

УДК 519.7

ОЦЕНИВАНИЕ НЕЛИНЕЙНОГО ФУНКЦИОНАЛА ОТ ПЛОТНОСТИ ВЕРОЯТНОСТИ ТРЁХМЕРНОЙ СЛУЧАЙНОЙ ВЕЛИЧИНЫ ДЛЯ ПОВЫШЕНИЯ ВЫЧИСЛИТЕЛЬНОЙ ЭФФЕКТИВНОСТИ НЕПАРАМЕТРИЧЕСКИХ РЕШАЮЩИХ ПРАВИЛ

© А. В. Лапко^{1,2}, В. А. Лапко^{1,2}

¹Институт вычислительного моделирования СО РАН,
660036, г. Красноярск, Академгородок, 50, стр. 44

²Сибирский государственный университет науки и технологий
им. академика М. Ф. Решетнева,
660037, г. Красноярск, просп. им. газеты «Красноярский рабочий», 31
E-mail: lapko@icm.krasn.ru

Предлагается методика оценивания нелинейного функционала от плотности вероятности трёхмерной случайной величины, которая является актуальной при реализации процедур быстрого выбора коэффициентов размытости в задаче оптимизации ядерных оценок плотности вероятности. Решение этой проблемы позволяет значительно повысить вычислительную эффективность непараметрических решающих правил. Основу предлагаемого подхода составляет анализ формулы оптимальных коэффициентов размытости ядерной оценки плотности вероятности. Коэффициенты размытости ядерных функций представляются в виде произведения неопределённого параметра и средних квадратических отклонений анализируемых случайных величин. Основная составляющая неопределённого параметра является нелинейным функционалом от плотности вероятности. Рассматриваемый функционал для семейства одномодальных законов распределения задаётся видом плотности вероятности и не зависит от параметров плотности. Он определяется аппроксимацией функциональной зависимости от коэффициентов контрэкссесса и асимметрии, которые оцениваются по исходным статистическим данным. Для упрощения задачи восстановления искомой зависимости коэффициенты контрэкссесса и асимметрии преобразуются в обобщённый параметр. Исходную информацию составляет семейство логнормальных законов распределения. Оцениваются ошибки аппроксимации рассматриваемого нелинейного функционала от плотности вероятности по значениям введённого обобщённого параметра для семейства трёхмерных логнормальных законов распределения независимых случайных величин. Исследуется возможность применения предлагаемой методики при оценивании нелинейных функционалов от плотностей вероятностей, отличающихся от логнормальных законов распределения. Проводится анализ влияния возникающих ошибок аппроксимации на средние квадратические критерии восстановления непараметрической оценки плотности вероятности трёхмерной случайной величины.

Ключевые слова: ядерная оценка плотности вероятности, нелинейный функционал от плотности вероятности, трёхмерная случайная величина, быстрый выбор коэффициентов размытости, коэффициент контрэкссесса, коэффициент асимметрии, выборки большого объёма, логнормальный закон распределения.

DOI: 10.15372/AUT20220211

Введение. При синтезе алгоритмов обработки информации в условиях априорной неопределённости о закономерностях изучаемых объектов широко используются непараметрические оценки плотности вероятности $\bar{p}(x)$ типа Розенблатта — Парзена [1–4]. Аппроксимационные свойства этих оценок плотностей вероятностей зависят от эффективного выбора коэффициента размытости ядерных функций. Традиционные методы оптимизации непараметрических оценок плотности вероятности основываются на выборе коэффициентов размытости из условия максимума функции правдоподобия либо минимума оценки

среднего квадратического отклонения $\bar{p}(x)$ от восстанавливаемой плотности вероятности $p(x)$ [5–8]. При увеличении объёма исходных данных наблюдений случайных величин реализация этих методов требует значительных временных затрат, что характерно, например, при обработке данных дистанционного зондирования. Поэтому возникает необходимость разработки новых методов, обеспечивающих значительное повышение вычислительной эффективности процедур выбора коэффициентов размытости ядерных функций в непараметрической оценке плотности вероятности.

