

УДК 004.942

UDC 004.942

ИСПОЛЬЗОВАНИЕ АППРОКСИМАЦИИ РОЗЕНБЛАТТА-ПАРЗЕНА ДЛЯ ВОССТАНОВЛЕНИЯ ФУНКЦИИ РАСПРЕДЕЛЕНИЯ НЕПРЕРЫВНОЙ СЛУЧАЙНОЙ ВЕЛИЧИНЫ С ОГРАНИЧЕННЫМ ОДНОМОДАЛЬНЫМ ЗАКОНОМ РАСПРЕДЕЛЕНИЯ

USING ROZENBLATT-PARZEN APPROXIMATION FOR RECOVERING A CUMULATIVE DISTRIBUTION FUNCTION OF CONTINUOUS RANDOM VARIABLE WITH A BOUNDED SINGLE-MODE DISTRIBUTION RULE

Поршнеv Сергей Владимирович
д.т.н., профессор

Porshnev Sergey Vladimirovich
Dr.Sc.(Tech.), Prof.

Копосов Александр Сергеевич
аспирант
Уральский федеральный университет имени первого Президента России Б.Н. Ельцина, Екатеринбург

Koposov Alexander Sergeevich
Postgraduate student
Ural Federal University named after the first President of Russia B.N. Yeltsin, Ekaterinburg

В статье изложены результаты сравнительного анализа погрешностей аппроксимации функции распределения непрерывной случайной величины с ограниченной областью, вычисляемого методом Парзена-Розенблата и методом мнимых источников, полученные при их сравнении теоретическим законом распределения, и обоснована целесообразность использования в рассматриваемой задаче метода мнимых источников

In this report we have compared an approximation accuracy of distribution function of continuous random variable in a bounded region calculated with Rozenblatt-Parzen method and imaginary sources method. The comparison has been performed with theoretical distribution rule. We have explained the reasonability of using imaginary sources method in case of continuous random variable with single-mode distribution rule

Ключевые слова: НЕПАРАМЕТРИЧЕСКАЯ СТАТИСТИКА, ВОССТАНОВЛЕНИЕ ФУНКЦИИ РАСПРЕДЕЛЕНИЯ, АППРОКСИМАЦИЯ РОЗЕНБЛАТТА-ПАРЗЕНА, ПАРАМЕТР РАЗМЫТОСТИ, МЕТОД МНИМЫХ ИСТОЧНИКОВ

Keywords: NON-PARAMETRIC STATISTICS, DISTRIBUTION FUNCTION RECOVERY, PARZEN-ROZENBLATT APPROXIMATION, FUZZINESS PARAMETER, IMAGINARY SOURCES METHOD

1. Введение

Восстановление функции распределения по выборке случайных данных, полученных в результате проведения тех или иных экспериментов, является основной задачей математической статистики [1]. Данная задача имеет важное практическое значение, например, при решении задач прочностной надежности элементов и объектов нефтегазового оборудования [2]. Данная задача имеет следующую постановку: по экспериментальной выборке, полученной из генеральной

совокупности значений случайной величины $X_i, i = \overline{1, N}$, найти функцию распределения $F(y) = \Pr\{X \leq y\}$, связанную с плотностью распределения $f(y)$ интегральным соотношением

$$F(y) = \int_{-\infty}^y f(\xi) d\xi, \quad (1)$$

Известны два основных подхода к решению этой задачи: параметрический и непараметрический. Параметрический подход предусматривает выбор на основе имеющейся априорной информации вида функции распределения случайной величины $F(y)$, зависящей в общем случае от некоторого набора параметров, и получении оценок значений этих параметров по имеющейся выборке данных. При этом искомые оценки значений параметров должны обеспечивать максимальную близость теоретической функции распределения $F(y)$ и эмпирической функций распределения

$$F_N(y) = \frac{1}{N} \sum_{i=1}^N \Theta(y - x_i), \quad (2)$$

где функция Хэвисайда

$$\Theta(y - x_i) = \begin{cases} 1, & \text{при } y - x_i \geq 0, \\ 0, & \text{при } y - x_i < 0, \end{cases}$$

в соответствие с выбранной мерой близости, зависящей, вообще говоря, от вида распределения [3].

