*ISSN 1814-1196*
*Научный вестник НГТУ*
*том 57, № 4, 2014, с. 69–78*

*http://journals.nstu.ru/vestnik*
*Science Bulletin of the NSTU*
*Vol. 57, No. 4, 2014, pp. 69–78*



УДК 519.213:519.23

# Оценивание параметров регрессионных моделей с использованием моментов, восстановленных на основе характеристической функции*

**ТИМОФЕЕВ В.С.**

*630073, РФ, г. Новосибирск, пр. Карла Маркса, 20, Новосибирский государственный технический университет, доктор технических наук, доцент. E-mail: v.timofeev@corp.nstu.ru*

В данной статье рассмотрена задача адаптивного оценивания параметров регрессионных моделей, решение которой проводится на основе техники максимально правдоподобного оценивания, а также одного из универсальных семейств распределений, а именно кривых Пирсона. Использование универсальных семейств распределений позволяет осуществлять восстановление регрессионных зависимостей, достаточно гибко подстраиваясь как к хорошо известным теоретическим распределениям, так и к очень широкому множеству практически реализуемых распределений. Для повышения устойчивости оценивания неизвестных параметров регрессионных моделей по отношению к грубым ошибкам наблюдения предложено осуществлять идентификацию кривых Пирсона на основе оценок моментов, вычисленных через эмпирическую характеристическую функцию. Представлена вычислительная схема нового алгоритма адаптивного оценивания неизвестных параметров регрессионных моделей. С помощью технологии статистического моделирования проведен ряд вычислительных экспериментов, направленных на исследование точности оценивания неизвестных параметров регрессионных моделей при различных условиях засорения исходных данных, а также разных объемах выборки. Показано, что при малом уровне засорения исходных данных грубыми ошибками наблюдений точность оценивания неизвестных параметров регрессионных моделей предложенным алгоритмом существенно повышается по сравнению с разработанным ранее алгоритмом, основанным на классических оценках моментов. С повышением объема выборки преимущество становилось более ощутимым. Кроме того, проведено сравнение точности оценивания неизвестных параметров регрессионных моделей представленным алгоритмом с одним из методов устойчивого оценивания, в качестве которого взят знаковый метод. По результатам всех проведенных исследований сделан ряд достаточно интересных выводов и даны рекомендации.

**Ключевые слова:** уравнение регрессии, оценивание параметров, выбросы, моменты случайной величины, характеристическая функция, универсальные распределения, метод максимального правдоподобия, кривые Пирсона

DOI: 10.17212/1814-1196-2014-4-69-78

## ВВЕДЕНИЕ

Опыт решения реальных задач, связанных с построением регрессионных моделей, показывает, что применение классических методов крайне редко позволяет получить статистически корректные выводы и результаты. Основная причина заключается в качестве исходных данных, которые, как правило, не соответствуют теоретическим (идеальным) предположениям, лежащим в основе классических методов [4]. В частности, распределение случайной ошибки в большинстве случаев нельзя считать нормальным. Кроме того, исходные данные могут содержать некоторое количество аномальных значений (выбросов). Формально наличие малого числа выбросов не противоречит упомянутым теоретическим предположениям, но имеет решающее влияние на качество получаемых результатов, оценок и выводов.

* *Статья получена 5 августа 2014 г.*

*Работа выполнена при финансовой поддержке Министерства образования и науки РФ по государственному заданию № 2014/138, проект № 1689.*

В связи с этим актуальной является задача создания универсальных алгоритмов построения регрессионных зависимостей, обеспечивающих получение корректных результатов для широкого спектра практически реализуемых ситуаций. На взгляд автора, перспективным является подход, основанный на использовании универсальных распределений, а именно кривых Пирсона, обобщенного лямбда-распределения, устойчивых распределений [3, 11, 13]. В составе этих семейств представлены многие хорошо известные законы распределения, такие как бета-распределение, гамма-распределение, распределение Стьюдента, нормальное распределение и др., что гарантирует автоматический переход к классическим результатам при появлении такой ситуации. Это позволило разработать ряд алгоритмов адаптивного оценивания параметров регрессионных моделей, обеспечивающих получение корректных результатов для большого числа практически реализуемых ситуаций, включая ситуации, характеризующиеся большой или бесконечной дисперсией случайной ошибки.

