

Модифицированный алгоритм быстрого выбора коэффициентов размытости ядерных оценок многомерных плотностей вероятностей

А. В. Лапко^{1,2}, В. А. Лапко^{1,2}

¹Институт вычислительного моделирования Сибирского отделения РАН, Красноярск, Российская Федерация, e-mail: lapko@icm.krasn.ru

²Сибирский государственный университет науки и технологий имени академика М. Ф. Решетнева, Красноярск, Российская Федерация, e-mail: valapko@yandex.ru

Обоснована оригинальная методика быстрого выбора коэффициентов размытости ядерных функций в непараметрической оценке многомерной плотности вероятности типа Розенблатта–Парзена. Предлагаемая методика позволяет значительно повысить вычислительную эффективность процедуры оптимизации ядерных оценок плотности вероятности в условиях статистических данных большого объема по сравнению с традиционным подходом. Оригинальная методика основана на анализе формулы оптимального параметра коэффициентов размытости многомерной ядерной оценки плотности вероятности. Обнаружены зависимости между нелинейным функционалом от плотности вероятности и её производных до второго порядка включительно от коэффициентов контрэксцесса случайных величин. Коэффициенты размытости по каждой случайной величине представлены в виде произведения неопределённого параметра на их среднее квадратическое отклонение. Определено влияние ошибки восстановления установленных функциональных зависимостей на аппроксимационные свойства ядерной оценки плотности вероятности. Полученные результаты реализованы в виде методики синтеза и анализа быстрого алгоритма выбора коэффициентов размытости ядерной оценки двумерной плотности вероятности независимых случайных величин. При реализации предлагаемой методики использованы данные о количественных характеристиках семейства логнормальных законов распределения.

Ключевые слова: ядерная оценка плотности вероятности, многомерная случайная величина, быстрый выбор коэффициентов размытости, коэффициент контрэксцесса, выборки большого объема, логнормальный закон распределения, производные от плотности вероятности.

Введение. Ядерные оценки плотности вероятности типа Розенблатта–Парзена $\bar{p}(x)$ являются основой синтеза алгоритмов принятия решений при априорной неопределённости о виде законов распределения в задачах восстановления стохастических зависимостей и распознавания образов [1, 2]. Их аппроксимационные свойства зависят от выбора коэффициентов размытости ядерных функций. Традиционный подход к оптимизации непараметрических оценок восстанавливаемой плотности вероятности $p(x)$, например для одномерной случайной величины, осуществляется из условия минимума оценки среднего квадратического отклонения (СКО) $\bar{p}(x)$ от $p(x)$ [3, 4]:

$$\bar{W}(c) = \int_{-\infty}^{\infty} \bar{p}^2(x) dx - \frac{2}{n} \sum_{i=1}^n \bar{p}(x^i),$$

где $x^i, i = \overline{1, n}$ – наблюдения случайной величины x ; n – количество наблюдений (объём статистических данных).

С увеличением размерности случайной величины x и объёма n статистических данных при поиске коэффициента размытости с ядерных функций при минимизации $\bar{W}(c)$ необходимы значительные затраты времени. Требовалось найти пути обхода возникающей проблемы. Первые решения основывались на анализе выражения оптимального коэффициента размытости ядерных функций в непараметрической оценке плотности вероятности [5–9]. Развитие полученных результатов представлено в работе [10].

Цель настоящей статьи – обоснование методики быстрого выбора коэффициентов размытости ядерной оценки плотности вероятности многомерной случайной величины. Предлагаемый подход основан на результатах анализа оптимальных коэффициентов размытости ядерных функций и зависимостей между нелинейными преобразованиями плотности вероятности, её производных от коэффициентов контрэксцесса случайных величин. Полученные результаты реализованы в виде методики быстрого выбора

коэффициентов размытости ядерной оценки плотности вероятности двумерных независимых случайных величин.

