

# ПРИНЦИПЫ ВЫБОРА АППРОКСИМАЦИОННОГО БАЗИСА И ЯДЕРНОЙ ФУНКЦИИ В ВЫЧИСЛИТЕЛЬНОМ АЛГОРИТМЕ ПРИБЛИЖЕНИЯ ВЕРОЯТНОСТНОЙ ПЛОТНОСТИ ПО ЗАДАННОЙ ВЫБОРКЕ

Нурлыбай Хамдуллаевич Шлымбетов<sup>1</sup>  
Убайдулла Пахтамурат улы Сейтмуратов<sup>2</sup>  
Антон Вацлавович Войтишек<sup>3</sup>

<sup>1,2</sup>Новосибирский Государственный Университет

<sup>3</sup>Институт Вычислительной Математики и Математической Геофизики  
СО РАН

<sup>1,2,3</sup>Новосибирск, Россия,

<sup>1</sup>nurlibay\_xamdullaevich@mail.ru; <https://orcid.org/0009-0004-4417-0842>

<sup>2</sup>u.seitmuratov@g.nsu.ru; <https://orcid.org/0009-0001-0189-6625>

<sup>3</sup>vav@osmf.sccc.ru; <https://orcid.org/0000-0001-8294-6948>

## *Аннотация*

В данной работе сформулированы два важных принципа – мультипликативность вида (в многомерном случае) и финитность (сосредоточенность носителей около узлов аппроксимационной сетки) – при выборе аппроксимационного базиса и ядерной функции для конструирования экономичного вычислительного (компьютерного) алгоритма приближения вероятностной плотности по заданной выборке. С учетом особенностей рассматриваемой сеточной вычислительной схемы, предложены новые критерии для оптимального выбора ядерной функции, включающие правильные сочетания вторых моментов и интегралов от квадрата этих функций, определяющих одновременно компоненты смещения и стохастические компоненты среднеквадратической погрешности рассматриваемого ядерного алгоритма. Особо выделен важный частный случай – многомерный аналог полигона частот (здесь выбираемая ядерная функция является кусочно-постоянной), для которого удастся найти параметры обеспечивающие минимальность затрат (при заданном уровне погрешности). На тестовом примере показано, что выбор известных типов ядерных функций, отличных от кусочно-постоянных, не позволяет проводить оптимизацию алгоритма и увеличивает время вычислений (при заданном уровне погрешности).

*Ключевые слова и фразы*

вычислительный (компьютерный) функциональный ядерный алгоритм, выбор аппроксимационного базиса, выбор ядерной функции, мультипликативный вид выбираемых функций, финитность выбираемых функций, среднеквадратическая погрешность, многомерный аналог полигона частот.

*Источник финансирования*

Исследования выполнены в рамках государственного задания ИВМиМГ СО РАН FWNM-2025-0002

*Для цитирования*

Шлымбетов Н. Х., Сейтмуратов У. П., Войтишек А. В. Принципы выбора аппроксимационного базиса и ядерной функции в вычислительном алгоритме приближения вероятностной плотности по заданной выборке // *Математические труды*, 2026, Т. 29, № 2, С. 200-218. DOI 10.25205/1560-750X-2026-29-2-200-218

## Principles of choosing an approximation basis and a kernel function in a computational algorithm for approximating the probability density from a given sample

Nurlibay Kh. Shlimbetov<sup>1</sup>, Ubaydulla P. Seytmuratov<sup>2</sup>, Anton V. Voytishek<sup>3</sup>,

<sup>1,2</sup>Novosibirsk State University,

<sup>3</sup>Institute of Computational Mathematics and Mathematical Geophysics (ICM MG) SB RAS, Novosibirsk, Russia

<sup>1</sup>nurlibay\_xamdullaevich@mail.ru;<https://orcid.org/0009-0004-4417-0842>

<sup>2</sup>u.seitmuratov@g.nsu.ru;<https://orcid.org/0009-0001-0189-6625>

<sup>3</sup>vav@osmf.sccc.ru;<https://orcid.org/0000-0001-8294-6948>

*Abstract*

This paper formulates two important principles – multiplicativity of the form (in the multivariate case) and finiteness (concentration of supports near the nodes of the approximation grid) – when choosing the approximation basis and the kernel function for constructing a cost-effective computational (computer) algorithm for approximating a probability density from a given sample. Taking into account the features of the considered grid-based computational scheme, new criteria for the optimal choice of the kernel function are proposed. These include appropriate combinations of the second moments and the integrals of the squares of these functions, which simultaneously determine the bias components and the stochastic components of the root-mean-square error of the considered kernel algorithm. A particularly important special case is highlighted – the multivariate analog of the frequency polygon (where the chosen kernel function is piecewise constant) – for which it is possible to find parameters that ensure minimal computational cost (for a given error level). A test example demonstrates that the choice of known types of kernel functions, different from piecewise constant ones, does not allow for algorithm optimization and increases computation time (for a given error level).

*Keywords*

computational (computer) functional kernel algorithm, choice of approximation basis, choice of kernel function, multiplicative form of chosen functions, finiteness of chosen functions, root-mean-square error, multivariate analog of the frequency polygon.

*Funding*

The research was carried out within the framework of the state assignment of the ICM MG of the SB RAS FWNM–2025–0002

*For citation*

*Shlimbetov N. Kh., Seytmuratov U. P., Voytishchik A. V., Principles of choosing an approximation basis and a kernel function in a computational algorithm for approximating the probability density from a given sample // Mat. Trudy, 2026, V. 29, N. 2, P. 200-218. DOI 10.25205/1560-750X-2026-29-2-200-218*

## § 1. Введение: вычислительный функциональный ядерный алгоритм

ISSN 1560-750X (Print) ISSN 3033-8271 (Online)

Математические труды, 2026, Том 29, № 2, С. 200-218

Mat. Trudy, 2026, V. 29, N. 2, P. 200-218

В нашей работе [1] отмечено, что при обработке больших данных (например, при применении технологий машинного обучения) актуальным оказывается использование экономичных алгоритмов компьютерного приближения неизвестной вероятностной плотности  $f_{\boldsymbol{\xi}}(\mathbf{x})$  по заданной выборке  $\{\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n\}$  с заданным уровнем погрешности  $\Delta > 0$  на всей компактной области распределения  $X \subset \mathbb{R}^d$ . Здесь хорошими свойствами обладает следующий вычислительный функциональный ядерный алгоритм.

Не ограничивая общности, предположим, что конечные границы  $a^{(k)}, b^{(k)}$ ,  $k = 1, \dots, d$  области  $X$  можно ввести таким образом, чтобы в  $X$  можно было бы ввести равномерную сетку с шагом  $h$  вида

$$\mathbf{y}_i = \mathbf{y}_{\mathbf{J}_i} = \left( a^{(1)} + j_i^{(1)}h, \dots, a^{(d)} + j_i^{(d)}h \right); \quad i = 1, \dots, M; \quad \mathbf{J}_i = \left( j_i^{(1)}, \dots, j_i^{(d)} \right), \quad (1)$$

$$j_i^{(m)} = 0, 1, \dots, M_m; \quad m = 1, \dots, d; \quad M = (M_1 + 1) \times \dots \times (M_d + 1);$$

с шагом  $h = (b^{(m)} - a^{(m)})/M_m$ ;  $m = 1, \dots, d$ .