Существование решения обсуждаемой задачи обеспечивает центральная теорема математической статистики, согласно которой с

ростом объема выборки N функция $F_N(y)$ с вероятностью, равной единице, равномерно приближается к $F(y)$:

$$\Pr \left\{ \limsup_{N \rightarrow \infty} |F_N(y) - F(y)| = 0 \right\} = 1.$$

В основе непараметрической статистики лежит подход, позволяющий получать адаптивные оценки эмпирических распределений в виде некоторых функционалов, независящих от вида неизвестного априорного распределения [6]. Для восстановления неизвестной функции распределения в непараметрической статистике известен ряд методов и алгоритмов [6]: метод гистограмм, «гребенка», метод ближайших соседей, метод разложения по базисным функциям, аппроксимация Розенблатта-Парзена и ряд других.

Работоспособность методов непараметрической статистики и целесообразность их применения при анализе экспериментальных данных подтверждается результатами, полученными различными исследователями (см., например, список литературы к разделу «Введение» в [6]). Отметим, что аппроксимация Розенблатта-Парзена, как показано в [2], что оказывается эффективной в задаче оценки долговечности нефте- и газопроводов на основе анализа накопленной статистической информации. При этом оказывается, что распределение экспериментальных данных, представляющих собой выборку из генеральной совокупности случайной величины, закон распределения которой не может быть описан ни одним из известных теоретических законов.

Напомним, следуя [6], что обсуждаемый метод восстановления плотности распределения экспериментальной выборки основан на

предположении о том, что функция плотности распределения оценивается локально в каждой точке x_i с помощью элементов обучающей выборки из некоторой окрестности x_i . При этом общая функция вероятности $F(y)$ есть сумма локальных функций

$$F(y) = \frac{1}{N} \sum_{i=1}^N K\left(\frac{y-x_i}{h}\right) \quad (3)$$

где $K(t)$, $t = (y - x_i)/h$ – ядерная функция, удовлетворяющая следующим условиям:

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

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

в) $h_N \rightarrow 0$ при $N \rightarrow \infty$;

h – параметр «размытости», определяющий гладкость получаемой оценки.

Соответственно, плотности вероятности $f(y)$ вычисляется по формуле

$$f(y) = \frac{1}{N \cdot h} \sum_{i=1}^N k\left(\frac{y-x_i}{h}\right) \quad (4)$$

где $k(y) = \frac{d}{dy} K(y)$.

Оценки плотности вероятности (4) предложены Розенблаттом [7] и исследованы Парзенем [8]. На практике наиболее часто в качестве ядерных функций $k(y)$ используются функции, представленные в таблице 1 [2].

Таблица 1. Ядерные функции, наиболее часто используемые на практике

№	Ядро	Формула
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	$k(t) = \frac{3 \cdot \left(1 - \frac{t^2}{5}\right)}{4\sqrt{5}}, t \leq \sqrt{5}$
7	Равномерное	$k(t) = \frac{1}{2}, t \leq 1$
8	Треугольное	$k(t) = 1 - t , t \leq 1$
9	Квадратичное	$k(t) = \frac{3 \cdot (1-t^2)}{4}, t \leq 1$

Оптимальные значения ядерной функции и параметра h находятся из условия достижения функционалом

$$J = \int \ln k(t) \cdot f(t) dt \tag{5}$$

максимального значения, которое, как очевидно, выполняется при $k(t) = f(t)$.

При анализе экспериментальных данных, представляющих собой некоторую дискретную случайную последовательность $\{x_i\}, i = \overline{1, N}$, в [2] рекомендовано находить оптимальные значения параметра «размытости» в

два этапа. На первом этапе для каждой из ядерных функций, представленных в табл. 1, находить значения $h_m^*, m = \overline{1,9}$ из условия

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

и далее выбирать ту ядерную функцию $k_m(y)$, для которой значение функции

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

в точке $h_m = h_m^*$ будет наибольшим.

При этом качество аппроксимации функции распределения в рассматриваемом методе оказывается зависящим от вида ядерной функции $k(t)$ и значения параметра размытости h [6], что иллюстрирует рис. 1.

