ISSN 1814-1196
Научный вестник НГТУ
том 77, № 4, 2019, с. 71–84

http://journals.nstu.ru/vestnik
Science Bulletin of the NSTU
Vol. 77, No. 4, 2019, pp. 71–84

ИНФОРМАТИКА,
ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА
И УПРАВЛЕНИЕ

INFORMATICS,
COMPPUTER ENGINEERING
AND MANAGEMENT

УДК 519.254

DOI: 10.17212/1814-1196-2019-4-71-84

Об оценивании функции плотности распределения случайной величины с использованием вейвлетов*

В.С. ТИМОФЕЕВ[a], **Е.В. ИСАЕВА**[b]

630073, РФ, г. Новосибирск, пр. К. Маркса, 20, Новосибирский государственный технический университет

[a] *v.timofeev@corp.nstu.ru* [b] *isaeva@corp.nstu.ru*

Проблема оценивания плотности вероятности распределения возникает во многих областях статистического анализа данных. Существуют различные подходы к ее решению, но в последнее время становятся популярными методы, основанные на вейвлет-оценивании. Суть методов состоит в разложении неизвестной функции в ряд по некоторому конечному набору ортонормированных базисных функций. В качестве такого набора используется система функций, определенная на отрезке $[0,1]$, который не всегда совпадает с областью значений случайной величины. Поэтому для вычисления вейвлет-оценки плотности распределения случайной величины, заданной на произвольной области, требуется выполнить переход к системе функций, определенной на том же отрезке, что значения случайной величины. В рамках данной работы будут рассмотрены вейвлет-оценки плотности распределения случайной величины с помощью вейвлетов Хаара и «Мексиканская шляпа». Выполнено исследование влияния на качество вейвлет-оценок таких параметров, как объем выборки, число коэффициентов разложения функции в ряд в выражении для вейвлет-оценки плотности. В ходе работы установлено, что качество вейвлет-оценки существенно зависит от параметра сглаживания и существует ее наилучшее значение, которое, в свою очередь, меняется от выбора материнского вейвлета. Авторами предложен новый вариант оценки параметра сглаживания. Для количественной оценки степени близости функции плотности распределения и ее вейвлет-оценки была проведена проверка согласия по критерию χ^2 и сформулированы выводы о том, какой базовый вейвлет обеспечивает наиболее качественное восстановление функции плотности.

Ключевые слова: вейвлет-оценка, вейвлет-анализ, вейвлет Хаара, вейвлет «Мексиканская шляпа», функция плотности, оценка функции плотности, критерий χ^2, параметр сглаживания, вычислительный эксперимент

* *Статья получена 03 сентября 2019 г.*

ВВЕДЕНИЕ

Решение практических задач, связанных со статистическим анализом данных, как правило, предполагает наличие информации о виде распределения признаков стохастической природы. Так, проведение регрессионного анализа с использованием метода максимального правдоподобия невозможно без знания закона распределения случайной ошибки. Априорно такая информация бывает доступной крайне редко, но ее можно получать из имеющихся статистических данных [1, 2]. Одной из характеристик, содержащих полную информацию о распределении исследуемых признаков, является функция плотности. Оценка функции плотности по статистическим данным в условиях полного отсутствия информации о виде распределения может быть проведена с помощью непараметрических методов. Среди таких методов наибольшую известность получила ядерная оценка Розенблатта–Парзена [3], которая использовалась авторами при построении регрессионных зависимостей [4]. Однако ее качество сильно зависит от значения параметра сглаживания, неоптимальный выбор которого может приводить к резким флуктуациям графика восстанавливаемой функции плотности, что особенно критично на малых выборках. Это обстоятельство заставляет исследователей быть осторожными при использовании ядерных оценок.

В настоящей работе исследуется другой подход, который также следует отнести к непараметрическим. Он основан на применении теории вейвлетов, которая активно развивается в последнее время и широко используется для аппроксимации различных функций. Цель работы состоит в выявлении такого базового вейвлета, который бы обеспечивал наилучшее качество восстановления функции плотности в различных условиях, включая малые объемы выборок, и в построении универсального алгоритма оценки функции плотности.

1. ПОСТАНОВКА ЗАДАЧИ

Пусть $\left\{x_j \middle| j=\overline{1,n}\right\}$ представляет собой выборку независимых значений случайной величины ξ, заданных на произвольном отрезке $[c,d]$, где $\left[\begin{array}{l} c=\min x_j \\ d=\max x_j \end{array}\right.$. Закон, а также функция плотности распределения $f(t)$ случайной величины ξ являются неизвестными. Требуется построить вейвлет-оценку $\hat{f}_n(t)$ плотности распределения $f(t)$ с использованием выборочных данных.

