

Быстрый алгоритм выбора коэффициентов размытости в многомерных ядерных оценках плотности вероятности

А. В. Лапко^{1,2}, В. А. Лапко^{1,2}

¹Институт вычислительного моделирования Сибирского отделения РАН, Красноярск, Россия

²Сибирский государственный университет науки и технологий им. акад. М. Ф. Решетнёва, Красноярск, Россия, e-mail: lapko@icm.krasn.ru

Предложена методика быстрого выбора коэффициентов размытости ядерных функций в непараметрической оценке многомерной плотности вероятности типа Розенблатта–Парзена. Методика основана на результатах анализа асимптотических свойств многомерной оценки плотности вероятности. Исследованы свойства быстрого алгоритма выбора коэффициентов размытости ядерной оценки плотности вероятности.

Ключевые слова: непараметрическая оценка многомерной плотности вероятности, выбор коэффициентов размытости, оценка Розенблатта–Парзена, быстрый алгоритм оптимизации, асимптотические свойства, многомерный анализ данных.

Непараметрические оценки плотности вероятности типа Розенблатта–Парзена $\bar{p}(x_1, \dots, x_k)$ используются при синтезе алгоритмов принятия решений и проверке гипотез о распределениях случайных величин в условиях априорной неопределённости [1]. Аппроксимационные свойства $\bar{p}(x_1, \dots, x_k)$ зависят от выбора коэффициентов размытости ядерных функций. При оценивании плотности вероятности по выборке $V = (x_1^i, \dots, x_k^i, i = 1, \dots, n)$ независимых наблюдений случайных величин применяют традиционные методы оптимизации непараметрических оценок плотности вероятности парзеновского типа из условия минимума критерия

$$W(c) = \int \dots \int \bar{p}^2(x_1, \dots, x_k) dx_1 \dots dx_k - \frac{2}{n} \sum_{j=1}^n \bar{p}(x_1^j, \dots, x_k^j). \quad (1)$$

Обоснование критерия (1) получено в результате анализа среднего квадратического отклонения (СКО) ядерной оценки плотности вероятности $\bar{p}(x_1, \dots, x_k)$ от $p(x_1, \dots, x_k)$ [2–7]. Здесь и далее бесконечные пределы интегрирования опускаются. Вычислительная эффективность процедуры минимизации критерия (1) значительно снижается при увеличении объёма n статистических данных. Поэтому возникает необходимость разработки новых методик, обеспечивающих сокращения временных затрат для определения коэффициентов размытости ядерных функций в многомерном случае. В настоящее время этой проблеме уделяется значительное внимание [8–11]. В частности, при оценивании одномерной плотности вероятности по выборке случайной величины объёма n для нормальных законов распределения с СКО σ и гауссовой ядерной функцией предложено правило выбора коэффициентов размытости [11]:

$$\tilde{c}^* = 1,06 \sigma n^{-1/5}.$$

Этот подход развит и при выборе коэффициентов размытости ядерных функций в непараметрических оценках плот-

ности вероятности многомерных случайных величин с нормальными законами распределения [12, 13]:

$$\tilde{c}_v^* = [4/(k+2)]^{1/(k+4)} \sigma_v n^{-1/(k+4)}, \quad v = \overline{1, k},$$

где σ_v – СКО случайной величины x_v , $v = \overline{1, k}$.

Алгоритмы быстрого выбора коэффициентов размытости ядерных функций способствуют широкому использованию методов непараметрической статистики в различных прикладных областях.

В данной работе развит и обоснован быстрый алгоритм выбора коэффициентов размытости ядерных функций в непараметрической оценке плотности вероятности многомерных случайных величин, основанный на принципах коллективного оценивания с применением информации тестового семейства плотностей вероятности.

