

СТАТЬИ

УДК 519.213:519.23

**ОЦЕНИВАНИЕ ФУНКЦИИ ПЛОТНОСТИ РАСПРЕДЕЛЕНИЯ
С ИСПОЛЬЗОВАНИЕМ ВЕЙВЛЕТА МОРЛЕ****Исаева Е.В.***ФГБОУ ВО «Новосибирский государственный технический университет», Новосибирск,
e-mail: isaeva@corp.nstu.ru*

Для решения большинства задач, которые возникают в результате статистической обработки данных, требуется информация о виде распределения случайной величины. Решить вопрос об оценивании плотности вероятности распределения случайной величины можно разными методами, в том числе и методом, основанным на вейвлет-анализе. Использование вейвлетов позволяет разложить неизвестную функцию в ряд по некоторому конечному набору ортонормированных базисных функций. В качестве такого набора можно использовать систему функций, определённую на том же отрезке, что и область значений случайной величины. В рамках данной работы будет рассмотрена вейвлет-оценка плотности распределения случайной величины, построенная с помощью вейвлета Морле. Установлено, что материнский вейвлет Морле является ненормированным, а значит, система базисных функций, построенная с его помощью, так же не обладает данным свойством. Поэтому при построении вейвлет-оценки плотности распределения случайной величины с его помощью требуется ввести нормировочный множитель, который позволит нивелировать данный недостаток материнского вейвлета. Полученное в работе значение нормировочного множителя позволяет улучшить качество восстановления функции плотности. Исследовано влияние на качество вейвлет-оценки объема выборки и числа коэффициентов разложения функции в ряд в выражении для оценки плотности распределения. Установлено, что качество вейвлет-оценки существенно зависит от параметра сглаживания и существует его наилучшее значение. Для количественной оценки степени близости функции плотности распределения и её вейвлет-оценки была проведена проверка согласия по критерию хи квадрат, которая показала, что выбор данного вейвлета в качестве базового обеспечивает качественное восстановление функции плотности.

Ключевые слова: вейвлет-оценка, вейвлет-анализ, вейвлет Морле, параметр сглаживания, функция плотности, оценка функции плотности, критерий хи квадрат, случайная величина, вычислительный эксперимент

**ESTIMATION OF THE DISTRIBUTION DENSITY FUNCTION USING
THE MORLET WAVELET****Isaeva E.V.***Novosibirsk State Technical University, Novosibirsk, e-mail: isaeva@corp.nstu.ru*

To solve most of the problems that arise as a result of statistical data processing, information about the type of distribution of a random variable is required. The problem of estimating the probability density of the distribution of a random variable can be solved by different methods, including the method based on wavelet analysis. The use of wavelets allows one to expand an unknown function into a series in terms of some finite set of orthonormal basis functions. As such a set, one can use a system of functions defined on the same interval as the range of values of a random variable. Within the framework of this work, a wavelet estimate of the distribution density of a random variable constructed using the Morlet wavelet will be considered. It has been established that the Morlet of the mother wavelet is not normalized, which means that the system of basic functions built with its help also does not have this property. Therefore, when constructing a wavelet for estimating the distribution density of a random variable with its help, it is required to introduce a normalization factor that will allow leveling this disadvantage of the mother wavelet. The value of the normalization factor obtained in the work makes it possible to improve the quality of the reconstruction of the density function. The influence on the quality of wavelet estimates of the sample size and the number of coefficients of the expansion of a function into a series in the expression for estimating the distribution density is studied. It has been established that the quality of the wavelet estimate significantly depends on the smoothing parameter and there is its best value. To quantify the degree of closeness of the distribution density function and its wavelet estimate, a fit test was carried out using the chi-square criterion, which showed that the choice of this wavelet as the base one provides a qualitative restoration of the density function.

Keywords: assessment wavelet, wavelet-analysis, wavelet Morlet, function of density, assessment of function of density wavelet, hi-square criteria, smoothing parameter, a computing experiment

Статистическая обработка данных зачастую связана с исследованием свойств случайных величин, характеризующих те или иные особенности изучаемого объекта. Важной характеристикой, которая несет в себе максимально возможный объем полезной информации об изучаемых случайных величинах, является функция плотности распределения. Следовательно,

задача восстановления функции плотности по имеющимся статистическим данным является очень актуальной. Существуют различные методы для ее решения, в частности метод, основанный на вейвлет-анализе [1, 2], который на сегодняшний день динамично развивается и широко применяется для аппроксимации различных функций [3, 4]. Ранее такой подход был реализован

с использованием вейвлетов Хаара, DOG, «Мексиканская шляпа» и Литлвуда – Пэли [5–7], а проведенные исследования подтвердили его эффективность. В данной работе исследуются вейвлет-оценки функции плотности распределения, построенные на основе вейвлета Морле.