АЛГОРИТМ. Выберем ядерную функцию  $\kappa^{(\mathbf{x})}(\mathbf{z})$  (см., например, [2–4]) такую, что  $f_{\boldsymbol{\xi}}(\mathbf{x}) \approx \int_X f_{\boldsymbol{\xi}}(\mathbf{z})\kappa^{(\mathbf{x})}(\mathbf{z}) d\mathbf{z} = \mathbf{E}\kappa^{(\mathbf{x})}(\boldsymbol{\xi})$  для всех  $\mathbf{x} \in X$ . Реализуем на компьютере соответствующие приближения  $f_{\boldsymbol{\xi}}(\mathbf{y}_i) \approx \tilde{f}_{\boldsymbol{\xi}}^{(\mathbf{y}_i)}(n) = \frac{1}{n} \sum_{j=1}^n \kappa^{(\mathbf{y}_i)}(\boldsymbol{\xi}_j)$ ;  $i = 1, \dots, M$  (здесь  $\{\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n\}$  – заданная выборка или ее подмножество) и строим окончательную классическую компьютерную аппроксимацию (см., например, главы 2 и 4 книги [5]) плотности  $f_{\boldsymbol{\xi}}(\mathbf{x})$ :

$$f_{\boldsymbol{\xi}}(\mathbf{x}) \approx L^{(M,n)} \tilde{f}_{\boldsymbol{\xi}}(\mathbf{x}) = \sum_{i=1}^M w^{(i)} \left[ \tilde{\mathbf{f}}_{\boldsymbol{\xi}}(n) \right] \chi^{(i)}(\mathbf{x}), \quad (2)$$

где  $\tilde{\mathbf{f}}_{\boldsymbol{\xi}}(n) = \left( \tilde{f}_{\boldsymbol{\xi}}^{(\mathbf{y}_1)}(n), \dots, \tilde{f}_{\boldsymbol{\xi}}^{(\mathbf{y}_M)}(n) \right)$ .

В формуле (2)  $\chi^{(M)} = \{\chi^{(1)}(\mathbf{x}), \dots, \chi^{(M)}(\mathbf{x})\}$  – заданный набор функций (аппроксимационный базис), а  $\mathbf{W}^{(M)} = \left\{ w^{(1)}(\mathbf{f}_{\boldsymbol{\xi}}), \dots, w^{(M)}(\mathbf{f}_{\boldsymbol{\xi}}) \right\}$  – набор аппроксимационных коэффициентов, являющихся, как правило, линейными комбинациями значений  $\mathbf{f}_{\boldsymbol{\xi}} = \left( f_{\boldsymbol{\xi}}(\mathbf{y}_1), \dots, f_{\boldsymbol{\xi}}(\mathbf{y}_M) \right)$  приближаемой функции  $f_{\boldsymbol{\xi}}(\mathbf{x})$  в узлах сетки (1) вида

$$w^{(i)}(\mathbf{f}_{\boldsymbol{\xi}}) = a_1^{(i)} f_{\boldsymbol{\xi}}(\mathbf{y}_{q_1}) + \dots + a_{s^{(i)}}^{(i)} f_{\boldsymbol{\xi}}(\mathbf{y}_{q_{s^{(i)}}}); \quad i = 1, \dots, M; \quad (3)$$

здесь среди номеров  $q_1, \dots, q_{s^{(i)}}$  присутствует номер  $i$ , а также выполнено соотношение  $a_1^{(i)} + \dots + a_{s^{(i)}}^{(i)} = 1$ .

Оригинальность представленного алгоритма и его теории по сравнению с широко известными «точечными» конструкциями (см., например, [2–4]) состоит в рассмотрении для этой задачи схемы из вычислительной математики (также широко известной) – разложения приближаемой функции по аппроксимационному базису (см. формулу (2)). В результате при оптимизации алгоритма (с помощью выбора параметров  $n$ ,  $M$  и функций  $\chi^{(M)}$ ,  $\kappa^{(\mathbf{x})}(\mathbf{z})$ ) кроме традиционных двух компонент среднеквадратической погрешности – компоненты смещения  $\delta_{bias}^{(\mathbb{L}_2)}(M) = \left\| L^{(M)} f_{\boldsymbol{\xi}} - L^{(M)} \bar{f}_{\boldsymbol{\xi}} \right\|_{\mathbb{L}_2(X)}$  и стохастической компоненты  $\delta_{stoch}^{(\mathbb{L}_2)}(M, n) = \int_X \mathbf{D}L^{(M,n)} \tilde{f}_{\boldsymbol{\xi}}(\mathbf{x}) d\mathbf{x}$  (см. [2–4], а также раздел 4 данной работы) – следует рассматривать соответствующую компоненту аппроксимации  $\delta_{appr}^{(\mathbb{L}_2)}(M) = \left\| f_{\boldsymbol{\xi}} - L^{(M)} f_{\boldsymbol{\xi}} \right\|_{\mathbb{L}_2(X)}$ :

$$\left[ \mathbf{E} \left\| f_{\boldsymbol{\xi}} - L^{(M,n)} \tilde{f}_{\boldsymbol{\xi}} \right\|_{\mathbb{L}_2(X)} \right]^2 \leq 2 \left[ \delta_{appr}^{(\mathbb{L}_2)}(M) \right]^2 + 2 \left[ \delta_{bias}^{(\mathbb{L}_2)}(M) \right]^2 + \delta_{stoch}^{(\mathbb{L}_2)}(M, n) \quad (4)$$

(подробности см. в нашей работе [6]); здесь  $L^{(M)} f_{\boldsymbol{\xi}}(\mathbf{x}) = \sum_{i=1}^M w^{(i)} \left[ \mathbf{f}_{\boldsymbol{\xi}} \right] \chi^{(i)}(\mathbf{x})$  – аппроксимация функции  $f_{\boldsymbol{\xi}}(\mathbf{x})$  по точным значениям

$\mathbf{f}_{\boldsymbol{\xi}} = \left( f_{\boldsymbol{\xi}}(\mathbf{y}_1), \dots, f_{\boldsymbol{\xi}}(\mathbf{y}_M) \right)$  в узлах сетки (1) и  $L^{(M)} \bar{f}_{\boldsymbol{\xi}}(\mathbf{x}) = \sum_{i=1}^M w^{(i)} \left[ \bar{\mathbf{f}}_{\boldsymbol{\xi}} \right] \chi^{(i)}(\mathbf{x})$  для  $\bar{\mathbf{f}}_{\boldsymbol{\xi}} = \left( \mathbf{E}\kappa^{(\mathbf{y}_1)}(\boldsymbol{\xi}), \dots, \mathbf{E}\kappa^{(\mathbf{y}_M)}(\boldsymbol{\xi}) \right)$ . Кроме того, при оптимизации алгоритма следует рассматривать проблему минимизации его затрат.

В данной работе показано, что для получения оптимизированных версий сформулированного ядерного алгоритма целесообразным является применение двух – вполне «естественных» – принципов при выборе функций  $\chi^{(M)}$  и  $\kappa^{(\mathbf{x})}(\mathbf{z})$ : мультипликативность вида (представимость в виде произведения однотипных функций одного переменного для каждой компоненты) – в многомерном случае и финитность (сосредоточенность носителей около узлов сетки (1)). Кроме того, в данной работе предложены новые критерии для оптимального выбора ядерной функции, включающие правильные сочетания вторых моментов и интегралов от квадрата этих функций, определяющих одновременно компоненты смещения и стохастические компоненты среднеквадратической погрешности рассматриваемого ядерного алгоритма. На тестовом примере показана целесообразность использования многомерного аналога полигона частот [1] (здесь выбираемая ядерная функция является кусочно-постоянной), для которого, в том числе, удастся найти параметры, обеспечивающие минимальность затрат (при заданном уровне погрешности  $\Delta > 0$ ).

