



ОБ ОСОБЕННОСТЯХ АППРОКСИМАЦИЙ РОЗЕНБЛАТТА-ПАРЗЕНА ЭМПИРИЧЕСКИХ ФУНКЦИЙ РАСПРЕДЕЛЕНИЙ И ПЛОТНОСТЕЙ РАСПРЕДЕЛЕНИЙ СЛУЧАЙНЫХ ВЫБОРОК

В статье обсуждаются результаты сравнительного анализа эмпирических функций распределения (ФР) и плотностей распределения (ПР) случайных выборок и их аппроксимаций Розенблатта-Парзена, зависящих от случайной выборки и параметра размытости ядерной функции h .

Для оценки качества аппроксимаций Розенблатта-Парзена были использованы накопленные среднеквадратические погрешности разностей между эмпирическими ФР и ПР и соответствующими аппроксимациями Розенблатта-Парзена.

Продемонстрировано, что при следовании существующим рекомендациям использовать для вычисления аппроксимации Розенблатта-Парзена значение параметра размытости h^* , равного абсциссе локального максимума информационного функционала, не удастся обеспечить минимального возможного значения выбранного критерия качества аппроксимации эмпирических ФР и ПР.

Предложено использовать в качестве h^* абсциссу минимума зависимости накопленной среднеквадратической погрешности разностей между эмпирической ПР и соответствующей аппроксимацией Розенблатта-Парзена от параметра размытости.

Ключевые слова: случайная выборка, функция распределения, плотность распределения, эмпирическая функция распределения, эмпирическая плотность распределения, аппроксимация Розенблатта-Парзена, критерий типа Колмогорова-Смирнова.

THE FEATURES OF THE ROSENBLATT-PARSEN APPROXIMATIONS OF EMPIRICAL DISTRIBUTION FUNCTIONS AND DISTRIBUTION DENSITIES OF RANDOM SAMPLES

The article discusses the results of a comparative analysis of empirical Cumulative Distribution Functions (CDF) and Probability Density Function (PDF) of random samples and their Rosenblatt-Parsen approximations that depending on the random sample and the fuzzy parameter of the kernel function h .

For estimation the quality of the Rosenblatt-Parsen approximations were used the accumulated root-mean-square errors (ARMSE) of the differences between the CDF and PDF and the corresponding Rosenblatt-Parsen approximations.

It has been demonstrated that following the existing recommendations to use the value of the fuzzy parameter h^ for the Rosenblatt-Parsen approximation calculating equal to the abscissa of the local maximum of the information functional, it is not possible to provide the minimum possible value of the selected criterion of the quality of the empirical CDF and PDF approximation.*

It is proposed to use as h^ the minimum abscissa of the ARMSE of the difference between the empirical PDF and the corresponding Rosenblatt-Parsen approximation dependence on fuzzy parameter.*

Keywords: *Random Sample, Distribution Function, distribution density, Empirical Distribution Function, Empirical distribution density, Rosenblatt-Parsen approximation, Kolmogorov-Smirnov test*

Введение

Задача аппроксимации функций распределений (ФР) и плотностей распределений (ПР) случайных выборок теми или иными аналитическими моделями возникает в различных отраслях наук, например, в радиотехнике при анализе случайных сигналов [1,2], при проектировании телекоммуникационных систем аудиообмена, систем подвижной связи и систем телекоммуникации [3,4], при оптимизации параметров процедуры обратимого сжатия цифровых данных [5], при создании математических моделей систем управления [6], при моделировании случайных процессов [7], при анализе количественных характе-

ристики Интернет-трафика [8,9], при оценке прочностной надежности магистральных газо- и нефтепроводов [10].

Напомним, что в наиболее общем виде данная задача имеет следующую формулировку. Имеется одномерная выборка независимых упорядоченных в порядке возрастания случайных величин $\{x_1, x_2, \dots, x_N\}$, извлеченных из некоторого распределения, в общем случае, с неограниченной областью рассеяния с неизвестными непрерывными ФР $F(x)$ и плотностью распределения ПР $f(x)$ такими что

$$\int_{-\infty}^x f(\xi) d\xi = F(x),$$

где $F(x) = \Pr[\xi \leq x]$,

здесь $\Pr[\xi \leq x]$ – вероятность того, что случайная величина $\xi \leq x$. Требуется на основе выборки $\{x_1, x_2, \dots, x_N\}$ оценить вид ФР $F(x)$ и ПР $f(x)$

Методы решения рассматриваемой задачи являются объектом исследования параметрической и непараметрической статистики. Напомним, что в параметрической статистике на основе априорной информации о свойствах случайного процесса, из которого извлечена данная выборка, выбирается один из известных законов распределения (число которых на сегодняшний день превышает 100) $F^{(\text{theory})}(x, a, b, \dots)$, a, b, \dots – параметры распределения. Далее по анализируемой случайной выборке $\{x_1, x_2, \dots, x_N\}$ находят оценки параметров $\tilde{a}, \tilde{b}, \dots$ выбранного закона распределения $F^{(\text{theory})}(x, a, b, \dots)$, например, с помощью метода максимального правдоподобия. Далее проверяют статистическую гипотезу о соответствии найденной ФР $F^{(\text{theory})}(x, \tilde{a}, \tilde{b}, \dots)$ случайной выборки $\{x_1, x_2, \dots, x_N\}$ априори выбранному теоретическому закону распределения $F^{(\text{theory})}(x, a, b, \dots)$, используя один из многочисленных известных статистических критериев (см. [11, Глава 3]), например, критерий типа Колмогорова–Смирнова.

В непараметрической статистике вычисляют те или иные аппроксимации эмпирической ФР распределения вероятностей $F_N^{(\text{emp})}(x)$:

$$F_N^{(\text{emp})}(x) = \frac{1}{N} \sum_{i=1}^N \Theta(x - x_i),$$

где $\Theta(s)$ – функция Хэвисайда,

$$\Theta(s) = \begin{cases} 1, & s \geq 0, \\ 0, & s < 0, \end{cases}$$

случайной выборки $\{x_1, x_2, \dots, x_N\}$, такой, что

$$\Pr \left[\limsup_{N \rightarrow \infty} \left| F_N^{(\text{emp})}(x) - F^{(\text{theory})}(x) \right| = 0 \right] = 1,$$

которая, в этой связи при $N \rightarrow \infty$ является оптимальной эмпирической оценкой теоретической функции распределения $F^{(\text{theory})}(x)$. Существование данной функции гарантируется центральной теоремой математической статистики (теорема Гливленко) [12]. Также в прикладной статистике рассматривают эмпирическую ПР случайной выборки $\{x_1, x_2, \dots, x_N\}$

$$f_N^{(\text{emp})}(k\Delta x) = \frac{n_k}{N\Delta x},$$

где n_k – число членов случайной выборки, попавших в k -ый интервал $[x_1 + (k-1)\Delta x, x_1 + k\Delta x]$, $k = 1, N_g$, N_g – число интервалов гистограммы, $\Delta x = (x_N - x_1)/N_g$, представляющую собой кусочно-постоянную

аппроксимацию непрерывной функции $F^{(\text{theory})}(x)$.

Для аппроксимации функции распределения вероятностей $F_N^{(\text{emp})}(x)$ на сегодняшний день разработано достаточно большое число различных методов, основанных на использовании некоторых линейных комбинаций (ядерных) функций, выбор которых не требует априорной информации о функции $F^{(\text{theory})}(x)$ (метод Розенблатта-Парзена [13,14]); ортогональных многочленов [15], в том числе, полиномов Чебышева-Эрмита [16]; многочленов Берштейна (см. [17] и приведенные там ссылки). При этом использование любого из методов сопряжено с решением проблем, связанных с выбором критериев для оценки качества аппроксимации функции распределения $F_N^{(\text{emp})}(x)$ и выбора таких параметров используемого метода, что обеспечивается оптимальное качество аппроксимации. Таким параметром в методе Розенблатта-Парзена, одном из наиболее активно используемых сегодня непараметрических методов аппроксимации ФР и ПР случайных выборок, является, так называемый, параметр размытости h .

В статье проведен критический анализ известного подхода к оценке оптимального значения данного параметра.

Метод Розенблатта-Парзена

Напомним, следуя [13,14], что в соответствии с методом Розенблатта-Парзена аппроксимация ФР случайной выборки $\{x_1, x_2, \dots, x_N\}$ $F_N^{(\text{approx})}(x, \{x_1, x_2, \dots, x_N\}, h)$ вычисляется локально в любой точке x как сумма

$$F_N^{(\text{approx})}(x, \{x_1, x_2, \dots, x_N\}, h) = \frac{1}{N} \sum_{i=1}^N K\left(\frac{x - x_i}{h}\right), \quad (1)$$

где

$K(t)$ – так называемая ядерная функция, удовлетворяющая следующим условиям:

а) $K(t)$ – монотонная неубывающая функция, область значений которой принадлежит интервалу $[0,1]$;

б) $K(t) = 1 - K(-t)$ – ядерная функция, симметричная относительно 0;

h – размытости параметр (называемый также полосой пропускания ядра), что $h \rightarrow 0$ при $N \rightarrow \infty$.