Впервые решение данной проблемы было предложено в работе [9], которое основано на результатах анализа асимптотических свойств непараметрической оценки плотности вероятности типа Розенблатта — Парзена. Полученный алгоритм быстрого выбора коэффициентов размытости был развит в [10–17].

Цель данной работы — предложить эффективную методику оценивания нелинейного функционала от плотности вероятности в формуле оптимального коэффициента размытости ядерных функций по значениям коэффициентов контрэксцесса и асимметрии компонент трёхмерной случайной величины.

Формула выбора оптимального коэффициента размытости. Рассмотрим непараметрическую оценку плотности вероятности $p(x_1, x_2, x_3)$ трёхмерной случайной величины

$$\bar{p}(x_1, x_2, x_3) = \left(n \prod_{v=1}^3 c\sigma_v \right)^{-1} \sum_{i=1}^n \prod_{v=1}^3 \Phi \left(\frac{x_v - x_v^i}{c\sigma_v} \right), \quad (1)$$

которая восстанавливается по выборке $V = (x_1^i, x_2^i, x_3^i, i = \overline{1, n})$ статистически независимых наблюдений $x = (x_1, x_2, x_3)$ объёма n . Предполагается, что условия измерений случайных величин не изменяются. В статистике (1) символом σ_v обозначено среднее квадратическое отклонение случайной величины x_v .

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

$$0 \leq \Phi(u_v) < \infty, \quad \Phi(u_v) = \Phi(-u_v), \quad \int \Phi(u_v) du_v = 1, \quad \int u_v^2 \Phi(u_v) du_v = 1,$$

$$\int u_v^m \Phi(u_v) du_v < \infty, \quad 0 \leq m < \infty, \quad v = \overline{1, 3}.$$

Здесь и далее бесконечные пределы интегрирования опускаются. Последовательно коэффициенты размытости ядерных функций $c_v(n) = c(n)\sigma_v \rightarrow 0$, а значения

$$n \prod_{v=1}^3 c_v(n) \rightarrow \infty \text{ при } n \rightarrow \infty.$$

Асимптотическое выражение среднего квадратического отклонения $\bar{p}(x_1, x_2, x_3)$ от восстанавливаемой плотности вероятности $p(x_1, x_2, x_3)$ представляется в виде [18]

$$\begin{aligned} W(c) &= M \iiint (\bar{p}(x_1, x_2, x_3) - p(x_1, x_2, x_3))^2 dx_1 dx_2 dx_3 \sim \\ &\sim \frac{1}{nc^3} \prod_{v=1}^3 (\sigma_v^{-1} \|\Phi(u_v)\|^2) + \frac{Bc^4}{4}, \end{aligned} \quad (2)$$

где

$$B = \iiint \left(\sum_{v=1}^3 \sigma_v^2 p_v^{(2)}(x_1, x_2, x_3) \right)^2 dx_1 dx_2 dx_3; \quad \|\Phi(u_v)\|^2 = \int \Phi^2(u_v) du_v;$$

M — знак математического ожидания, а $p_v^{(2)}(x_1, x_2, x_3)$ — вторая производная функции $p(x_1, x_2, x_3)$ по x_v .

Оптимальный параметр c коэффициентов размытости $c_v = c\sigma_v$, $v = \overline{1, 3}$, ядерных функций определяется из минимума критерия (2) и соответствует значению

$$c^* = \left(3(\|\Phi(u)\|^2)^3 / \left(nB \prod_{v=1}^3 \sigma_v \right) \right)^{1/7} \quad (3)$$

при условии, что ядерные функции имеют вид $\Phi(u_v) = \Phi(u)$, $v = \overline{1, 3}$.

Тогда с учётом принятых предположений оптимальные коэффициенты размытости ядерных функций запишем в виде

$$c_v^* = c^* \sigma_v = \beta \sigma_v n^{-1/7}, \quad v = \overline{1, 3},$$

где

$$\beta = \left(3(\|\Phi(u)\|^2)^3 / \left(B \prod_{v=1}^3 \sigma_v \right) \right)^{1/7}. \quad (4)$$

Значение функционала β от $p(x_1, x_2, x_3)$ для семейства одномодальных плотностей вероятностей независимых случайных величин является константой, которая не зависит от параметров $p(x_1, x_2, x_3)$. Этому семейству соответствуют, например, гауссовский, Лапласа, логистический и экспоненциальный законы распределения. Так, для нормальной плотности вероятности при $k = 3$ и ядерной функции Епанечникова [18] значение β равно 0,945 и не зависит от математических ожиданий и средних квадратических отклонений случайных величин.