## § 2. О выборе аппроксимационного базиса $\chi^{(M)}$

Согласно общей теории компьютерных сеточных приближений функций (см., например, главы 2 и 4 книги [5]), базис  $\chi^{(M)}$  и аппроксимационные коэффициенты  $\mathbf{W}^{(M)}$  следует выбирать таким образом, чтобы обеспечить определенный (например,  $r$ -й) порядок аппроксимации по шагу  $h$  сетки (1), т. е. для компоненты аппроксимации из соотношения (4) должно выполняться неравенство вида

$$\delta_{appr}^{(\mathbb{L}_2)}(M) \leq H_r h^r = \tilde{H}_r \times M^{-r/d} \quad (5)$$

для некоторых констант  $H_r$  и  $\tilde{H}_r$ . Отметим также, что для минимизации компоненты смещения  $\delta_{bias}^{(\mathbb{L}_2)}(M)$  и стохастической компоненты  $\delta_{stoch}^{(\mathbb{L}_2)}(M, n)$  из соотношения (4) выбор функций  $\chi^{(M)}$  и коэффициентов  $\mathbf{W}^{(M)}$  должен обеспечивать относительную малость т. н. константы устойчивости  $K^{(stab)} = K(\chi^{(M)}) \times K(\mathbf{W}^{(M)})$ , где  $K(\chi^{(M)}) = \sup_{\mathbf{x} \in X} \sum_{i=1}^M |\chi^{(i)}(\mathbf{x})|$  – т. н. «константа Лебега» (см., например, раздел 3.1 книги [7]), а  $K(\mathbf{W}^{(M)}) = \max_{i=1, \dots, M} \left( |a_1^{(i)}| + \dots + |a_{s^{(i)}}^{(i)}| \right)$  (см. формулу (3)) – т. н. «константа компактности коэффициентов» [1].

Следуя нашей работе [1], можно отметить, что наилучшими свойствами для выполнения сформулированных выше принципов и соотношений обладает т. н. «аппроксимации Стренга–Фикса» (см. главу 2 книги [8]) с базисными функциями

$$\chi^{(i)}(\mathbf{x}) = \hat{\chi}^{(i, r-1)}(\mathbf{x}) = B^{(r-1)} \left[ \frac{x^{(1)} - a^{(1)}}{h} - j_i^{(1)} \right] \times \dots \times B^{(r-1)} \left[ \frac{x^{(d)} - a^{(d)}}{h} - j_i^{(d)} \right] \quad (6)$$

для сетки (1); здесь функция  $B^{(r-1)}(w)$  обозначает  $B$ -сплайн  $(r-1)$ -го порядка, который определяется рекуррентным соотношением

$$B^{(r-1)}(w) = B^{(r-2)} * B^{(0)}(w) = \int_{-\infty}^{\infty} B^{(r-2)}(w-v) B^{(0)}(v) dv; \quad r = 2, 3, \dots,$$

$$\text{где } B^{(0)}(w) = \begin{cases} 1 & \text{при } -1/2 \leq w \leq 1/2; \\ 0 & \text{при } |w| > 1/2. \end{cases}$$

Мультипликативная конструкция (6) позволяет обосновывать целый ряд свойств функций аппроксимационного базиса, используя метод математической индукции по размерности  $d$ . Так, используя этот метод (а также дополнительную индукцию по порядку  $(r-1)$  используемых сплайнов) несложно показать, что базис (6) является разложением единицы, т. е.  $\sum_i \hat{\chi}^{(i, r-1)}(\mathbf{x}) \equiv 1$  для всех  $r = 1, 2, 3, \dots$  и  $\mathbf{x} \in \mathbb{R}^d$  (см., например, [1]). Из

последнего утверждения следует, что для базиса (6) выполнено равенство  $K(\mathbf{X}^{(M)}) = 1$ .

В работе [1] показана целесообразность использования базиса (6) для  $r = 2$ . Здесь индукцией по  $d$  удалось получить соотношение (5) с константой  $H_2 = \frac{\sqrt{\text{mes } X}}{8} \times \sum_{s=1}^d \max_{\mathbf{x} \in X} \left| \frac{\partial^2}{\partial (x^{(s)})^2} f_{\boldsymbol{\xi}}(\mathbf{x}) \right|$ ; при этом выполнение неравенства обеспечивают простейшие коэффициенты  $w^{(i)}(\mathbf{f}_{\boldsymbol{\xi}}) = f_{\boldsymbol{\xi}}(\mathbf{y}_i)$ ;  $i = 1, \dots, M$ , для которых  $K(\mathbf{W}^{(M)}) = 1$ . Еще одним аргументом в пользу применения базиса (6) для  $r = 2$  является то, что компонента смещения  $\delta_{bias}^{(L_2)}(M)$  из соотношения (4) имеет (как и компонента  $\delta_{appr}^{(L_2)}(M)$  из соотношения (5) для  $r = 2$ ) порядок  $h^2$  по шагу сетки (1) (см. далее формулу (15)).

Для случая  $r > 2$ , по-видимому, верно следующее утверждение.

**Теорема.** Если функция  $f_{\boldsymbol{\xi}}(\mathbf{x})$  принадлежит соболевскому пространству  $\mathbb{W}_2^r(X)$ , и в приближении по точным значениям  $L^{(M)} f_{\boldsymbol{\xi}}(\mathbf{x})$  используются функции аппроксимационного базиса (6), то найдутся такие коэффициенты  $\mathbf{W}^{(M)}$ , что справедливо неравенство (5), где  $(r - 1)$  – порядок сплайнов в формуле (6).

Во всяком случае, для  $r = 4$  (т.е. для кубических сплайнов (6)) для размерности  $d = 1$  оценка вида (5) для  $f_{\boldsymbol{\xi}} \in \mathbb{W}_2^4[a, b]$  получена в гл. 9 книги [9] для коэффициентов

$$w^{(i)}(\mathbf{f}_{\boldsymbol{\xi}}) = -\frac{1}{6}f_{\boldsymbol{\xi}}(y_i - h) + \frac{4}{3}f_{\boldsymbol{\xi}}(y_i) - \frac{1}{6}f_{\boldsymbol{\xi}}(y_i + h); \quad i = 2, \dots, M - 1$$

(это комбинация значений функции  $f_{\boldsymbol{\xi}}(x)$  в трех узлах сетки, включая узел  $y_i = a + (i - 1)h$ , т.е. в формуле (3) имеем  $s^{(i)} = 3$ ). Этот результат обобщен в [10] на случай произвольного  $d$ , и получена оценка (5) с константой  $H_4 = 0.030382 \times \text{mes } X \times \sum_{s=1}^d \max_{\mathbf{x} \in X} \left| \frac{\partial^4}{\partial (x^{(s)})^4} f_{\boldsymbol{\xi}}(\mathbf{x}) \right|$  для коэффициентов вида

$$w^{(i)}(\mathbf{f}_{\boldsymbol{\xi}}) = \left(\frac{4}{3}\right)^d \sum_{\mathbf{k} \in \{-1, 0, +1\}^d} \frac{f_{\boldsymbol{\xi}}(\mathbf{y}_{\mathbf{J}_i + \mathbf{k}})}{(-8)^{|\mathbf{k}^{(1)}| + \dots + |\mathbf{k}^{(d)}|}}, \quad \text{где } \mathbf{k} = (k^{(1)}, \dots, k^{(d)});$$

здесь задействованы уже  $s^{(i)} = 3^d$  значений функции  $f_{\boldsymbol{\xi}}(\mathbf{x})$ .