Для этого рассмотрим выборку $\{x_j | j = \overline{1, n}\}$, заданную на произвольном отрезке $[c, d]$, где $\begin{cases} c = \min x_j \\ d = \max x_j \end{cases}$. Выборка

состоит из независимых значений случайной величины ξ . Отметим, что информации о функции плотности распределения $f(t)$ случайной величины ξ не имеется. Необходимо построить вейвлет-оценку $\hat{f}_n(t)$ плотности распределения на основе имеющихся данных.

Вейвлет-оценки функции плотности распределения

Вейвлет-оценка функции плотности распределения $\hat{f}_n(t)$ случайной величины, согласно [1, 2] может быть представлена в виде следующего ряда:

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

где $\psi_i(t)$ – ортонормированные базисные функции, N – параметр сглаживания (количество членов ряда), \hat{c}_i – коэффициенты разложения, которые выражаются следующим соотношением:

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

Если подставить (2) в выражение (1), то получим вейвлет-оценку функции плотности распределения в следующем виде

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

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

В роли базисных функций $\psi_i(t)$ можно выбрать предложенную в [1, 2] систему функций, которая ортонормирована на отрезке $[0, 1]$ и определяется следующим образом:

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

где $\psi(t)$ – материнский вейвлет, $k \geq 0, 1 \leq j \leq 2^k$ и $i = 2^k + j$. Отметим, что выражение (4) выполняется для всех $i > 1$, а в случае $i = 1$

полагают, что $\psi_1(t) = 1$ для всех $t \in [0, 1]$, иначе $\psi_1(t) = 0$.

Так как выбранная система базисных функций $\psi_i(t)$ определена на отрезке $[0, 1]$, необходимо выполнить переход к ортонормированной системе функций $\tilde{\psi}_i(t)$ заданной на отрезке $[c, d]$, который соответствует области определения случайной величины ξ .

Согласно [5–7] переход от одной системы базисных функций к другой выражается следующим соотношением

$$\tilde{\psi}_i(t) = \frac{1}{\sqrt{d-c}} \psi_i\left(\frac{t-c}{d-c}\right). \quad (5)$$

Тогда вейвлет-оценка функции плотности распределения $\hat{f}_n(t)$ на произвольном отрезке $[c, d]$ определяется выражением (3), где в качестве базисных функций используются функции $\tilde{\psi}_i(t)$.

Согласно определению ортонормированной системы функций [1, 2] система базисных функций $\psi_i(t)$ должна удовлетворять двум условиям:

$$1) (\psi_i(t), \psi_j(t)) = \int_{-\infty}^{\infty} \psi_i(t) \psi_j(t) dt = 0,$$

для всех $i \neq j$;

$$2) \|\psi_i(t)\| = \sqrt{(\psi_i(t), \psi_i(t))} = \sqrt{\int_{-\infty}^{\infty} \psi_i(t) \psi_i(t) dt} = 1.$$

Следовательно, из соотношения (4) и свойств нормы материнского вейвлета [2] получаем

$$\begin{aligned} \|\psi_i(t)\| &= \left\| 2^{k/2} \psi(2^k t - (j-1)) \right\| = \\ &= \left\{ 2^{k/2} \int_{-\infty}^{+\infty} \left| \psi(2^k t - (j-1)) \right|^2 dt \right\}^{1/2} = \|\psi(t)\|. \end{aligned} \quad (6)$$

где $\psi(t) \in L^2(R)$.

Таким образом, выполнение этих условий семейством $\psi_i(t)$ зависит от выбора порождающего их материнского вейвлета $\psi(t)$. Если материнский вейвлет ортонормированный, то и система функций $\psi_i(t)$ также будет являться ортонормированной.

Перейдем к исследованию вейвлет-оценки функции плотности (3), где в качестве материнского вейвлета выбран вейвлет Морле [8–10].

Оценивание функции плотности распределения с помощью материнского вейвлета Морле

Аналитическая запись материнского вейвлета Морле [8–10] определяется соотношением

$$\psi(t) = e^{-\frac{t^2}{2}} \cos 5t. \quad (7)$$

Тогда система функций (4) на отрезке $[0, 1]$ выражается следующим образом:

$$\psi_i(t) = 2^{\frac{k}{2}} e^{-\frac{(2^k t - (j-1))^2}{2}} \cos\left(5\left(2^k t - (j-1)\right)\right), \quad (8)$$

где i, k, j такие же, как в (4). Известно, что материнский вейвлет Морле является ортогональным [2, 8–10]. Покажем, что вейвлет Морле не является нормированным, как следствие, система функций (8) тоже.