Непараметрическая оценка плотности вероятности и её свойства. Для восстановления плотности вероятности многомерной случайной величины $p(x_1, \dots, x_k)$ по выборке $(x_v^i, v = \overline{1, k}, i = \overline{1, n})$ её независимых наблюдений рассмотрим статистику [14]:

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

Ядерные функции $\Phi(u_v)$ удовлетворяют условиям:

$$\Phi(u_v) = \Phi(-u_v), \quad 0 \leq \Phi(u_v) < \infty;$$

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

$$\int u_v^m \Phi(u_v) du_v < \infty, \quad 0 \leq m < \infty, \quad v = \overline{1, k}.$$

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

$n \rightarrow \infty$. Асимптотические свойства $\bar{p}(x_1, \dots, x_k)$ в условиях $c_v = c$, $v = \overline{1, k}$ исследованы в работах [15–17].

Т а б л и ц а 1
Зависимость отношения $G(\alpha, k)$ от параметра α ошибки оценивания B и размерности k случайной величины

α	Размерность случайной величины k				
	1	2	3	4	5
0,2	1,305	1,365	1,362	1,342	1,318
0,4	1,082	1,105	1,109	1,107	1,102
0,6	1,023	1,031	1,033	1,033	1,032
0,8	1,004	1,006	1,006	1,006	1,006
1,0	1,000	1,000	1,000	1,000	1,000
1,2	1,003	1,004	1,004	1,004	1,004
1,4	1,008	1,012	1,014	1,014	1,014
1,6	1,016	1,023	1,027	1,028	1,028
1,8	1,025	1,036	1,041	1,043	1,044
2,0	1,034	1,050	1,057	1,061	1,061
2,2	1,043	1,064	1,074	1,079	1,080
2,4	1,052	1,079	1,091	1,097	1,099
2,6	1,062	1,093	1,109	1,116	1,119
2,8	1,071	1,107	1,126	1,135	1,139
3,0	1,080	1,122	1,144	1,155	1,159

Применив технологию доказательства, изложенную в работе [16], определим смещение и СКО непараметрической оценки плотности вероятности (2) при достаточно больших объемах n статистических данных. Выполним преобразования в выражении усреднённой непараметрической оценки плотности вероятности

$$M[\bar{p}(x_1, \dots, x_k)] = \left(n \prod_{v=1}^k c\sigma_v \right)^{-1} \times \\ \times \int \dots \int \prod_{v=1}^k \Phi\left(\frac{x_v - t_v}{c\sigma_v}\right) p(t_1, \dots, t_k) dt_1 \dots dt_k.$$

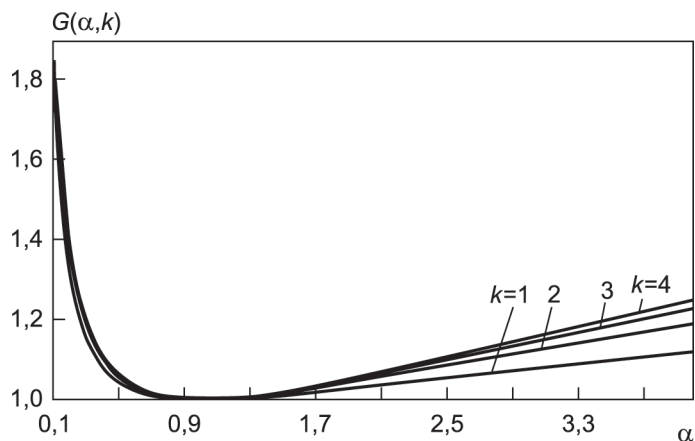


Рис. 1. Зависимость отношения $G(\alpha, k)$ от параметра α ошибки оценивания B и размерности k случайной величины $x = (x_1, \dots, x_k)$

После преобразований запишем выражение

$$M[\bar{p}(x_1, \dots, x_k)] = \int \dots \int \prod_{v=1}^k \Phi\left(\frac{x_v - t_v}{c\sigma_v}\right) \times \\ \times p(x_1 - c\sigma_1 u_1, \dots, x_k - c\sigma_k u_k) du_1 \dots du_k,$$

где M — знак математического ожидания.