Однако переход к универсальным семействам распределения системно решает только одну из рассмотренных проблем – проблему отклонения фактически реализуемого распределения от нормального. Оценивание параметров регрессионных моделей предложенными методами в условиях засорения исходных данных единичными выбросами оказывается второстепенной задачей, которая в ряде случаев решается не очень хорошо. В связи с этим автором ставится задача повышения устойчивости оценок, полученных адаптивными методами. В настоящей работе представлено решение данной задачи на примере оценок, основанных на использовании кривых Пирсона. В дальнейшем задача может быть решена и для других универсальных семейств распределений, построение которых основывается на моментах.

### 1. ПОСТАНОВКА ЗАДАЧИ И ОСНОВНЫЕ ПРЕДПОЛОЖЕНИЯ

Рассмотрим регрессионное уравнение вида

$$y = X\theta + \varepsilon, \tag{1}$$

где $X = \begin{bmatrix} f_1(x_{11}) & \cdots & f_p(x_{1p}) \\ \cdots & \cdots & \cdots \\ f_1(x_{N1}) & \cdots & f_p(x_{Np}) \end{bmatrix}$ – матрица значений регрессионных функций, имеющая полный столбцовый ранг, т. е. $rg(X) = p$, $\theta = (\theta_1,...,\theta_p)^T$ – вектор неизвестных параметров, подлежащих оцениванию; $p$ – количество неизвестных параметров; $N$ – количество проведенных экспериментов; $f_i(x)$ – известные действительные функции вещественного аргумента $x$; $x_{ij}$ – заданные значения входных факторов в $N$ наблюдениях; $y = (y_1,...,y_N)^T$ – вектор значений отклика; $\varepsilon = (\varepsilon_1,...,\varepsilon_N)^T$ – вектор ошибок наблюдений.

Будем предполагать, что ошибки $\varepsilon_i$ наблюдений являются независимыми одинаково распределенными случайными величинами с унимодальной функцией плотности $\psi(x)$, для которых верно, что

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

Также будем предполагать, что существуют третий ($\mu_3$) и четвертый ($\mu_4$) центральные моменты данных случайных величин. Задача состоит в том, чтобы по имеющимся исходным данным (значениям отклика и всех входных факторов) как можно точнее оценить вектор неизвестных параметров уравнения регрессии (1).

**2. КРИВЫЕ ПИРСОНА И ОЦЕНКИ МОМЕНТОВ**

Введенное К. Пирсоном еще на рубеже XX века семейство кривых, по всей видимости, следует считать первым универсальным семейством распределений [8]. Оно состоит из 12 основных типов распределений*, полностью определяется первыми четырьмя моментами, а функция плотности $\psi(x)$ является решением следующего дифференциального уравнения:

$$\frac{d\psi(x)}{dx} = \frac{(x-a)\psi(x)}{b_0 + b_1 x + b_2 x^2},$$

где $a, b_0, b_1, b_2$ – неизвестные параметры, значения которых определяются на основе первых четырех начальных моментов ($m_1, m_2, m_3, m_4$) изучаемой случайной величины.

В настоящее время данное семейство не очень популярно у исследователей. Одна из причин этого обстоятельства состоит в необходимости использования выборочных моментов высоких порядков (до 4-го порядка включительно). Действительно, вычисление по классическим соотношениям сопряжено с накоплением вычислительных погрешностей. Тем не менее применение кривых Пирсона при построении алгоритма адаптивного оценивания параметров регрессионных моделей [11] позволило автору получить достаточно хорошие результаты [12]. При этом более высокая точность достигалась с использованием несмещенных оценок моментов [7]. Дальнейшее развитие этой идеи позволило автору перейти к построению устойчивых оценок моментов. В качестве инструмента для решения этой задачи выбрана характеристическая функция.

Хорошо известно [6, 10], что характеристическая функция $\varphi(t)$ некоторой случайной величины $\xi$ с плотностью $\psi(x)$ определяется следующим образом:

$$\varphi(t) = E\left[e^{itx}\right] = \int_{-\infty}^{\infty} e^{itx}\psi(x)\,dx, \tag{2}$$

где $t \in R$, $i = \sqrt{-1}$ – так называемая мнимая единица. Поскольку

$$\left|e^{itx}\right| = 1, \quad \forall t \in R,$$

то характеристическая функция существует для любой действительной случайной величины. Данная функция содержит всю информацию о распределении случайной величины и обладает целым рядом важных свойств [6, 10]. Непосредственно из определения характеристической функции (2) следует, что

$$\varphi(0) = 1, \quad \left|\varphi(t)\right| \le 1, \quad \varphi(-t) = \overline{\varphi(t)}.$$

На основе имеющейся реализации $x_1, \ldots, x_N$ случайной величины $\xi$ можно определить выборочную оценку характеристической функции [14]:

$$\hat{\varphi}(t) = \frac{1}{N}\sum_{j=1}^{N} e^{itx_j} = \frac{1}{N}\sum_{j=1}^{N}\left(\cos(tx_j) + i\sin(tx_j)\right). \tag{3}$$

Отметим, что в соответствии с законом больших чисел [1] оценка (3) состоятельна.