Утверждение 1. Пусть $\psi(t) = e^{-\frac{t^2}{2}} \cos 5t$ – материнский вейвлет Морле, порождающий систему функций $\psi_i(t) \in L^2(R)$, которая принимает вид

$$\psi_i(t) = 2^{\frac{k}{2}} e^{-\frac{(2^k t - (j-1))^2}{2}} \cos\left(5\left(2^k t - (j-1)\right)\right),$$

где $k \geq 0, 1 \leq j \leq 2^k$ и $i = 2^k + j$. Тогда для любых i, k, j $\|\psi_i(t)\| = \frac{\sqrt[4]{\pi}}{\sqrt{2}} \left(1 + e^{-25}\right)^{1/2}$.

Доказательство.

Аналитическая запись материнского вейвлета Морле выражается соотношением

$$\psi(t) = e^{-\frac{t^2}{2}} \cos 5t.$$

Тогда по определению нормы в $L_2(R)$ получаем

$$\begin{aligned} \|\psi(t)\|^2 &= \int_{-\infty}^{\infty} \left(e^{-\frac{t^2}{2}} \cos 5t \right)^2 dt = \int_{-\infty}^{\infty} e^{-t^2} \cos^2 5t dt = \frac{1}{2} \int_{-\infty}^{\infty} e^{-t^2} (1 + \cos 10t) dt = \\ &= \frac{1}{2} \int_{-\infty}^{\infty} e^{-t^2} dt + \int_0^{\infty} e^{-t^2} \cos 10t dt = \frac{\sqrt{\pi}}{2} + \frac{\sqrt{\pi}}{2e^{25}} = \frac{\sqrt{\pi}}{2} \left(1 + e^{-25}\right). \end{aligned}$$

Поскольку для любых $k \geq 0, 1 \leq j \leq 2^k$, где $i = 2^k + j$ выполняется равенство (6), то норма $\|\psi_i(t)\| = \frac{\sqrt[4]{\pi}}{\sqrt{2}} \left(1 + e^{-25}\right)^{1/2}$. Утверждение доказано.

Отсюда следует, что базис (4) на основе материнского вейвлета Морле не нормирован: $\|\psi_i\| = \frac{\sqrt[4]{\pi}}{\sqrt{2}} \left(1 + e^{-25}\right)^{1/2} \neq 1$. В соответствии с [2, 8] такая система базисных функций может быть использована для построения (3), если ввести нормировочный множитель $z = \frac{1}{\|\psi_i(t)\|}$.

Тогда система функций (8) становится ортонормированной и принимает вид

$$\psi_i(t) = z 2^{\frac{k}{2}} e^{-\frac{(2^k t - (j-1))^2}{2}} \cos\left(5\left(2^k t - (j-1)\right)\right), \quad (9)$$

где i, k, j аналогичны (4), $z = \frac{\sqrt{2}}{\sqrt[4]{\pi}} \left(1 + e^{-25}\right)^{-1/2}$.

С учетом (9) переход от отрезка $[0, 1]$ к отрезку $[c, d]$, который соответствует области определения наблюдаемой случайной величины, дает результат

$$\tilde{\psi}_i(\tau) = \frac{z}{\sqrt{d-c}} 2^{\frac{k}{2}} e^{-\frac{(2^k \tau - (j-1))^2}{2}} \cos\left(5(2^k \tau - (j-1))\right), \quad (10)$$

где i, k, j аналогичны (4), $\tau = \frac{t-c}{d-c}$.

Тогда выражение (3), где в качестве базисных функций используются функции (10), является вейвлет-оценкой функции плотности распределения $\hat{f}_n(t)$ на произвольной области определения $[c, d]$ на основе материнского вейвлета Морле.