Оценивание значений нелинейного функционала β от плотности вероятности трёхмерной случайной величины. Обозначим через δ_v, γ_v , $v = \overline{1, 3}$, коэффициенты контрэкссеса и асимметрии независимых случайных величин x_v , $v = \overline{1, 3}$, соответственно. Имеются данные $V_1 = (\delta_v^i, \gamma_v^i, v = \overline{1, 3}, \beta^i, i = \overline{1, N})$ для восстановления зависимости $\beta = \varphi(\delta_v, \gamma_v, v = \overline{1, 3})$, которые формируются на основе семейства плотностей вероятности

$$p(x_1, x_2, x_3) = \prod_{v=1}^3 p(x_v).$$

Для упрощения поставленной задачи будем восстанавливать зависимость $\beta = \varphi'(\alpha)$, где α определим выражением:

$$\alpha = (\delta_1^2 + \delta_2^2 + \delta_3^2 + \gamma_1^2 + \gamma_2^2 + \gamma_3^2)^{1/2}.$$

При оценивании искомой зависимости будем использовать семейство логнормальных законов распределения

$$p(x_v) = \frac{1}{s_v x_v \sqrt{2\pi}} \exp \left(- \frac{(\ln x_v - a_v)^2}{2s_v^2} \right), \quad v = \overline{1, 3}, \quad (5)$$

Таблица 1

Значения коэффициентов контрэксцесса δ , асимметрии γ и их нелинейного преобразования для семейства логнормальных законов распределения

$LN(a; s_v)$	δ_v	γ_v	$\delta_v^2 + \gamma_v^2$
$LN(1; 0,05)$	0,574	0,15	0,3520
$LN(1; 0,1)$	0,562	0,302	0,4070
$LN(1; 0,2)$	0,521	0,614	0,6484
$LN(1; 0,3)$	0,464	0,95	1,1178
$LN(1; 0,4)$	0,4	1,322	1,9077
$LN(1; 0,5)$	0,335	1,75	3,1747
$LN(1; 0,6)$	0,274	2,26	5,1827
$LN(1; 0,7)$	0,219	2,888	8,3885
$LN(1; 0,8)$	0,171	3,689	13,6380
$LN(1; 0,9)$	0,129	4,745	22,5317
$LN(1; 1,0)$	0,094	6,185	38,2631
$LN(1; 1,1)$	0,066	8,213	67,4577
$LN(1; 1,2)$	0,044	11,164	124,6368
$LN(1; 1,3)$	0,028	15,598	243,2984

которыми определяются, например, спектральные признаки данных дистанционного зондирования. Параметры a_v и s_v меняют вид рассматриваемых плотностей вероятности $p(x_v)$, $v = \overline{1, 3}$. Значения β задаются параметром s_v и не зависят от величины a_v . Поэтому для трёхмерной случайной величины будем считать $a = a_v = 1$, $v = \overline{1, 3}$. Параметры логнормального закона распределения для одномерной случайной величины приведены в табл. 1.

В табл. 1 символами $LN(a; s_v)$ обозначены логнормальные законы распределения вида (5) с параметрами $a = 1$, s_v , $v = \overline{1, 14}$. По данным таблицы сформированы количественные характеристики логнормальных законов распределения независимых компонент трёхмерной случайной величины, которые представлены в табл. 2.