* Интересно, что на самом деле типов тринадцать, но последний, тринадцатый, тип – это нормальное распределение, которому К. Пирсон не придавал особого практического значения.

Переход от характеристической функции к функции плотности осуществляется посредством преобразования Фурье [9]:

$$\psi\left(x_j\right)=\frac{1}{2\pi}\int\limits_{-\infty}^{\infty}\varphi(t)\,e^{-itx_j}\,dt\,,\quad j=1,...,N\,. \tag{4}$$

Искомая непараметрическая оценка $\hat{\psi}\left(x\right)$ получается после замены $\varphi(t)$ в (4) на ее эмпирический аналог (3) и замены интеграла (4) конечной суммой.

Отметим также, что характеристическая функция $\varphi_\eta(t)$ случайной величины $\eta$, полученной в результате линейного преобразования $\eta=a_0+a_1\xi$ ($a_0, a_1$ – константы), связана с характеристической функцией $\varphi(t)$ случайной величины $\xi$:

$$\varphi_\eta(t)=e^{ita_0}\varphi(a_1t)\,.$$

Имеет место разложение характеристической функции в ряд по моментам. Если для случайной величины $\xi$ существуют начальные моменты $m_r$ до $\nu$-го порядка включительно, то они выражаются через $\varphi(t)$:

$$m_r=i^{-r}\varphi^{(r)}(0)\,,\quad r=1,...,\nu\,, \tag{5}$$

где $\varphi^{(r)}(t)$ – производная характеристической функции порядка $r$. Тогда имеет место разложение Маклорена [7, 10]:

$$\varphi(t)=1+\sum_{r=1}^{\nu}\frac{i^r m_r}{r!}t^r+R_\nu\,, \tag{6}$$

где $R_\nu$ – остаточный член. При необходимости можно записать разложение, аналогичное (4), но по центральным моментам [6, 10].

Использование соотношения (5) требует знания производных характеристической функции. При решении реальных задач вычислить характеристическую функцию по (2), как правило, не представляется возможным из-за отсутствия информации о виде распределения рассматриваемых случайных величин. Поэтому было осуществлено численное дифференцирование выборочной характеристической функции. Поскольку (3) есть комплекснозначная функция действительного аргумента, то ее можно записать в виде

$$\hat{\varphi}(t)=u(t)+iv(t)\,.$$

Следовательно, согласно [7] ее производная равна

$$\hat{\varphi}'(t)=u'(t)+iv'(t)\,. \tag{7}$$

Далее были использованы стандартные формулы численного дифференцирования [7] отдельно для действительной и мнимой частей характеристической функции. Значение шага дискретизации $h$ выбрано равным $\frac{\pi}{32}$. Приведем выражения для первых четырех производных действительной части, вычисленных при нулевом значении аргумента:

$$u'(0)=\frac{1}{2h}\left(u(h)-u(-h)\right), \tag{8}$$

$$u''(0) = \frac{1}{h^2}\left(u(h) - 2u(0) + u(-h)\right), \tag{9}$$

$$u^{(3)}(0) = \frac{1}{2h^3}\left(u(2h) - 2u(h) + 2u(-h) - u(-2h)\right), \tag{10}$$

$$u^{(4)}(0) = \frac{1}{h^4}\left(u(2h) - 4u(h) + 6u(0) - 4u(-h) + u(-2h)\right). \tag{11}$$

Производные мнимой части вычислялись аналогичным образом.

Автором проведен ряд вычислительных экспериментов, направленных на исследование точности восстановления моментов в различных условиях засорения. Некоторые из полученных результатов можно найти в [15]. В целом результаты подтвердили возможность проведения устойчивого оценивания моментов случайных величин с помощью характеристической функции. Это позволило использовать данную идею при оценивании неизвестных параметров регрессионных моделей.

