяется соотношением

$$z = \left(\frac{\sqrt{\pi}}{n} (2 - \sqrt{2}) \sum_{j=1}^n \sum_{i=1}^N 2^{-k/2} \tilde{\psi}_i(x_j) \right)^{-1/2}. \quad (15)$$

Вейвлет-оценка функции плотности $\hat{h}_n(t)$ с использованием DOG вейвлета выражается разложением (4) по ортонормированным базисным функциям (14). Функция правдоподобия (7) на основе DOG вейвлета принимает вид

$$L(e_1, \dots, e_n, \hat{\theta}) = \frac{1}{n^n} \prod_{s=1}^n \sum_{j=1}^n \sum_{i=1}^N \frac{z^2 2^k}{d-c} \left(e^{\frac{-\tau_s^2}{2}} - \frac{1}{2} e^{\frac{-\tau_s^2}{8}} \right) \left(e^{\frac{-\tau_j^2}{2}} - \frac{1}{2} e^{\frac{-\tau_j^2}{8}} \right), \quad (16)$$

где $\tau_s = \frac{2^k}{d-c} (y_s - X_s \hat{\theta} - c) - (j-1)$, $\tau_j = \frac{2^k}{d-c} (y_j - X_j \hat{\theta} - c) - (j-1)$, а i, k, j такие же, как в (5).

Логарифмическое представление (16) выражается соотношением

$$\ln(L(e_1, \dots, e_n, \hat{\theta})) = -n \ln n + \sum_{s=1}^n \ln \left(\sum_{j=1}^n \sum_{i=1}^N \frac{z^2 2^k}{d-c} \left(e^{\frac{-\tau_s^2}{2}} - \frac{1}{2} e^{\frac{-\tau_s^2}{8}} \right) \left(e^{\frac{-\tau_j^2}{2}} - \frac{1}{2} e^{\frac{-\tau_j^2}{8}} \right) \right). \quad (17)$$

Ортонормированная система функций $\tilde{\psi}_i(t)$ на произвольном отрезке $[c, d]$ с использованием вейвлета «Мексиканская шляпа» (таблица) была получена авторами в работе [6] и выражается соотношением

$$\tilde{\psi}_i(\tau) = \frac{z}{\sqrt{d-c}} 2^{k/2} (1 - \tau^2) e^{-\tau^2/2}, \quad (18)$$

где $\tau = \frac{2^k}{d-c} (t - c) - (j-1)$, а значения i, k, j определяются как в (5). Значение параметра z выражается соотношением (15).

Разложение (4) по ортонормированным базисным функциям (18) определяет вейвлет-оценку функции плотности $h_n(t)$ случайной величины с помощью вейвлета «Мексиканская шляпа». Функция правдоподобия (7), построенная с учетом соотношения (18) и того, что ошибки наблюдений независимы, принимает вид

$$L(e_1, \dots, e_n, \hat{\theta}) = \frac{1}{n^n} \prod_{s=1}^n \sum_{j=1}^n \sum_{i=1}^N \frac{z^2 2^k}{d-c} (1 - \tau_s^2) e^{\frac{-\tau_s^2}{2}} (1 - \tau_j^2) e^{\frac{-\tau_j^2}{2}}, \quad (19)$$

где $\tau_s = \frac{2^k}{d-c} (y_s - X_s \hat{\theta} - c) - (j-1)$, $\tau_j = \frac{2^k}{d-c} (y_j - X_j \hat{\theta} - c) - (j-1)$, а i, k, j такие же, как в (5), z из (15).