Свойства ядерной оценки многомерной плотности вероятности. Рассмотрим непараметрическую оценку плотности вероятности $p(x_1, \dots, x_k)$ многомерной случайной величины размерностью k :

$$\bar{p}(x_1, \dots, x_k) = \left(n \prod_{v=1}^k c_v \right)^{-1} \sum_{i=1}^n \prod_{v=1}^k \Phi \left(\frac{x_v - x_v^i}{c_v} \right), \quad (1)$$

которая восстанавливается по выборке статистически независимых наблюдений $(x_1^i, \dots, x_k^i, i = \overline{1, n})$ объёмом n .

Ядерные функции $\Phi(u_v)$ в статистике (1) удовлетворяют условиям:

$$\Phi(u_v) = \Phi(-u_v);$$

$$0 \leq \Phi(u_v) < \infty;$$

$$\int \Phi(u_v) du_v = 1;$$

$$\int u_v^2 \Phi(u_v) du_v = 1;$$

$$\int u_v^m \Phi(u_v) du_v < \infty;$$

$$0 \leq m < \infty; v = \overline{1, k}.$$

Здесь и далее бесконечные пределы интегрирования опустим. Коэффициенты размытости ядерных функций c_v зависят от интервалов изменения значений $x_v, v = \overline{1, k}$. Поэтому считаем, что $c_v = c \sigma_v$, где σ_v – СКО случайных величин $x_v, v = \overline{1, k}$; c – неопределённый параметр. Тогда ядерную оценку плотности вероятности (1) запишем в виде

$$\bar{p}(x_1, \dots, x_k) = \left(n \prod_{v=1}^k c \sigma_v \right)^{-1} \sum_{i=1}^n \prod_{v=1}^k \Phi \left(\frac{x_v - x_v^i}{c \sigma_v} \right). \quad (2)$$

Последовательности коэффициентов размытости ядерных функций $c_v(n) = c(n) \sigma_v \rightarrow 0$, значения $n \prod_{v=1}^k c_v(n) \rightarrow \infty$ при $n \rightarrow \infty$.

Асимптотические свойства $\bar{p}(x_1, \dots, x_k)$ рассмотрены в работах [10, 11]. Асимптотическое выражение СКО $\bar{p}(x_1, \dots, x_k)$ от восстанавливаемой плотности вероятности $p(x_1, \dots, x_k)$ имеет вид

$$W(c) = M \int \dots \int [\bar{p}(x_1, \dots, x_k) - p(x_1, \dots, x_k)]^2 dx_1 \dots dx_k \sim \\ \sim (nc^k)^{-1} \prod_{v=1}^k (\sigma_v^{-1} \|\Phi(u_v)\|^2) + \frac{c^4}{4} B, \quad (3)$$

где

$$B = \int \dots \int \left(\sum_{v=1}^k \sigma_v^2 p_v^{(2)}(x_1, \dots, x_k) \right)^2 dx_1 \dots dx_k;$$

$$\|\Phi(u_v)\|^2 = \int \Phi^2(u_v) du_v;$$

M – знак математического ожидания; $p_v^{(2)}(x_1, \dots, x_k)$ – вторая производная функции $p(x_1, \dots, x_k)$ по компоненте x_v многомерной случайной величины.

Оптимальное значение параметра c коэффициентов размытости ядерных функций $c_v, v = \overline{1, k}$ непараметрической оценки плотности вероятности (2) определим из условия минимума критерия $W(c)$ (см. формулу (3)):

$$c^* = \left(k \|\Phi(u)\|^2 \right)^k / \left(n B \prod_{v=1}^k \sigma_v \right)^{\frac{1}{k+4}} \quad (4)$$

для ядерных функций вида $\Phi(u_v) = \Phi(u), v = \overline{1, k}$.