Тем не менее, следуя работе [1], отметим, что при использовании сформулированного выше ядерного алгоритма рассмотрение случаев  $r > 2$  в формуле (6), а также базисов, отличных от (6), приводит к существенному усложнению выбора подходящих (обеспечивающих оптимальный порядок погрешности по шагу  $h$  сетки (1)) коэффициентов  $\mathbf{W}^{(M)}$  и радикальному возрастанию обоих множителей в выражении для соответствующей константы устойчивости  $K^{(stab)} = K(\mathbf{X}^{(M)}) \times K(\mathbf{W}^{(M)})$ .

### § 3. О выборе ядерной функции $\kappa^{(\mathbf{x})}(\mathbf{z})$

По аналогии с работой [2] будем выбирать ядерную функцию в виде

$$\kappa^{(\mathbf{x})}(\mathbf{z}) = \prod_{s=1}^d \frac{1}{\hat{h}} K\left(\frac{x^{(s)} - z^{(s)}}{\hat{h}}\right); \mathbf{z} = (z^{(1)}, \dots, z^{(d)}), \quad (7)$$

где положительное число  $\hat{h}$  определяют носитель («коэффициент размытости») ядерной функции  $\kappa^{(\mathbf{x})}(\mathbf{z})$ , а ограниченная неотрицательная функция  $K(y)$  является симметричной относительно нуля (четной) вероятностной плотностью:  $K(y) = K(-y)$  и  $\int_{-\infty}^{+\infty} K(y) dy = 1$ .

Обратим внимание на то, что конструкция ядерной функции (7) аналогична мультипликативной конструкции (6). Такая конструкция позволяет перенести свойства одномерных ядерных функций, определяемых теми или иными образующими функциями  $K(y)$ , на многомерный случай, реализуя соответствующую индукцию по размерности  $d$ . Поэтому далее – для простоты изложения – рассмотрим одномерный случай функции (7) вида

$$\kappa^{(x)}(z) = \frac{1}{\hat{h}} K\left(\frac{x - z}{\hat{h}}\right).$$

Заметим также, что оптимальная (с точки зрения минимальной трудоемкости) компьютерная реализация сформулированного выше ядерного алгоритма получается в случае, когда коэффициент размытости  $\hat{h}$  равен шагу сетки  $h$  (см. формулу (1)), и, кроме того, функция  $K(y)$  должна быть положительна только на интервале  $|y| < 1/2$ :

$$\kappa^{(x)}(z) = \frac{1}{h} K\left(\frac{x - z}{h}\right); K(y) > 0 \text{ при } |y| < \frac{1}{2} \text{ и } K(y) = 0 \text{ при } |y| > \frac{1}{2}. \quad (8)$$

В этом случае каждое из используемых выборочных значений  $\xi_j$ ;  $j = 1, \dots, n$  попадает только в один из кубов  $D^{(y_i)}$ ;  $i = 1, \dots, M$ , где

$$D^{(\mathbf{x})} = \{\mathbf{z} = (z^{(1)}, \dots, z^{(d)}) : x^{(s)} - h/2 \leq z^{(s)} < x^{(s)} + h/2; s = 1, \dots, d\}.$$

Пусть  $t$  – время определения того, в какой из кубов  $D^{(y_i)}$  попадает очередное выборочное значение  $\xi_j$ . Несложно добиться того, чтобы время  $t$  не зависело от числа  $M$  кубов  $D^{(y_i)}$ .

Действительно, если выборочное значение  $\xi_j$  имеет компоненты  $\xi_j = (\xi_j^{(1)}, \dots, \xi_j^{(d)})$ , а  $M = M_1 \times \dots \times M_d$  (где  $M_m$  – число разбиений отрезка  $[a^{(m)}, b^{(m)}]$  для  $X = [a^{(1)}, b^{(1)}] \times \dots \times [a^{(d)}, b^{(d)}]$ ; т. е.  $h = \frac{b^{(m)} - a^{(m)}}{M_m}$ ;  $m = 1, \dots, d$

– см. формулу (1)), то определить номер  $s_m$  полуинтервала разбиения отрезка  $[a^{(m)}, b^{(m)}]$ , в который попадает случайная компонента  $\xi_j^{(m)}$ , можно по формуле

$$s_m = \left[ \frac{\xi_j^{(m)} - a^{(m)}}{h} \right] + 1; \quad m = 1, \dots, d,$$

где  $[A]$  – целая часть числа  $A$ . Таким образом, затраты оптимизированных версий ядерного алгоритма явно не зависят от  $M$ .

#### § 4. Использование $\mathbb{L}_2$ -подхода

По аналогии с работами [2–4] рассмотрим следующие рассуждения о среднеквадратической погрешности для «точечной» ядерной оценки  $\tilde{f}_\xi^{(x)}(n) = \frac{1}{n} \sum_{j=1}^n \kappa^{(x)}(\xi_j)$  (здесь  $\{\xi_1, \dots, \xi_n\}$  – заданная одномерная выборка) – для нее сеточное приближение (2) не рассматривается. Применяя неравенство Коши–Буняковского и теорему Фубини, можно получить представление этой погрешности в виде суммы компоненты смещения и стохастической компоненты:

$$\begin{aligned} \left[ \mathbf{E} \left\| f_\xi(x) - \tilde{f}_\xi^{(x)}(n) \right\|_{\mathbb{L}_2(\mathbb{R})} \right]^2 &\leq \int_{-\infty}^{+\infty} \mathbf{E} \left[ f_\xi(x) - \tilde{f}_\xi^{(x)}(n) \right]^2 dx = \\ &= \int_{-\infty}^{+\infty} \left[ f_\xi(x) - \mathbf{E} \tilde{f}_\xi^{(x)}(n) \right]^2 dx + \int_{-\infty}^{+\infty} \mathbf{D} \tilde{f}_\xi^{(x)}(n) dx. \end{aligned}$$

Далее, с учетом соотношения (8), в рассматриваемом одномерном случае имеем:

$$\begin{aligned} \mathbf{E} \tilde{f}_\xi^{(x)}(n) &= \frac{1}{n} \sum_{j=1}^n \frac{1}{h} \mathbf{E} K \left( \frac{x - \xi_j}{h} \right) = \\ &= \frac{1}{h} \int_{-\infty}^{+\infty} K \left( \frac{x - u}{h} \right) f_\xi(u) du = \int_{-1/2}^{+1/2} K(v) f_\xi(x - hv) dv; \end{aligned}$$

здесь использована замена  $v = (x - u)/h$ . Разлагая функцию  $f_\xi(x - hv)$  в ряд Тейлора по переменной  $s = hv$  при  $h \downarrow 0$ , имеем

$$f_\xi(x - hv) = f_\xi(x) - hv f'_\xi(x) + \frac{1}{2} (hv)^2 f''_\xi(x) + o(h^2); \quad (9)$$

и поэтому, учитывая то, что  $\int_{-1/2}^{+1/2} K(v) dv = 1$ , справедливо соотношение

$$\mathbf{E} \tilde{f}_\xi^{(x)}(n) = f_\xi(x) \int_{-1/2}^{+1/2} K(v) dv - h f'_\xi(x) \int_{-1/2}^{+1/2} v K(v) dv +$$

$$+\frac{h^2}{2}f''_{\xi}(x)\int_{-1/2}^{+1/2}v^2K(v)dv+o(h^2)=f_{\xi}(x)+\frac{h^2}{2}f''_{\xi}(x)\int_{-1/2}^{+1/2}v^2K(v)dv+o(h^2). \tag{10}$$