Разложим плотность вероятности $p(x_1 - c\sigma_1 u_1, \dots, x_k - c\sigma_k u_k)$ в ряд Тейлора по всем x_v в точке $x = (x_v, v=1, k)$. Проинтегрируем полученное выражение. Тогда при достаточно больших значениях n имеем

$$M[\bar{p}(x_1, \dots, x_k)] \sim p(x_1, \dots, x_k) + \frac{c^2}{2} \sum_{v=1}^k \sigma_v^2 p_v^{(2)}(x_1, \dots, x_k),$$

где $p_v^{(2)}(x_1, \dots, x_k)$ — вторая производная $p(x_1, \dots, x_k)$ по x_v .

Используя приведённую выше технологию преобразований, получим асимптотическое выражение СКО аппроксимации $p(x_1, \dots, x_k)$:

$$W(c) = M \int \dots \int [\bar{p}(x_1, \dots, x_k) - p(x_1, \dots, x_k)]^2 dx_1 \dots dx_k \sim$$

$$\sim (nc^k)^{-1} \prod_{v=1}^k (\sigma_v^{-1} \|\Phi(u_v)\|^2) + \frac{c^4}{4} B, \quad (3)$$

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

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

Из условия минимума $W(c)$ определим оптимальное значение параметра c коэффициентов размытости $c_v = c\sigma_v$, $v=1, k$ ядерных функций:

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

$$\text{где } \|\Phi(u)\|^2 = \|\Phi(u_v)\|^2, v=1, k.$$

Результаты исследований зависимости значений

$$W = W(c^*) / \left[\int \dots \int p^2(x_1, \dots, x_k) dx_1 \dots dx_k \right]$$

от объёма n и размерности k статистических данных V рассмотрены в работе [16]. Предполагается, что восстанавливаемая плотность вероятности соответствует многомерному нормальному закону распределения, а статистика $\bar{p}(x_1, \dots, x_k)$ формируется на основе гауссовых ядерных функций. Проведём анализ зависимости значений $W(c)$ от ошибки оценивания функционала B в выражении (3). Будем полагать, что в результате оценивания B получено значение, равное произведению αB , причём коэффициент α характеризует ошибку оценивания. Тогда выражение (4) представим в виде

$\bar{c}^* = \alpha^{-1/(k+4)} c^*$. Подставив значение $\bar{c}^*$ в формулу (3), получим выражение

$$W(\bar{c}^*) = W(c^*) (4\alpha + k) / [(4 + k) \alpha^{4/(k+4)}],$$

где

$$W(c^*) = \frac{4+k}{4k^{k/(k+4)}} \left\{ \left[\frac{(\|\Phi(u)\|^2)^k}{n \prod_{v=1}^k \sigma_v} \right]^4 B^k \right\}^{(k+4)^{-1}}. \quad (5)$$

Проведём анализ отношения

$$G(\alpha, k) = \frac{W(\bar{c}^*)}{W(c^*)} = \frac{4\alpha + k}{(4 + k) \alpha^{4/(k+4)}},$$

значения которого приведены в табл. 1 и представлены на рис. 1.

Отношение $G(\alpha, k) = 1$, если размерность k случайной величины фиксирована и $\alpha = 1$, что соответствует известному значению B . В условиях $\alpha > 1$ либо $\alpha < 1$ значения отношения $G(\alpha, k)$ увеличиваются. Эта тенденция характерна для $\alpha < 1$, когда значения оценки коэффициента $\bar{c}^*$ больше его оптимального значения c^* . В этом случае отношение $G(\alpha, k)$ практически не зависит от размерности k случайных величин. Если $\alpha > 1$, то значения отношения $G(\alpha, k)$ увеличиваются с ростом размерности $k \in [1; 4]$. При $k > 4$, $\alpha > 1$ и дальнейшем увеличении k наблюдается незначительное снижение значений $G(\alpha, k)$. Например, при $\alpha = 0,5$ и увеличении k от четырёх до десяти отношение $G(0,5; k)$ принимает значения от 1,06 до 1,045. Если $\alpha = 2$, то в условиях $k > 4$ значения $G(2; k) \in [1,06; 1,055]$. Обнаруженные закономерности позволяют обосновать разработку быстрых алгоритмов оптимизации ядерных оценок плотности вероятности по коэффициентам размытости.