При вычислении элементов табл. 2 для $a = 1$ учитывается принятое условие независимости случайных величин x_v , $v = \overline{1, 3}$:

$$LN(s_1, s_2, s_3) = LN(a, s_1) LN(a, s_2) LN(a, s_3).$$

В табл. 2 приведены частичные данные вычислительных экспериментов. Полные сведения составляют результаты вычислительных экспериментов в количестве $N = 80$. Из табл. 2 определена аппроксимация функциональной зависимости β (4) в виде

$$\bar{\beta} = 1,406 \exp(-0,426\alpha^{0,81}) + 0,0475. \quad (6)$$

Средняя относительная ошибка аппроксимации искомой зависимости (6)

$$\bar{\rho} = \frac{1}{N} \sum_{j=1}^N \frac{|\beta^j - \bar{\beta}^j|}{\beta^j}$$

функционала β соответствует значению 0,0563 при использовании параболической ядерной функции Епанечникова $\Phi_2(u)$ [18]. Здесь β^j — значения β , полученные по результатам вычислительных экспериментов, а $\bar{\beta}^j$ — оценка β , вычисленная по модели (6). Принятый вид обобщённого параметра α обеспечивает минимальное значение $\bar{\rho}$ по сравнению с другими его вариантами.

Таблица 2

Значения нелинейного функционала β (4) и обобщённого параметра α при использовании ступенчатой $\Phi_1(u)$, параболической $\Phi_2(u)$ и гауссовской $\Phi_3(u)$ ядерных функций

Вид закона распределения $LN(s_1, s_2, s_3)$			Значение β для ядерных функций			α
s_1	s_2	s_3	$\Phi_1(u)$	$\Phi_2(u)$	$\Phi_3(u)$	
0,05	0,05	0,05	0,974	0,944	0,964	0,711
0,05	0,05	0,4	0,857	0,83	0,848	1,475
0,05	0,05	0,8	0,432	0,418	0,427	3,741
0,05	0,05	1,3	0,082	0,079	0,081	15,610
0,05	0,2	0,2	0,929	0,901	0,92	1,086
0,05	0,2	0,6	0,658	0,638	0,652	2,420
0,05	0,2	0,8	0,429	0,416	0,425	3,788
0,05	0,2	1,3	0,081	0,079	0,081	15,621
0,05	0,4	0,4	0,785	0,761	0,777	1,985
0,05	0,8	0,4	0,421	0,408	0,417	3,964
0,05	0,4	1,3	0,08	0,077	0,079	15,665
0,05	0,6	0,6	0,57	0,552	0,564	3,252
0,05	0,6	1,0	0,231	0,224	0,229	6,612
0,05	0,8	0,8	0,354	0,343	0,35	5,251
0,05	0,8	1,3	0,075	0,072	0,074	16,039
0,05	1,0	1,3	0,071	0,069	0,07	16,790
0,05	1,3	1,3	0,058	0,056	0,057	22,067
0,2	0,2	0,2	0,91	0,882	0,901	1,215
0,2	0,2	0,6	0,653	0,633	0,647	2,481
0,2	0,2	0,8	0,427	0,414	0,422	3,826
0,2	0,2	1,3	0,081	0,078	0,08	15,631
0,2	0,4	0,4	0,775	0,751	0,767	2,042
0,2	0,4	0,8	0,419	0,406	0,414	4,001
0,2	0,4	1,3	0,08	0,077	0,079	15,675
0,2	0,6	0,6	0,566	0,548	0,56	3,297
0,2	0,6	1,0	0,23	0,223	0,228	6,634
0,2	0,8	0,8	0,352	0,341	0,348	5,279
0,2	0,8	1,3	0,074	0,072	0,073	16,049
0,2	1,0	1,0	0,189	0,183	0,187	8,784
0,2	1,0	1,3	0,07	0,068	0,07	16,799
0,2	1,3	1,3	0,057	0,056	0,057	22,074
0,4	0,4	0,4	0,732	0,71	0,725	2,330
0,4	0,4	0,8	0,411	0,398	0,407	4,156
0,4	0,4	1,3	0,078	0,076	0,077	15,715
0,4	0,6	0,6	0,551	0,534	0,545	3,483
0,4	0,6	1,0	0,226	0,219	0,224	6,728
0,4	0,8	0,8	0,345	0,335	0,342	5,397
0,4	0,8	1,3	0,073	0,071	0,072	16,088
0,4	1,0	1,0	0,186	0,18	0,184	8,855
0,4	1,0	1,3	0,069	0,067	0,068	16,836
0,4	1,3	1,3	0,056	0,055	0,056	22,102
0,6	0,6	0,6	0,508	0,492	0,502	3,925
0,6	0,6	1,0	0,22	0,213	0,217	6,967
0,6	0,8	0,8	0,333	0,323	0,33	5,692
0,4	0,8	1,3	0,071	0,069	0,07	16,189
0,6	1,0	1,0	0,18	0,175	0,179	9,038
0,6	1,3	1,3	0,055	0,053	0,054	22,176
0,8	0,8	0,8	0,302	0,293	0,299	6,392
0,8	0,8	1,3	0,068	0,066	0,067	16,448
1,0	1,0	1,0	0,154	0,15	0,153	10,713
1,3	1,3	1,3	0,042	0,041	0,042	27,017