С учётом ранее принятых предположений запишем оптимальные коэффициенты размытости ядерных функций как

$$c_v^* = c^* \sigma_v = \beta \sigma_v n^{-1/(k+4)}, v = \overline{1, k},$$

где

$$\beta = \left(k \|\Phi(u)\|^2 \right)^k / \left(B \prod_{v=1}^k \sigma_v \right)^{\frac{1}{k+4}}. \quad (5)$$

Основная проблема при определении оптимальных параметров c^* (см. формулу (4)) коэффициентов размытости $c_v^*, v = \overline{1, k}$, ядерных функций состоит в оценивании нелинейного функционала

$$\lambda = B \prod_{v=1}^k \sigma_v$$

от вторых производных плотности вероятности $p(x_1, \dots, x_k)$ по компонентам $x_v, v = \overline{1, k}$ многомерной случайной величины. Предположим, что λ является аппроксимацией функционала λ по значениям количественных характеристик плотности вероятности $p(x_1, \dots, x_k)$, $\rho = |\lambda - \bar{\lambda}| / \lambda$ – относительная ошибка оценивания λ . В этих условиях с учётом выражения (4) оценку $\bar{c}^*$ оптимального параметра c^* определим как

$$\bar{c}^* = c^* (1 \pm \rho)^{-1/(k+4)} = c^* / \alpha^{1/(k+4)},$$

где $\alpha = (1 \pm \rho)$.

Подставим значения $\bar{c}^*, c^*$ в критерий (3) и получим отношение СКО

$$R = W(\bar{c}^*) / W(c^*) = (4\alpha + k) / [(4 + k) \alpha^{4/(k+4)}]. \quad (6)$$

Проанализируем отношение R , определяемое формулой (6). С увеличением размерности k случайной величины $x = (x_1, \dots, x_k)$ значения числителя R увеличиваются. При $k=2$ значения числителя $R(1)$ отношения R принадлежат

интервалу $R(1) \in [8; 4]$, а при $k=6$ значения $R(1) \in [12; 8]$ в условиях $\alpha \in [1,5; 0,5]$. При $k=2$ значения знаменателя $R(2)$ отношения R принадлежат интервалу $[7,86; 3,78]$ с уменьшением α в интервале $[1,5; 0,5]$. С увеличением размерности x до $k=6$ значения $R(2) \in [1,76; 7,58]$. В этих условиях, независимо от размерности $k=2, 6$ случайной величины и значений $\alpha \in [1,5; 0,5]$, критерий R находится в пределах $[1; 1,06]$. С уменьшением $\alpha < 1$ наблюдается увеличение значений отношения R . При $\alpha > 1$ отмеченная закономерность проявляется менее значимо. Данный факт объясняется увеличением дисперсии непараметрической оценки плотности вероятности $\bar{p}(x)$ при $\bar{c}^* < c^*$, значение которой меньше чем для условий $\bar{c}^* > c^*$. Отмеченная асимметрия зависимости $R(\alpha)$ для интервалов значений $\alpha < 1$ и $\alpha > 1$ показывает, что превышение значений $\bar{c}^*$ над c^* в меньшей степени сказывается на аппроксимационных свойствах $\bar{p}(x)$ по сравнению со значением $\bar{c}^*$.

Оценивание значений функционала λ для двухмерных независимых случайных величин. В качестве примера рассмотрим значение $\lambda = B \prod_{v=1}^k \sigma_v$ в выражении (5) при $k=2$ для независимых случайных величин x_1, x_2 :

$$\begin{aligned} \lambda &= \sigma_1 \sigma_2 \iint (\sigma_1^2 p_1^{(2)}(x_1, x_2) + \sigma_2^2 p_2^{(2)}(x_1, x_2)) dx_1 dx_2 = \\ &= \sigma_1^5 \sigma_2 \int (p_1^{(2)}(x_1))^2 dx_1 \int p^2(x_2) dx_2 + \\ &+ \sigma_1 \sigma_2^5 \int (p_2^{(2)}(x_2))^2 dx_2 \int p^2(x_1) dx_1 + \\ &+ 2\sigma_1^3 \sigma_2^3 \int p_1^{(2)}(x_1) p(x_1) dx_1 \int p_2^{(2)}(x_2) p(x_2) dx_2. \end{aligned} \quad (7)$$

Существуют плотности вероятности $p(x_1), p(x_2)$, для которых составляющие полученного выражения (7) являются константами при $v=1, 2$:

$$\lambda_v(1) = \sigma_v^5 \int (p_v^{(2)}(x_v))^2 dx_v; \quad (8)$$

$$\lambda_v(2) = \sigma_v \int p^2(x_v) dx_v; \quad (9)$$

$$\lambda_v(3) = \sigma_v^3 \int p_v^{(2)}(x_v) p(x_v) dx_v. \quad (10)$$

Эти константы определяются видом плотностей вероятности $p(x_1), p(x_2)$ и не зависят от их параметров (табл. 1).