2. ВЕЙВЛЕТ-ОЦЕНКИ ВЕРОЯТНОСТНЫХ ПЛОТНОСТЕЙ

Согласно [5–7] вейвлет-оценка функции плотности $\hat{f}_n(t)$ случайной величины выражается разложением в ряд по некоторым ортонормированным базисным функциям $\psi_i(t)$

$$\hat{f}_n(t)=\sum_{i=1}^{N}\hat{c}_i\psi_i(t)\,, \tag{1}$$

где N – число членов ряда (параметр сглаживания); $\hat{c}_i$ – оценки коэффициентов разложения по данному базису, которые определяются по имеющимся статистическим данным

$$\hat{c}_i = \frac{1}{n}\sum_{j=1}^{n} \psi_i(x_j) . \tag{2}$$

С учетом (2) вейвлет-оценку функции плотности можно записать в следующем виде

$$\hat{f}_n(t) = \frac{1}{n}\sum_{j=1}^{n} W_N(t, x_j) , \tag{3}$$

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

В качестве ортонормированной системы $\psi_i(t)$, согласно [2], рассмотрим набор ортонормированных на $[0,1]$ базисных функций:

$$\psi_i(t) = 2^{k/2}\psi(2^k t - (j-1)) , \tag{4}$$

где $k \ge 0$, $1 \le j \le 2^k$ такие, что $i = 2^k + j$, $\psi(t)$ – материнский вейвлет или вейвлет-функция. Выражение (4) справедливо для $i > 1$, в случае $i = 1$ функция $\psi_1(t) = 1$ для всех $t \in [0,1]$ и $\psi_1(t) = 0$ в противном случае. Примеры вейвлет-функций приведены в табл. 1 [8–12].

Таблица 1

Table 1

Вейвлет-функции

Wavelet functions

Вейвлет-функция	Аналитическая запись
WAVE-вейвлет (гауссов первого порядка)	$\psi(t) = -te^{-t^2/2}$
«Мексиканская шляпа» (гауссов второго порядка)	$\psi(t) = (1-t^2)e^{-t^2/2}$
Гауссов *n*-го порядка	$\psi(t) = (-1)^n \frac{d^n}{dt^n}\left(e^{-t^2/2}\right)$
DOG-вейвлет	$\psi(t) = e^{-t^2/2} - \frac{1}{2}e^{-t^2/8}$
Вейвлет Хаара	$\psi(t) = \begin{cases} 1, & t \in [0, 0.5) \\ -1, & t \in [0.5, 1), \\ 0, & t \notin (0, 1], \end{cases}$

Отметим, что рассмотренная выше ортонормированная система функций $\psi_i(t)$ задана на отрезке $[0,1]$, который может не совпадать с фактической областью значений ξ. Поэтому для вычисления $\hat{f}_n(t)$ требуется перейти к ортонормированной системе функций $\tilde{\psi}_i(t)$, определенной на том же отрезке $[c,d]$, что и значения случайной величины. Чтобы выполнить такой переход, воспользуемся соотношением [13, 14]

$$\tilde{\psi}_i(t) = \lambda_1 \psi_i(\lambda_2 t + \lambda_3), \tag{5}$$

где $\lambda_1, \lambda_2, \lambda_3 \in R$. Согласно определению, ортонормированная система функций $\tilde{\psi}_i(t)$ должна удовлетворять двум условиям:

$$(\tilde{\psi}_i(t),\ \tilde{\psi}_j(t)) = \int_c^d \tilde{\psi}_i(t)\tilde{\psi}_j(t)dt = 0 \text{ для всех } i \neq j;$$

$$\|\tilde{\psi}_i(t)\| = \sqrt{(\tilde{\psi}_i(t),\ \tilde{\psi}_i(t))} = \sqrt{\int_c^d \tilde{\psi}_i(t)\tilde{\psi}_i(t)dt} = 1.$$

Из этих условий и с учетом (5) построим систему уравнений для нахождения коэффициентов $\lambda_1, \lambda_2, \lambda_3$:

$$\begin{cases} \lambda_2 c + \lambda_3 = 0, \\ \lambda_2 d + \lambda_3 = 1, \\ \lambda_1 = \sqrt{\lambda_2}. \end{cases} \tag{6}$$

Решая систему (6), получаем

$$\lambda_1 = \frac{1}{\sqrt{d-c}}, \qquad \lambda_2 = \frac{1}{d-c}, \qquad \lambda_3 = \frac{-c}{d-c}. \tag{7}$$

Подставляя (7) в (5), получаем преобразование базисных функций при переходе от одной системы функций к другой:

$$\tilde{\psi}_i(t) = \frac{1}{\sqrt{d-c}} \psi_i\left(\frac{t-c}{d-c}\right). \tag{8}$$

Итоговая вейвлет-оценка функции плотности $\hat{f}_n(t)$ случайной величины на произвольном отрезке $[c,d]$ выражается разложением (3), где вместо базисных функций $\psi_i(t)$ используется система функций $\tilde{\psi}_i(t)$.

Далее в работе будет более подробно рассмотрена и исследована оценка (3), построенная на основе материнских вейвлетов Хаара и «Мексиканская шляпа».