### 3. АЛГОРИТМ ОЦЕНИВАНИЯ ПАРАМЕТРОВ РЕГРЕССИОННЫХ МОДЕЛЕЙ

Как и ранее [11, 12], при решении задачи оценивания параметров регрессионного уравнения (1) на основе универсальных семейств распределений (в данном случае кривых Пирсона) будет использован метод максимального правдоподобия. Применение в алгоритме устойчивых оценок моментов, основанных на характеристической функции, предположительно повысит устойчивость оценок параметров регрессионных моделей к наличию в исходных данных некоторой доли выбросов. В силу предположений о независимости случайных ошибок и истинности структуры рассматриваемого регрессионного уравнения (1) значения остатков $e_i = y_i - x_i\hat{\theta}$ ($x_i$ – $i$-я строка матрицы $X$ из (1)) также будут статистически независимыми случайными величинами с плотностью распределения $\psi(z_i, \theta)$. Тогда для оценивания параметров уравнения (1) можно воспользоваться методом максимального правдоподобия [4]. Учитывая тот факт, что остатки наблюдаемы, т. е. их значения определяются на основе имеющихся исходных данных, запишем логарифмическую функцию правдоподобия

$$l\left(e_1,...,e_N,\hat{\theta}\right) = \ln\left(\prod_{i=1}^{N}\psi\left(e_i,\hat{\theta}\right)\right) = \sum_{i=1}^{N}\ln\left(\psi\left(e_i,\hat{\theta}\right)\right). \tag{12}$$

Модифицированный итерационный алгоритм оценивания неизвестных параметров уравнения регрессии состоит в следующем.

Шаг 1. Определение начального приближения ($k := 0$) вектора неизвестных параметров уравнения (1), в качестве которого можно использовать оценку метода наименьших квадратов, что по сравнению с произвольным начальным приближением позволит сократить число итераций и время вычислений.

Шаг 2. Вычисление остатков регрессионного уравнения.

Шаг 3. Вычисление выборочной характеристической функции (3) и на ее основе оценок моментов (5) с использованием соотношений (7)–(11).

Шаг 4. Определение типа кривой Пирсона и осуществление идентификации распределения выбранного типа по соотношениям из [11].

Шаг 5. Вычисление значения логарифмической функции правдоподобия (12).

Шаг 6. Поиск очередного значения оценки неизвестных параметров $\hat{\theta}^{k+1}$:

$$\hat{\theta}^{k+1} = \arg\max_{\theta} l\left(e_1, e_2, ..., e_N, \hat{\theta}^k\right).$$

Шаг 7. Если $\left\|\hat{\theta}^{k+1}-\hat{\theta}^{k}\right\|<\varepsilon$, то происходит завершение процесса, в противном случае $k:=k+1$ и переход на шаг 2 ($\varepsilon$ – заданная погрешность вычисления).

#### 4. РЕЗУЛЬТАТЫ ВЫЧИСЛИТЕЛЬНЫХ ЭКСПЕРИМЕНТОВ

Для исследования предложенного алгоритма оценивания вектора неизвестных параметров $\theta$ уравнения (1) автором проводились многочисленные вычислительные эксперименты. Приведем лишь некоторые из полученных результатов. В качестве исследуемой зависимости рассмотрим следующее уравнение регрессии:

$$y=\theta_0+\theta_1 x+\theta_2 x^2+\varepsilon, \tag{13}$$

где количество регрессоров $p=3$; значения входных факторов $x_{ij}$ выбирались из отрезка $[-1,1]$; истинные значения неизвестных параметров: $\theta_0=50$, $\theta_1=25$, $\theta_2=10$. Случайные ошибки $\varepsilon_i$ моделировались независимыми и одинаково распределенными с функцией распределения вида

$$F(x)=(1-\lambda)F_1(x,m_1,\sigma_1)+\lambda F_2(x,m_2,\sigma_2), \tag{14}$$

где $F_i(x,m_i,\sigma_i)$ – функция нормального распределения с математическим ожиданием, равным $m_i$, и дисперсией $\sigma_i^2$; $i=1,2$, $\lambda\in[0,1]$ – параметр смеси. Во всех проведенных вычислительных экспериментах $m_1=m_2=0$.