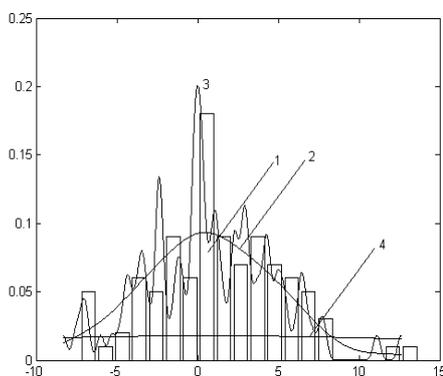


Рисунок 1 – Плотность распределения случайной последовательности $x_i, i = \overline{1,100}$, сгенерированной в соответствие с нормальным законом распределения $N(1,4):1$ – гистограмма случайной последовательности; 2 – нормальное ядро, $h = h_1^*$, где h_1^* пока не определено; 3 – нормальное ядро, $h < h_1^* h = h_1^*$, где h_1^* пока не определено; 4 – нормальное ядро, $h > h_1^* h < h_1^*$, где h_1^* пока не определено.

Отметим, что формально метод Розенблатта-Парзена позволяет, построить аппроксимацию функции распределения любой конечной случайной последовательности, которая при условии правильного выбора

параметра h , оказывается достаточно гладкой (см. [6]). При $N \rightarrow \infty, h_N \rightarrow 0$ оценка (3) является непараметрической и асимптотически оптимальной оценкой функции распределения в классе непрерывных функций [6]. Однако на практике случайные последовательности $X_i, i = \overline{1, N}$, для которых ищется функция распределения, имеют конечную длину. В этой ситуации метод Розенблатта-Парзена, вообще говоря, не гарантирует близости аппроксимирующей функции, найденной с его помощью, и истинной функции распределения. В этой связи представляет практический интерес провести анализ точности восстановления методом Розенблатта-Парзена функции распределения случайных последовательностей конечной длины, теоретический закон распределения которых известен.

В статье изложены результаты анализа точности восстановления функции распределения случайной величины с ограниченной областью рассеяния, основные свойства которой описаны в [3], методом Розенблатта-Парзена [7,8] и методом мнимых источников, в котором параметры аппроксимирующей функции находятся с помощью генетических алгоритмов [9].

2. Методика исследования

Для анализа точности восстановления функции распределения случайной величины с ограниченной областью выбранными методами мы использовали методику, реализующуюся следующей последовательностью действий.

1. В соответствие с заданным законом распределения случайной величины с ограниченной областью рассеяния

$$F_{teor}(x; x_0, \sigma, l) = \int_{x_{min}}^x f_{teor}(\xi; x_0, \sigma, l) d\xi,$$

где:

$$f_{LIM}(x; x_0, \sigma, l) = A \left[\varphi(x; x_0, \sigma, l) + \sum_{n=0}^{\infty} \varphi_{2n+1}^{\pm}(x; x_0, \sigma, l) + \sum_{n=1}^{\infty} \varphi_{2n}^{\pm}(x; x_0, \sigma, l) \right],$$

x_0 – положение центра рассеяния в системе координат с началом в центре отрезка $[x_{\min}, x_{\max}]$,

σ – СКО случайного процесса при отсутствии ограничения,

$l = x_{\max} - x_{\min}$ – размаха области рассеяния,

A – нормировочный коэффициент, определяемый из условия:

$$\int_{x_{\min}}^{x_{\max}} f_{teor}(\xi; x_0, \sigma, l) d\xi = 1,$$

$$\varphi(x; x_0, \sigma, l) = \exp\left[-(x - x_0)^2 / 2\sigma^2\right],$$

$$\varphi_{2n+1}^{\pm}(x; x_0, \sigma, l) = \exp\left[-(x - x_{2n+1}^{\pm})^2 / 2\sigma^2\right],$$

$$\varphi_{2n}^{\pm}(x; x_0, \sigma, l) = \exp\left[-(x - x_{2n}^{\pm})^2 / 2\sigma^2\right],$$

здесь $x_{2n+1}^{\pm}, x_{2n}^{\pm}$ вычисляются в по формулам:

$$x_{2n}^{\pm} = \pm 4nl + x_0, \quad x_{2n+1}^{\pm} = \pm(4n + 2)l - x_0,$$

генерировались методом обратного преобразования (см., например, [2]) случайные последовательности $\{x_i\}, i = \overline{1, N}$.