Методика быстрого выбора коэффициентов размытости ядерной оценки многомерной плотности вероятности. Рассмотрим алгоритм быстрого выбора коэффициентов размытости, которые с учётом ранее принятых предположений и выражения (4) представим в виде

$$c_v^* = c^* \sigma_v = \beta \sigma_v n^{-1/(k+4)}, \quad v = \overline{1, k}, \quad (6)$$

где

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

Если случайные величины x_v , $v = \overline{1, k}$ независимы, а составляющие $p(x_v)$ плотности вероятности $p(x_1, \dots, x_k)$ определяются математическим ожиданием и дисперсией, то при конкретных ядерных функциях значения β являются константами. В этих условиях значения β не зависят от параметров $p(x_1, \dots, x_k)$ (табл. 2).

С учётом (6) оптимальные значения коэффициентов размытости ядерных функций при восстановлении плотности

Т а б л и ц а 2

Зависимость значений коэффициента β от законов распределения независимых двумерных случайных величин и вида ядерных функций

Закон распределения	Вид ядерной функции		
	ступенчатая	Епанечникова	Гаусова
Нормальный	1,008	0,983	1
Логистический	0,88	0,859	0,873
Лапласа	0,661	0,645	0,656
Экспонента	0,742	0,724	0,736
Односторонний нормальный	1,327	1,295	1,317
логистический	1,082	1,056	1,073

вероятности $p_j(x_1, \dots, x_k)$ с помощью статистики (2) запишем в виде

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

где N – количество тестовых плотностей вероятностей, представленных, например, в табл. 2.

В условиях неопределённости о виде восстанавливаемой плотности вероятности $p(x_1, \dots, x_k)$ оценим коэффициенты размытости ядерных функций как

$$\bar{c}_v^* = \frac{1}{N} \sum_{j=1}^N \beta(j) \bar{\sigma}_v n^{-1/(k+4)}, \quad v = \overline{1, k}, \quad (7)$$

где $\bar{\sigma}_v$ – статистическая оценка СКО случайной величины x_v , вычисляемая по исходной выборке её наблюдений $(x_v^i, i = \overline{1, n})$.

Проанализируем результаты применения значений $\bar{c}_v^*$, $v = \overline{1, k}$ при восстановлении $p(x_1, \dots, x_k)$ с помощью статистики (2). Для этого с учётом выражений (4), (7) определим значение $\bar{B}$,

соответствующее значению $\bar{\beta} = \frac{1}{N} \sum_{j=1}^N \beta(j)$ и являющееся оценкой B :

$$\bar{B} = k \left(\|\Phi(u)\|^2 \right)^k / \left(\bar{\beta}^{k+4} \prod_{v=1}^k \bar{\sigma}_v \right).$$

В этих условиях оценку СКО аппроксимации $p(x_1, \dots, x_k)$ при использовании статистики (2) определим следующим образом:

$$\bar{W}(\bar{c}^*) = \frac{4+k}{4k^{k/(k+4)}} \left\{ \left[\left(\|\Phi(u)\|^2 \right)^k / \left(n \prod_{v=1}^k \bar{\sigma}_v \right) \right]^4 \bar{B}^k \right\}^{(k+4)^{-1}}. \quad (8)$$

Рассмотрим применение методики быстрого выбора коэффициента размытости статистики (2). Пусть известны значения $\beta(j)$, $j = \overline{1, N}$, например, представленные в табл. 2. Тогда реализация предложенной методики сводится к выполнению следующих действий.

1. Оценить по выборке V значения СКО σ_v случайных величин x_v , $v = \overline{1, k}$.

2. Определить по данным табл. 2 значения $\beta(j)$, $j = \overline{1, N}$ при известном виде ядерной функции.