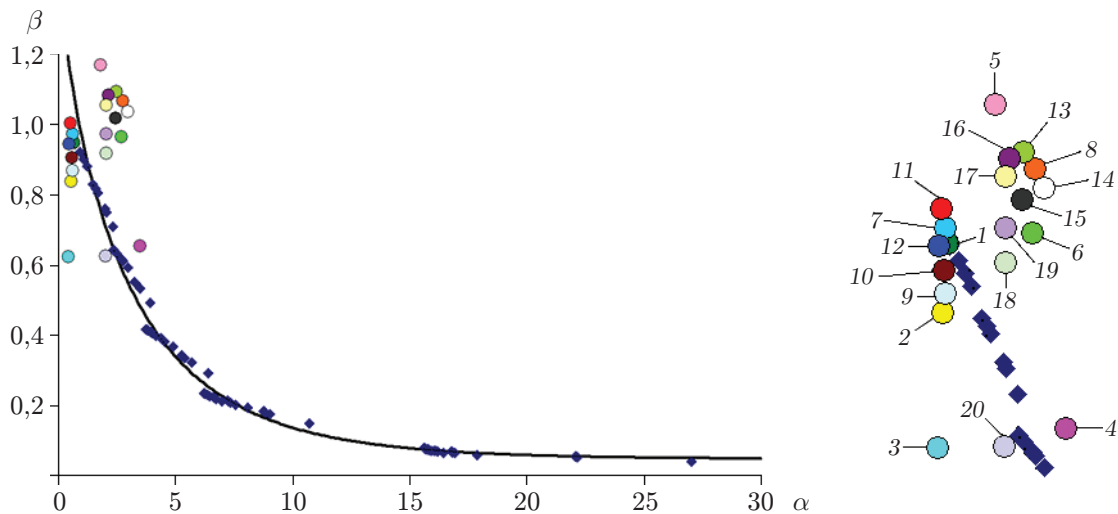


Рис. 1. Зависимость оценки функционала $\bar{\beta}$ (6) от обобщённого параметра α : точки (без номеров) — результаты вычислительных экспериментов, которые частично представлены в табл. 2; точки 1–20 — данные контрольных законов распределения (1–20) из табл. 3

Графическая иллюстрация функциональной зависимости (6) приведена на рис. 1.

Вид зависимости (6) справедлив при оценивании функционала β для одномерной β_1 и двумерной β_2 случайных величин из семейства логнормальных законов распределения. В этих условиях

$$\bar{\beta}_1 = 1,49 \exp(-0,589 \alpha^{0,857}) + 0,021, \quad \bar{\beta}_2 = 1,498 \exp(-0,524 \alpha^{0,809}) + 0,0356.$$

Соответствующие им средние относительные ошибки аппроксимации функционала β определяются значениями $\bar{\rho}_1 = 0,0436$, $\bar{\rho}_2 = 0,048$. Полученные результаты уточняют модели $\bar{\beta}_1$, $\bar{\beta}_2$ из [17].

Методика оценивания значений β .