3. ОЦЕНКА ПЛОТНОСТИ РАСПРЕДЕЛЕНИЯ С ПОМОЩЬЮ ВЕЙВЛЕТА ХААРА

Рассмотрим ортогональный вейвлет Хаара [15–19], определяемый соотношением из табл. 1, для которого ортонормированная система функций (4) на отрезке $[0,1]$ (система Хаара) принимает вид

$$\psi_i(t) = \begin{cases} 2^{k/2}, & t \in \left[2^{-k}(j-1),\ 2^{-(k+1)}(2j-1)\right), \\ -2^{k/2}, & t \in \left[2^{-(k+1)}(2j-1),\ 2^{-k}j\right), \\ 0, & t \notin \left(2^{-k}(j-1),\ 2^{-k}j\right], \end{cases} \qquad (9)$$

где i, k, j такие же, как в (4).

Переход от системы функций (9) к ортонормированной системе $\tilde{\psi}_i(t)$ на произвольном отрезке $[c,\ d]$ с учетом (8) дает результат

$$\tilde{\psi}_i(t) = \frac{1}{\sqrt{d-c}} \begin{cases} 2^{k/2}, & t \in \left[2^{-k}(j-1)(d-c)+c,\ 2^{-(k+1)}(2j-1)(d-c)+c\right), \\ -2^{k/2}, & t \in \left[2^{-(k+1)}(2j-1)(d-c)+c,\ 2^{-k}j(d-c)+c\right), \\ 0, & t \notin \left(2^{-k}(j-1)(d-c)+c,\ 2^{-k}j(d-c)+c\right], \end{cases} \qquad (10)$$

где i, k, j такие же, как в (4).

В результате вейвлет-оценка функции плотности $\hat{f}_n(t)$ случайной величины на произвольном отрезке $[c,d]$ с использованием материнского вейвлета Хаара выражается разложением (3) по ортонормированным базисным функциям (10).

4. ОЦЕНКА ПЛОТНОСТИ РАСПРЕДЕЛЕНИЯ С ПОМОЩЬЮ ВЕЙВЛЕТА «МЕКСИКАНСКАЯ ШЛЯПА»

Материнский вейвлет «Мексиканская шляпа» [18–20] определяется соотношением из табл. 1. В этом случае ортонормированная система функций (4) на отрезке $[0,1]$ принимает вид

$$\psi_i(t) = z2^{k/2}\left(1-(2^k t-(j-1))^2\right)e^{-(2^k t-(j-1))^2}, \qquad (11)$$

где i, k, j такие же, как в (4). Множитель $z = \frac{1{,}031}{\sqrt{2}}$ находится из условий определения ортонормированной системы функций [18].

Переход от системы функций (11) к ортонормированной системе $\tilde{\psi}_i(t)$ на произвольном интервале $[c,d]$ с учетом (8) дает результат

$$\tilde{\psi}_i(t) = \frac{1}{\sqrt{d-c}} 2^{k/2} z(1-\tau^2)e^{-\tau^2/2}, \qquad (12)$$

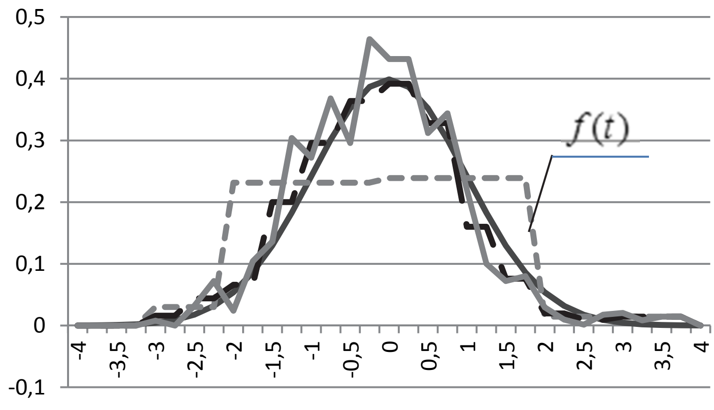
где $\tau = \frac{2^k}{d-c}(t-c)-(j-1)$, а значения i, k, j такие же, как в (4).

В результате вейвлет-оценка функции плотности $\hat{f}_n(t)$ случайной величины на произвольном отрезке $[c,d]$ с использованием материнского вейвлета «Мексиканская шляпа» выражается разложением (3) по ортонормированным базисным функциям (12).

5. ИССЛЕДОВАНИЕ ВЕЙВЛЕТ-ОЦЕНКИ ПЛОТНОСТИ РАСПРЕДЕЛЕНИЯ

Для сравнения качества восстановления функции плотности с использованием вейвлетов Хаара и «Мексиканская шляпа» был проведен ряд вычислительных экспериментов. Исследования проводились при разных условиях формирования выборки и различных значениях параметра сглаживания.