Логарифмическое представление (19) выражается соотношением

$$\ln(L(e_1, \dots, e_n, \hat{\theta})) = -n \ln n + \sum_{s=1}^n \ln \left(\sum_{j=1}^n \sum_{i=1}^N \frac{z^2 2^k}{d-c} (1 - \tau_s^2) e^{\frac{-\tau_s^2}{2}} (1 - \tau_j^2) e^{\frac{-\tau_j^2}{2}} \right). \quad (20)$$

Покажем, что представленная вычислительная схема алгоритма оценивания параметров регрессионных моделей, где логарифмическая функция правдоподобия выражается одним из предложенных выше соотношений, соответствующих различным вейвлетам, обеспечивает получение качественных результатов при различных условиях.

Результаты вычислительных экспериментов

Исследуем предложенный выше алгоритм оценивания неизвестных параметров θ регрессионной зависимости (1) на основе различных материнских вейвлетов с помощью методов статистического моделирования. Рассмотрим регрессионную зависимость, выражющуюся в следующем виде:

$$y = \theta_1 + \theta_2 x + \theta_3 x^2 + \varepsilon. \quad (21)$$

Отметим, что значения входных факторов x_{ij} регрессионного уравнения (21) определялись из отрезка $[-4, 4]$, количество неизвестных параметров равно 3, а их истинные значения: $\theta_1 = 4$, $\theta_2 = -9$, $\theta_3 = 2$. Значения вектора ошибок наблюдений ε моделировались независимыми с функцией распределения вида

$$F(x) = (1 - \mu)F_1(x, m_1, \sigma_1) + \mu F_2(x, m_2, \sigma_2), \quad (22)$$

где $F_i(x, m_i, \sigma_i)$ – функция нормального распределения с математическим ожиданием m_i и дисперсией σ_i^2 ; $i = 1, 2$, $\mu \in [0, 1]$ – параметр смеси. При выполнении вычислительных экспериментов будем считать $m_1 = m_2 = 0$.

Отметим, что моделирование ошибки выполняется при различной степени отклонения от нормального распределения. Доли наблюдений с дисперсиями σ_1^2 и σ_2^2 в выборке определяются с помощью параметра μ , где при $\mu = 0$ и $\mu = 1$ распределение ошибки будет нормальным. В процессе моделирования предполагалось, что $\sigma_2^2 \geq \sigma_1^2$, а значения дисперсий σ_1^2 и σ_2^2 задавались через значения уровня шума [11], который определяется соотношением

$$\rho = \frac{\delta}{\delta_c} \cdot 100\%, \quad (23)$$

где δ – дисперсия ошибки,

$$\delta_c^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 \text{ – интенсивность сигнала.}$$

Определение точности оценивания параметров выполнялось с помощью L_1 нормы отклонений оценок неизвестных параметров от истинных значений

$$\nu_1 = \frac{1}{R} \sum_{i=1}^R \sum_{j=1}^3 \frac{|\theta_j - \hat{\theta}_{ij}|}{|\theta_j|}, \quad (24)$$

где R – число выполненных вычислительных экспериментов, $\hat{\theta}_{ij}$ – оценка j -го параметра регрессионной зависимости (21) в i -ом вычислительном эксперименте.

Количество вычислительных экспериментов для различных комбинаций μ и ρ варьировалось от 100 до 500. Варьирование значения параметра смеси μ в диапазоне от 0 до 0,5 с шагом 0,05 позволило исследовать точность оценивания неизвестных параметров уравнения (21) при разной степени отклонения распределения случайной ошибки от нормального распределения. Моделирование исходных данных выполнялось в соответствии с регрессионной зависимостью (21), а оценивание неизвестных параметров этой модели разработанным алгоритмом на основе различных материнских вейвлетов. Кроме того, вычисление оценок выполнялось методом наименьших квадратов и методом полупараметрического восстановления функции плотности на основе обобщенного лямбда-распределения [12]. Точность оценивания определялась значением показателя ν_1 .

Результаты исследования точности оценивания неизвестных параметров уравнения (21) представлены на рис. 1 для объема выборки 200 и на рис. 2 для объема выборки 500, при фиксированном уровне шума $\rho_1 = 10\%$ и $\rho_2 = 100\%$. Вычисление нормы отклонения ν_1 проводилось с усреднением по 100 вычислительным экспериментам.