Возникает задача оценивания обнаруженных констант (8)–(10) по значениям количественных характеристик восстанавливаемой плотности вероятности [12]. Поставленная задача отличается от задач, приведённых в работе [10]. Результаты решения этой задачи позволят повысить вычислительную эффективность метода быстрого выбора коэффициентов размытости при оценивании плотности вероятности по статистическим данным наблюдений случайной величины большого объёма.

При восстановлении зависимостей между составляющими $\lambda(1), \lambda(2), \lambda(3)$ выражения (7) от коэффициента контрэксцесса δ

Таблица 1

Значения констант $\lambda(1), \lambda(2), \lambda(3)$, параметра эксцесса η и коэффициента контрэксцесса δ семейства одномодальных плотностей вероятностей

Закон распределения	$\lambda(1)$	$\lambda(2)$	$\lambda(3)$	η	δ
Гауссовый	0,211	0,282	−0,141	3,0	0,577
Логистический	0,467	0,302	−0,199	4,2	0,488
Лапласа	1,414	0,354	0,707	6,0	0,408
Экспоненциальный	0,500	0,500	0,500	9,0	0,333

используем семейство логнормальных законов распределения

$$p(x_v) = (s x_v \sqrt{2\pi})^{-1} \exp[-(\ln x_v - a)^2 / (2s^2)], \quad v=1, 2,$$

где параметры a, s изменяют вид плотностей вероятности $p(x_v), v=1, 2$. Значения $\lambda(1), \lambda(2), \lambda(3)$ определяются параметром s и не зависят от значения a . Логнормальными законами распределения представляют, например, спектральные признаки при обработке данных дистанционного зондирования (табл. 2).

По данным табл. 2 определим функциональные зависимости между установленными константами (8)–(10) и коэффициентом контрэксцесса δ одномерной случайной величины x для семейства логнормальных законов распределения:

$$\bar{\lambda}(1) = \begin{cases} 0,0079\delta^{-5,764} \vee \delta > 0,22 \\ 0,0033\delta^{-6,51} \vee \delta \leq 0,22 \end{cases}; \quad (11)$$

$$\bar{\lambda}(2) = 0,1966\delta^{-0,585}; \quad (12)$$

$$\bar{\lambda}(3) = -0,0254\delta^{-2,817}. \quad (13)$$

Зависимостям (11)–(13) соответствуют средние относительные ошибки аппроксимации

$$\bar{\rho}(t) = \frac{1}{N} \sum_{j=1}^N \frac{|\lambda^j(t) - \bar{\lambda}^j(t)|}{\lambda^j(t)}, \quad t=1, 3,$$

равные 0,088; 0,021; 0,08. Здесь N – количество законов распределения случайных величин, представленных в табл. 2; $\lambda^j(t)$ – оценки значений констант $\lambda^j(t), t=1, 3$, которые вычисляются по формулам (11)–(13) при конкретном значении коэффициента контрэксцесса δ^j .

Определим влияние ошибки оценивания функционала λ (7) на аппроксимационные свойства непараметрической оценки плотности вероятности (2). Обозначим через $(1 \pm \rho)\lambda$ результат оценивания λ с помощью моделей (11)–(13).