Соответственно, аппроксимация ПР $f_N^{(\text{approx})}(x, \{x_1, x_2, \dots, x_N\})$ вычисляется по формуле

$$f_N^{(\text{approx})}(x, \{x_1, x_2, \dots, x_N\}, h) = \frac{1}{Nh} \sum_{i=1}^N k\left(\frac{y - x_i}{h}\right), \quad (2)$$

где

$$k(x) = \frac{d}{dx} K(x). \quad (3)$$

Ядерные функции, используемые в методе Розенблатта-Парзена

№	Ядро	Функция
1	Нормальное	$k(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$
2	Лапласа	$k(t) = \frac{1}{2} e^{- t }$
3	Фишера	$k(t) = \frac{1}{2\pi} \left(\frac{\sin\left(\frac{t}{2}\right)}{\frac{t}{2}} \right), \left \frac{t}{2} \right \leq \pi$
4	Коши	$k(t) = \frac{1}{\pi} \left(\frac{1}{1+t^2} \right)$
5	Логистическое	$k(t) = \frac{e^{-t}}{(1+e^{-t})^2}$
6	Епанечникова	$k(t) = \frac{3 \cdot \left(1 - \frac{t^2}{5}\right)}{4\sqrt{5}}, \quad t \leq \sqrt{5}$
7	Равномерное	$k(t) = \frac{1}{2}, \quad t \leq 1$
8	Треугольное	$k(t) = 1 - t , \quad t \leq 1$
9	Квадратичное	$k(t) = \frac{3 \cdot (1-t^2)}{4}, \quad t \leq 1$

Функции $k(x)$, связанные с функцией $K(x)$ соотношением (3), называемые ядерными функциями представлены в таблице 1.

При практическом использовании метода Розенблатта-Парзена приходится решать задачи:

1) выбора оптимальной ядерной функции $k(x)$;

2) нахождение для выбранной ядерной функции оценки оптимального значения параметра размытости h^* , обеспечивающего на выбранному критерию минимальное отклонение между аппроксимациями Розенблатта-Парзена и соответствующими эмпирическими ФР и ПР.

Отметим, что универсальных методов выбора ядерных функций и нахождения соответствующего оптимального значения размытости h^* , определяющего вид ФР (1) и ПР (2) (см., например, [10] и рисунки 2,3, обсуждаемые далее), несмотря на непрерывно продолжающийся поиск данных методов (см. например, [18–24]), в настоящее время не создано. Основная проблема здесь связана с тем, что в теоретические формулы, предложенные для расчета оптимального значения параметра размытости h^* , входит неизвестная теоретическая ПР $f^{(\text{theory})}(x)$ анализируемой выборки $\{x_1, x_2, \dots, x_N\}$. В качестве примера

приведем формулу, полученную как результат минимизации среднего накопленного квадрат ошибки (англ. – Mean Integrated Squared Error, MISE):

$$h^* = \arg \min_h \left[\int (f_N^{(\text{approx})}(\xi, h) - f^{(\text{theory})}(\xi))^2 d\xi \right],$$

называемую в этой связи асимптотической MISE оценкой оптимального значения параметра размытости:

$$h^* = \frac{R(k)^{1/5}}{m_2(k)^{2/5} R\left(\left(f^{(\text{theory})}\right)''\right)^{1/5} N^{1/5}}, \quad (4)$$

где

$$R(k) = \int k(\xi)^2 d\xi,$$

$$m_2(k) = \int \xi^2 k(\xi) d\xi.$$

Из (4) видно, что асимптотическая MISE-оценка оптимального значения параметра размытости h^* зависит от неизвестной плотности распределения $f^{(\text{theory})}(x)$. Отметим, что получить законченное аналитическое выражение для вычисления h^* :

$$h^* = \left(\frac{4\hat{\sigma}}{3N} \right)^{1/5} = 1.06\hat{\sigma}N^{-1/5},$$

где $\hat{\sigma}$ – среднеквадратическое отклонение анализируемой выборки $\{x_1, x_2, \dots, x_N\}$, в даже в предположении о виде теоретической ПР $f^{(\text{theory})}(x)$ удастся в единственном случае, когда удастся функции $f^{(\text{theory})}(x)$ $k(x)$ гауссовы. В остальных случаях найти аналогичные аналитические выражения, минимизируя средний накопленный квадрат ошибки разности теоретической ПР $f^{(\text{emp})}(x)$ и ее аппроксимации Розенблатта-Парзена $f_N^{(\text{approx})}(x)$

оказывается невозможным.

В этой связи на практике применяется следующий метод оценки эффективного значения параметра размытости h^* аппроксимации Розенблатта-Парзена ФР и ПР случайной выборки $\{x_1, x_2, \dots, x_N\}$ [25,10]. Для каждой из ядерных функций $k_m(t)$, $m = \overline{1,9}$, представленных в таблице 1, находят значения h_m^* , обеспечивающие максимальные значения информационных функционалов

$$J_m(k_m(h_m)) = \int \ln(k_m(h_m, \xi)) k_m(h_m, \xi) d\xi.$$

Далее сравнивают максимальные значения информационных функционалов $\max(J(h_m)) = J_m(k_m(h_m^*))$, и выбирают для вычисления аппроксимации Розенблатта-Парзена ту m_0 -ую ядерную функцию, у которой значение информационного функционала $J_{m_0}(k_{m_0}(h_{m_0}^*))$ оказывается наибольшим. Для

дискретных данных информационный функционал принимает вид:

$$J(k_m(h_m)) = \frac{1}{N} \sum_{i=1}^N \ln \left[\frac{1}{(N-1) \cdot h_m} \sum_{j \neq i}^{N-1} k_m \left(\frac{x_i - x_j}{h_m} \right) \right], \quad (6)$$

соответственно, задачи нахождения абсциссы его максимального значения и выбора оптимальной ядерной функции записываются в виде:

$$h_m^* = \arg \max_{h_m} [J(k_m(h_m))] = \arg \max \left\{ \frac{1}{N} \sum_{i=1}^N \ln \left[\frac{1}{(N-1) \cdot h_m} \sum_{j \neq i}^{N-1} k_m \left(\frac{x_i - x_j}{h_m} \right) \right] \right\}, \quad (7)$$

$$\{h_{m_0}^*, k_{m_0}\} = \arg \max_{h_m^*, k_m(x)} (J_m(k_m(h_m^*))). \quad (8)$$

Из (7) видно, что нахождение оптимального значения параметра размытости h_m^* для каждой из ядерных базисных функций $k_m(t)$ сводится к решению сложного нелинейного уравнения

$$\frac{\partial J(k_m(h_m))}{\partial h_m} = \frac{\partial}{\partial h_m} \sum_{i=1}^N \ln \left[\frac{1}{(N-1) \cdot h_m} \sum_{j \neq i}^{N-1} k_m \left(\frac{x_i - x_j}{h_m} \right) \right] = 0, \quad (9)$$

найти которое оказывается возможным только численно с помощью какого-либо известного итерационного метода. Так как сходимость итерационных методов решения нелинейных уравнений к искомому решению зависит от выбора начального приближения, который, априори, не очевиден, на практике ищут не решение (9), но вычисляют на выбранном интервале $[h_m^{(\min)}, h_m^{(\max)}]$ значения информационного функционала (6) и далее находят значение аргумента h_m^* , соответствующее максимальному значению информационного функционала $J(k_m(h_m))$ [10]. Однако, при использовании обсуждаемого подхода возникает целый ряд проблем, в том числе:

1) выбора интервала $[h_m^{(\min)}, h_m^{(\max)}]$, состоящая в том, что в условиях отсутствия правил обоснованного выбора границ интервала поиска максимального значения информационного функционала (6), искомый локальный минимум может отсутствовать на выбранном интервале;

2) наличия у вычисленной зависимости $J(k_m(h_m))$ разрывов первого и второго (пример подобной зависимости представлен на рисунке 1).