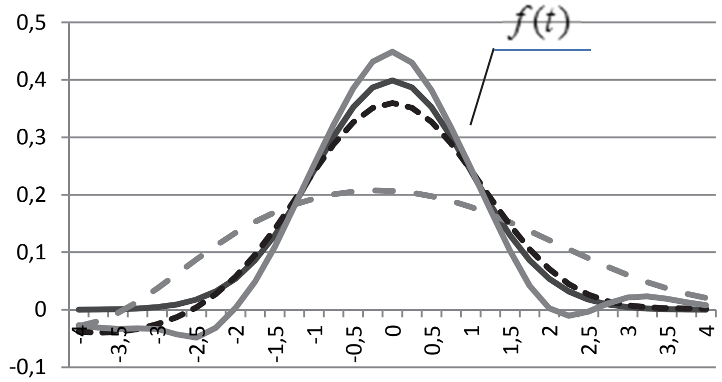
Прежде всего исследуем качество восстановления функции плотности стандартного нормального распределения $\hat{f}_n(t)$ от количества членов N ряда (3). Для этого смоделируем выборку объема $n = 1000$, состоящую из независимых значений случайной величины ξ, распределенной по нормальному закону с параметрами $(0,1)$. Построим для данной выборки при каждом значении N, начиная с 5 и до 50, вейвлет-оценки функции плотности (3) с помощью вейвлета Хаара, а затем вейвлета «Мексиканская шляпа». Результаты восстановления $f(t)$ для $N = 5$, $N = 16$, $N = 50$ представлены на рис. 1 и 2 соответственно. Вместе с полученными оценками на рисунках представлен график истинной плотности распределения стандартного нормального закона $f(t)$.



$\hat{f}_n(t)$, ---- $N = 5$, ----- $N = 16$, — $N = 50$

Рис. 1. Вейвлет-оценка функции плотности распределения на основе материнского вейвлета Хаара

Fig. 1. The wavelet estimate of the distribution density function based on the Haar mother wavelet



$\hat{f}_n(t)$, ---- $N=5$, ----- $N=8$, — $N=50$

Рис. 2. Вейвлет-оценка плотности распределения на основе материнского вейвлета «Мексиканская шляпа»

Fig. 2. The wavelet estimate of the distribution density based on the Mexican Hat mother wavelet

Прежде всего отметим, что вид оценки $\hat{f}_n(t)$ существенно зависит от выбора материнского вейвлета. Так оценка, построенная с помощью вейвлета Хаара, имеет ступенчатый характер (см. рис. 1). А оценка, построенная с помощью вейвлета «Мексиканская шляпа», лишена этого недостатка (см. рис. 2).

Кроме того, из рисунков видно, что значение параметра сглаживания N также влияет на величину отклонения вейвлет-оценки от плотности распределения закона, по которому моделировалась исходная выборка. При малом значении N имеет место существенное отклонение оценки $\hat{f}_n(t)$ от истинной плотности. При достаточно большом N оценки функции плотности имеют дополнительные экстремумы, которые ухудшают ее качество. Количество таких экстремумов зависит от вида вейвлета, который используется при построении оценки (см. рис. 1 и 2). В результате оценка $\hat{f}_n(t)$ дает ложное представление о поведении истинной функции распределения $f(t)$. Результаты исследования позволили сделать вывод, что наилучшее значение параметра сглаживания для вейвлета Хаара – это $N=16$, а для вейвлета «Мексиканская шляпа» – это $N=8$.

Также были проведены исследования зависимости качества получаемой оценки от величины параметра сглаживания N с меньшими объемами выборки $n=50,\ n=100,\ n=500$. Полученные результаты качественно совпали с приведенными выше. Отметим, что наилучшее значение N меняется при изменении вида материнского вейвлета. Анализ литературы показал, что этот факт можно учесть при оценке параметра N. Для этого доопределим набор ортонормированных базисных функций (4) для $(N+1)\le i\le 2N.$ Согласно [21], оценка параметра состоит в следующем:

$$N_{\min} = \underset{1<N<n}{\arg\min}\,\tau_n, \tag{13}$$

где $\tau_n = \sum\limits_{j=N+1}^{2N}\left(\frac{1}{n}\sum\limits_{i=1}^{n}\psi_j(x_i)\right)^2$.

Результаты оценки параметра сглаживания с использованием соотношения (13) приведены в табл. 2.

Таблица 2

Table 2

Оценка параметра сглаживания

A smoothing parameter estimate

n	Хаар		«Мексиканская шляпа»	
	$N_{\min}$	N	$N_{\min}$	N
100	0,00245	16	0,044	8
500	0,00045		0,004	
1000	0		0,002	

Наилучшее значение N, установленное в результате моделирования, не совпадает с предполагаемым значением, вычисленным по формуле (13). Проводя численные эксперименты, удалось установить, что наилучшее значение N должно быть не меньше значения вычисленного следующим образом:

$$N_{\min} = \arg\min_{1<N<25}\left[1000\sum_{j=N+1}^{2N}\left(\frac{1}{T}\sum_{i=1}^{T}\psi_j(t_i)\right)^2\right], \qquad (14)$$

где T – количество интервалов разбиения выборки; точка t_i – середина i-го интервала. Результаты оценки параметра сглаживания с использованием соотношения (14) приведены в табл. 3.

Таблица 3

Table 3

Оценка параметра сглаживания с учетом количества интервалов разбиения выборки

A smoothing parameter estimate into account the number of sampling intervals

n	«Мексиканская шляпа», $N = 8$			
	$N_{\min}$	T	$N_{\min}$	T
100	7,863	34	7,635	11
500	7,863		8,62	
1000	7,863		8,654	

Для вейвлета «Мексиканская шляпа» при $T = 34$ и $T = 11$ получаем $N_{\min} = 8$. Значения $N_{\min}$ совпадают с наилучшим значением N, полученным при моделировании.