1. На основе исходной выборки $V = (x_1^i, x_2^i, x_3^i, i = \overline{1, n})$ при восстановлении трёхмерной плотности вероятности $p(x_1, x_2, x_3) = \prod_{v=1}^3 p(x_v)$ оценить соответствующие им коэффициенты контрэксцесса

$$\bar{\delta}_v = 1/\sqrt{\bar{\eta}_v},$$

где

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

и асимметрии случайных величин x_1, x_2, x_3 :

$$\bar{\gamma}_v = \frac{1}{n} \sum_{i=1}^n (x_v^i - \bar{x}_v)^3 / \left(\frac{1}{n} \sum_{i=1}^n (x_v^i - \bar{x}_v)^2 \right)^{3/2}, \quad \bar{x}_v = \frac{1}{n} \sum_{i=1}^n x_v^i, \quad v = \overline{1, 3}.$$

2. По полученным данным вычислить оценку обобщённого параметра $\bar{\alpha} = (\bar{\delta}_1^2 + \bar{\delta}_2^2 + \bar{\delta}_3^2 + \bar{\gamma}_1^2 + \bar{\gamma}_2^2 + \bar{\gamma}_3^2)^{1/2}$ трёхмерного закона распределения независимых случайных величин.

Таблица 3

Значения нелинейного функционала β (4) и обобщённого параметра α при использовании ступенчатой $\Phi_1(u)$, параболической $\Phi_2(u)$ и гауссовской $\Phi_3(u)$ ядерных функций

№	Вид закона распределения $p(x_1)p(x_2)p(x_3)$	Значение β для ядерных функций			α
		$\Phi_1(u)$	$\Phi_2(u)$	$\Phi_3(u)$	
1	Гаусса	0,978	0,948	0,968	0,666
2	Логистический	0,865	0,838	0,856	0,543
3	Лапласа	0,643	0,623	0,636	0,441
4	Экспоненциальный	0,675	0,655	0,669	3,482
5	Односторонний нормальный	1,206	1,169	1,194	1,815
6	Односторонний логистический	0,996	0,965	0,986	2,700
7	Гаусса, логистический, Лапласа	1,004	0,973	0,994	0,610
8	Односторонний Гаусса, логистический, Лапласа	1,101	1,067	1,09	2,763
9	Гаусса, логистический, логистический	0,896	0,869	0,887	0,624
10	Логистический, Гаусса, Гаусса	0,933	0,905	0,924	0,591
11	Лапласа, Гаусса, Гаусса	1,036	1,004	1,026	0,527
12	Лапласа, логистический, логистический	0,974	0,944	0,964	0,472
13	Экспоненциальный, односторонний Гаусса, односторонний Гаусса	1,128	1,093	1,117	2,481
14	Экспоненциальный, односторонний логистический, односторонний логистический	1,07	1,037	1,06	2,979
15	Односторонний Гаусса, логистический, логистический	1,051	1,019	1,041	2,452
16	Односторонний логистический, Гаусса, Гаусса	1,119	1,084	1,107	2,138
17	Экспоненциальный, Гаусса, Гаусса	1,089	1,055	1,078	2,055
18	Экспоненциальный, логистический, логистический	0,946	0,917	0,937	2,041
19	Экспоненциальный, логистический, Гаусса	1,005	0,974	0,995	2,047
20	Экспоненциальный, Лапласа, Лапласа	0,644	0,625	0,638	2,034

Таблица 4

**Результаты анализа контрольных законов распределения
при использовании предлагаемой методики**

Номера законов распределения из табл. 3	Интервалы значений		
	α^j	$\bar{\beta}^j$	$\bar{\rho}^j$
Группа 1: 1, 4, 7, 9–12, 18–20	0,47–3,48	0,625–1,004	0,13–0,277
Группа 2: 2, 5, 6, 15–17	0,54–2,7	0,84–1,17	0,33–0,39
Группа 3: 8, 13, 14	2,48–2,98	1,037–1,093	0,43–0,47

3. Используя модель (6), вычислить оценку $\bar{\beta}$ нелинейного функционала β (4).

Значение $\bar{\beta}$ является результатом применения предлагаемой методики, которая необходима для формирования быстрой процедуры оценивания оптимального параметра c^* (3) коэффициента размытости ядерных функций в статистике (1).

Определим эффективность предлагаемой методики при оценивании функционала β в формуле оптимального коэффициента размытости ядерных функций статистики (3) при $k = 3$. В качестве контрольных законов распределения будем использовать плотности вероятностей случайных величин, близких и отличающихся от информации табл. 2. Параметры контрольных законов распределения приведены в табл. 3.