Аналогично, используя разложение (9), имеем

$$\begin{aligned} \mathbf{D}\tilde{f}_{\xi}^{(x)}(n) &= \frac{1}{nh^2}\mathbf{D}K\left(\frac{x-\xi}{h}\right) = \frac{1}{nh^2}\left[\int_{-\infty}^{+\infty}K^2\left(\frac{x-u}{h}\right)f_{\xi}(u)du - \right. \\ &- \left.\left(\int_{-\infty}^{+\infty}K\left(\frac{x-u}{h}\right)f_{\xi}(u)du\right)^2\right] = \frac{1}{nh}\left[\int_{-\infty}^{+\infty}K^2(v)f_{\xi}(x-hv)dv - \right. \\ &- \left.h\left(\int_{-1/2}^{+1/2}K(v)f_{\xi}(x-hv)dv\right)^2\right] = \frac{f_{\xi}(x)}{nh}\int_{-1/2}^{+1/2}K^2(v)dv+o(h). \end{aligned} \tag{11}$$

Используя соотношения (10), (11), получаем следующий одномерный аналог оценки (4):

$$\begin{aligned} \int_{-\infty}^{+\infty}\mathbf{E}\left[f_{\xi}(x)-\tilde{f}_{\xi}^{(x)}(n)\right]^2dx &= \int_{-\infty}^{+\infty}\mathbf{E}\left[\left(f_{\xi}(x)-\mathbf{E}\tilde{f}_{\xi}^{(x)}(n)\right)+ \right. \\ &+ \left.\left(\mathbf{E}\tilde{f}_{\xi}^{(x)}(n)-\tilde{f}_{\xi}^{(x)}(n)\right)\right]^2dx = \int_{-\infty}^{+\infty}\left(f_{\xi}(x)-\mathbf{E}\tilde{f}_{\xi}^{(x)}(n)\right)^2dx + \\ &+ \int_{-\infty}^{+\infty}\mathbf{D}\tilde{f}_{\xi}^{(x)}(n)dx \approx C_1(h)\left[\int_{-1/2}^{+1/2}v^2K(v)dv\right]^2 + C_2(n,h)\int_{-1/2}^{+1/2}K^2(v)dv, \end{aligned} \tag{12}$$

где  $C_1(h) = \frac{h^4 \int_{-\infty}^{+\infty} [f''_{\xi}(x)]^2 dx}{4}$  и  $C_2(n, h) = \frac{1}{nh}$ ; здесь учтены очевидные соотношения

$$2\mathbf{E}\left[\left(f_{\xi}(x)-\mathbf{E}\tilde{f}_{\xi}^{(x)}(n)\right)\times\left(\mathbf{E}\tilde{f}_{\xi}^{(x)}(n)-\tilde{f}_{\xi}^{(x)}(n)\right)\right]=0$$

и  $\int_{-\infty}^{+\infty}f_{\xi}(x)dx=1$ .

### § 5. Выбор показателя качества приближения

Отметим прежде всего работу [2], в которой автор минимизирует величину  $\int K^2(y)dy$  с неопределенными изначально пределами интегрирования в предположении, что  $\int y^2K(y)dy=1$ , и получает вариационную

задачу, решением которой является знаменитое квадратичное ядро («ядро Епанечникова»), положительное на интервале  $(a, b) = (-\sqrt{5}, +\sqrt{5})$ .

Для рассматриваемых в данной работе экономичных версий вычислительного функционального ядерного алгоритма из Раздела 1 следует использовать коэффициент размытости, равный  $h$  (для него  $(a, b) = (-1/2, +1/2)$ ), и, в силу соотношения (12), при выборе функции  $K(y)$  нужно одновременно учитывать и интеграл от квадрата  $E_1^{(K)} = \int_{-1/2}^{+1/2} K^2(y) dy$ , и «дисперсию»  $E_2^{(K)} = \int_{-1/2}^{+1/2} y^2 K(y) dy$ .

Следуя известной технологии из математической статистики, можно предложить приравнять (при фиксированных  $h$  и  $n$ ) слагаемые суммы в правой части неравенства (12), соответствующие компоненте смещения и стохастической компоненте; при этом возникает пропорция вида

$$\int_{-1/2}^{+1/2} y^2 K(y) dy = C_3 \times \sqrt{\int_{-1/2}^{+1/2} K^2(y) dy}; \quad \text{или} \quad E_2^{(K)} = C_3 \sqrt{E_1^{(K)}}; \quad (13)$$

здесь  $C_3 = \sqrt{\frac{C_2(n, h)}{C_1(h)}}$ . При этом нужно учитывать, что, с учетом выбранного нами интервала положительности  $(-\frac{1}{2}, \frac{1}{2})$  функции  $K(y)$ , должно быть выполнено неравенство Коши–Буняковского

$$\int_{-1/2}^{+1/2} y^2 K(y) dy \leq \sqrt{\frac{1}{80}} \times \sqrt{\int_{-1/2}^{+1/2} K^2(y) dy}; \quad \text{или} \quad E_2^{(K)} \leq \sqrt{\frac{1}{80}} \sqrt{E_1^{(K)}}. \quad (14)$$

Таким образом, при применении технологии приравнивания компонент погрешности (13) в ядерном алгоритме при выборе параметров  $n$  (это количество используемых выборочных значений  $\{\xi_1, \dots, \xi_n\}$ ) и  $h$  (это шаг сетки (1)) нужно позаботиться о том, чтобы было выполнено неравенство

$$\frac{C_2(n, h)}{C_1(h)} = \frac{4}{\int_{-\infty}^{+\infty} [f''_{\xi}(x)]^2 dx} \times \frac{1}{nh^5} \leq \frac{1}{80}.$$

В связи с соотношениями (12)–(14) неудачным видится предложение из статьи [11], где в качестве показателя малости погрешности приближения  $\tilde{f}_{\xi}^{(x)}(n)$  плотности  $f_{\xi}(x)$  предложено брать величину  $E_3^{(K)} = E_1^{(K)} \times \sqrt{E_2^{(K)}}$ .

Еще раз подчеркнем, что для исследуемого в данной статье ядерного алгоритма мы должны рассматривать ядра  $K(y)$ , имеющие отличные от нуля (положительные) значения на интервале  $(-\frac{1}{2}, \frac{1}{2})$ . Кроме того, с учетом соотношений (12)–(14) мы предлагаем использовать при выборе функции  $K(y)$  показатели малости погрешности приближения  $\tilde{f}_{\xi}^{(x)}(n)$  плотности  $f_{\xi}(x)$  вида  $E_4^{(K)} = \sqrt{E_1^{(K)}} \times E_2^{(K)}$ , а лучше даже  $E_5^{(K)} = \sqrt{E_1^{(K)} + E_2^{(K)}}$ .

$K^*(y)$	$E_1^{(K)}$	$E_2^{(K)}$	$E_3^{(K)}$	$E_4^{(K)}$	$E_5^{(K)}$
$K_1^*(y) \equiv 1$	1.000	0.083	0.289	0.083	1.083
$K_2^*(y) = 3/2 - 6y^2$	1.200	0.050	0.268	0.055	1.145
$K_3^*(y) = \frac{\pi}{2} \cos(\pi y)$	1.234	0.047	0.268	0.053	1.158
$K_4^*(y) = 2 - 4y$	1.333	0.042	0.273	0.048	1.197
$K_5^*(y) = \frac{140}{81} (1 - 8y^3)^3$	1.417	0.036	0.269	0.043	1.226
$K_6^*(y) = \frac{15}{8} (1 - 4y^2)^2$	1.429	0.036	0.270	0.043	1.231

**Таблица 1.** Сравнение показателей малости погрешности приближения  $\tilde{f}_\xi^{(x)}(n)$  плотности  $f_\xi(x)$  для разных ядер

В Таблице 1 показаны наиболее употребимые ядра из статьи [11], взятые на интервале  $(-\frac{1}{2}, \frac{1}{2})$  (см. соотношение (8)) – вместо интервала  $(-1, 1)$  из [11]. В первом столбце Таблицы 1 показаны различные типы функции  $K^*(y)$ , определенной на полуинтервале  $[0, 1/2)$  и такой, что  $K(y) = K^*(y)$  при  $y \geq 0$  и  $K(y) = K^*(-y)$  при  $y \leq 0$ . В остальных столбцах приведены показатели малости погрешности приближения  $\tilde{f}_\xi^{(x)}(n)$  плотности  $f_\xi(x)$ .