Далее найденные наилучшие значения параметра N были использованы для исследования зависимости качества оценки функции плотности $\hat{f}_n(t)$ от объема выборки n. Для этого были смоделированы выборки независимых значений случайной величины ξ, распределенной по нормальному закону с параметрами $(0,1)$, объемом $n = 100$, $n = 500$, $n = 1000$. Для каждой из них была построена вейвлет-оценка функции плотности (3) с помощью материнского вейвлета «Мексиканская шляпа». Результаты представлены на рис. 3. Вместе с графиками оценок на рисунке представлен график истинной плотности распределения стандартного нормального закона $f(t)$.

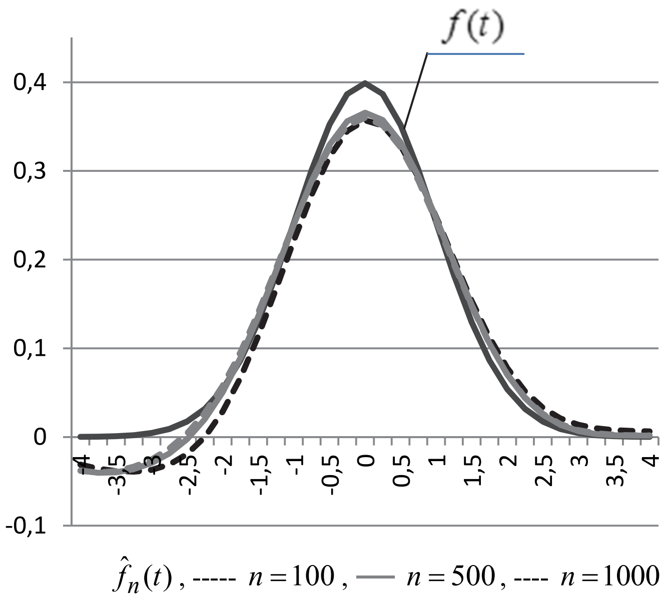


Рис. 3. Вейвлет-оценка плотности распределения на основе материнского вейвлета «Мексиканская шляпа»

Fig. 3. The wavelet estimate of the distribution density based on the Mexican Hat mother wavelet

Полученные результаты позволяют сделать вывод, что объем исходной выборки мало влияет на качество вейвлет-оценки и величину отклонения от истинной функции плотности распределения, что показано на рис. 3. Подобное исследование было проведено для оценки функции плотности (3), построенной с учетом вейвлета Хаара, и установлено, что оценка $\hat{f}_n(t)$ обладает таким же качеством.

Аналогичным образом проводилось исследование выборок, смоделированных по другим законам распределения. В частности, рассматривался экспоненциальный закон с параметром $\lambda = 2$ и гамма-распределение. Полученные результаты качественно совпали с приведенными выше.

Для количественной оценки степени близости $\hat{f}_n(t)$ и $f(t)$ была проведена проверка согласия по критерию χ^2 [22–25]. Рассматривались следующие гипотезы:

гипотеза H_0: $\hat{f}_n(t) = f(t)$,

гипотеза H_1: $\hat{f}_n(t) \neq f(t)$.

Для имеющихся выборок объемом $n=100$, $n=500$, $n=1000$, состоящих из множества независимых значений случайной величины ξ, распределенной по нормальному закону с параметрами $(0,1)$, построим вейвлет-оценки функции плотности (3) с использованием вейвлетов Хаара и «Мексиканская шляпа» при найденных выше наилучших значениях параметра N. Вычислим значения статистики $\chi^2_{\text{эмп}}$:

$$\chi^2_{\text{эмп}} = \sum_{i=1}^{T} \frac{\left(\hat{f}_n(t_i) - f(t_i)\right)^2}{f(t_i)}, \tag{15}$$

где T – количество интервалов разбиения выборки, полученное по формуле Стерджеса [23]; точка t_i – середина *i*-го интервала.

Критические значения критерия $\chi^2_{\text{кр}}$ Пирсона при числе степеней свободы $\nu = T - 1$ для $\alpha \le 0,05$ и $\alpha \le 0,01$ и результаты расчета согласия приведены в табл. 4 [23], из которой видно, что значения величины $\chi^2_{\text{эмп}}$ попали в зону незначимости, причем они намного меньше $\chi^2_{\text{кр}}$ для 5 %-го уровня значимости. Следовательно, гипотеза H_0 не отвергается. Вейвлет-оценка $\hat{f}_n(t)$ и теоретическая функция плотности $f(t)$ очень близки.