Из рисунка 1 видно, что при $h \in [10^{-4}, 10^{-1}]$ функция в точках отрезка $[1.1643 \cdot 10^{-3}, 1.1645 \cdot 10^{-3}]$ имеет разрывы первого рода, в точках отрезка $[2.6196 \cdot 10^{-3}, 2.6230 \cdot 10^{-3}]$ – разрывы второго рода. Для значений параметра размытости h , не принадлежащих данным отрезкам, функ-

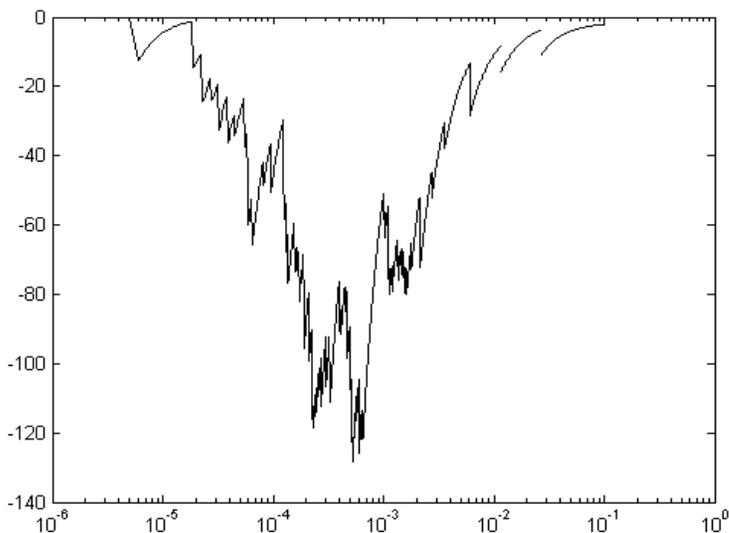


Рис. 1. График зависимости $\ln[J(k_1(h_1))] = f[\ln(h_1)]$, вычисленной для случайной последовательности $x_i, i = 1, 100$, сгенерированной в соответствии с нормальным законом распределения $N(1,4)$, в пакете MATLAB

ция $\ln[J(k_1(h_1))] = f[\ln(h_1)]$ оказывается непрерывной. Обсуждаемая особенность функции $J(k_1(h_1))$, которая у ядерных функций с ограниченной областью определения ($m = 3, 6, 7 - 9$) отсутствует, обусловлена конечной точностью машинной арифметики в (см., [7, Глава 3]). Также в [7]:

- обоснован выбор интервалов поиска оптимальных значений параметров размытости h_m^* и предложен алгоритм его поиска для ядерных функций с неограниченной областью определения (1, 2, 4, 5);

- получены аналитические выражения для вычисления соответствующих оптимальных значений параметров размытостей для ядерных функций с ограниченной областью определения ($m = 3, 6, 7, - 9$):

$$h_3^* = \frac{\max(x_i - x_j)}{2\pi}, \quad h_6^* = \frac{\max(x_i - x_j)}{\sqrt{5}},$$

$$h_{7,8,9}^* \geq \max(x_i - x_j);$$

- предложена модифицированная формула для вычисления значений информационного функционала $J(k(h_m))$ целочисленных случайных последовательностей:

$$J(k(h_m)) = \frac{1}{N} \sum_{i=1}^N \ln \left[\frac{1}{(N-1) \cdot h_m} \sum_{x_i \neq x_j}^{N-1} k_m \left(\frac{x_i - x_j}{h_m} \right) \right].$$

Отметим, что рассмотренный выше алгоритм реализации метода Розенблатта-Парзена имеет очевидный недостаток, связанный с тем, что здесь, де-факто, отсутствуют количественные критерии позволяющие количественно оценить качество аппроксимации

эмпирической ФР и ПР. Отмеченную проблему иллюстрирует рисунок 2, на котором представлены аппроксимации Розенблатта-Парзена с нормальной ядерной функцией случайной выборки $\{x_1, x_2, \dots, x_{50}\}$, сгенерированной в соответствии с логнормальным законом распределения

$$f_{LN}(x, \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}},$$

у которого $\mu = 0.3, \sigma = 0.03$.

Из рисунка 1 видно, что форма аппроксимаций ПР и ФР случайной выборки анализируемой случайной выборки $\{x_1, x_2, \dots, x_{50}\}$ определяется значением параметра размытости:

- график функция $F_{50}^{(\text{approx})}(x, \{x_1, x_2, \dots, x_{50}\}, 0.0029)$ при выбранном масштабе координатных осей визуально неотличима от соответствующего графика эмпирической ФР, одновременно, функция $f_{50}^{(\text{approx})}(x, \{x_1, x_2, \dots, x_{50}\}, 0.0029)$, имеющая 17 локальных максимумов, оказывается существенно отличной от гистограммы анализируемой случайной выборки $\{x_1, x_2, \dots, x_{50}\}$,

- по мере увеличения h функция $F_{50}^{(\text{approx})}(x, \{x_1, x_2, \dots, x_{50}\}, h)$ стремится к ФР случайной величины с равномерным законом распределения на интервале $[x_1, x_{50}]$ при этом, одновременно, уменьшается число локальных максимумов функции $f_{50}^{(\text{approx})}(x, \{x_1, x_2, \dots, x_{50}\}, h)$, которое оказывается равным 1 при $h=0.0349$ и 0 при $h=0.4028$.

Следовательно, при выборе оптимального значения параметра размытости h^* необходимо обеспечивать компромисс между

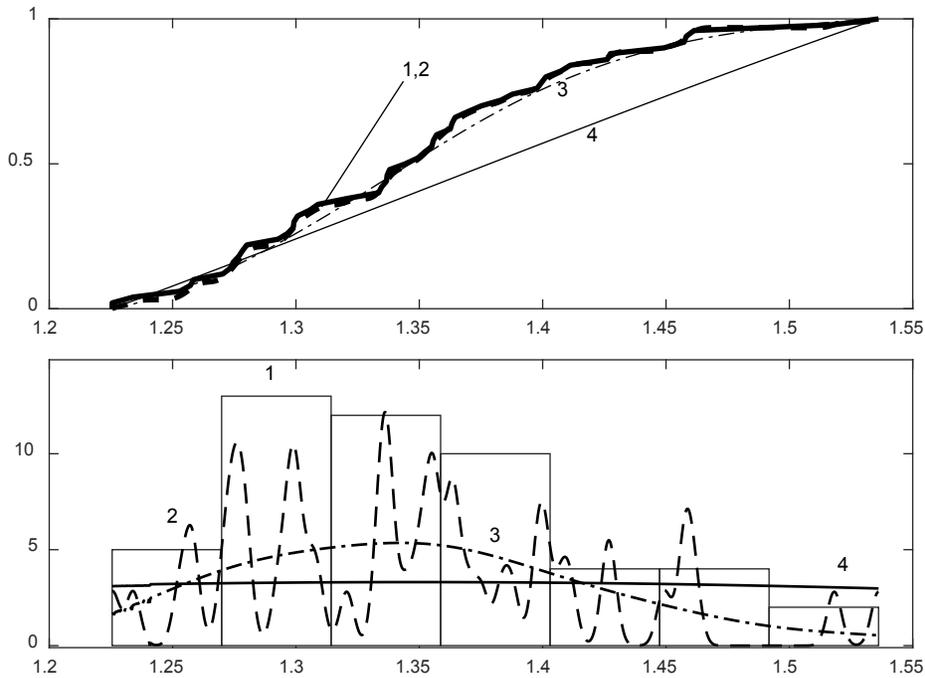


Рис. 2. Случайная выборка $\{x_1, x_2, \dots, x_{50}\}$, сгенерированная в соответствие с логнормальным законом распределения $f_{LN}(x, 0.3, 0.03)$: сверху 1 – эмпирическая ПР, 2 – график функции $F_{50}^{(approx)}(x, \{x_1, x_2, \dots, x_{50}\}, 0.0029)$, 3 – график функции $F_{50}^{(approx)}(x, \{x_1, x_2, \dots, x_{50}\}, 0.0349)$, 4 – график функции $F_{50}^{(approx)}(x, \{x_1, x_2, \dots, x_{50}\}, 0.4028)$; 1 – гистограмма случайной выборки $\{x_1, x_2, \dots, x_{50}\}$; 2 – график функции $f_{50}^{(approx)}(x, \{x_1, x_2, \dots, x_{50}\}, 0.0029)$, 3 – график функции $f_{50}^{(approx)}(x, \{x_1, x_2, \dots, x_{50}\}, 0.0349)$, 4 – график функции $f_{50}^{(approx)}(x, \{x_1, x_2, \dots, x_{50}\}, 0.4028)$

требованием «похожести» функции $F_{50}^{(approx)}(x, \{x_1, x_2, \dots, x_{50}\}, h)$ и эмпирической ФР, а также «похожести» функции $f_{50}^{(approx)}(x, \{x_1, x_2, \dots, x_{50}\}, h)$ и эмпирической ФР, в то время как устоявшаяся точка зрения рекомендует находить h^* как решение задачи (7), в постановке которой отсутствуют и эмпирическая ПР и эмпирическая ФР.