Исследование точности вейвлет-оценки на основе вейвлета Морле

С помощью серии вычислительных экспериментов была исследована точность вейвлет-оценки на основе материнского вейвлета Морле. Для этого были смоделированы выборки из независимых значений случайной величины ξ объемом $n = 300; 500; 1000$. Распределение случайной величины ξ соответствует стандартному нормальному закону. Для данных выборок было выполнено оценивание функции плотности распределения с помощью вейвлета Морле при изменении числа членов ряда N от 4 до 40. На рис. 1 представлены результаты оценивания функции плотности для выборки объемом $n = 1000$ при $N = 5; 15; 25$. Качество оценивания $f(t)$ напрямую зависит от параметра сглаживания. Если количество членов ряда оказывается недостаточным,

то вейвлет-оценка не соответствует исходной функции плотности распределения из-за существенного отклонения. Если количество членов ряда велико, то оценка функции $\hat{f}_n(t)$ имеет дополнительные максимумы и минимумы, что никак не соответствует модельной плотности распределения. Как в первом, так и во втором случае восстановленная функция плотности достаточно сильно отличается от истинной. Таким образом, в ходе вычислительных экспериментов было получено наилучшее значение параметра сглаживания для вейвлета Морле $N = 5$. Результаты моделирования для выборок с меньшими объемами $n = 300; 500$ подтвердили приведенные выше выводы.

Найденное выше значение параметра сглаживания $N = 5$ было использовано для исследования точности восстановления функции плотности от объема выборки n . Для этого были построены вейвлет-оценки (3) на основе материнского вейвлета Морле для рассматриваемых ранее выборок объема $n = 300; 500; 1000$. На рис. 2 представлены полученные результаты.

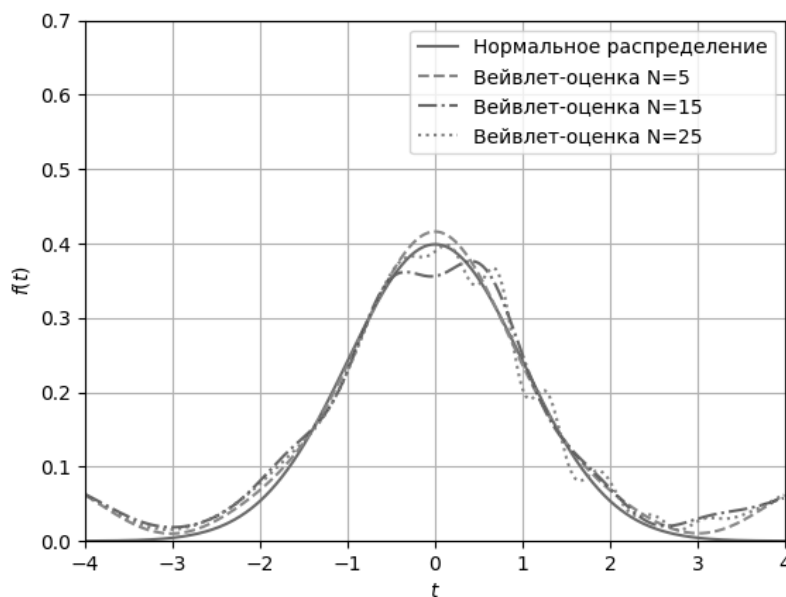


Рис. 1. Результаты оценивания функции плотности с помощью вейвлета Морле

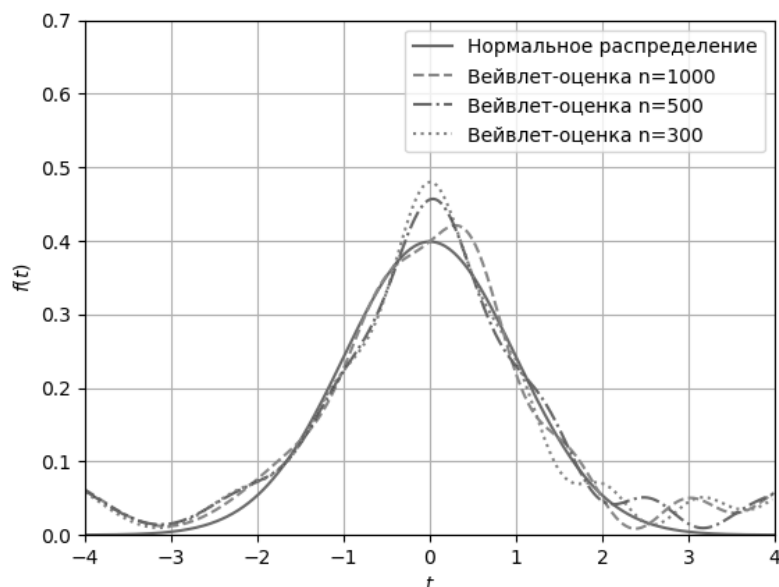


Рис. 2. Результаты оценивания функции плотности с помощью вейвлета Морле

Значение статистики χ^2

n	T	$\chi_{кр}^2$		χ^2
		$\alpha \leq 0,05$	$\alpha \leq 0,01$	
300	8	14,067	18,475	10,589
500	10	19,919	21,666	11,413
1000	11	18,307	23,209	11,884

Эти вычислительные эксперименты подтвердили предположение о том, что объем данной выборки влияет на точность восстановления функции плотности. Из рис. 2 видно, что выборки с наименьшим объемом имеют наибольшую величину отклонения истинной функции плотности распределения от её оценки.