2. Далее для каждой последовательности $\{x_i\}$ находились эмпирические оценки функций распределения и плотностей распределения методом Розенблатта-Парзена ($F_{RP}(x)$ и $f_{RP}(x)$, соответственно) и методом мнимых источников ($F_{IMS}(x)$ и $f_{RP}(x)$, соответственно), которые сравнивались далее с известными теоретическими функциями распределения $F_{Teor}(x)$ и плотностью распределения $f_{teor}(x)$.

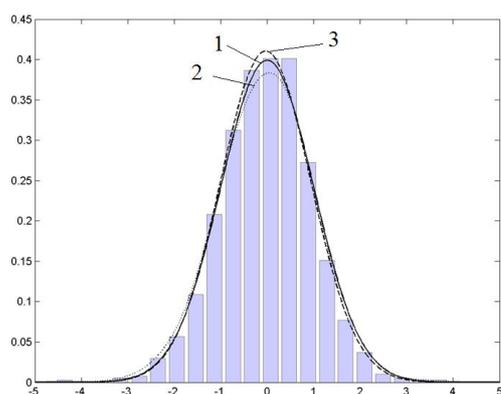
Для реализации описанной выше методики в пакете MATLAB было создано соответствующее программное обеспечение.

3. Анализ качества аппроксимации функции распределения случайной последовательности методом Парзена-Розенблатта

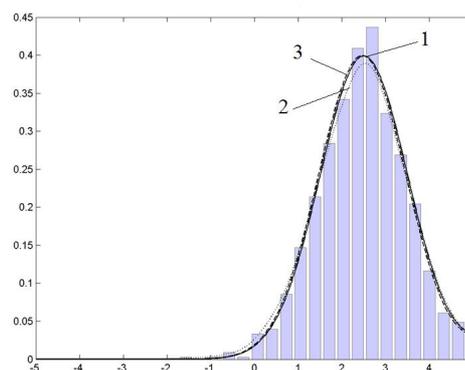
Рассмотрим типичные результаты, полученные для следующих параметров блуждания:

- количество элементов в случайной последовательности $N = 1000$;
- размер области рассеяния $[-5; 5]$;
- координаты центра распределения $\mu = [-5, -2.5, 0, 2.5, 5]$;
- СКО случайной величины $\sigma = [1, 3, 5, 7, 10]$,

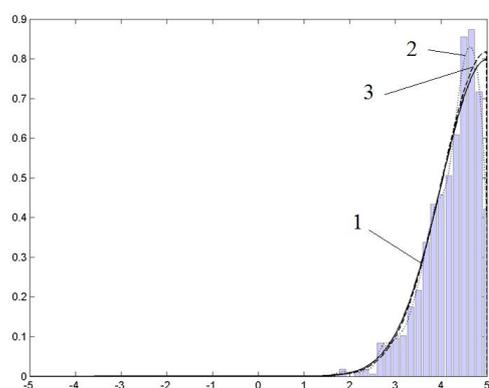
представленные на рис. 2, 3.



а)

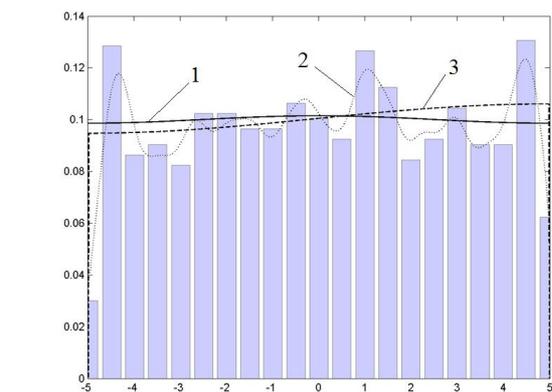


б)

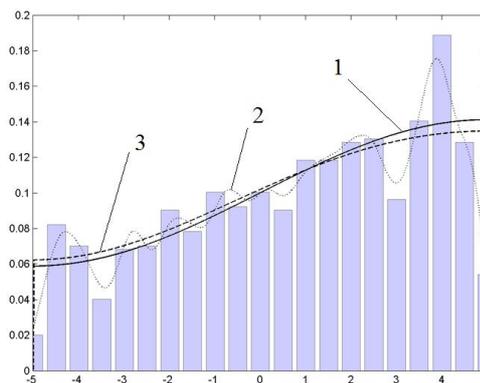


в)