В этой связи авторы провели самостоятельные исследования и получили количественные качества аппроксимаций Розенблатта-Парзена ПР и ФР случайных выборок, обсуждаемые далее.

Методика анализа качества аппроксимаций Розенблатта-Парзена ПР и ФР случайных выборок

В качестве критериев качества аппроксимации Розенблатта-Парзена были выбраны следующие критерии, традиционно используемые в задачах аппроксимации функций, заданных таблично, и функций распределений независимых случайных выборок с законами распределения $F_1(x)$ и $F_2(x)$:

– накопленная среднеквадратическая ошибка разности между функциями $F_1(x)$ и $F_2(x)$, вычисляемая по формуле

$$M_1 = \left[\frac{1}{x_N - x_1} \int_{x_1}^{x_N} (F_1(\xi) - F_2(\xi))^2 d\xi \right]^{1/2},$$

которая для дискретных выборок записывается в виде:

$$M_1 = \left[\frac{1}{N} \sum_{j=1}^N (F_1(x_j) - F_2(x_j))^2 \right]^{1/2}; \quad (10)$$

– критерий типа Колмогорова-Смирнова:

$$M_2 = \max \left(|F_1(x_i) - F_2(x_j)| \right), \quad (11)$$

где $i = 1, N, j = 1, N_h$, используемый для проверки статистических гипотез вида:

$$H_0 : F_1(x) = F_2(x),$$

$$H_1 : F_1(x) \neq F_2(x).$$

Исследование особенностей качества аппроксимаций ПР и ФП случайных выборок было проведено в соответствие с методикой, реализующейся выполнением следующей последовательности действий.

1. Выбор закона распределения случайной выборки $F^{(theory)}(x, \alpha)$, α – вектор параметров.

2. Выбор объема случайной выборки N .

3. Генерация случайной выборки $\{\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_N\}$ в соответствие в выбранным законом распределения $F^{(theory)}(x, \alpha)$.

4. Вычисление в соответствии с методом максимального правдоподобия $\tilde{\alpha}$ – оценок параметров закона распределения $F^{(estimation)}(x, \tilde{\alpha})$ случайной выборки $\{\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_N\}$.

5. Упорядочивание случайной выборки $\{\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_N\}$ в порядке возрастания значений $\tilde{x}_i, i = 1, N$ (далее – случайная выборка $\{\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_N\}$).

6. Выбор диапазона изменения значений параметра размытости $h, h \in [h_{\min}, h_{\max}]$.

7. Выбор N_h – числа узлов сетки по параметру размытости h .

8. Вычисление координат узлов сетки по параметру размытости h :

$$h_j = h_{\min} + \frac{h_{\max} - h_{\min}}{N_h - 1}(j - 1), \quad j = \overline{1, N_h}.$$

9. Выбор N_x – числа узлов сетки по переменной x , используемой для вычисления значений аппроксимаций Розенблатта-Парзена ПР случайной выборки $\{x_1, x_2, \dots, x_N\}$.

10. Вычисление координат узлов сетки по переменной x :

$$y_k = x_1 + \frac{x_N - x_1}{N_x - 1}(k - 1), \quad k = \overline{1, N_x}. \quad (12)$$

11. Выбор ядерной функции ядерной функции $k(t)$.

12. Вычисление аппроксимаций Розенблатта-Парзена с использованием выбранной ядерной функции ПР случайной выборки $\{x_1, x_2, \dots, x_N\}$ в каждом узле сетки по параметру

$h f^{(approx)}(y_k, h_j), j = \overline{1, N_h}, y_k$ – координаты узлов сетки (12).

13. Вычисление численным интегрированием аппроксимаций Розенблатта-Парзена ФР случайной выборки $\{x_1, x_2, \dots, x_N\}$ в

каждом узле сетки по параметру h $F^{(approx)}(y_k, h_j), j = \overline{1, N_h}$:

$$F^{(approx)}(y_k, h_j) = \int_{x_j}^{y_k} f^{(approx)}(\xi, h_j) d\xi.$$

14. Нормировка аппроксимаций Розенблатта-Парзена ФР случайной выборки $\{x_1, x_2, \dots, x_N\}$ $F^{(approx)}(y_k, h_j)$, вычисленных в каждом узле сетки по параметру $h, k = \overline{1, N_x}, j = \overline{1, N_h}$:

$$F^{(norm approx)}(y_k, h_j) = F^{(approx)}(y_k, h_j) / F^{(approx)}(x_N, h_j).$$

15. Вычисление на основе использования зависимости $F^{(norm approx)}(y_k, h_j), k = \overline{1, N_x}, j = \overline{1, N_h}$ с помощью линейной интерполяции значений аппроксимации Розенблатта-Парзена ФР случайной выборки $\{x_1, x_2, \dots, x_N\}$ в точках $x_1, x_2, \dots, x_N - F^{(norm approx)}(x_i, h_j), i = \overline{1, N}, j = \overline{1, N_h}$.

16. Вычисление эмпирической ФР случайной выборки $\{x_1, x_2, \dots, x_N\}$ $F^{(emp)}(x_i, h_j), i = \overline{1, N}, j = \overline{1, N_h}$.

17. Вычисление зависимости $M_1(h)$ в соответствии с (10) для следующих пар зависимостей $F_1(x), F_2(x)$, представленных в таблице 2.

Таблица 2

Пары ФР и ПР, использованных для вычисления зависимостей критерия оценки качества аппроксимации Розенблатта-Парзена M_1 от параметра размытости h

$F_1(x)$	$F^{(theory)}(x_i, \alpha)$	$F^{(estimation)}(x_i, \tilde{\alpha})$	$f^{(theory)}(x_i, \alpha)$	$f^{(estimation)}(x_i, \tilde{\alpha})$
$F_2(x)$	$F^{(norm approx)}(x_i, h_j)$	$F^{(norm approx)}(x_i, h_j)$	$f^{(norm approx)}(x_i, h_j)$	$f^{(norm approx)}(x_i, h_j)$

Таблица 3

Пары зависимостей, использованных для вычисления зависимостей критерия оценки качества аппроксимации Розенблатта-Парзена M_1 от параметра размытости h

$F_1(x)$	$F^{(theory)}(x_i, \alpha)$	$F^{(estimation)}(x_i, \tilde{\alpha})$
$F_2(x)$	$F^{(norm approx)}(x_i, h_j)$	$F^{(norm approx)}(x_i, h_j)$

18. Вычисление зависимостей $M_2(h_j)$ в соответствии с (11) на основе результатов проверки статистических гипотез для каждого из узлов выбранной сетки по параметру размытости h :

$$H_0 : F_1(x) = F_2(x), H_1 : F_1(x) \neq F_2(x),$$

если верна H_0 $H(h_j) = 0$ иначе $H(h_j) = 1$ для следующих пар зависимостей $F_1(x), F_2(x)$, представленных в таблице 3.

19. Вычисление для выбранной ядерной функции в соответствие (6) зависимости информационного функционала $J(h)$ от параметра размытости h .

20. Вычисление для выбранного ядра оптимального в смысле (7) значения параметра размытости h^* .

21. Анализ зависимостей $M_1(h_j)$, $M_2(h_j)$, $f^{(approx)}(x_k, h_j)$, $F^{(norm approx)}(x_k, h_j)$, $H(h_j)$, $i = \overline{1, N}$, $j = \overline{1, N_h}$.

Анализ качества аппроксимаций Розенблатта-Парзена ПР и ФР случайных выборок, вычисляемых на основе анализа зависимости

$$J = J(h)$$

Рассмотрим типичные зависимости $M_1(h_j)$, $M_2(h_j)$ для случайной выборки $\{x_1, x_2, \dots, x_{50}\}$, сгенерированной в соответствие с нормальным законом распределения $N(1,3)$, вычисленные для аппроксимаций Розенблатта-Парзена эмпирических ПР и ФР с нормальной ядерной функцией (см. рисунок 3).

Зависимости

$$J = J(h),$$

$$M_1^{(1)} = M_1(F^{(theory)}(x_i, \alpha), F^{(approx)}(x_i, h_i), h_i),$$

$$M_1^{(2)} = M_1(F^{(estimation)}(x_i, \tilde{\alpha}), F^{(approx)}(x_i, h_i), h_i),$$

$$M_1^{(3)} = M_1(f^{(theory)}(x_i, \alpha), f^{(approx)}(x_i, h_i), h_i),$$

$$M_1^{(4)} = M_1(f^{(estimation)}(x_i, \tilde{\alpha}), f^{(approx)}(x_i, h_i), h_i),$$

представлены на рисунке 4.