Результатом применения методики являются значения $\bar{\beta}$, которые рекомендуются для реализации быстрой процедуры оценивания параметра c^* (3) — коэффициента размытости ядерных функций $\bar{c}_v^* = \bar{\beta}\bar{\sigma}_v n^{-1/7}$, $v = \bar{1}, \bar{3}$, в статистике (1). Отметим, что аналитические данные в табл. 1–3 не являются случайными. Они определяются в соответствии с принятыми законами распределения случайных величин. Рассмотрим эффективность предлагаемой методики оценивания значений β в формуле (4) для закона распределения независимых случайных величин, представленных в табл. 3. В формуле (4) будем использовать оптимальную ядерную функцию Епанечникова $\Phi_2(u)$. Полученные результаты представлены в табл. 4.

Проведём анализ информации табл. 4, где значения относительных ошибок аппроксимации $\bar{\rho}^j = |\beta^j - \bar{\beta}^j|/\beta^j$, $j = \bar{1}, \bar{20}$. В первой группе контрольных распределений трёхмерной случайной величины $\bar{\rho}^j \in [0,13; 0,277]$, $j = (1, 4, 7, 9–12, 18–20)$. Минимальные относительные ошибки $\bar{\rho}^j$ аппроксимации нелинейного функционала β для контрольных примеров (см. табл. 3) соответствуют трёхмерным случайным величинам с номерами $j = 1, 7, 11, 20$. Этим сочетаниям соответствуют значения $\bar{\rho}^j \in [0,13–0,14]$. Случайные величины с законами распределения Гаусса, логистическим и экспоненциальным соответствуют малым значениям ошибки аппроксимации значений функционала β . Эти случайные величины имеют вторые производные плотностей вероятностей, которые являются составляющими функционала β . Нарушение этого условия снижает аппроксимационные свойства модели $\bar{\beta}$ (6), синтез которой основан на использовании данных табл. 2. Наличие экспоненциального, двух логистических, логистического и Лапласа законов распределения независимых компонент трёхмерной случайной величины увеличивает относительную ошибку аппроксимации β до значения 0,26. Этим значениям оценки относительной ошибки аппроксимации соответствуют наборы случайных величин под номерами: 4, 9, 10, 12, 18, 19.

Для наборов компонент трёхмерной случайной величины из группы 2 (см. табл. 4) ошибка аппроксимации нелинейного функционала β от плотностей вероятностей принадлежит интервалу $\bar{\rho}^j \in [0,33–0,39]$, $j = (2, 5, 6, 15–17)$. Эти наборы x_v , $v = \bar{1}, \bar{3}$, формируются в основном из односторонних законов распределения анализируемых случайных величин. В рассматриваемой группе присутствуют наборы из трёх логистических и односторонних нормальных законов распределения, односторонних логистических и нормальных плотностей вероятностей.

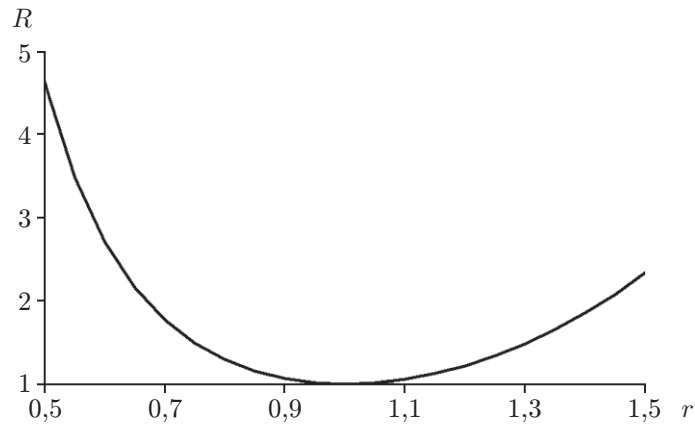


Рис. 2. Зависимость отношения R (7) от параметра r

Третью группу наборов составляют трёхмерные случайные величины, которые характеризуются в основном сочетаниями односторонних плотностей вероятностей, включающих экспоненциальные и Лапласа законы распределения. В этих условиях $\bar{\rho}^j \in [0,43-0,47]$, $j = (3, 8, 13, 14)$. Максимальная относительная ошибка аппроксимации функционала β соответствует набору из законов распределения Лапласа. Данный вывод является ожидаемым, так как распределение Лапласа не имеет производной при значении математического ожидания случайной величины.