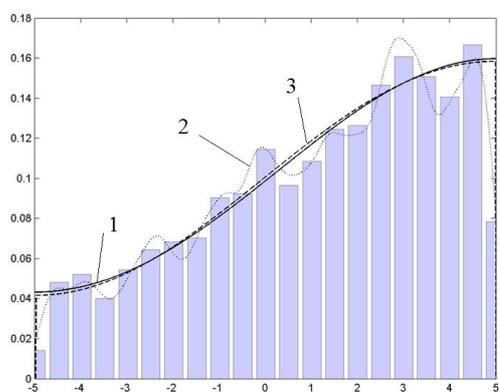
Рис. 2. Результаты аппроксимации плотности функции распределения случайной последовательности $\{x_i\}$: а) $\mu = 0, \sigma = 1$; б) $\mu = 2.5, \sigma = 1$; в) $\mu = 5, \sigma = 1$; 1 – $f_{teor}(x)$; 2 – $f_{RP}(x)$, ядерная функция № 5, №2, №1 (см. табл. 1), соответственно; 3 – $f_{IMS}(x)$



а)



б)



в)

Рис. 3. Результаты аппроксимации плотности распределения случайной последовательности $\{x_i\}$: а) $\mu=0, \sigma=5$; б) $\mu=2.5, \sigma=5$; в) $\mu=5, \sigma=5$; 1 – $f_{теор}(x)$; 2 – $f_{RP}(x)$, ядерная функция №3, №1, №3 (см. табл. 1), соответственно; 3 – $f_{IMS}(x)$

В качестве количественной меры отличия эмпирических функций распределения $F_{RP}(x)$, $F_{IMS}(x)$ от теоретической функции распределения $F_{Теор}(x)$ было использовано среднеквадратическое отклонение остатков моделей от теоретической плотности распределения:

$$\Delta_{RP} = \frac{1}{N-1} \sum_{i=1}^N (F_{Теор}(x_i) - F_{RP}(x_i))^2,$$

$$\Delta_{IMS} = \frac{1}{N-1} \sum_{i=1}^N (F_{Теор}(x_i) - F_{IMS}(x_i))^2.$$

представленные в табл. 2, табл. 3.

Таблица 2. Погрешность оценки функции распределения методом Розенблатта-Парзена

Метод Парзена-Розенблатта		Центр распределения, μ				
		-5	-2,5	0	2,5	5
СКО, σ	1	$1,432 \cdot 10^{-3}$	$3,531 \cdot 10^{-5}$	$1,389 \cdot 10^{-4}$	$7,225 \cdot 10^{-5}$	$1,428 \cdot 10^{-3}$
	3	$2,274 \cdot 10^{-4}$	$2,608 \cdot 10^{-4}$	$8,821 \cdot 10^{-5}$	$1,752 \cdot 10^{-4}$	$5,532 \cdot 10^{-4}$
	5	$2,792 \cdot 10^{-4}$	$1,855 \cdot 10^{-4}$	$1,638 \cdot 10^{-4}$	$1,616 \cdot 10^{-4}$	$2,011 \cdot 10^{-4}$
	7	$2,004 \cdot 10^{-4}$	$1,440 \cdot 10^{-4}$	$1,298 \cdot 10^{-4}$	$1,093 \cdot 10^{-4}$	$1,518 \cdot 10^{-4}$
	10	$1,431 \cdot 10^{-4}$	$1,316 \cdot 10^{-4}$	$1,379 \cdot 10^{-4}$	$1,727 \cdot 10^{-4}$	$1,470 \cdot 10^{-4}$