Из рисунка 4 видно, что на рассматриваемом интервале зависимость $J=J(h)$ имеет локальный максимум в точке $h^*=2.34$; зависимости $M_1^{(1)}(h_i)$, $M_1^{(2)}(h_i)$ достигают локального минимума в точках $h_{min}^{(1)} = 3.01$, $h_{min}^{(2)} = 3.65$, соответственно; зависимости $M_1^{(3)}(h_i)$, $M_1^{(4)}(h_i)$ – в точках $h_{min}^{(3)} = 3.09$, $h_{min}^{(4)} = 2.67$. Значения зависимостей $J = J(h)$, $M_1^{(1)}(h_i)$, $M_1^{(2)}(h_i)$, $M_1^{(3)}(h_i)$, $M_1^{(4)}(h_i)$ для указанных выше значений аргумента h^* , $h_{min}^{(1)}$, $h_{min}^{(2)}$, $h_{min}^{(3)}$, $h_{min}^{(4)}$ представлены в таблице 3.

Из таблицы 3 видно, что $M_1^{(1)}(h^*) > M_1^{(1)}(h_{min}^{(1)})$, $M_1^{(2)}(h^*) > M_1^{(2)}(h_{min}^{(2)})$.

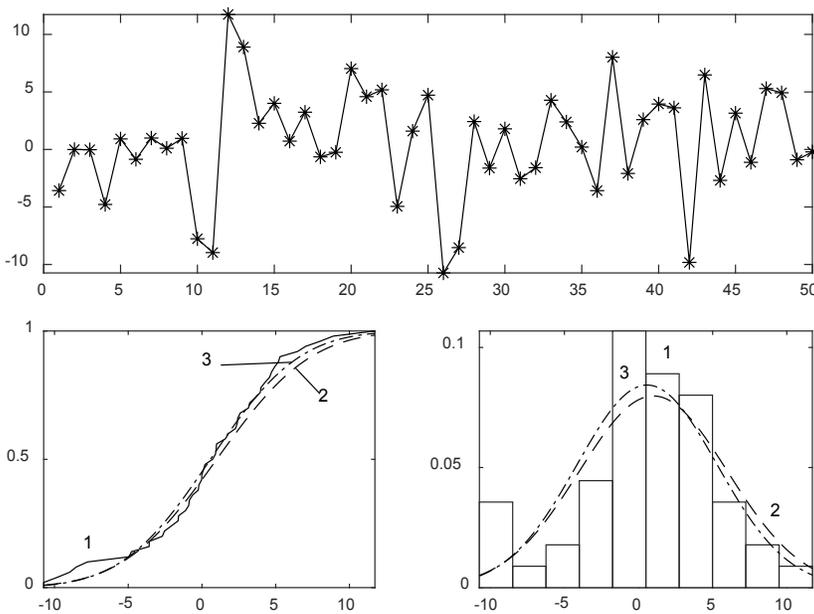


Рис. 3. Случайная выборка $\{X_1, X_2, \dots, X_{50}\}$, сгенерированная в соответствие с нормальным законом распределения $N(1,4)$, и ее статистические характеристики (на рисунке внизу слева: 1 – $F^{(emp)}(x)$, 2 – $F^{(theory)}(x, \alpha)$, 3 – $F^{(estimation)}(x, \tilde{\alpha})$; на рисунке внизу справа: 1 – гистограмма случайной выборки $\{X_1, X_2, \dots, X_{50}\}$, 2 – $f^{(theory)}(x, \alpha)$, 3 – $f^{(estimation)}(x, \tilde{\alpha})$)

$$M_1^{(3)}(h^*) > M_1^{(3)}(h_{min}^{(3)}), M_1^{(4)}(h^*) > M_1^{(4)}(h_{min}^{(4)}).$$

При этом минимальное значение вероятности принятия гипотезы H_0 для значений параметра размытости $h_{min}^{(1)}$, $h_{min}^{(2)}$, $h_{min}^{(3)}$, $h_{min}^{(4)}$, представленных в таблице 3, оказалось равным 0.91.

Аналогично результаты были получены и случайных выборок, генерируемых в соответствие с другими известными законами распределения, что иллюстрируют, например, результаты аппроксимации случайной выборки объемом $N = 50$, сгенерированной в соответствие с логнормальным законом распределения

$$f_{LN}(x, \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}},$$

в котором $\mu = 0.3$, $\sigma = 0.03$, представленные на рисунках 5, 6 и таблице 4.

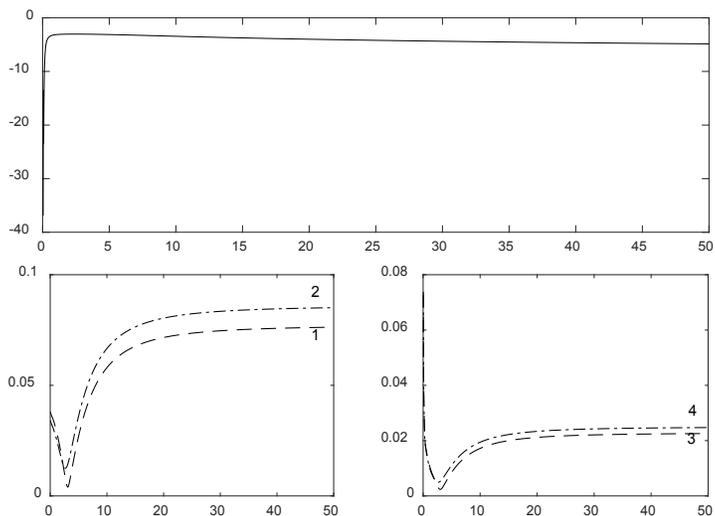


Рис. 4. Случайная выборка, сгенерированная в соответствие с нормальным законом $N(1,4)$. Графики зависимостей: $J = J(h)$ (сверху); $M_1^{(1)}(h_i)$, $M_1^{(2)}(h_i)$ (внизу слева); $M_1^{(3)}(h_i)$, $M_1^{(4)}(h_i)$ (внизу справа)

Таблица 3

Значения зависимостей $M_1^{(1)}(h_i)$, $M_1^{(2)}(h_i)$, $M_1^{(3)}(h_i)$, $M_1^{(4)}(h_i)$ в выбранных точках (случайная выборка, сгенерированная в соответствие с нормальным законом $N(1,4)$ распределения

hi	$M_1^{(1)}(h_i)$	$M_1^{(2)}(h_i)$	$M_1^{(3)}(h_i)$	$M_1^{(4)}(h_i)$
2.34	0.0129	0.0137	0.0042	0.0052
3.01	0.0040	—	—	—
3.65	—	0.0125	—	—
3.09	—	—	0.0024	—
2.67	—	—	—	0.0050

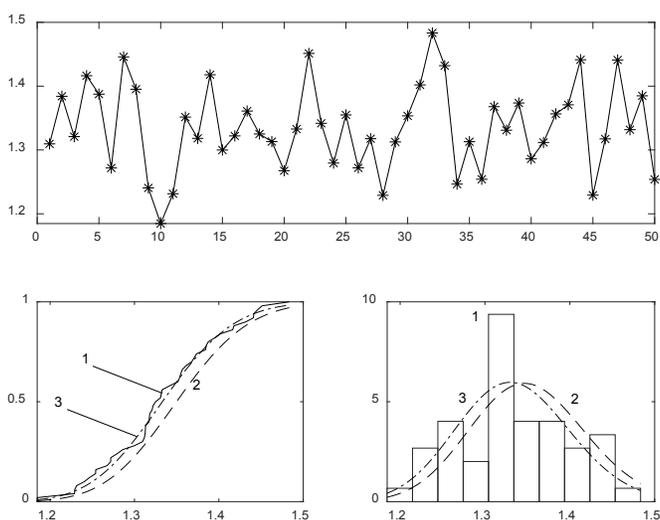


Рис. 5. Случайная выборка $\{X_1, X_2, \dots, X_{50}\}$, сгенерированная в соответствие с логнормальным законом распределения $f_{LN}(x, 0.3, 0.03)$, и ее статистические характеристики (на рисунке внизу слева: 1 – $F^{(emp)}(x)$, 2 – $F^{(theory)}(x, \alpha)$, 3 – $F^{(estimation)}(x, \hat{\alpha})$; на рисунке внизу справа: 1 – гистограмма случайной выборки $\{X_1, X_2, \dots, X_{50}\}$, 2 – $f^{(theory)}(x, \alpha)$, 3 – $f^{(estimation)}(x, \hat{\alpha})$)