Таблица 4

Table 4

Значение статистики χ^2

The chi-square statistic value

n	T	$\chi^2_{\text{кр}}$		$\chi^2_{\text{эмп}}$ (Хаар)	$\chi^2_{\text{эмп}}$ («Мексиканская шляпа»)
		$\alpha \le 0,05$	$\alpha \le 0,01$		
100	8	14,067	18,475	3,192	12,243
500	10	16,919	21,666	2,986	11,923
1000	11	18,307	23,209	2,592	11,746

Проверка согласия по критерию χ^2 выполнялась для 1000 выборок объемом $n = 100$, $n = 500$, $n = 1000$. Во всех случаях результат совпал с приведенным в табл. 4. Так же проверялось согласие для случайных величин, распределенных по экспоненциальному закону с параметром $\lambda = 2$ и по гамма-распределению. Значения величины $\chi^2_{\text{эмп}}$ всегда оказывались меньше критической величины для 5 %-го уровня значимости, что позволило сделать вывод: вейвлет-оценка $\hat{f}_n(t)$ и теоретическая функция плотности $f(t)$ очень близки.

ЗАКЛЮЧЕНИЕ

В работе были рассмотрены оценки функции плотности распределения случайной величины на основе вейвлетов Хаара и «Мексиканская шляпа». Установлено, что объем исходной выборки мало влияет на качество вейвлет-оценки, а вот выбор материнского вейвлета, наоборот, очень важен. От выбора материнского вейвлета зависит значение параметра сглаживания, который, в свою очередь, существенно влияет на качество оценки функции плотности распределения случайной величины. В ходе вычислительных экспериментов удалось установить, что наилучшее значение параметра N для вейвлета Хаара равно 16, а для вейвлета «Мексиканская шляпа» – равно 8. Эти значения не совпадают со значениями, найденными с помощью существующих методов оценивания. В работе предложен новый способ оценивания, который позволяет получать значения параметра, близкие к тем, что получены в результате экспериментов. Показан ряд преимуществ каждой из построенных оценок. Выполнена поверка согласия по критерию χ^2 количественной оценки степени близости $\hat{f}_n(t)$ и $f(t)$, в результате чего была подтверждена эффективность рассмотренных вариантов оценивания.

СПИСОК ЛИТЕРАТУРЫ

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

2. *Тимофеев В.С., Хайленко Е.А.* Робастные оценки моментов при идентификации лямбда-распределения в рамках адаптивного оценивания // Доклады Академии наук высшей школы Российской Федерации. – 2017. – № 4 (37). – С. 101–111.

3. *Fan J., Gijbels I.* Local polynomial modelling and its application. – London: CRC Press, 1996. – 360 p.

4. *Тимофеев В.С.* Ядерные оценки плотности при идентификации уравнений регрессии. // Научный вестник НГТУ. – 2010. – № 3 (40). – С. 41–50.

5. *Добеши И.* Десять лекций по вейвлетам. – Ижевск: РХД, 2001. – 464 с.

6. Wavelets on irregular point sets / I. Daubechies, I. Guskov, P. Schro¨der, W. Sweldens // Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. – 1999. – Vol. 357. – P. 2397–2413.

7. *Silverman B.* Density estimation for statistics and data analysis. – London: Chapman and Hall, 1986. – 176 p.

8. *Смоленцев Н.К.* Введение в теорию вейвлетов. – М.; Ижевск: РХД, 2010. – 282 с.

9. *Захарова Т.В., Шестаков О.В.* Теория вейвлетов и ее применение в обработке сигналов: учебное пособие / Московский государственный университет им. М.В. Ломоносова, Факультет вычислительной математики и кибернетики, Кафедра математической статистики. – М.: МастерПринт, 2018. – 178 с.

10. *Витязев В.В.* Вейвлет-анализ временных рядов. – СПб.: Изд-во С.-Петерб. ун-та, 2001. – 58 с.

11. *Малашкевич И.А.* Вейвлет-анализ сигналов: от теории к практике: учебное пособие / Поволжский государственный технологический университет. – Йошкар-Ола: ПГТУ, 2016. – 276 с.

12. *Блаттер К.* Вейвлет-анализ. Основы теории / пер. с нем. Т.Э. Кренкеля; под ред. А.Г. Кюркчана. – М.: Техносфера, 2004. – 273 с.

13. *Шестаков О.В.* Вероятностно-статистические методы анализа и обработки сигналов на основе вейвлет-алгоритмов. – М.: Аргамак-Медиа, 2016. – 200 с.

14. *Фрейзер М.* Введение в вэйвлеты в свете линейной алгебры / пер. с англ. Я.М. Жилейкина. – Москва: Бином. Лаборатория знаний, 2007. – 487 с.