Таблица 3. Погрешность оценки функции распределения методом мнимых источников

Метод мнимых источников		Центр распределения, μ				
		-5	-2,5	0	2,5	5
СКО, σ	1	$6,412 \cdot 10^{-4}$	$3,423 \cdot 10^{-5}$	$1,498 \cdot 10^{-5}$	$7,267 \cdot 10^{-5}$	$1,125 \cdot 10^{-4}$
	3	$7,934 \cdot 10^{-5}$	$2,924 \cdot 10^{-5}$	$2,274 \cdot 10^{-5}$	$2,917 \cdot 10^{-5}$	$9,254 \cdot 10^{-5}$
	5	$2,315 \cdot 10^{-5}$	$3,184 \cdot 10^{-5}$	$4,368 \cdot 10^{-5}$	$6,254 \cdot 10^{-5}$	$4,132 \cdot 10^{-5}$
	7	$3,157 \cdot 10^{-5}$	$8,193 \cdot 10^{-5}$	$2,141 \cdot 10^{-5}$	$5,432 \cdot 10^{-5}$	$6,778 \cdot 10^{-5}$
	10	$5,682 \cdot 10^{-5}$	$6,647 \cdot 10^{-5}$	$2,798 \cdot 10^{-5}$	$1,782 \cdot 10^{-5}$	$1,167 \cdot 10^{-5}$

Из таблиц 2 и 3 видно, что аппроксимация функции распределения непрерывной случайной величины в ограниченной области рассеяния методом мнимых источников для большинства рассмотренных значений параметров оказывается более точной, чем аппроксимация Розенблатта-Парзена. Данный результат, с нашей точки зрения, объясняется следующим: аппроксимация Розенблатта-Парзена при правильном выборе значения параметра размытости h гарантирует получение гладких оценок функции распределения. Однако данный метод является, в известной мере, формальным, поскольку выбор ядерной функции оказывается независимым от математической модели изучаемого процесса. В то время как в методе мнимых источников математическая модель распределения является решением уравнения случайных блужданий броуновской частицы в ограниченной области рассеяния, а потому более адаптированным к

особенностям распределения изучаемых случайных последовательностей $\{x_i\}$.

Также было проведено сравнение времен, затрачиваемых для вычисления аппроксимаций функций распределения случайных последовательностей каждым из методов. Оказалось, что в методе мнимых источников программной реализации генетического алгоритма, использующегося для поиска параметров распределения (1), одна итерация выполняется ≈ 5 с. При этом параметры распределения, обеспечивающие минимальное значение целевой функции генетического алгоритма, находится за 4 итерации. Таким образом, общее время нахождения решения методом мнимых источников составляет ≈ 20 с. В методе Розенблатта-Парзена наиболее затратной по времени оказывается операция нахождения максимума информационного функционала (6) для ядер с неограниченной областью определения № 1, 2, 4, 5. Здесь для нахождения значения h_m^* в соответствие с (6) приходится вычислять значения функции $\varphi(h_m)$ в соответствие с (7) для нескольких значений h_m , т.е. для каждого значения h_m приходится вычислять в (7) двойную сумму по индексам i и j , поэтому в зависимости от типа ядра затрачивается от 140 до 200 с (при условии удачного выбора интервала поиска h_m^* (см. подробнее [10]). Для ограниченных ядер в [10] получены простые с вычислительной точки зрения аналитические выражения для вычисления оптимального значения параметра размытости искомого значения параметра размытости h_m^* , поэтому время вычисления является незначительным. Таким образом, общее время нахождения аппроксимации функции распределения случайной последовательности $\{x_i\}$ методом Розенблатта-Парзена составляет ≈ 640 – 800 секунд. Таким образом, в рассматриваемой задаче временные затраты при использовании метода

мнимых источников оказывают в $\approx 32-40$ раз меньше, чем у метода Парзена-Розенבלата.

4. Выводы

Проведено сравнение точности аппроксимации функции распределения непрерывной случайной величины с ограниченной областью, вычисляемого методом Парзена-Розенבלата и методом мнимых источников, с теоретическим законом распределения. Результаты сравнения показывают, что в рассматриваемой задаче метод мнимых источников, основанный на физической модели изучаемого процесса, оказывается предпочтительнее как с точностной, так и вычислительной точек зрения.