## § 6. О преимуществах использования многомерного аналога полигона частот

Из Таблицы 1 видно, что из-за относительной малости интервала  $(-\frac{1}{2}, \frac{1}{2})$  величина «дисперсии» (второго момента)  $E_2^{(K)} = \int_{-1/2}^{+1/2} y^2 K(y) dy$  ядерной функции  $K(y)$  более чем на порядок меньше интеграла от квадрата этой функции  $E_1^{(K)} = \int_{-1/2}^{+1/2} K^2(y) dy$  (что вполне соответствует неравенству Коши–Буняковского (14)), так что в данной ситуации по-своему оправдан подход из работы [2], в которой предлагалось игнорировать изменение значения  $E_2^{(K)}$ . В связи с последним замечанием более перспективным для использования в рассматриваемом в данной работе вычислительном ядерном алгоритме является предлагаемый нами новый показатель  $E_5^{(K)}$ . К слову, по этому показателю (как и по «более весомому» – по сравнению с  $E_2^{(K)}$  – показателю  $E_1^{(K)}$ ) кусочно-постоянное ядро  $K_1^*(y)$  имеет определенное преимущество. Ядерный алгоритм с такой функцией  $K^*(v)$  и с базисными функциями (6) для  $r = 2$  и с коэффициентами  $w^{(i)}(\mathbf{f}_\xi) = f_\xi(\mathbf{y}_i)$ ;  $i = 1, \dots, M$  назван нами в [1] *многомерным аналогом полигона частот*.

Особо отметим, что пока только для этого алгоритма удалось получить выражения для условно-оптимальных параметров: числа  $n$  используемых выборочных значений и числа  $M$  узлов сетки (1), которые позволяют минимизировать затраты алгоритма при заданном уровне среднеквадратичной погрешности  $\Delta$ .

Действительно, из рассуждений раздела 2 следует, что для многомерного аналога полигона частот можно построить следующую оценку сверху для компоненты аппроксимации из соотношения (4):

$$\delta_{appr}^{(\mathbb{L}_2)}(M) \leq \frac{H_{appr}^{(\mathbb{L}_2)}}{M^{2/d}}, \quad \text{где } H_{appr}^{(\mathbb{L}_2)} = \frac{[H^{(h \rightarrow M)}]^2 \sqrt{\text{mes } X}}{8} \times \sum_{s=1}^d \max_{\mathbf{x} \in X} \left| \frac{\partial^2}{\partial (x^{(s)})^2} f_{\boldsymbol{\xi}}(\mathbf{x}) \right| \quad (15)$$

и  $\text{mes } X = (b^{(1)} - a^{(1)}) \times \dots \times (b^{(d)} - a^{(d)})$ , а константа  $H^{(h \rightarrow M)}$ , такова, что  $h = H^{(h \rightarrow M)} M^{-1/d}$ . Несложно также получить оценки сверху для компоненты смещения и для стохастической компоненты (подробности см. в работе [12])

$$\delta_{bias}^{(\mathbb{L}_2)}(M) \leq \frac{H_{bias}^{(\mathbb{L}_2)}}{M^{2/d}}, \quad \text{где } H_{bias}^{(\mathbb{L}_2)} = \frac{[H^{(h \rightarrow M)}]^2 \sqrt{\text{mes } X}}{24} \times \sum_{s=1}^d \max_{\mathbf{x} \in X} \left| \frac{\partial^2}{\partial (x^{(s)})^2} f_{\boldsymbol{\xi}}(\mathbf{x}) \right|; \quad (16)$$

$$\delta_{stoch}^{(\mathbb{L}_2)}(M, n) \leq \frac{H_{stoch}^{(\mathbb{L}_2)} M}{n}, \quad \text{где } H_{stoch}^{(\mathbb{L}_2)} = \frac{f_{max} \times \text{mes } X}{[H^{(h \rightarrow M)}]^d} \text{ и } f_{max} = \max_{\mathbf{x} \in X} f_{\boldsymbol{\xi}}(\mathbf{x}). \quad (17)$$

Заметим, что получение соотношений вида (16) и (17) для ядер, отличных от  $K_1^*(v) \equiv 1$  (например, для ядер  $K_i^*(v)$ ;  $i = 2 - 6$  из таблицы 1), затруднено.

Далее в теории условной оптимизации (см., например, [1, 12]) нужно решать такую оптимизационную задачу: найти значения  $M_{opt}^{(\mathbb{L}_2)}(\Delta)$  и  $n_{opt}^{(\mathbb{L}_2)}(\Delta)$ , для которых достигается минимум

$$\min_{M, n} S(M, n) = \min_{M, n} [t \times n(M, \Delta) + S_0] \quad (18)$$

(здесь  $t$  – как и выше – время определения того, в какой из кубов  $D^{(y_i)}$  попадает очередное выборочное значение  $\boldsymbol{\xi}_j$ , а  $S_0$  – время вычисления  $2^d$  ненулевых слагаемых  $\tilde{f}_{\boldsymbol{\xi}}^{(y_i)}(n) \chi^{(i)}(\mathbf{x})$  в сумме (2) для заданной точки  $\mathbf{x} \in X$ ) при условии

$$\frac{A_1^{(\mathbb{L}_2)}}{M^{4/d}} + \frac{A_2^{(\mathbb{L}_2)} M}{n} = \Delta^2, \quad (19)$$

где

$$A_1^{(\mathbb{L}_2)} = 2 (H_{appr}^{(\mathbb{L}_2)})^2 + 2 (H_{bias}^{(\mathbb{L}_2)})^2 = \frac{[H^{(h \rightarrow M)}]^4 \times \text{mes } X}{18} \times \left( \sum_{s=1}^d \max_{\mathbf{x} \in X} \left| \frac{\partial^2}{\partial (x^{(s)})^2} f_{\boldsymbol{\xi}}(\mathbf{x}) \right| \right)^2$$

и

$$A_2^{(\mathbb{L}_2)} = H_{stoch}^{(\mathbb{L}_2)} = \frac{f_{max} \times \text{mes } X}{[H^{(h \rightarrow M)}]^d}$$

– см. соотношения (15)–(17).

Из равенств (18) и (19) имеем

$$n = \frac{A_2^{(\mathbb{L}_2)} M}{\Delta^2 - A_1^{(\mathbb{L}_2)} / M^{4/d}} \text{ и } \tilde{S}^{(\mathbb{L}_2, \Delta)}(M) = \frac{t A_2^{(\mathbb{L}_2)} M}{\Delta^2 - A_1^{(\mathbb{L}_2)} / M^{4/d}} + S_0.$$

Найдем теперь минимум функции  $\tilde{S}^{(\mathbb{L}_2, \Delta)}(M)$  по  $M$ . Продифференцируем эту функцию

$$\frac{\partial \tilde{S}^{(\mathbb{L}_2, \Delta)}(M)}{\partial M} = t A_2^{(\mathbb{L}_2)} \left[ \Delta^2 - \frac{A_1^{(\mathbb{L}_2)}}{M^{4/d}} - \frac{4 A_1^{(\mathbb{L}_2)}}{d \times M^{4/d}} \right] / \left[ \Delta^2 - \frac{A_1^{(\mathbb{L}_2)}}{M^{4/d}} \right]^2$$

и найдем критическое значение (точку минимума), а затем и *условно-оптимальные параметры для  $\mathbb{L}_2$ -подхода*:

$$M_{opt}^{(\mathbb{L}_2)}(\Delta) = \left[ A_1^{(\mathbb{L}_2)} \right]^{d/4} \left( \frac{d+4}{d} \right)^{d/4} \Delta^{-d/2}$$

и

$$n_{opt}^{(\mathbb{L}_2)}(\Delta) = \frac{A_2^{(\mathbb{L}_2)} \left[ A_1^{(\mathbb{L}_2)} \right]^{d/4} (d+4)^{d/4+1}}{4d^{d/4}} \Delta^{-2-d/2}. \quad (20)$$

В следующем разделе на тестовом примере показано, что многомерный аналог полигона частот дает как малую погрешность, так и наименьшее время вычислений по сравнению с другими вариантами реализации рассматриваемого в данной работе ядерного алгоритма.