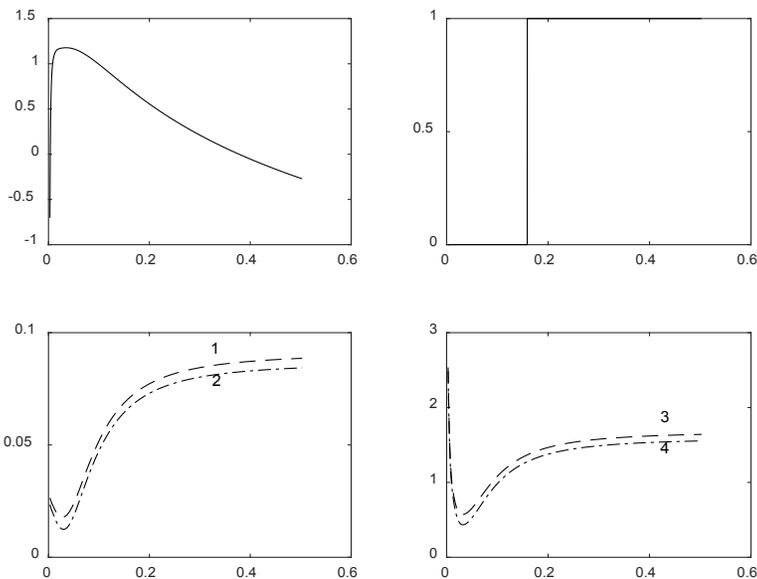


Рис. 6. Случайная выборка $\{x_1, x_2, \dots, x_{50}\}$, сгенерированная в соответствии с логнормальным законом распределения $f_{LN}(x, 0.3, 0.03)$. Графики зависимостей: $J = J(h)$ (сверху слева); $M_2(h_i)$ (сверху справа); $M_1^{(1)}(h_i)$, $M_1^{(2)}(h_i)$ (внизу слева); $M_1^{(3)}(h_i)$, $M_1^{(4)}(h_i)$ (внизу справа)

Таблица 4

Значения зависимостей $M_1^{(1)}(h_i)$, $M_1^{(2)}(h_i)$, $M_1^{(3)}(h_i)$, $M_1^{(4)}(h_i)$ в выбранных точках (случайная выборка $\{x_1, x_2, \dots, x_{50}\}$, сгенерированная в соответствии с логнормальным законом распределения $f_{LN}(x, 0.3, 0.03)$)

h_i	$M_1^{(1)}(h_i)$	$M_1^{(2)}(h_i)$	$M_1^{(3)}(h_i)$	$M_1^{(4)}(h_i)$
0.0361	0.0299	0.0154	0.7102	0.4194
3.01	0.0293	–	–	–
3.65	–	0.0095	–	–
3.09	–	–	0.7065	–
2.67	–	–	–	0.3588

Из рисунка 6 видно, что значения параметров размытости h^* , $h_{\min}^{(1)}$, $h_{\min}^{(2)}$, $h_{\min}^{(3)}$, $h_{\min}^{(4)}$ таковы, что они попадают в диапазон значений параметра размытости $[0, 0.1589]$, в котором принимается гипотеза H_0 .

Из таблицы 4 видно, что, как и в предыдущем случае, $M_1^{(1)}(h^*) > M_1^{(1)}(h_{\min}^{(1)})$, $M_1^{(2)}(h^*) > M_1^{(2)}(h_{\min}^{(2)})$, $M_1^{(3)}(h^*) > M_1^{(3)}(h_{\min}^{(3)})$, $M_1^{(4)}(h^*) > M_1^{(4)}(h_{\min}^{(4)})$. При этом минимальное значение вероятности принятия гипотезы H_0 для значений параметра размытости $h_{\min}^{(1)}$, $h_{\min}^{(2)}$, $h_{\min}^{(3)}$, $h_{\min}^{(4)}$, представленных в таблице 3, оказалось более 0.78.

Таким образом, обсуждаемые выше результаты исследований опровергают устоявшуюся точку зрения о том, что максимально возможная точность аппроксимации Розенблатта-Парзена ПР и ФР случайной выборки $\{x_1, x_2, \dots, x_N\}$, оцениваемая накопленным сред-

неквадратическим отклонением разностей $F^{(theory)}(x_i, \alpha) - F^{(norm approx)}(x_i, h_j)$, $F^{(estimation)}(x_i, \tilde{\alpha}) - F^{(norm approx)}(x_i, h_j)$, $f^{(theory)}(x_i, \alpha) - f^{(norm approx)}(x_i, h_j)$, $f^{(estimation)}(x_i, \tilde{\alpha}) - f^{(norm approx)}(x_i, h_j)$, достигается при использовании оптимального значения параметра размытости h^* , являющегося решением задачи (7).

В этой связи возникает необходимость использования альтернативных методов оценивания оптимального значения параметра размытости h^* , основанных на сравнении доступных исследователю эмпирических ФР и ПР с их аппроксимациями Розенблатта-Парзена, например, на основе анализа накопленной среднеквадратической ошибки $M_1^{(5)}(h) = M_1(f^{(approx)}(x_j, h), f_N^{(emp)}(x_j))$,

где $f_N^{(emp)}(x_i)$ - кусочно-постоянная аппроксимация эмпирической ПР случайной выборки $\{x_1, x_2, \dots, x_N\}$,

$$f_N^{(emp)}(y) = \frac{F_N^{(emp)}(y+h) - F_N^{(emp)}(y)}{h} = \frac{1}{Nh} \sum_{i=1}^N [\theta(y+h-x_i) - \theta(y-x_i)] = \frac{v_y}{Nh}, \quad (13)$$

здесь v_y - количество значений случайной выборки $\{x_1, x_2, \dots, x_N\}$, попавших в интервал $[y, y+h]$, обсуждаются в следующем разделе.

Их выбор обусловлен тем, что исследователю статистических характеристик случайных выборок оказываются доступными исключительно эмпирические ФР и ПР анализируемых выборок.

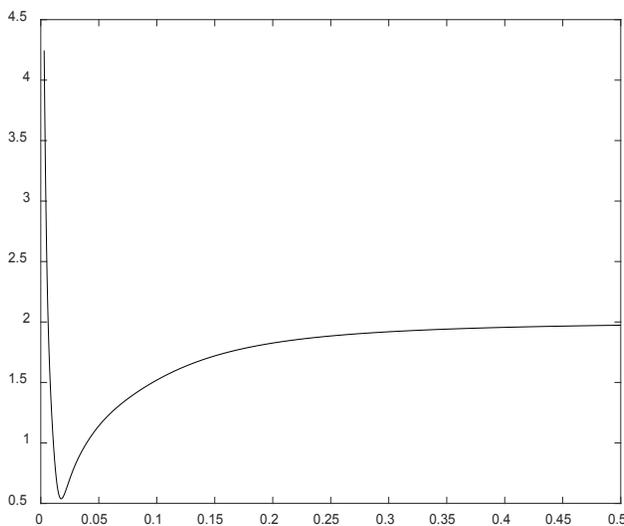


Рис. 6. Случайная выборка $\{x_1, x_2, \dots, x_{50}\}$, сгенерированная в соответствии с логнормальным законом распределения $f_{LN}(x, 0.3, 0.03)$. График зависимости $M_1^{(s)} = M_1^{(s)}(h)$

зависимость достигает $M_1^{(s)}(h)$ своего минимального 0.5385 в точке $h_{\min}^{(s)} = 0.0176$. Оценка оптимального значения параметра размытости для данной выборки, вычисленная как решение задачи (7), оказалась равной $h^* = 0.0156$, значения анализируемых зависимостей в выбранных точках $M_1^{(s)}(h^*) = 0.9208$, $M_1^{(s)}(h_{\min}^{(s)}) = 0.5385$. Так как $M_1^{(s)}(h^*) > M_1^{(s)}(h_{\min}^{(s)})$, качество аппроксимации Розенблатта-Парзена эмпирической ПР для значения $h = h_{\min}^{(s)}$ выше качества обсуждаемых аппроксимаций при $h = h^*$. Соответствующие графики функций $F^{(emp)}(x)$, $F^{(norm approx)}(x, h_{\min}^{(s)})$, $F^{(norm approx)}(x, h^*)$, $f^{(emp)}(x)$, $f^{(approx)}(x, h_{\min}^{(s)})$, $f^{(approx)}(x, h^*)$ представлены на рисунке 7.

Аналогичные свойства имеют аппроксимации Розенблатта-Парзена эмпирических ФР и ПР с иными другими ядерными функциями случайных выборок, сгенерированных в

Анализ качества аппроксимаций Розенблатта-Парзена ПР и ФР случайных выборок на основе анализа зависимости $M_1^{(s)}(h)$

Рассмотрим типичную зависимость $M_1^{(s)}(h)$ случайной выборки $\{x_1, x_2, \dots, x_{50}\}$, сгенерированной в соответствии с логнормальным законом распределения $f_{LN}(x, \mu, \sigma)$ $\mu = 0.3$, $\sigma = 0.03$, представленную на рисунке 6. Здесь число интервалов кусочно-постоянной аппроксимации эмпирической ПР (гистограммы) N_g выбиралось в соответствии с правилом Стерджеса:

$$N_g = \lceil 1 + \log_2 N \rceil,$$

где $\lceil \cdot \rceil$ - целая часть выражения.