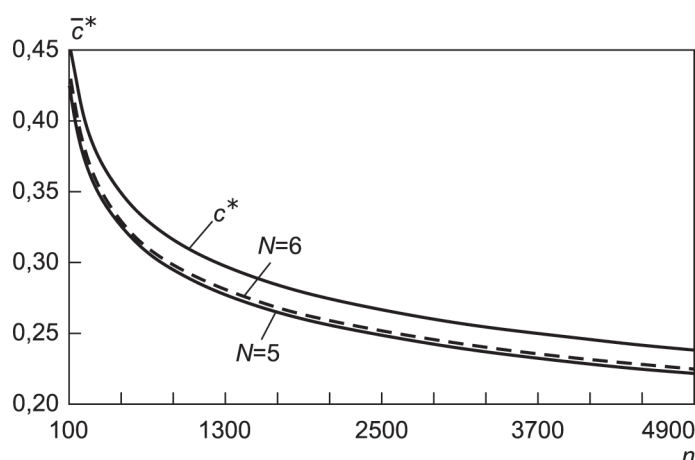


Рис. 2. Зависимость параметра $\bar{c}^*$ от объёма n исходных данных при восстановлении двумерной плотности вероятности независимых случайных величин с нормальными законами распределения при $N=6; 5$

3. Определить по формуле (7) оценки $\bar{c}_v^*$, $v = \overline{1, k}$ оптимальных коэффициентов размытости ядерных функций в статистике (2) при известных значениях $n, k, \bar{\sigma}_v, v = \overline{1, k}, \beta(j), j = \overline{1, N}$.

4. Оценить по приведённым выше рекомендациям СКО $\bar{W}(\bar{c}^*)$ непараметрической оценки плотности вероятности (2) в соответствии с выражением (8).

Для иллюстрации работы предложенной методики быстрого выбора коэффициентов размытости исследуем зависимость значений параметров $c^*, \bar{c}^*$ и соответствующих им значений $W(c^*), \bar{W}(\bar{c}^*)$ от объёма n статистических данных. Будем восстанавливать плотность вероятности $p(x_1, x_2)$ по выборке $V = (x_1^i, x_2^i, i = \overline{1, n})$, формируемой с помощью датчика независимых случайных величин с нормальными законами распределения $p(x_1) = N(0; 1), p(x_2) = N(0; 2)$. В этих условиях произведение $B\sigma_1\sigma_2 = 0,159$ и не зависит от параметров закона распределения. Синтез непараметрической оценки плотности вероятности (2) при $\Phi(u_1) = \Phi(u_2) = \Phi(u)$ осуществим на

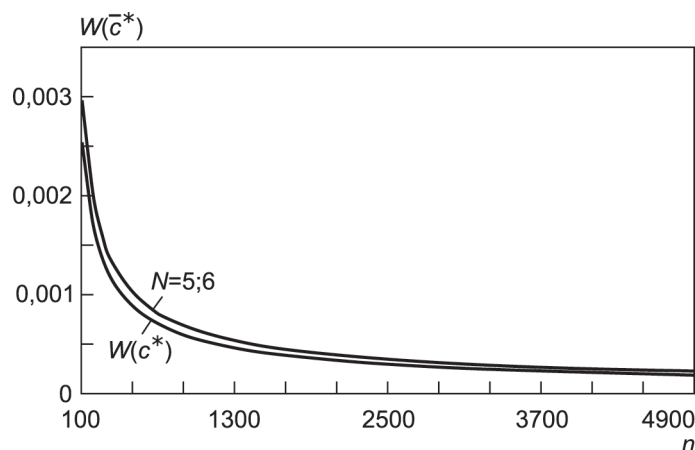


Рис. 3. Зависимость значений $W(\bar{c}^*)$ от объёма n исходных данных при $N=6; 5$

основе ядерных функций В. А. Епанечникова [16]:

$$\Phi(u) = \begin{cases} \frac{3}{4\sqrt{5}} - \frac{3}{20\sqrt{5}}u^2 & \forall |u| < \sqrt{5}; \\ 0 & \forall |u| \geq \sqrt{5}. \end{cases}$$

Тогда значение $\|\Phi(u)\|^2 = 3/(5\sqrt{5})$.