15. *Lepik Ü.* Numerical solution of evolution equations by the Haar wavelet method // Applied Mathematics and Computation. – 2007. – Vol. 185 (1). – P. 695–704.
16. *Lepik Ü., Hein H.* Haar wavelets: with applications. – Cham: Springer, 2014. – 207 p. – (Mathematical engineering).
17. *Stankovic R., Falkowski B.* The Haar wavelet transform: its status and achievements // Computers and Electrical Engineering. – 2003. – Vol. 29 (1). – P. 25–44.
18. Вейвлет-анализ в примерах: учебное пособие / О.В. Нагорнов, В.Г. Никитаев, В.М. Простокишин, С.А. Тюфлин, А.Н. Проничев, Т.И. Бухарова, К.С. Чистов, Р.З. Кашафутдинов, В.А. Хоркин. – М.: НИЯУ МИФИ, 2010. – 120 с.
19. *Mallat S.* A wavelet tour of signal processing. – New York, NY: Academic Press, 1999. – 851 p.
20. *Чернов А.В.* О применении квадратичных экспонент для дискретизации задач оптимального управления // Вестник Удмуртского университета. Математика. Механика. Компьютерные науки. – 2017. – Т. 27, вып. 4. – С. 558–575.
21. *Антонов А.В.* Системный анализ: учебник для вузов. – М.: Высшая школа, 2004. – 454 с.
22. *Геворкян П.С., Потемкин А.В., Эйсымонт И.М.* Теория вероятностей и математическая статистика. – М.: Физматлит, 2016. – 176 с.
23. *Гмурман В.Е.* Теория вероятностей и математическая статистика: учебник для прикладного бакалавриата. – Люберцы: Юрайт, 2016. – 479 с.
24. *Ивченко Г.И., Медведев Ю.И.* Математическая статистика. – М.: Либроком, 2019. – 352 с.
25. *Максимов Ю.Д.* Математическая статистика: опорный конспект. – М.: Проспект, 2016. – 104 с.

Тимофеев Владимир Семенович, доктор технических наук, декан факультета прикладной математики и информатики Новосибирского государственного технического университета. Имеет более 70 публикаций. E-mail: v.timofeev@corp.nstu.ru.

Исаева Елена Валерьевна, старший преподаватель кафедры высшей математики Новосибирского государственного технического университета. Имеет более 20 публикаций. E-mail: isaeva@corp.nstu.ru.

Timofeev Vladimir Semenovich, D.Sc. (Eng.), dean at the applied mathematics and computer science department of Novosibirsk State Technical University. He is the author of more than 70 publications. E-mail: v.timofeev@corp.nstu.ru.

Isaeva Elena Valerievna, senior lecturer higher mathematics department of Novosibirsk State Technical University. He is the author of more than 20 publications. E-mail: isaeva@corp.nstu.ru.

DOI: 10.17212/1814-1196-2019-4-71-84

*On the estimation of the distribution density function of a random variable using wavelets**

V.S. TIMOFEEV[a], *E.V. ISAEVA*[b]

Novosibirsk State Technical University, 20 K. Marx Prospekt, Novosibirsk, 630073, Russian Federation

[a] *v.timofeev@corp.nstu.ru* [b] *isaeva@corp.nstu.ru*

Abstract

The problem of estimating the probability density of distribution arises in many statistical data analysis. There are different approaches to solving it, but recently methods based on wavelet assessment have become popular. The essence of the methods is to decompose an unknown

* *Received 03 September 2019.*

function into a series over some finite set of orthonormal basis functions. A system of functions defined on the $[0,1]$ line which does not always coincide with the range of values of a random quantity is used as such a set. Therefore, in order to calculate a wavelet estimate of the distribution density of a random quantity given in an arbitrary region, it is necessary to perform a transition to a system of functions defined on the same segment as the random quantity values. Within the framework of this work, the wavelet estimates of the distribution density of a random value will be considered with the help of the Haar and Mexican Hat wavelets. The study of the effect on the quality of wavelet estimates of such parameters as the sample size, the number of coefficients of expansion of a function in a series in the expression for wavelet density estimation was performed. In the course of the work, it has been found that the quality of the wavelet estimation depends significantly on the smoothing parameter and that its best value exists, which in turn depends on the choice of the mother wavelet. The authors propose a new version of smoothing parameter estimation. In order to quantify the degree of proximity of the distribution density function and its wavelets estimation, agreement on the chi-square criterion was tested and conclusions were made as to which basic wavelet provides the highest quality recovery of the density function.

Keywords: wavelet estimate, wavelet-analysis, Haar wavelet, Mexican Hat wavelet, function of density, estimation of density wavelet function, chi-square criterion, smoothing parameter, computing experiment

REFERENCES

1. 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.

2. Timofeev V.S., Khailenko E.A. Robastnye otsenki momentov pri identifikatsii lyambda-raspredeleniya v ramkakh adaptivnogo otsenivaniya [Robust estimates of moments in the identification of generalized lambda-distribution within the adaptive regression model estimation]. *Doklady Akademii nauk vysshei shkoly Rossiiskoi Federatsii – Proceedings of the Russian higher school Academy of sciences*, 2017, no. 4 (37), pp. 101–111.

3. Fan J., Gijbels I. *Local polynomial modelling and its application*. London, CRC Press, 1996. 360 p.

4. Timofeev V.S. Yadernye otsenki plotnosti pri identifikatsii uravnenii regressii [The Kernel estimation of density function in the regression identification problem]. *Nauchnyi vestnik Novosibirskogo gosudarstvennogo tekhnicheskogo universiteta – Science bulletin of the Novosibirsk state technical university*, 2010, no. 3 (40), pp. 41–50.

5. Daubechies I. *Desyat' lektsii po veivletam* [Ten lectures on weevletam]. Izhevsk, Regular and Chaotic Dynamics Publ., 2001. 464 p. (In Russian).