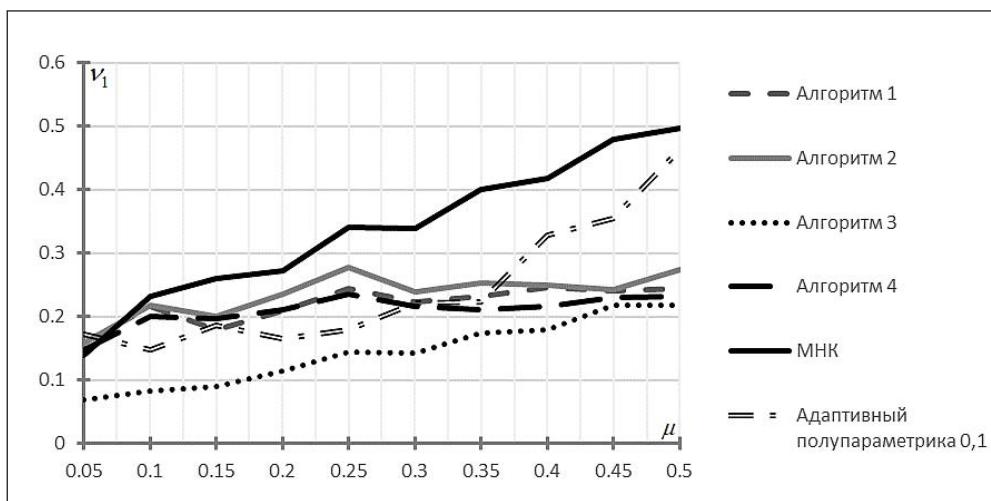
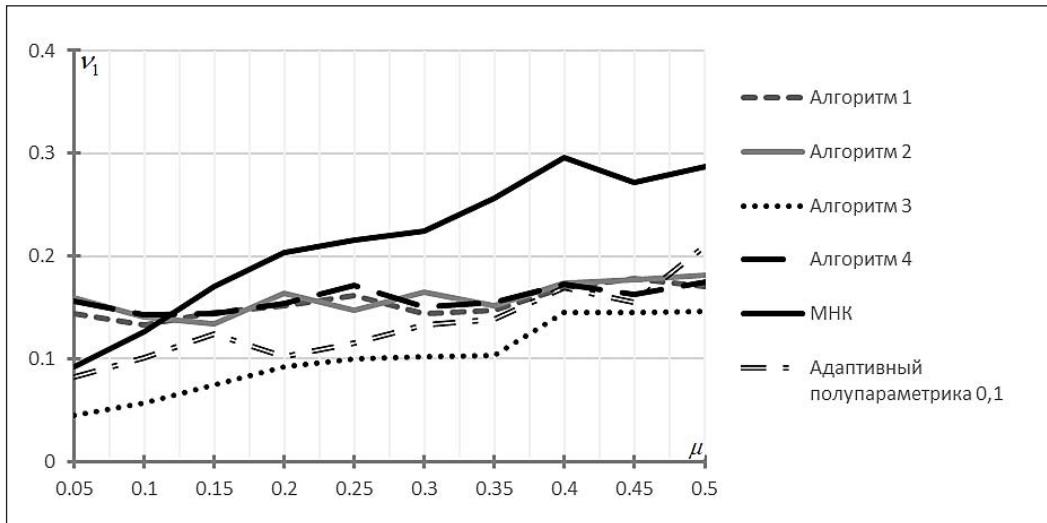