Таблица 2

Значения констант $\lambda(1)$, $\lambda(2)$, $\lambda(3)$, параметра эксцесса η и коэффициента контрэксцесса δ для семейства логнормальных законов распределения

$LN(a; s)$	$\lambda(1)$	$\lambda(2)$	$\lambda(3)$	η	δ
$LN(1; 0,05)$	0,218	0,283	-0,143	3,040	0,574
$LN(1; 0,1)$	0,238	0,285	-0,148	3,162	0,562
$LN(1; 0,2)$	0,337	0,294	-0,172	3,678	0,521
$LN(1; 0,3)$	0,602	0,309	-0,221	4,645	0,464
$LN(1; 0,4)$	1,351	0,331	-0,313	6,260	0,400
$LN(1; 0,5)$	3,797	0,363	-0,491	8,898	0,335
$LN(1; 0,6)$	13,32	0,405	-0,848	13,273	0,274
$LN(1; 0,7)$	58,165	0,463	-1,617	20,791	0,219
$LN(1; 0,8)$	315,402	0,540	-3,402	34,368	0,171
$LN(1; 0,9)$	2120	0,643	-7,902	60,411	0,129
$LN(1; 1)$	17660	0,783	-20,263	113,937	0,094
$LN(1; 1,1)$	182100	0,975	-57,391	232,633	0,066
$LN(1; 1,2)$	2326000	1,242	-179,66	518,169	0,044
$LN(1; 1,3)$	36850000	1,620	-622,2	1266,000	0,028

Тогда с учётом выражения (6) при $k=2$ запишем отношение асимптотических выражений СКО:

$$W(\bar{c}^*)/W(c^*) = (2\alpha + 1)/(3\alpha^{2/3}).$$

Если ошибка ρ оценивания функционала λ равна нулю, то отношение $W(\bar{c}^*)/W(c^*)=1$. Это результат подтверждает корректность выполненных преобразований. При средней относительной ошибке $\rho=0,1$ и $0,3$ отношения $W(\bar{c}^*)/W(c^*)$ равны 1,001 и 1,015 соответственно.

Методика быстрого выбора коэффициентов размытости ядерных функций. Методика состоит из следующих шагов.

1. По исходной выборке $V = (x_1^i, x_2^i, i=1, n)$, которая существует при восстановлении двумерной плотности вероятности независимых случайных величин $p(x_1, x_2) = p(x_1)p(x_2)$ с использованием статистики (2), оценивают СКО σ_1 , σ_2 случайных величин x_1 , x_2 и соответствующие им коэффициенты контрэксцесса $\bar{\delta}_v = 1/\sqrt{\eta_v}$, где

$$\eta_v = \frac{1}{n} \sum_{i=1}^n (x_v^i - \bar{x}_v)^4 / \left(\frac{1}{n} \sum_{i=1}^n (x_v^i - \bar{x}_v)^2 \right)^2;$$

$$\bar{x}_v = \frac{1}{n} \sum_{i=1}^n x_v^i, \quad v=1, 2.$$

2. С использованием моделей (11)–(13) зависимости между константами $\lambda_v(1)$, $\lambda_v(2)$, $\lambda_v(3)$ и коэффициентами контрэксцесса $\bar{\delta}_v$, вычисляют значения $\bar{\lambda}_v(1)$, $\bar{\lambda}_v(2)$, $\bar{\lambda}_v(3)$, $v=1, 2$.

3. В соответствии с формулой (7) по вычисленным значениям $\bar{\lambda}_v(j)$, $j=1, 3$, $\bar{\sigma}_v$, $v=1, 2$ находят константу

$$\bar{\lambda} = \bar{\lambda}_1(1)\bar{\lambda}_2(2) + \bar{\lambda}_2(1)\bar{\lambda}_1(2) + 2\bar{\lambda}_1(3)\bar{\lambda}_2(3).$$

4. С учётом результатов, полученных в п. 3, для ядерной функции $\Phi(u)$ выбранного вида определяют значение $\bar{\beta}$ по формуле (5) при $k=2$.