Из рисунка 6 видно, что анализируемая

соответствие с законами распределений, отличными от рассмотренными в статье.

Заключение

Результаты проведенных исследований особенностей аппроксимации Розенблатта-Парзена с помощью нормальной ядерной функции $k(t)$ ПР и ФР упорядоченных случайных выборок $\{x_1, x_2, \dots, x_n\}$, сгенерированных в соответствии с нормальным и логнормальным распределениями опровергают устоявшуюся точку зрения, рекомендующую использовать в качестве оптимального значения параметра размытости h^* решение задачи (7).

Оценку данного параметра следует находить как значение параметра размытости $h_{\min}^{(s)}$, которое обеспечивает достижение минимального значения накопленной случайной ошибки разности между эмпирической

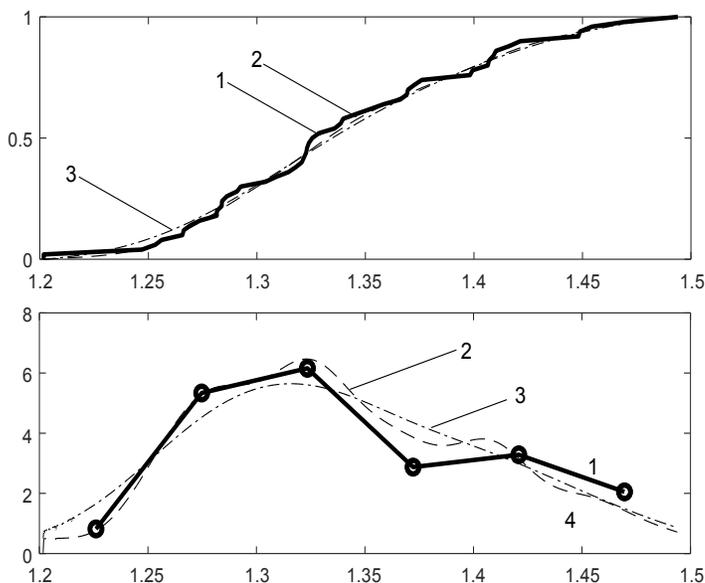


Рис. 7. Графики функций: сверху – 1 – $F^{(emp)}(x)$, 2 – $F^{(norm approx)}(x, h_{min}^{(s)})$, 3 – $F^{(norm approx)}(x, h^*)$; снизу – 1 – $f^{(emp)}(x)$, 2 – $f^{(approx)}(x, h_{min}^{(s)})$, 3 – $f^{(approx)}(x, h^*)$

ПР $f^{(emp)}(x)$ и аппроксимацией Розенблатта-Парзена ПР $f^{(approx)}(x, h)$ анализируемой случайной выборки $\{x_1, x_2, \dots, x_N\}$. При этом следует проводить проверку с помощью критерия типа Колмогорова-Смирнова статистической гипотезы о том, что $H_0: F^{(emp)}$

$(x) = F^{(norm approx)}(x, h_{min}^{(s)})$. При отклонении гипотезы H_0 в качестве оптимального значения параметра использовать левую границу интервала значений параметра размытости $h - h^{(s)}$, на котором гипотеза H_0 оказывается верной.

Литература

1. Левин Б. Р. Вероятностные модели и методы в системах связи и управления/Б. Р. Левин, В. Шварц// – М.: Радио и связь, 1985. – 312 с.
2. Тихонов В. И. Статистическая радиотехника/В.И. Тихонов// –М.: Радио и связь, 1982. – 624 с.
3. Кропотов Ю. А Методы оценивания моделей плотности вероятностей акустических сигналов в телекоммуникациях аудиообмена/ Ю.А. Кропотов// Системы управления, связи и безопасности, 2017. № 1. С. 26–39.
4. Кропотов Ю.А. Моделирование и методы исследования акустических сигналов, шумов и помех в системах телекоммуникаций/ Ю.А. Кропотов, В.А. Ермолаев// – Берлин: Директ-Медиа, 2016. – 251 с.
5. Сушко Д.В. Оптимальная аппроксимация частотных вероятностей/Д.В. Сушко//Информационные процессы, 2018. Том 18. № 1. С. 40–54.
6. Новоселов А.А. Параметризация моделей управляемых систем // Вестник государственного аэрокосмического университета им. академика М.Ф. Решетнева, 2010. № 5 (31). С. 52–56.
7. Поршнев С.В., Копосов А.С. Случайные величины с ограниченной областью рассеяния: математическое и алгоритмическое обеспечение для оценивания плотностей вероятностей и функций распределений/С.В. Поршнев, А.С. Копосов// –М.: Горячая линия-Телеком, 2018. –184 с.
8. Киреева Н.В., Чупахина Л.Р. Сравнение возможностей использования различных методов аппроксимации для анализа трафика с самоподобным распределением/Н.В. Киреев, Л.Н. Чупахина// Международный журнал прикладных и фундаментальных исследований, 2016. № 12. С. 1287–1289.
9. Поршнев С.В., Божалкин Д.А. Математическое и алгоритмическое обеспечение для анализа характеристик информационных потоков в магистральных интернет-каналах/С.В. Поршнев, Д.А. Божалкин// –М.: Горячая линия-Телеком, 2021. –214 с.
10. Сызранцев В.Н. Расчет прочностной надежности изделий на основе методов непараметрической статистики/В.Н. Сызранцев, Я.Н. Невелев, С.Л. Голофаст// –Новосибирск: Наука, 2008. –128 с.
11. Кобзарь А.И. Прикладная математическая статистика. Для инженеров и научных работников/ А.И. Кобзарь// –М.: ФИЗМАТЛИТ, 2006. –816 с.

12. Rosenblatt M. Remarks on Some Nonparametric Estimates of a Density Function // *The Annals of Mathematical Statistics.*, 1956. – Т. 27. Vol. 3. –Р. 832–837.
13. Боровков А.А. Математическая статистика. – М.: Наука, 1984, 472 с.
14. Parzen E. On the estimation of probability density function and the mode // *Ann. Math. Stat.*, 1962. – Vol. 33. – Р. 1065–1076.
15. Суетин П.К. Классические ортогональные многочлены. – М.: Наука, 1976. – 328 с.
16. Димаки А.В., Светлаков А.А. Аппроксимация плотностей распределений случайных величин с применением ортогональных полиномов Чебышева–Эрмита/А.В. Димаки, А.А. Светлаков// *Известия Томского политехнического университета*, 2006. Т. № 8. С. 6–11.
17. Голик Ф. В. Аппроксимация эмпирических распределений вероятностей полиномами Бернштейна/ Ф.В. Голик// *Журнал радиоэлектроники [Электронный журнал]*. 2018. № 7. Режим доступа: <http://jre.cplire.ru/jre/jul18/5/text.pdf> (дата обращения 16 октября 2021 г.) DOI 10.30898/1684-1719.2018.7.5.
18. Лемешко Б.Ю. О нахождении параметра размытости непараметрических оценок функции плотности/ Б.Ю. Лемешко, С.Н. Постовалов, А.В. Французов// *Труды V международной конференции «Актуальные проблемы электронного приборостроения», АПЭП-2000. Новосибирск 26–29 сентября, 2000. В 7 тт.* –Новосибирск: Новосибирский государственный технический университет, 2000. –Т. 6. –С. 17–20.
19. Botev. Z.I., Grotowski. J.F., Kroese. D.P. Kernel density estimation via diffusion, *Annal Statistic*, October 2010, Vol. 38 (5), pp. 2916–2957, DOI: <https://doi.org/10.1214/10-AOS799>.
20. Silverman B.W. *Density Estimation for Statistics and Data Analysis.* –Chapman and Hall/CRC, 1986. –175 p.
21. Leiva-Murillo J.M., Artés-Rodríguez A. Algorithms for maximum-likelihood bandwidth selection in kernel density estimators. *Pattern Recognition Letters*, Volume 33, Issue 13, 2012, pp. 1717–1724.
22. Xu Xiaoyuan, Yan Zheng, Xu Shaolun Estimating wind speed probability distribution by diffusion-based kernel density method, *Electric Power Systems Research*, Volume 121, April 2015, pp. 28–37.
23. Ouarda T.B.M.J., Charron C., Shin J.-Y., Marpu P.R., Al-Mandoos A.H., Al-Tamimi, Ghedira H., Al Hosary T.N, Probability distributions of wind speed in the UAE, *Energy Conversion and Management*, Volume 93, 2015, pp. 414-434.
24. Jones M.C., Marron J.S., Sheather S.J. A brief survey of bandwidth selection for density estimation, *Journal of the American Statistical Association*, Volume 91, № 433, March 1996, pp.401–4071.
25. Симахин В.А. Робастные непараметрические оценки: адаптивные оценки взвешенного максимального правдоподобия в условиях статистической априорной неопределенности/ В.А. Симахин // –Saarbrucken, Germany: LAP LAMBERT Academic Publishing GmbH & Co. KG, 2011. –292 с.