При расчётах значений β информация о нормальном законе распределения случайных величин присутствует и исключается из табл. 2. В первом случае примем $N=6$, во втором – $N=5$. Для определения зависимости $\bar{W}(\bar{c}^*)$ от n проведём вычислительный эксперимент. При конкретных заданных условиях значения $\bar{W}(\bar{c}^*)$ рассчитаем 50 раз, полученные результаты усредним.

Параметры c^* коэффициентов размытости ядерных функций, вычисленные по формуле (4), превышают значения оценок

$$\bar{c}^* = \frac{1}{N} \sum_{j=1}^N \beta(j) n^{-1/(k+4)}$$

на всём диапазоне изменения объёма n исходных статистических данных (рис. 2).

Эта закономерность сохраняется и при исключении сведений о нормальном законе распределения случайных величин из табл. 2. В этом случае $N=5$. Зависимости $\bar{c}^*$ от объёма n статистических данных при $N=6, N=5$ отличаются незначительно. При этом зависимости СКО аппроксимации $W(c^*)$ из формулы (5) и $\bar{W}(\bar{c}^*)$ от объёма n статистических данных практически не отличаются (рис. 3). Полученные результаты подтверждают слабую зависимость отношения $W(\bar{c}^*)/W(c^*)$ от ошибки оценивания функционала B из формулы (3) (см. табл. 1). Например, значения $B(n)$ на всём диапазоне изменения n принимают значения в пределах 0,08, а их оценки $B(n) \approx 0,115; 0,124$ при $N=6; 5$, соответственно. Таким образом, если существует значительная ошибка при оценивании функционала B из формулы (3) СКО $W(c^*), \bar{W}(\bar{c}^*)$ отличаются незначительно. Данный факт подтверждает возможность разработки быстрых алгоритмов выбора коэффициентов размытости ядерных функций в статистике $\bar{p}(x_1, \dots, x_k)$ (2) путём оценивания функционала B в формуле (4).

Похожие результаты получены при восстановлении двумерной плотности вероятности независимых случайных величин с экспоненциальными законами распределения. В этих условиях максимальное отклонение $W(c^*)$ от $\bar{W}(\bar{c}^*)$ не превышает значение 0,005 при отклонении значений $B, \bar{B}$ в два раза.

Закключение. Оптимальные значения параметра c^* коэффициентов размытости $c_v^* = c^* \sigma_v, v = \overline{1, k}$ ядерной оценки многомерной плотности вероятности определяются значением

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

Зависимость отношения СКО $W(\bar{c}^*)/W(c^*)$ от параметра α ошибки оценивания B имеет экстремальный характер. Если оценка параметра $\bar{c}^*$ коэффициентов размытости ядерных функций больше его оптимального значения, то наблюдается более значимое увеличение $W(\bar{c}^*)$ по сравнению с условием $\bar{c}^* < c^*$. При двукратном увеличении

ошибки оценивания функционала B отношение $W(\bar{c}^*)/W(c^*)$ не превышает значения 1,1 для $k \in [1; 5]$, что обосновывает возможность разработки быстрых алгоритмов оптимизации ядерных оценок плотности вероятности по коэффициентам размытости ядерных функций.

Существует семейство многомерных плотностей вероятности независимых случайных величин, для которых значение

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

не зависит от её параметров. На этой основе разработана методика быстрого выбора коэффициентов размытости ядерных функций для непараметрической оценки многомерной плотности вероятности. Идея предложенного подхода состоит в вычислении оптимальных значений коэффициентов размытости ядерных функций для непараметрических оценок составляющих тестового семейства многомерных плотностей вероятности. Они обобщаются при определении $\bar{c}^*$ в процедуре коллективного типа. Для найденного значения $\bar{c}^*$ можно определить оценку СКО $W(\bar{c}^*)$ статистики $\bar{p}(x_1, \dots, x_k)$ от $p(x_1, \dots, x_k)$.