Такое представление позволяет моделировать ошибку с различной степенью отклонения от нормального распределения, в том числе появление отдельных, довольно грубых засоряющих наблюдений – «выбросов». Параметр $\lambda$ определяет соответствующие доли наблюдений с дисперсиями $\sigma_1^2$ и $\sigma_2^2$ в выборке. Очевидно, что при $\lambda=0$ и $\lambda=1$ ошибка будет иметь нормальное распределение. В проведенных вычислительных экспериментах полагалось, что $\sigma_2^2\geq\sigma_1^2$. Однако при моделировании задавались не сами значения дисперсий $\sigma_1^2$ и $\sigma_2^2$, а им соответствующие значения уровня шума. Уровень шума введен в [5] и определяется как отношение шум/сигнал в %:

$$\rho=\frac{\sigma}{c}100,$$

где $\sigma$ – дисперсия ошибки; $c^2=\dfrac{1}{n-1}\sum_{i=1}^{n}(y_i^0-\bar{y}^0)^2$ – интенсивность сигнала (не зашумленных измерений $y_i^0$).

В качестве показателей точности оценивания параметров использовались $L_1$ нормы отклонений оценок неизвестных параметров от истинных значений

$$S_1=\frac{1}{T}\sum_{i=1}^{T}\left|\frac{\theta_i^{\text{ист}}-\hat{\theta}_i}{\theta_i^{\text{ист}}}\right| \text{ и } S_2=\frac{1}{T}\sum_{i=1}^{T}\left|\theta_i^{\text{ист}}-\hat{\theta}_i\right|,$$

где $T$ – число проведенных вычислительных экспериментов.

Для различных комбинаций $\lambda$ и $\rho$ проводилось по 500 вычислительных экспериментов. Каждый такой эксперимент заключался в моделировании выборки исходных данных в соответствии с моделью (13) и в последующем оценивании параметров этой модели разработанным

алгоритмом, основанным на использовании кривых Пирсона с классическими оценками моментов [11], а также методом наименьших квадратов (МНК) [4]. Кроме того, вычислялись оценки знаковым методом [2], который относится к методам устойчивого оценивания. В качестве итоговых показателей точности оценивания использовались усредненные по 500 проведенным вычислительным экспериментам значения показателей $S_1$ и $S_2$.

Рассмотрим результаты исследования точности оценивания неизвестных параметров уравнения (13) при разной степени отклонения распределения случайной ошибки от нормального распределения. Для этого изменению подвергался параметр смеси $\lambda$. При малых значениях $\lambda$ в выборке будет появляться небольшое число выбросов, а при значениях $\lambda$, близких к 0.5, можно говорить об изменении формы распределения. Было зафиксировано $\rho_1 = 5\,\%$, $\rho_2 = 50\,\%$, а доля выбросов $\lambda$ изменялась от 0 до 0.5 с шагом 0.02. Результаты оценивания представлены на рис. 1 и 2, причем на рис 1. показано изменение показателя $S_2$ для объема выборки 200 элементов, а на рис. 2 – изменение показателя $S_2$ для объема выборки 500 элементов.

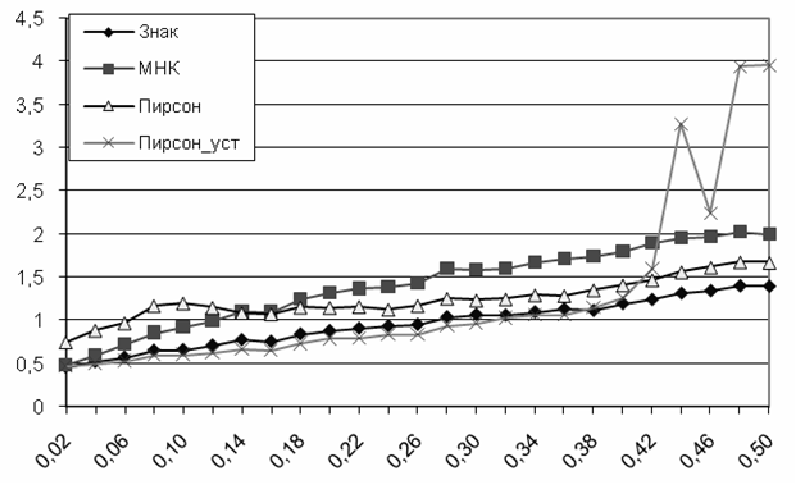


*Рис. 1.* Изменение показателя $S_2$ в зависимости от $\lambda$ ($N = 200$)