Рис. 1. Значение нормы отклонений в зависимости от μ , объем выборки – 200

Рис. 2. Значение нормы отклонений в зависимости от μ , объем выборки – 500

Для удобства обозначим адаптивный алгоритм на основе вейвлета LITTLEWOOD &PALEY как алгоритм 1, где наилучшее значение параметра сглаживания установлено в [8] и равно $N = 5$; адаптивный алгоритм на основе вейвлета Морле – как алгоритм 2, где наилучшее значение параметра сглаживания установлено в [9] и равно $N = 10$; адаптивный алгоритм на основе вейвлета DOG – как алгоритм 3, где наилучшее значение параметра сглаживания установлено в [7] и равно $N = 34$; адаптивный алгоритм на основе вейвлета «Мексиканская шляпа» – как алгоритм 4, где наилучшее значение параметра сглаживания установлено в [6] и равно $N = 8$.

Из рис. 1 и 2 видно, что при значениях $\mu \in [0.05, 0.1]$, алгоритмы 1, 2, 4, оценивания неизвестных параметров регрессионной зависимости, основанные на вейвлет-анализе, имеют не очень высокую точность, несколько уступая методу наименьших квадратов и методу полупараметрического восстановления функции плотности на основе обобщенного лямбда-распределения. При $\mu = 0.05$ и объеме выборки 500 значение показателя ν_1 , полученное в результате применения метода наименьших квадратов, на 40% меньше значений ν_1 , полученных при использовании алгоритмов 2 и 4, и на 36% меньше значения ν_1 , полученного при использовании алгоритма 1. Аналогичный результат при $\mu = 0.05$ и объеме выборки 500 наблюдается для значения ν_1 , полученного при использовании метода полупараметрического восстановления функции плотности на основе обобщенного лямбда-распределения, 47% и 42% со-

ответственно. Отметим, что для выборки объемом 200 при $\mu = 0.05$ значение показателей ν_1 , полученных в результате применения метода наименьших квадратов и метода полупараметрического восстановления функции плотности на основе обобщенного лямбда-распределения, меньше значений ν_1 , полученных при использовании алгоритмов 1, 2 и 4 (рис. 1). С ростом значений μ , то есть с ростом числа грубых ошибок наблюдений в исходных данных, алгоритмы на основе вейвлетов дают более точную оценку параметров регрессионной модели в сравнении с методом наименьших квадратов и методом полупараметрического восстановления функции плотности на основе обобщенного лямбда-распределения. При $\mu = 0.5$ и объеме выборки 500 нормы отклонений ν_1 , полученные при использовании алгоритмов 1 и 4, меньше на 43% нормы отклонения ν_1 , полученной при оценивании параметров методом наименьших квадратов, и на 24% меньше нормы отклонения ν_1 , полученной при применении метода полупараметрического восстановления функции плотности на основе обобщенного лямбда-распределения. Значение нормы отклонения ν_1 при $\mu = 0.5$, объеме выборки 500 и использовании алгоритма 2 на 39% меньше значения ν_1 , полученного при оценивании параметров регрессионной модели методом наименьших квадратов, и на 19% меньше нормы отклонения ν_1 , полученной при применении метода полупараметрического восстановления функции плотности на основе обобщенного лямбда-распределения. Для выборки объемом 200 при $\mu = 0.5$ значение показателей