6. Daubechies I., Guskov I., Schro¨der P., Sweldens W. Wavelets on irregular point sets. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 1999, vol. 357, pp. 2397–2413.

7. Silverman B. *Density estimation for statistics and data analysis*. London, Chapman and Hall, 1986. 176 p.

8. Smolentsev N.K. *Vvedenie v teoriyu veivletov* [Introduction to the theory of weevlets]. Moscow, Izhevsk, Regular and chaotic dynamics Publ., 2010. 282 p.

9. Zakharova T.V., Shestakov O.V. *Teoriya veivletov i ee primenenie v obrabotke signalov* [The theory of weevlets and its application in signal processing]. Moscow, MasterPrint Publ., 2018. 178 p.

10. Vityazev V.V. *Veivlet-analiz vremennykh ryadov* [Wevlet-analysis of time series]. St. Petersburg, St. Petersburg University Publ., 2001. 58 p.

11. Malashkevich I.A. *Veivlet-analiz signalov: ot teorii k praktike* [Weivlet-analysis of signals: from theory to practice]. Yoshkar-Ola, Volga state technological university Publ., 2016. 276 p.

12. Blatter C. *Veivlet-analiz. Osnovy teorii* [Wavlet analysis. Foundations of theory]. Moscow, Tekhnosfera Publ., 2004. 273 p. (In Russian).

13. Shestakov O.V. *Veroyatnostno-statisticheskie metody analiza i obrabotki signalov na osnove veivlet-algoritmov* [Probabilistic-statistical methods of analysis and processing of signals on the basis of weevet-algorithm]. Moscow, Argamak-Media Publ., 2016. 200 p.

14. Frasier M. *An introduction to wavelets through linear algebra*. New York, Springer, 1999 (Russ. ed.: Freizer M. *Vvedenie v veivlety v svete lineinoi algebry*. Moscow, Binom Publ., 2007. 487 p.).

15. Lepik Ü. Numerical solution of evolution equations by the Haar wavelet method. *Applied Mathematics and Computation*, 2007, vol. 185 (1), pp. 695–704.

16. Lepik Ü., Hein H. *Haar wavelets: with applications*. Cham: Springer, 2014. 207 p.

17. Stankovic R., Falkowski B. The Haar wavelet transform: its status and achievements. *Computers and Electrical Engineering*, 2003, vol. 29 (1), pp. 25–44.

18. Nagornov O.V., Nikitaev V.G., Prostokishin V.M., Tyuflin S.A., Pronichev A.N., Bukharova T.I., Chistov K.S., Kashafutdinov R.Z., Khorkin V.A. *Veivlet-analiz v primerakh* [Weevlet analysis in examples]. Moscow, National Research Nuclear University MEPhI Publ., 2010. 120 p.

19. Mallat S. *A wavelet tour of signal processing*. New York, NY, Academic Press, 1999. 851 p.

20. Chernov A.V. O primenenii kvadratichnykh eksponent dlya diskretizatsii zadach optimal'nogo upravleniya [On the application of Gaussian functions for discretization of optimal control problems]. *Vestnik Udmurtskogo universiteta. Matematika Mekhanika. Komp'yuternye nauki – The Bulletin of Udmurt University. Mathematics. Mechanics. Computer Science*, 2017, vol. 27, iss. 4, pp. 558–575.

21. Antonov A.V. *Sistemnyi analiz* [System analysis]. Moscow, Vysshaya shkola Publ., 2004. 454 p.

22. Gevorkyan P.S., Potemkin A.V., Eisymont I.M. *Teoriya veroyatnostei i matematicheskaya statistika* [Probability theory and mathematical statistics]. Moscow, Fizmatlit Publ., 2016. 176 c.

23. Gmurman V.E. *Teoriya veroyatnostei i matematicheskaya statistika* [Probability theory and mathematical statistics]. Lubertsy, Yurait Publ., 2016. 479 p.

24. Ivchenko G.I., Medvedev Yu.I. *Matematicheskaya statistika* [Mathematical statistics]. Moscow, Librokom Publ., 2019. 352 p.

25. Maksimov Yu.D. *Matematicheskaya statistika* [Mathematical statistics]. Moscow, Prospekt Publ., 2016. 104 p.

Для цитирования:

Тимофеев В.С., Исаева Е.В. Об оценивании функции плотности распределения случайной величины с использованием вейвлетов // Научный вестник НГТУ. – 2019. – № 4 (77). – С. 71–84. – DOI: 10.17212/1814-1196-2019-4-71-84.

For citation:

Timofeev V.S., Isaeva E.V. Ob otsenivanii funktsii plotnosti raspredeleniya sluchainoi velichiny s ispol'zovaniem veivletov [On the estimation of the distribution density function of a random variable using wavelets]. *Nauchnyi vestnik Novosibirskogo gosudarstvennogo tekhnicheskogo universiteta – Science bulletin of the Novosibirsk state technical university*, 2019, no. 4 (77), pp. 71–84. DOI: 10.17212/1814-1196-2019-4-71-84.