## § 7. Тестовый пример: усеченное экспоненциальное распределение

Покажем преимущество выбора кусочно-постоянного ядра  $K_1^*(v)$  на примере одномерной выборки, которая моделировалась численно по формулам

$$\xi_j = -\frac{\ln [1 - \alpha_j (1 - e^{-\lambda L})]}{\lambda}; \quad j = 1, \dots, n; \quad L > 0; \lambda > 0; \quad (21)$$

здесь  $\alpha_j \in U(0, 1)$  – стандартные случайные числа. Ситуация здесь является тестовой по той причине, что соотношения (21) представляют собой формулы метода обратной функции распределения для численного

моделирования одномерной случайной величины  $\xi$ , имеющей плотность усеченного экспоненциального распределения

$$f_{\xi}(x) = \frac{\lambda e^{-\lambda x}}{1 - e^{-\lambda L}}; \quad 0 < x < L; \quad (22)$$

(см. пример 11.7 из раздела 11.6 учебника [13]).

Определим соответствующие плотности (22) оптимальные параметры  $M_{opt}^{(\mathbb{L}_2)}$  и  $n_{opt}^{(\mathbb{L}_2)}$  для соответствующего полигона частот. Используем выражения (20) для размерности  $d = 1$ , области распределения  $X = (0, L)$  (при этом  $\text{mes } X = L$ ), константы  $H^{(h \rightarrow M)} = L$  и получаем

$$f_{max} = f_{\xi}(0) = \frac{\lambda}{1 - e^{-\lambda L}}, \quad \max_{x \in X} \left| \frac{\partial^2}{\partial x^2} f_{\xi}(x) \right| = \left| \frac{\partial^2}{\partial x^2} f_{\xi}(0) \right| = \frac{\lambda^3}{1 - e^{-\lambda L}}$$

и константы

$$A_1^{(\mathbb{L}_2)} = 2 \times \frac{10}{9} \times \left[ \frac{L^2}{8} \times \frac{\lambda^3}{1 - e^{-\lambda L}} \times \sqrt{L} \right]^2 = \frac{5L^5 \lambda^6}{144(1 - e^{-\lambda L})^2} \quad \text{и} \quad A_2^{(\mathbb{L}_2)} = \frac{\lambda}{1 - e^{-\lambda L}}.$$

Отсюда получаем условно-оптимальные параметры (20) для рассматриваемого тестового примера:

$$M_{opt}^{(\mathbb{L}_2)}(\Delta) = \frac{\sqrt{5}}{2\sqrt{3}} \times \frac{\sqrt[4]{L^5} \sqrt{\lambda^3}}{\sqrt{\Delta(1 - e^{-\lambda L})}} \quad \text{и} \quad n_{opt}^{(\mathbb{L}_2)}(\Delta) = \frac{\sqrt{125}}{8\sqrt{3}} \times \frac{\sqrt[4]{L^5} \sqrt{\lambda^5}}{\sqrt{\Delta^5(1 - e^{-\lambda L})^3}}.$$

Опишем результаты тестовых экспериментов. В расчетах использовались следующие параметры:  $L = 1$ ,  $\lambda = 1$ ,  $\Delta = 0.004$ . С учетом того, что выбираемые параметры  $M$  и  $n$  должны быть целыми, получаем значения  $M_{opt}^{(\mathbb{L}_2)}(\Delta) \approx 12$  и  $n_{opt}^{(\mathbb{L}_2)}(\Delta) \approx 1\,594\,025$ .

$n$	Максимальная узловaя ошибка				Время вычисления коэффициентов, с			
	$K_1^*(y)$	$K_2^*(y)$	$K_3^*(y)$	$K_4^*(y)$	$K_1^*(y)$	$K_2^*(y)$	$K_3^*(y)$	$K_4^*(y)$
797012	0.0076791	0.0091869	0.0093371	0.0099639	0.070	0.097	0.368	0.098
1195519	0.0056726	0.0078557	0.0080067	0.0085447	0.102	0.159	0.557	0.162
1434622	0.0046682	0.0064523	0.0065404	0.0070954	0.131	0.201	0.691	0.200
1594025	0.00363541	0.0052640	0.0053694	0.0058959	0.155	0.221	0.742	0.226
1753428	0.0043278	0.0049060	0.0050128	0.0055214	0.177	0.260	0.844	0.251
1992531	0.0041707	0.0061136	0.0063294	0.0067802	0.229	0.325	1.237	0.280
2391038	0.0050853	0.0057335	0.0058441	0.0065181	0.292	0.347	1.120	0.328
3188050	0.0039062	0.0037557	0.0038880	0.0043688	0.298	0.486	1.651	0.760
4782075	0.0033518	0.0036725	0.0037671	0.0040358	1.560	1.832	2.883	1.110
7970125	0.0025030	0.0024311	0.0024971	0.0026357	1.173	1.849	5.143	1.761
11955188	0.0016869	0.0019120	0.0019397	0.0023534	1.867	3.031	6.584	2.738
15940250	0.0020310	0.0018295	0.0018569	0.0021032	2.555	4.083	8.867	3.847

**Таблица 2.** Максимальная узловaя ошибка и время работы ядерного алгоритма для  $M = 12$  и четырех ядер для усеченного экспоненциального распределения (21).

В таблице 2 приведены полученные значения максимальной узловой ошибки и времени вычисления для фиксированного  $M = 12$  при различных размерах выборки  $n$  (они выбирались как доли от найденного оптимального значения  $n_{opt}^{(L_2)}(\Delta)$ ) для различных ядер из таблицы 1 (конкретнее, для постоянного ядра  $K_1^*(y)$ , квадратичного ядра  $K_2^*(y)$ , тригонометрического ядра  $K_3^*(y)$ , и линейного ядра  $K_4^*(y)$ ).

Из таблицы 2 видно, что полигон частот (т.е. ядерный алгоритм с кусочно-постоянным ядром  $K_1^*(y)$ ) дает практически во всех случаях и минимальную погрешность, и наименьшее время реализации алгоритма.

Аналогичные результаты получились и при меняющемся параметре  $M$ ; к слову, здесь были подтверждены соображения из раздела 3 данной работы о том, что затраты ядерного алгоритма для случая, когда финитная ядерная функция  $K(v)$  имеет в качестве носителя интервал  $(-\frac{1}{2}, \frac{1}{2})$ , не зависят от числа  $M$  узлов сетки (1).

## § 8. Заключение

В данной работе сформулированы два важных принципа – мультипликативность вида (в многомерном случае) и финитность (сосредоточенность носителей около узлов аппроксимационной сетки) – при выборе аппроксимационного базиса и ядерной функции для конструирования экономичного вычислительного (компьютерного) алгоритма приближения вероятностной плотности  $f_{\xi}(\mathbf{x})$  по заданной выборке  $\{\xi_1, \dots, \xi_n\}$ .