5. При $k=2$ и известных значениях n , $\bar{\beta}$, $\bar{\sigma}_1$, $\bar{\sigma}_2$ вычисляют значения оценок оптимальных коэффициентов размытости в статистике (2):

$$\bar{c}_v^* = \bar{\beta} \bar{\sigma}_v n^{-1/6}, \quad v=1, 2.$$

Заключение. Быстрый выбор коэффициентов размытости ядерных оценок плотностей вероятностей позволяет значительно повысить вычислительную эффективность процедуры оптимизации этих оценок, что особенно актуально при обработке статистических данных большого объёма. Предложенный алгоритм основан на результатах анализа формулы расчёта оптимального параметра коэффициентов размытости ядерной оценки плотности вероятности многомерных случайных величин. При этом коэффициенты размытости ядерных функций представлены в виде произведения априори неопределённого параметра и СКО составляющих многомерной случайной величины. В качестве критерия оптимального выбора неопределённого параметра взят минимум СКО ядерной оценки плотности вероятности. Найден диапазон ошибок оценивания оптимального параметра коэффициентов размытости ядерных функций, при которых аппроксимационные свойства непараметрической оценки плотности вероятности изменяются незначительно. Разработанная на основе полученных результатов методика быстрого выбора коэффициентов размытости ядерной оценки плотности вероятности двумерных независимых случайных величин актуальна при обработке статистических данных большого объёма, характеристики которых близки к логнормальным законам распределения. Обнаруженные константы (составляющие оптимального параметра коэффициента размытости ядерных функций) определяются нелинейными функционалами от составляющих восстанавливаемой плотности вероятности и их производных. Зависимости обнаруженных констант от коэффициентов контрэксцесса компонент случайных величин свойственны семейству логнормальных законов распределения, а ошибки при определении этих констант незначительно влияют на аппроксимационные свойства ядерных оценок плотностей вероятностей.

Дальнейшее развитие предлагаемого подхода состоит в разработке методик быстрого выбора коэффициентов размытости ядерных функций для более широкого класса законов распределения независимых и зависимых многомерных случайных величин.

Список литературы

1. Лапко А. В., Лапко В. А., Им С. Т., Тубольцев В. П., Авдеев В. А. Непараметрический алгоритм выделения

классов, соответствующих одномерным фрагментам плотности вероятности многомерных случайных величин // Автометрия. 2019. Т. 55. № 3. С. 22–30. <https://doi.org/10.15372/AUT20190303>

2. Лапко А. В., Лапко В. А. Методика проверки гипотез о распределениях многомерных спектральных данных с использованием непараметрического алгоритма распознавания образов // Компьютерная оптика. 2019. Т. 43. № 2. С. 238–244. <https://doi.org/10.18287/2412-6179-2019-43-2-238-244>

3. Rudemo M., *Scandinavian J. Statistics*, 1982, no. 9, pp. 65–78.

4. Hall P., *Annals of Statistics*, 1983, vol. 11, pp. 1156–1174.

5. Silverman B. W., *Density estimation for statistics and data analysis*, London, Chapman & Hall, 1986, 175 p.

6. Sheather S., Jones M., *Journal of Royal Statistical Society Series B*, 1991, vol. 53, no. 3, pp. 683–690. <https://doi.org/10.1111/j.2517-6161.1991.tb01857.x>

7. Sheather S. J., *Statistical Science*, 2004, vol. 19, no. 4, pp. 588–597. <https://doi.org/10.1214/088342304000000297>

8. Terrell G. R., Scott D. W., *Journal of the American Statistical Association*, 1985, vol. 80, pp. 209–214.

9. Jones M. C., Marron J. S., Sheather S. J., *Journal of the American Statistical Association*, 1996, vol. 91, pp. 401–407.

10. Лапко А. В., Лапко В. А. Быстрый алгоритм выбора коэффициентов размытости в многомерных ядерных оценках плотности вероятности // Измерительная техника. 2018. № 10. С. 19–23. <https://doi.org/10.32446/0368-1025it.2018-10-19-23>

11. Епанечников В. А. Непараметрическая оценка многомерной плотности вероятности // Теория вероятности и её применения. 1969. Т. 14. № 1. С. 156–161.