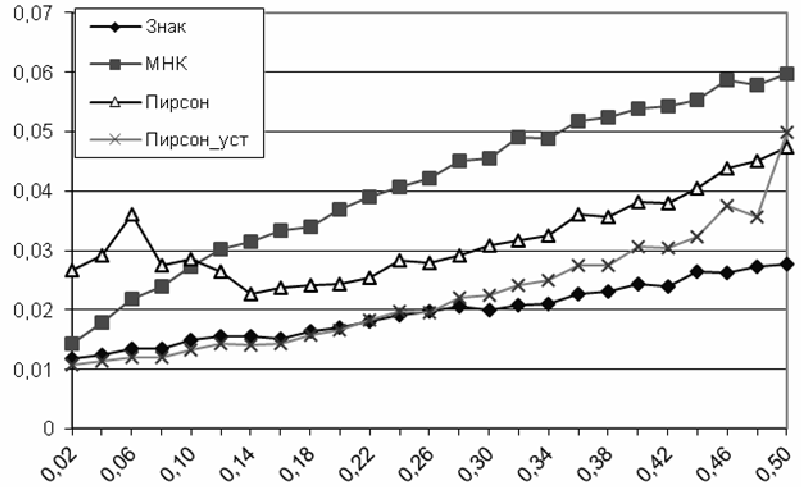


*Рис. 2.* Изменение показателя $S_1$ в зависимости от $\lambda$ ($N = 500$)

Из рисунков видно, что при малых значениях $\lambda$, т. е. при наличии в выборке небольшого числа выбросов, алгоритм адаптивного оценивания, основанный на классических оценках моментов, показывает не очень высокую точность, несколько уступая даже МНК. По мере увеличения в исходных данных доли грубых ошибок наблюдения начинает меняться форма фактически реализованного распределения случайной ошибки и алгоритм адаптивного оценивания, основанный на классических оценках моментов, начинает подстраиваться, что сразу приводит к увеличению точности оценивания по сравнению с МНК. Новый алгоритм, основанный на кривых Пирсона и устойчивых оценках моментов, при малых значениях $\lambda$ имеет очень хорошую точность оценивания, явно превосходящую МНК. Интересно, что и знаковый метод также немного уступает новому алгоритму. На больших засорениях ситуация меняется. При объеме выборки 200 элементов начиная с $\lambda = 0.42$ наблюдается резкое падение точности оценивания параметров регрессионного уравнения новым методом. Дело в том, что устойчивые оценки моментов, как и любые устойчивые методы, сокращают влияние грубых ошибок наблюдения, что в случае с изменением формы распределения приводит к потере некоторой доли полезной информации. Однако уже при объеме выборки 500 элементов падение точности становится не столь резким (рис. 2). Видимо, общий объем информации здесь уже достаточен для уверенной идентификации распределения, и отмеченные потери становятся не столь заметными. Это обстоятельство позволяет рекомендовать представленный в данной работе алгоритм к использованию на выборках с небольшим уровнем засорения выбросами, т. е. в качестве устойчивого метода оценивания. Его преимущество будет состоять не только в хорошей точности оценивания, но и в свойствах оценок, поскольку он основан на методе максимального правдоподобия.

## ЗАКЛЮЧЕНИЕ

В работе рассмотрена задача адаптивного оценивания параметров регрессионных зависимостей. Ее решение осуществляется на основе метода максимального правдоподобия с использованием универсальных семейств распределений, а именно кривых Пирсона. Для обеспечения более высокой устойчивости к наличию в исходных данных выбросов предложено использовать устойчивые оценки моментов, вычисление которых проведено на основе характеристической функции. Проведенные посредством вычислительных экспериментов исследования позволяют сделать вывод о возможности применения адаптивных методов к задаче устойчивого оценивания параметров регрессионных моделей. Это обстоятельство дает возможность рассматривать задачи устойчивого и адаптивного оценивания как единую группу задач, решение которых можно проводить с единых позиций.

## СПИСОК ЛИТЕРАТУРЫ

1. *Гнеденко Б.В.* Курс теории вероятностей. – М.: Едиториал УРСС, 2001. – 320 с.

2. *Денисов В.И., Тимофеев В.С.* Знаковый метод: преимущества, проблемы, алгоритмы // Научный вестник НГТУ. – 2001. – № 1 (10). – С. 21–35.

3. *Денисов В.И., Тимофеев В.С.* Устойчивые распределения и оценивание параметров регрессионных зависимостей // Известия Томского политехнического университета. – 2011. – Т. 318, № 2. – С. 10–15.

4. *Дрейпер Н.Р., Смит Г.* Прикладной регрессионный анализ: пер. с англ. – М.: Статистика, 1973. – 392 с.

5. *Ивахненко А.Г., Степашко В.С.* Помехоустойчивость моделирования. – Киев: Наукова думка, 1985. – 216 с.