Отметим, что подобные исследования вейвлет-оценок на основе вейвлета Морле проводились для выборок, состоящих из случайных величин, имеющих другие законы распределения, в частности рассматривалось гамма распределение и экспоненциальное с параметром $\lambda = 2$. Результаты моделирования также показали, что с увеличением объема выборки точность вейвлет-оценки повышается.

Проверка степени близости между истинной функцией плотности распределения и ее оценкой для 1000 выборок объема $n = 300; 500; 1000$ состоящих из множества независимых значений случайной величины ξ , была проведена по критерию χ^2 [11]:

$$\chi^2 = \sum_{i=1}^T \frac{(\hat{f}_n(t_i) - f(t_i))^2}{f(t_i)}, \quad (11)$$

где T – число интервалов, полученное с помощью формулы Стерджеса [11], t_i – середина i -го интервала. В таблице приведены критические значения критерия $\chi_{кр}^2$ при числе степеней свободы $\nu = T - 1$ для $\alpha \leq 0,05$ и $\alpha \leq 0,01$. Также в таблице представлены результаты, из которых видно, что значения величины χ^2 намного меньше $\chi_{кр}^2$, что подтверждает близость вейвлет-оценки $\hat{f}_n(t)$ и теоретической функции плотности $f(t)$.

Аналогичным образом была выполнена оценка степени близости $\hat{f}_n(t)$ и $f(t)$ для случайных величин, имеющих другие законы распределения, а именно гамма распределение и экспоненциальное с параметром $\lambda = 2$. Результаты моделирования подтвердили работоспособность данного метода оценивания.

Заключение

В работе была рассмотрена оценка функции плотности распределения случайной величины на основе вейвлета Морле. Сформулировано и доказано утверждение о том, что данный материнский вейвлет является ненормированным, следовательно, и система базисных функций, построенная с его помощью. Получено значение нормировочного множителя, которое позволяет улучшить качество восстановления функции плотности. Установлено, что качество оценки функции плотности существенно зависит от параметра сглаживания и в ходе вычислительных экспериментов получено его наилучшее значение. Показано, что объем выборки оказывает влияние на качество восстановления функции плотности распределения. В частности, в случае малого объема выборки наблюдается большее отклонение вейвлет-оценки плотности распределения от теоретической функции плотности. Проверка по критерию хи квадрат количественной степени близости между истинной функцией плотности распределения и ее оценкой показала, что выбор вейвлета Морле в роли базового обеспечивает качественное восстановление функции плотности.

Список литературы

1. Добеши И. Десять лекций по вейвлетам. Ижевск: НИЦ «Регулярная и хаотическая динамика», 2001. 464 с.
2. Чуи К. Введение в вейвлеты. М.: Мир, 2001. 412 с.
3. Hari Krishan Malhotra, Lalit Kumar Vashisht On scaling functions of non-uniform multiresolution analysis in $L_2(\mathbb{R})$. International Journal of Wavelets, Multiresolution and Information Processing. 2020. Vol. 18. No. 02. 14 p.
4. Emmanuel de Dieu N., Guy Martial N. Wavelet-based estimation in a semiparametric regression model. International Journal of Wavelets, Multiresolution and Information Processing. 2021. 30 p.
5. 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.
6. 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.
7. Тимофеев В.С., Исаева Е.В., Малышкина Е.Д., Слободчикова А.Э. Оценивание функции плотности распределения с использованием вейвлета Литлвуда – Пэли // Обработка информации и математическое моделирование: материалы Российской научно-технической конференции. 2020. С. 165–170.
8. Витязев В.В. Вейвлет-анализ временных рядов. СПб.: СПбГУ, 2001. 58 с.
9. Малашкевич И.А. Вейвлет-анализ сигналов: от теории к практике: учебное пособие. Йошкар-Ола: ПГТУ, 2016. 276 с.
10. Шестаков О.В. Вероятностно-статистические методы анализа и обработки сигналов на основе вейвлет-алгоритмов. М.: Аргатак-Медиа, 2016. 200 с.
11. Ивченко Г.И., Медведев Ю.И. Математическая статистика. М.: КД Либроком, 2019. 352 с.