12. Лапко А. В., Лапко В. А. Оценивание интеграла от квадрата производных симметричных плотностей вероятностей одномерных случайных величин // Метрология. 2020. № 1. С. 15–27. <https://doi.org/10.32446/0132-4713.2020-1-15-27>

References

1. Lapko A. V., Lapko V. A., Im S. T., Tuboltsev V. P., Avdeenok V. A., *Optoelectronics, Instrumentation and Data Processing*, 2019, vol. 55, no. 3, pp. 230–236.

<https://doi.org/10.3103/S8756699019030038>

2. Lapko A. V., Lapko V. A., *Computer Optics*, 2019, vol. 43, no. 2, pp. 238–244.

<https://doi.org/10.18287/2412-6179-2019-43-2-238-244>

3. Rudemo M., *Scandinavian J. Statistics*, 1982, no. 9, pp. 65–78.

4. Hall P., *Annals of Statistics*, 1983, vol. 11, pp. 1156–1174.

5. Silverman B. W. *Density estimation for statistics and data analysis*, London, Chapman & Hall, 1986, 175 p.

6. Sheather S., Jones M., *Journal of Royal Statistical Society Series B*, 1991, vol. 53, no. 3, pp. 683–690. <https://doi.org/10.1111/j.2517-6161.1991.tb01857.x>

7. Sheather S. J., *Statistical Science*, 2004, vol. 19, no. 4, pp. 588–597. <https://doi.org/10.1214/088342304000000297>

8. Terrell G. R., Scott D. W., *Journal of the American Statistical Association*, 1985, vol. 80, pp. 209–214.

9. Jones M. C., Marron J. S., Sheather S. J., *Journal of the American Statistical Association*, 1996, vol. 91, pp. 401–407.

10. Lapko A. V., Lapko V. A., *Measurement Techniques*, 2019, vol. 61, no. 10, pp. 979–986. <https://doi.org/10.1007/s11018-019-01536-x>

11. Epanechnikov V. A., *Theory of Probability & Its Applications*, 1969, vol. 14, no. 1, pp. 156–161.

12. Lapko A. V., Lapko V. A., *Measurement Techniques*, 2020, vol. 63, no. 3, pp. 171–176. <https://doi.org/10.1007/s11018-020-01768-2>

Дата поступления: 14.05.2020.

Дата принятия: 19.08.2020.

Modified algorithm for fast bandwidth selection for kernel estimates of multidimensional probability densities

Aleksandr V. Lapko^{1,2}, Vasily A. Lapko^{1,2}

¹Institute of Computational Modeling, Siberian Branch of the Russian Academy of Sciences (ICM SB RAS), Krasnoyarsk, Russian Federation, e-mail: lapko@icm.krasn.ru

²Reshetnev Siberian State University of Science and Technology, Krasnoyarsk, Russian Federation, e-mail: valapko@yandex.ru

An original technique has been justified for the fast bandwidths selection of kernel functions in a nonparametric estimate of the multidimensional probability density of the Rosenblatt–Parzen type. The proposed method makes it possible to significantly increase the computational efficiency of the optimization procedure for kernel probability density estimates in the conditions of large-volume statistical data in comparison with traditional approaches. The basis of the proposed approach is the analysis of the optimal parameter formula for the bandwidths of a multidimensional kernel probability density estimate. Dependencies between the nonlinear functional on the probability density and its derivatives up to the second order inclusive of the antikurtosis coefficients of random variables are found. The bandwidths for each random variable are represented as the product of an undefined parameter and their mean square deviation. The influence of the error in restoring the established functional dependencies on the approximation properties of the kernel probability density estimation is determined. The obtained results are implemented as a method of synthesis and analysis of a fast bandwidths selection of the kernel estimation of the two-dimensional probability density of independent random variables. This method uses data on the quantitative characteristics of a family of lognormal distribution laws.

Key words: kernel probability density estimation, multidimensional random variable, fast bandwidths selection, antikurtosis coefficient, large sampling, lognormal distribution law, derivatives of probability density.