6. *Кендалл М.Дж., Стьюарт А.* Теория распределений: пер. с англ. – М.: Наука, 1966. – 587 с.

7. *Корн Г.А., Корн Т.М.* Справочник по математике для научных работников и инженеров: пер. со 2 амер. перераб. изд. – М.: Наука, 1984. – 832 с.

8. *Митропольский А.К.* Техника статистических вычислений. – 2-е изд., перераб. и доп. – М.: Наука, 1971. – 576 с.

9. *Оппенгейм А.В., Шафер Р.В.* Цифровая обработка сигналов: пер. с англ. – М.: Связь, 1979. – 416 с.

10. *Пугачев В.С.* Теория вероятностей и математическая статистика. – М.: Наука, 1979. – 496 с.

11. *Тимофеев В.С.* Оценивание параметров регрессионных зависимостей с использованием кривых Пирсона. Ч. 1 // Научный вестник НГТУ. – 2009. – № 4 (37). – С. 57–66.

12. *Тимофеев В.С.* Оценивание параметров регрессионных зависимостей с использованием кривых Пирсона. Ч. 2 // Научный вестник НГТУ. – 2010. – № 1 (38). – С. 57–62.

13. *Тимофеев В.С., Хайленко Е.А.* Адаптивное оценивание параметров регрессионных моделей с использованием обобщенного лямбда-распределения // Доклады Академии наук высшей школы Российской Федерации. – 2010. – № 2 (15). – С. 25–36.

14. *Feuerverger A., Mureika R.A.* The empirical characteristic function and its applications // The Annals of Statistics. – 1977. – Vol. 5, N 1. – P. 88–97.

15. *Timofeev V.S.* Characteristic function in estimation of probability distribution moments [Electronic resource] // International Journal of Mathematical, Computational, Physical and Quantum Engineering. – 2014. – Vol. 8, N 8. – P. 1065–1067. – URL: http://waset.org/Publication/characteristic-function-in-estimation-of-probability-distribution-moments-/9999015 (accessed: 01.08.2014).

*Тимофеев Владимир Семенович*, доктор технических наук, профессор кафедры программных систем и баз данных Новосибирского государственного технического университета. Основное направление научных исследований – разработка и исследование устойчивых методов и алгоритмов анализа многофакторных объектов, в том числе с использованием непараметрической статистики. Имеет более 80 публикаций, в том числе один учебник. E-mail: v.timofeev@corp.nstu.ru

## *Estimation of regression model parameters using moments based on the characteristic function*[*]

*V.S. TIMOFEEV*

*Novosibirsk State Technical University, 20 K. Marks Prospekt, Novosibirsk, 630073, Russian Federation, D.Sc. (Eng.) associate professor. E-mail: v.timofeev@corp.nstu.ru*

The problem of adaptive estimation of the parameters of regression models is addressed in the paper. The solution of this problem is based on maximum likelihood estimation techniques and a universal distribution family, namely Pearson's curves. The use of universal distribution families allows building a relationship by flexible adjustment to both well-known theoretical distributions and to a very wide set of practically realizable distributions. To increase the estimation robustness of unknown parameters of regression models with respect to outliers it is proposed to carry Pearson's curves identification based on the estimates of the moments calculated by the empirical characteristic function. The computational scheme of a new algorithm of adaptive estimation of unknown parameters of regression models is presented. Using statistical modeling techniques, a number of computational experiments were designed to study the estimation accuracy of unknown parameters of regression models under different conditions of output variable contamination and different sample sizes. It is shown that at a low level of data contamination by outliers, the estimation accuracy of unknown parameters of regression models by using the proposed algorithm significantly increases compared to a previously developed algorithm based on classical moment estimates. With increasing the sample size the advantage becomes more tangible. Also, a comparison of the estimation accuracy of unknown parameters of regression models achieved by the proposed algorithm with one of the robust estimation methods, namely the sign method, is made. Based on the results of all the conducted studies a number of interesting conclusions are made and some recommendations are given.

**Keywords:** regression equation, parameter estimation, outlier, moments of random variable, characteristic function, universal distributions, maximum likelihood method, Pearson's curves

**REFERENCES**

1. Gnedenko B.V. *Kurs teorii veroyatnostei* [Course of the theory of probability]. Moscow, Editorial URSS Publ., 2001. 320 p.