Полученные результаты создают основу для широкого применения непараметрических оценок плотностей вероятности в прикладных задачах, характеризующихся наличием статистических данных большого объёма.

Работа выполнена при поддержке РФФИ (грант № 18-01-00251).

Л и т е р а т у р а

1. **Лапко А. В., Лапко В. А.** Многоуровневые непараметрические системы обработки информации. Красноярск: СибГАУ, 2013.
2. **Rudemo M.** Empirical choice of histogram and kernel density estimators // *Scandinavian Journal of Statistics*. 1982. № 9. P. 65–78.
3. **Hall P.** Large-sample optimality of least squares cross-validation in density estimation // *Annals of Statistics*. 1983. V. 11. P. 1156–1174.
4. **Jiang M., Provost S.B.** A hybrid bandwidth selection methodology for kernel density estimation // *Journal of Statistical Computation and Simulation*. 2014. V. 84. No. 3. P. 614–627.

5. **Dutta S.** Cross-validation Revisited // *Communications in Statistics-Simulation and Computation*. 2016. V. 45. No. 2. P. 472–490.

6. **Heidenreich N.-B., Schindler A., Sperlich S.** Bandwidth selection for kernel density estimation: a review of fully automatic selectors // *AStA Advances in Statistical Analysis*. 2013. V. 97, No. 4. P. 403–433.

7. **Лапко А. В., Лапко В. А.** Анализ методов оптимизации непараметрической оценки плотности вероятности по коэффициенту размытости ядерных функций // *Измерительная техника*. 2017. № 6. С. 3–8.

8. **Borrajó M. I., González-Manteiga W., Martínez-Miranda M. D.** Bandwidth selection for kernel density estimation with length-biased data // *Journal of Nonparametric Statistics*. 2017. V. 29. No. 3. P. 636–668.

9. **Chen S.** Optimal bandwidth selection for kernel density functionals estimation // *Journal of Probability and Statistics*. 2015. V. 2015. P. 1–21.

10. **Sheather S. J.** Density Estimation // *Statistical Science*. 2004. V. 19. No. 4. P. 588–597.

11. **Silverman B. W.** Density Estimation for Statistics and Data Analysis. London: Chapman and Hall, 1986.

12. **Scott D. W.** Multivariate Density Estimation: Theory, Practice, and Visualization. N. Y.: John Wiley, 1992.

13. **Bowman A. W., Azzalini A.** Applied Smoothing Techniques for Data Analysis. London: Oxford University Press, 1997.

14. **Parzen E.** On estimation of a probability density function and mode // *The Annals of Mathematical Statistics*. 1962. V. 33. No 3. P. 1065–1076.

15. **Деврой Л., Дьерфи Л.** Непараметрическое оценивание плотности. (L_1 -подход). М.: Мир, 1988.

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

17. **Лапко А. В., Лапко В. А.** Regression Estimate of the Multidimensional Probability Density and Its Properties // *Optoelectronics, Instrumentation and Data Processing*. 2014. V. 50. No. 2. P. 148–153.

Дата принятия: 13.04.2018.

Fast algorithm bandwidth selection for multivariate kernel density estimation

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

¹Institute of Computational Modeling, Siberian Branch of the Russian Academy of Sciences (ICM SB RAS), Krasnoyarsk, Russia

²Reshetnev Siberian State University of Science and Technology, Krasnoyarsk, Russia, e-mail: lapko@icm.krasn.ru

Proposed fast technique bandwidth selection for kernel function in multivariate nonparametric density estimation of the Rosenblatt–Parzen type. The method is based on the results of an analysis of the asymptotic properties of the multivariate density estimate. The properties of a fast algorithm bandwidth selection for multivariate kernel density estimation are investigated.

Key words: multivariate nonparametric density estimation, bandwidth selection, Rosenblatt–Parzen estimate, fast optimization algorithm, asymptotic properties, multivariate data analysis.