Список использованной литературы

1. Крамер Г. Математические методы статистики. –М.: Мир, 1975. –648 с.
2. Сызранцев В.Н. Расчет прочностной надежности изделий на основе методов непараметрической статистики / В.Н. Сызранцев, Я.П. Невелев, С.Л. Голофаст //–Новосибирск: Наука, 2008. – 218 с.
3. Поршнева С.В. Теория и алгоритмы аппроксимации эмпирических зависимостей и распределений / Е. В. Овечкина, В.Е. Каплан // –Екатеринбург: УрО РАН, 2006. –166 с.
4. Поршнева С.В. О выборе математических моделей распределений ограниченных случайных последовательностей / С.В. Поршнева, А.С. Копосов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2012. – №10(84). – Режим доступа: <http://ej.kubagro.ru/2012/10/pdf/53.pdf>, 1,000 у.п.л.
5. Поршнева С.В., Копосов А.С. Аналитическое исследование особенностей случайных блужданий броуновской частицы в ограниченной области рассеяния // Фундаментальные исследования. – 2013. – № 4 (часть 1). – стр. 57-64;
6. Симахин В.А. Робастные непараметрические оценки: адаптивные оценки взвешенного максимального правдоподобия в условиях статистической априорной неопределенности/ В.А. Симахин// –Saarbrucken, Germany: LAPLAMBERTAcademicPublishingGmbH&Co. KG, 2011. –292 с.
7. Rozenblatt M. Remarks on some nonparametric estimates of density function// Ann. Math. Statist., 1956, 27, 832–835 p.
8. Parzen E. On estimation of probability density function and mode // Ann. Math. Statist., 33, 3, 162, 1065–1076 p.
9. Гладков Л.А., Курейчик В.В., Курейчик В.М. Генетические алгоритмы / Под ред. В.М. Курейчика. – 2-е изд., исправл. и доп. – М.: ФИЗМАТЛИТ, 2010. – 368 с.

10. Поршнеv С.В., Копосов А.С. Об особенностях численных оценок параметра размытости плотностей распределений случайных последовательностей в аппроксимации Розенблатта-Парзена // В мире научных открытий. – 2013. – В печати

References

1. Kramer G. Matematicheskie metody statistiki. –M.: Mir, 1975. –648 s.
2. Syzrancev V.N. Raschet prochnostnoj nadezhnosti izdelij na osnove metodov neparametricheskoj statistiki / V.N. Syzrancev, Ja.P. Nevelev, S.L. Golofast // –Novosibirsk: Nauka, 2008. – 218 s.
3. Porshnev S.V. Teorija i algoritmy approksimacii jempiricheskix zavisimostej i raspredelenij / E. V. Ovechkina, V.E. Kaplan // –Ekaterinburg: UrO RAN, 2006. –166 s.
4. Porshnev S.V. O vybore matematicheskix modelej raspredelenij ogranichennyh sluchajnyh posledovatel'nostej / S.V. Porshnev, A.S. Kopusov // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2012. – №10(84). – Rezhim dostupa: <http://ej.kubagro.ru/2012/10/pdf/53.pdf>, 1,000 u.p.l.
5. Porshnev S.V., Kopusov A.S. Analiticheskoe issledovanie osobennostej sluchajnyh bluzhdanij brounovskoj chasticy v ogranichennoj oblasti rassejanija // Fundamental'nye issledovanija. – 2013. – № 4 (chast' 1). – str. 57-64;
6. Simahin V.A. Robastnye neparametricheskie ocenki: adaptivnye ocenki vzveshennogo maksimal'nogo pravdopodobija v uslovijah statisticheskoj apriornoj neopredelennosti/ V.A. Simahin// –Saarbrucken, Germany: LAPLAMBERTAcademicPublishingGmbH&Co. KG, 2011. –292 s.
7. Rozenblatt M. Remarks on some nonparametric estimates of density function// Ann. Math. Statist., 1956, 27, 832–835 p.
8. Parzen E. On estimation of probability density function and mode // Ann. Math. Statist., 33, 3, 162, 1065–1076 p.
9. Gladkov L.A., Kurejchik V.V., Kurejchik V.M. Geneticheskie algoritmy / Pod red. V.M. Kurejchika. – 2-e izd., ispravl. i dop. – M.: FIZMATLIT, 2010. – 368 s.
10. Porshnev S.V., Kopusov A.S. Ob osobennostjah chislennyh ocenok parametra razmytosti plotnostej raspredelenij sluchajnyh posledovatel'nostej v approksimacii Rozenblatta-Parzena // V mire nauchnyh otkrytij. – 2013. – V pechati