v_1 , полученных при использовании алгоритмов 1, 2 и 4 меньше значений v_1 , полученных в результате применения метода наименьших квадратов и метода полупараметрического восстановления функции плотности на основе обобщенного лямбда-распределения (рис. 1). Наилучшее качество оценивания неизвестных параметров уравнения (21) получается в результате использования в алгоритме в качестве базисного вейвлета DOG вейвлета (алгоритм 1), причем как при малых значениях параметра μ , так и близким к 0.5. При $\mu = 0.05$ и объеме выборки 500 значение показателя v_1 , полученного при использовании алгоритма 3, на 44% меньше значения v_1 , полученного при оценивании параметров регрессионной модели методом наименьших квадратов, и на 38% меньше нормы отклонения v_1 , полученной при применении метода полупараметрического восстановления функции плотности на основе обобщенного лямбда-распределения. При $\mu = 0.5$ и объеме выборки 500 значение показателя v_1 , полученного при использовании алгоритма 3, на 46% меньше значения v_1 , полученного при оценивании параметров регрессионной модели методом наименьших квадратов, и на 29% меньше нормы отклонения v_1 , полученной при применении метода полупараметрического восстановления функции плотности на основе обобщенного лямбда-распределения. Отметим, что для выборки объемом 200 при $\mu = 0.05$ и при $\mu = 0.5$ алгоритм 1 имеет более высокую точность по сравнению с методом наименьших квадратов и методом полупараметрического восстановления функции плотности на основе обобщенного лямбда-распределения (рис. 1). Таким образом, алгоритмы оценивания, основанные на таких материнских вейвлетах, как Морле, LITTLEWOOD&PALEY, DOG и «Мексиканская шляпа», показывают хорошую точность оценивания как для выборок с объемом, равным 200, так для выборки, объем которой равен 500.

В данной статье рассмотрено применение вейвлет-анализа при оценивании неизвестных параметров регрессионных моделей с помощью метода максимального правдоподобия.

Предложены алгоритмы оценивания неизвестных параметров θ регрессионной зависимости, основанные на различных материнских вейвлетах. Для каждого базового вейвлета получено выражение логарифмической функции правдоподобия. Установлено, что такой подход дает ряд преимуществ при отсутствии известных предположений о свойствах случайной компоненты. Результаты численного моделирования подтвердили возможность применения теории вейвлетов для оценивания неизвестных параметров регрессионных моделей.

Список литературы

1. Ивченко Г.И., Медведев Ю.И. Математическая статистика. М.: КД Либроком, 2019. 352 с.
2. Hari Krishan Malhotra, Lalit Kumar Vashisht. On scaling functions of non-uniform multiresolution analysis in $L_2(R)$. International Journal of Wavelets, Multiresolution and Information Processing. 2020. Vol. 18. No. 02. 14 p.
3. Emmanuel de Dieu Nkou, Guy Martial Nkiet Wavelet-based estimation in a semiparametric regression model. International Journal of Wavelets, Multiresolution and Information Processing. 2021. 30 p.
4. Добеши И. Десять лекций по вейвлетам. Ижевск: НИЦ Регулярная и хаотическая динамика, 2001. 464 с.
5. Чуи К. Введение в вейвлеты. М.: Мир, 2001. 412 с.
6. Тимофеев В.С., Исаева Е.В. Об оценивании функции плотности распределения случайной величины с использованием вейвлетов // Научный вестник Новосибирского государственного технического университета. 2019. № 4 (77). С. 71–84.
7. Timofeev V.S., Isaeva E.V. Estimating the distribution density function using a DOG wavelet. Journal of Physics: Conference Series. 2020. Vol. 1661. 6 p.
8. Тимофеев В.С., Исаева Е.В., Малышкина Е.Д., Слободчикова А.Э. Оценивание функции плотности распределения с использованием вейвлета Литлвуда – Пэли [Электронный ресурс] // Обработка информации и математическое моделирование: материалы Российской научно-технической конференции. 2020. С. 165–170.
9. Исаева Е.В. Оценивание функции плотности распределения с использованием вейвлета Морле // Международный журнал прикладных и фундаментальных исследований. 2022. № 2. С. 22–27.
10. Дрейпер Н.Р., Смит Г. Прикладной регрессионный анализ. М.: Статистика, 1973. 392 с.
11. Ивахненко А.Г., Степашко В.С. Помехоустойчивость моделирования. Киев: Наукова думка, 1985. 216 с.
12. Денисов В.И., Тимофеев В.С., Хайленко Е.А. Полупараметрическое восстановление функции плотности на основе обобщенного лямбда-распределения в задаче идентификации регрессионных моделей // Сибирский журнал индустриальной математики. 2014. Т. 17. № 3. С. 71–77.