Кроме того, в работе проведен анализ различных (в том числе, предложенных авторами статьи) показателей качества рассматриваемого приближения при выборе формы финитной ядерной функции  $K(y)$ , имеющей носитель на интервале  $(-\frac{1}{2}, \frac{1}{2})$ , для использования ее в вычислительном (компьютерном) функциональном ядерном алгоритме. Эти показатели выглядят как различные комбинации квадрата  $E_1^{(K)} = \int_{-1/2}^{+1/2} K^2(y) dy$  и «дисперсии»  $E_2^{(K)} = \int_{-1/2}^{+1/2} y^2 K(y) dy$  ядерной функции, наилучшим из которых предложено считать показатель  $E_5^{(K)} = \sqrt{E_1^{(K)} + E_2^{(K)}}$ .

Тестовые расчеты показали, что, несмотря на «средние» показатели качества приближения, кусочно-постоянная ядерная функция  $K_1^*(y)$  (соответствующая многомерному аналогу полигона частот [1]) дает приемлемую точность приближения плотности за гораздо меньшее время, чем другие виды ядерной функции  $K(y)$ .

## Список литературы

1. Войтишек А. В., Шлымбетов Н. Х. Выбор аппроксимационных базисов, используемых в компьютерных функциональных алгоритмах

- приближения вероятностных плотностей по заданной выборке // *Сибирский журнал вычислительной математики*. 2024. Т. 27, № 2. С. 147–164.
2. Епанечников В. А. Непараметрическая оценка многомерной плотности вероятности // *Теория вероятностей и ее применения*. 1969. Т. 14, № 1. С. 156–161.
  3. Devroye L., Lugosi G. *Combinatorial Methods in Density Estimation*. New York: Springer, 2001.
  4. Gramacki A. *Nonparametric Kernel Density Estimation and Its Computational Aspects*. Cham: Springer, 2018.
  5. Бахвалов Н. С. *Численные методы*. М.: Наука, 1975.
  6. Войтишек А. В., Шлымбетов Н. Х. Компоненты погрешности функциональных вычислительных алгоритмов приближения вероятностной плотности по заданной выборке:  $L_2$ -подход // *Информационные технологии и математическое моделирование (ИТММ-2024): Материалы XXIII Международной конференции имени А. Ф. Терпугова (20–26 октября 2024 года)*. Томск: Издательство Томского государственного университета, 2024. С. 475–480.
  7. Бахвалов Н. С., Корнеев А. А., Чижонков Е. В. *Численные методы. Решение задач и упражнения*. М.: Лаборатория знаний, 2016.
  8. Марчук Г. И., Агошков В. И. *Введение в проекционно-сеточные методы*. М.: Наука, 1981.
  9. Завьялов Ю. С., Квасов Б. И., Мирошниченко В. Л. *Методы сплайн-функций*. М.: Наука, 1980.
  10. Милосердов В. В. *Дискретно-стохастические численные алгоритмы со сплайн-восполнениями: дис. ... канд. физ.-мат. наук: 01.01.07*. Новосибирск: ИВМиМГ СО РАН, 2006.
  11. Ядро (статистика). Ядро (статистика) // *Википедия*. 2025. URL: <https://ru.wikipedia.org/?curid=7083878&oldid=146697015> (дата обращения: 14.10.2025).
  12. Булгакова Т. Е., Войтишек А. В. Условная оптимизация функционального вычислительного ядерного алгоритма приближения вероятностной плотности по заданной выборке // *Журнал вычислительной математики и математической физики*. 2021. Т. 61, № 9. С. 1431–1446.

13. *Войтишек А. В. Лекции по численным методам Монте-Карло.* Новосибирск: ИПЦ НГУ, 2018.

## References

1. Voytishchek A. V., Shlimbetov N. Kh. Choice of approximation bases used in computational functional algorithms for approximating probability densities on the basis of given sample // *Numerical Analysis and Applications*. 2024. V. 17, N 2. P. 116–131. DOI: 10.1134/S1995423924020022.
2. Epanechnikov V. A. Non-parametric estimation of a multivariate probability density // *Theory of Probability and its Applications*. 1969. V. 14, N 1. P. 153–158.
3. Devroye L., Lugosi G. *Combinatorial Methods in Density Estimation*. New York: Springer, 2001.
4. Gramacki A. *Nonparametric Kernel Density Estimation and Its Computational Aspects*. Cham: Springer, 2018.
5. Bakhvalov N. S. *Numerical Methods*. Moscow: Nauka, 1975. [in Russian].
6. Voytishchek A. V., Shlimbetov N. Kh. Components of Error in Functional Computational Algorithms for Approximating Probability Density from a Given Sample: An  $L_2$ -Approach // *Information Technologies and Mathematical Modeling (ITMM-2024): Proceedings of the XXIII International Conference named after A. F. Terpugov (October 20–26, 2024)*. Tomsk: Tomsk State University Publishing House, 2024. P. 475–480. [in Russian].
7. Bakhvalov N. S., Korneev A. A., Chizhonkov E. V. *Numerical Methods. Problems Solving and Exercises*. Moscow: Knowledge Laboratory, 2016. [in Russian].
8. Marchuk G. I., Agoshkov V. I. *Introduction to Projection–mesh Methods*. Moscow: Nauka, 1981. [in Russian].
9. Zavyalov Yu. S., Kvasov B. I., Miroshnichenko V. L. *Methods of Spline Functions*. Moscow: Nauka, 1980. [in Russian].
10. Miloserdov V. V. *Discrete-Stochastic Numerical Algorithms with Spline Reconstructions: Cand. Phys.-Math. Sci. Diss.: 01.01.07*. Novosibirsk: Institute of Computational Mathematics and Mathematical Geophysics SB RAS, 2006. [in Russian].
11. *Kernel (statistics)*. Kernel (statistics) // *Wikipedia*. 2025. URL: <https://en.wikipedia.org/?curid=7083878&oldid=146697015> (accessed: October 14, 2025). [in Russian].

12. Bulgakova T.E., Voytishek A.V. Conditional optimization of the computational kernel algorithm for approximating the probability density on the basis of a given sample // *Computational Mathematics and Mathematical Physics*. 2021. V. 61, N 9. P. 1401–1415.
13. Voytishek A.V. *Lectures on Numerical Monte Carlo Methods*. Novosibirsk: NSU Publishing and Printing Center, 2018. [in Russian].

### Информация об авторах

**Нурлыбай Хамдуллаевич Шлымбетов**, аспирант

Scopus Author ID 59003466000

**Убайдулла Пахтамурат улы Сейтмуратов**, магистрант

**Антон Вацлавович Войтишек**, доктор физико-математических наук, профессор, главный научный сотрудник

SPIN 7494-4885AuthorID: 3523

Scopus Author ID 6506412968      ResearcherID Web of Science: S-1911-2017

### Author Information

**Nurlibay Kh. Shlimbetov**, PhD Student

Scopus Author ID 59003466000

**Ubaydulla P. Seytmuratov**, Master's Student

**Anton V. Voytishek**, Doctor of Mathematics, Professor, Chief Researcher

SPIN 7494-4885AuthorID: 3523

Scopus Author ID 6506412968      ResearcherID Web of Science: S-1911-2017

*Статья поступила в редакцию 02.12.2025;  
одобрена после рецензирования 02.03.2026; принята к публикации  
01.04.2026*

*The article was submitted 02.12.2025;  
approved after reviewing 02.03.2026; accepted for publication 01.04.2026*