References

1. Levin B. R. Veroyatnostnyye modeli i metody v sistemakh svyazi i upravleniya/B. R. Levin, V. Shvarts// – М.: Radio i svyaz', 1985. – 312 с.
2. Tikhonov V. I. Statisticheskaya radiotekhnika/V.I. Tikhonov// –М.: Radio i svyaz', 1982. – 624 с.
3. Kropotov YU. A Metody otsenivaniya modeley plotnosti veroyatnostey akusticheskikh signalov v telekommunikatsiyakh audioobmena/YU.A. Kropotov// *Sistemy upravleniya, svyazi i bezopasnosti*, 2017. № 1. S. 26–39.
4. Kropotov YU.A. Modelirovaniye i metody issledovaniya akusticheskikh signalov, шумов i pomekh v sistemakh telekommunikatsiy/YU.A. Kropotov, V.A. Yermolayev// – Berlin: Direkt-Media, 2016. – 251 с.
5. Sushko D.V. Optimal'naya approksimatsiya chastotnykh veroyatnostey/D.V. Sushko// *Informatsionnyye protsessy*, 2018. Tom 18. № 1. S. 40–54.
6. Novoselov A.A. Parametrizatsiya modeley upravlyayemykh sistem // *Vestnik gosudarstvennogo aerokosmicheskogo universiteta im. akademika M.F. Reshetneva*, 2010. № 5 (31). S. 52–56.
7. Porshnev S.V., Koposov A.S. Sluchaynyye velichiny s ogranichennoy oblast'yu rasseyaniya: matematicheskoye i algoritmicheskoye obespecheniye dlya otsenivaniya plotnostey veroyatnostey i funktsiy raspredeleniy/S.V. Porshnev, A.S. Koposov// –М.: Goryachaya liniya-Telekom, 2018. –184 с.
8. Kireyeva N.V., Chupakhina L.R. Sravneniye vozmozhnostey ispol'zovaniya razlichnykh metodov approksimatsii dlya analiza trafika s samopodobnym raspredeleniyem/N.V. Kireyev, L.N. Chupakhina// *Mezhdunarodnyy zhurnal prikladnykh i fundamental'nykh issledovaniy*, 2016. № 12. S. 1287–1289.
9. Porshnev S.V., Bozhalkin D.A. Matematicheskoye i algoritmicheskoye obespecheniye dlya analiza kharakteristik informatsionnykh potokov v magistral'nykh internet-kanalakh/S.V. Porshnev, D.A. Bozhalkin// –М.: Goryachaya liniya-Telekom, 2021. –214 с.

10. Syzrantsev V.N. Raschet prochnostnoy nadezhnosti izdeliy na osnove metodov neparametricheskoy statistiki/V.N. Syzrantsev, YA.N. Nevelev, S.L. Golofast// –Novosibirsk: Nauka, 2008. –128 s.

11. Kobzar' A.I. Prikladnaya matematicheskaya statistika. Dlya inzhenerov i nauchnykh rabotnikov/A.I. Kobzar'// –M.: FIZMATLIT, 2006. –816 s.

12. Rosenblatt M. Remarks on Some Nonparametric Estimates of a Density Function // The Annals of Mathematical Statistics, 1956. – T. 27. Vol. 3. –P. 832–837.

13. Borovkov A.A. Matematicheskaya statistika. – M.: Nauka, 1984, 472 s.

14. Parzen E. On the estimation of probability density function and the mode // Ann. Math. Stat., 1962. – Vol. 33. – P. 1065–1076.

15. Suyetin P.K. Klassicheskoye ortogonal'nyye mnogochleny. – M.: Nauka, 1976. – 328 s.

16. Dimaki A.V., Svetlakov A.A. Approksimatsiya plotnostey raspredeleniy sluchaynykh velichin s primeneniym ortogonal'nykh polinomov Chebysheva–Ermita/A.V. Dimaki, A.A. Svetlakov// Izvestiya Tomskogo politekhnicheskogo universiteta, 2006. T. № 8. S. 6–11.

17. Golik F.V. Approksimatsiya empiricheskikh raspredeleniy veroyatnostey polinomami Bernshteyna/ F.V. Golik// Zhurnal radioelektroniki [elektronnyy zhurnal]. 2018. № 7. Rezhim dostupa: <http://jre.cplire.ru/jre/jul18/5/text.pdf> (data obrashcheniya 16 oktyabrya 2021 g.) DOI 10.30898/1684-1719.2018.7.5.

18. Lemeshko B.YU. O nakhozhenii parametra razmytosti neparametricheskikh otsenok funktsii plotnosti/ B.YU. Lemeshko, S.N. Postovalov, A.V. Frantsuzov// Trudy V mezhdunarodnoy konferentsii «Aktual'nyye problemy elektronnoy priborostroyeniya», APEP-2000. Novosibirsk 26–29 sentyabrya, 2000. V 7 tt. –Novosibirsk: Novosibirskiy gosudarstvennyy tekhnicheskyy universitet, 2000. –T. 6. –S. 17–20.

19. Botev. Z.I., Grotowski. J.F., Kroese. D.P. Kernel density estimation via diffusion, Annal Statistic, October 2010, Vol. 38 (5), pp. 2916–2957, DOI: <https://doi.org/10.1214/10-AOS799>.

20. Silverman B.W. Density Estimation for Statistics and Data Analysis. –Chapman and Hall/CRC, 1986. –175 p.

21. Leiva-Murillo J.M., Artés-Rodríguez A. Algorithms for maximum-likelihood bandwidth selection in kernel density estimators. Pattern Recognition Letters, Volume 33, Issue 13, 2012, pp. 1717–1724.

22. Xu Xiaoyuan, Yan Zheng, Xu Shaolun Estimating wind speed probability distribution by diffusion-based kernel density method, Electric Power Systems Research, Volume 121, April 2015, pp. 28–37.

23. Ouarda T.B.M.J., Charron C., Shin J.-Y., Marpu P.R., Al-Mandoos A.H., Al-Tamimi, Ghedira H., Al Hosary T.N, Probability distributions of wind speed in the UAE, Energy Conversion and Management, Volume 93, 2015, pp. 414-434.

24. Jones M.C., Marron J.S., Sheather S.J. A brief survey of bandwidth selection for density estimation, Journal of the American Statistical Association, Volume 91, № 433, March 1996, pp.401–4071.

25. Simakhin V.A. Robastnyye neparametricheskiye otsenki: adaptivnyye otsenki vzveshennogo maksimal'nogo pravdopodobiya v usloviyakh statisticheskoy apriornogo neopredelennosti/ V.A. Simakhin // –Saarbrucken, Germany: LAP LAMBERT Academic Publishing GmbH & Co. KG, 2011. –292 s.

ПОРШНЕВ Сергей Владимирович, доктор технических наук, профессор, директор Учебно-научного центра «Информационная безопасность» федерального государственного автономного образовательного учреждения высшего образования «Уральский федеральный университет им. первого Президента России Б.Н. Ельцина». 620002, г. Екатеринбург, ул. Мира, 32. E-mail: s.v.porshnev@urfu.ru

РЯБКО Николай Юрьевич, аспирант федерального государственного автономного образовательного учреждения высшего образования «Уральский федеральный университет им. первого Президента России Б.Н. Ельцина». 620002, г. Екатеринбург, ул. Мира, 32. E-mail: N.Yu.Ryabko@urfu.ru

PORSHNEV Sergey Vladimirovich, Doctor of Technical Sciences, Professor, Director of the Educational and Scientific Center «Information Security» of the Federal State Autonomous Educational Institution of Higher Education «Ural Federal University named after the first President of Russia B.N. Yeltsin». 620002, Yekaterinburg, st. Mira, 32. E-mail: s.v.porshnev@urfu.ru

RYABKO Nikolay Yurievich, post-graduate student of the Federal State Autonomous Educational Institution of Higher Education “Ural Federal University named after the first President of Russia B.N. Yeltsin”. 620002, Yekaterinburg, st. Mira, 32. E-mail: N.Yu.Ryabko@urfu.ru