2. Denisov V.I., Timofeev V.S. Znakovyi metod: preimushchestva, problemy, algoritmy [Sign methods: advantages, problems, algorithms]. *Nauchnyi vestnik NGTU – Science Bulletin of Novosibirsk State Technical University*, 2001, no. 1 (10), pp. 21–35.

3. Denisov V.I., Timofeev V.S. Ustoichivye raspredeleniya i otsenivanie parametrov regressionnykh zavisimostei [Stable distributions and parameter estimation of a regression]. *Izvestiya Tomskogo politekhnicheskogo universiteta – Bulletin of the Tomsk Polytechnic University,* 2011, vol. 318, no. 2, pp. 10–15.

4. Draper N.R., Smith H. *Applied regression analysis*. New York, John Wiley&Sons, 1966. 407 p. (Russ. ed.: Dreiper N.R., Smit G. *Prikladnoi regressionnyi analiz*. Moscow, Statistika Publ., 1973. 392 p.).

[*] *Received 5 August 2014.*

*This research has been supported by the Ministry of Education and Science of the Russian Federation as part of the state task № 2014/138 (project № 1689).*

5. Ivakhnenko A.G., Stepashko V.S. *Pomekhoustoichivost' modelirovaniya* [Noise immunity modeling]. Kiev, Naukova dumka Publ., 1985. 216 p.

6. Kendall M.G., Stuart A. *The advanced theory of statistics*. Vol. 1. *Distribution theory*. London, Ch. Griffin & Company, 1960. 590 p. (Russ. ed.: Kendall M.Dzh., St'yuart A. *Teoriya raspredelenii*. Moscow, Nauka Publ., 1966. 587 p.).

7. Korn G.A., Korn T.M. *Mathematical handbook for scientists and engineers: definitions, theorems and formulas* for reference and review. 2 nd enl. and rev. ed. New York, McGraw-Hill, 1968. xix, 1130 p. (Russ. ed.: Korn G.A., Korn T.M. *Spravochnik po matematike dlya nauchnykh rabotnikov i inzhenerov*. Moscow, Nauka Publ., 1984. 832 p.).

8. Mitropol`skii A.K. *Tekhnika statisticheskikh vychislenii* [The technique of statistical calculations]. 2nd ed., rev. and add. Moscow, Nauka Publ., 1971. 576 p.

9. Oppenheim A.V., Schafer R.W. *Digital Signal Processing*. New Jersey, Prentice Hall, 1975, 420 p. (Russ. ed.: Oppengeim A.V., Shafer R.V. *Tsifrovaya obrabotka signalov*. Moscow, Svyaz', 1979. 416 p.).

10. Pugachev V.S. *Teoriya veroyatnostei i matematicheskaya statistika* [Probability theory and mathematical statistics]. Moscow, Nauka Publ., 1979. 496 p.

11. Timofeev V.S. Otsenivanie parametrov regressionnykh zavisimostei s ispol`zovaniem krivykh Pirsona. Ch. 1 [The Pirson's curves in parameter estimation problem for regression model. Pt. 1]. *Nauchnyi vestnik NGTU – Science Bulletin of Novosibirsk State Technical University,* 2009, no. 4 (37), pp. 57–66.

12. Timofeev V.S. Otsenivanie parametrov regressionnykh zavisimostei s ispol`zovaniem krivykh pirsona. Ch. 2 [The Pirson's curves in parameter estimation problem for regression model. Pt. 2]. *Nauchnyi vestnik NGTU – Science Bulletin of Novosibirsk State Technical University,* 2010, no. 1 (38), pp. 57–62.

13. Timofeev V.S., Khailenko E.A. Adaptivnoe otsenivanie parametrov regressionnykh modelei s ispol'zovaniem obobshchennogo lyambda-raspredeleniya [Adaptive estimation of regression models parameters using generalized lambda-distribution]. *Doklady Akademii nauk vysshei shkoly Rossiiskoi Federatsii – Proceedings of the Russian higher school Academy of sciences,* 2010, no. 2 (15), pp. 25–36.

14. Feuerverger A., Mureika R.A. The empirical characteristic function and its applications. *The Annals of Statistics,* 1977, vol. 5, no. 1, pp. 88–97.

15. Timofeev V.S. Characteristic function in estimation of probability distribution moments. *International Journal of Mathematical, Computational, Physical and Quantum Engineering*, 2014, vol. 8, no. 8, pp. 1065–1067. Available at: http://waset.org/Publication/characteristic-function-in-estimation-of-probability-distribution-moments-/9999015 (accessed 01.08.2014).