Определим зависимость аппроксимационных свойств непараметрической оценки плотности вероятности от ошибки оценивания функционала β при $k = 3$. Предположим, что $\bar{\beta} = r\beta$, где $r = 1 \pm \bar{\rho}$. Значение $\bar{\beta}$ вычисляется по формуле (6). Тогда с учётом выражений (3), (4) значения $\bar{c}^* = rc^*$, где $c^* = (\beta/n)^{1/7}$.

Рассмотрим отношение средних квадратических ошибок аппроксимации плотности вероятности $p(x_1, x_2, x_3) = p(x_1)p(x_2)p(x_3)$ при значениях \bar{c}^* и c^* . Подставляя значения c^* и \bar{c}^* в критерий (2), получим

$$R = \frac{W(\bar{c}^*)}{W(c^*)} = \frac{3r^7 + 4}{7r^3}. \tag{7}$$

При ошибке $\bar{\rho}$ оценивания функционала β , равной нулю, значение параметра $r = 1$. Тогда отношение R равно 1, что подтверждает корректность выполненных преобразований. Зависимость отношения R от параметра r представлена на рис. 2.

При оценивании значений β (4) с использованием её модели (6) средняя относительная ошибка аппроксимации $\bar{\rho}$ составляет 0,0563, а значение r равно $1 \pm 0,0563$. В этих условиях максимальное отношение $R = 1,015$. Если принять значение относительной ошибки оценивания функционала β , например, $\bar{\rho} = 0,1$, то максимальное отношение $R = 1,065$. При относительных ошибках аппроксимации β (4), равных $\bar{\rho} = 0,2; 0,3$, отношение (7) соответствует максимальным значениям $R = 1,29; 1,76$.

Проведём подобный анализ для контрольных плотностей вероятностей независимых случайных величин, представленных в табл. 3, 4. Для первой группы независимых компонент трёхмерной случайной величины $R \in [1,15; 1,35]$, а минимальное их значение $R \in [1,09; 1,3]$. Использование модели (6) для второй группы плотностей вероятностей при оценивании функционала β приводит к значительному снижению аппроксимационных свойств непараметрических оценок плотностей вероятностей. В этих условиях минимальное отношение $R \in [1,8; 2,7]$. Очевидно, что для второй и третьей групп законов распределений случайных величин использование модели оценивания значений β по формуле (6)

не является обоснованным. Поэтому применение модели (6) для оценивания функционала β рекомендуется использовать для законов распределения независимых трёхмерных случайных величин, составляющих их первую группу.

Результаты исследований, представленные в [17] и полученные в данной работе, открывают возможность их обобщения с использованием метода аналогий при оценивании функционала от плотности вероятности многомерной случайной величины при построении быстрых процедур оптимизации непараметрической оценки плотности вероятности.

Заключение. Нелинейные функционалы от плотности вероятности являются основными составляющими формул расчёта оптимальных коэффициентов размытости ядерных функций непараметрической оценки плотности вероятности и показателей её аппроксимационных свойств. Поэтому задача оценивания нелинейного функционала от плотности вероятности приобретает актуальное значение особенно для условий больших объёмов статистических данных. Существует зависимость нелинейного функционала от плотности вероятности в формуле оптимального коэффициента размытости её ядерной оценки от значения обобщённого параметра $\bar{\alpha}$, определяемого нелинейной зависимостью от коэффициентов контрэкссесса и асимметрии, вид которой близок к одномерной, двухмерной и трёхмерной случайным величинам. Обнаруженная зависимость характерна для семейства логнормальных законов распределения трёхмерной случайной величины. Она может быть использована при анализе случайных величин с различающимися сочетаниями плотностей вероятностей компонент трёхмерной случайной величины. К таким плотностям вероятностей относятся гауссовские, экспоненциальные и логистические законы распределения. Наличие в наборах трёхмерных независимых случайных величин с плотностями вероятностей, для которых не существуют вторые производные, сопряжено со значительным снижением аппроксимационных свойств их непараметрических оценок.

Полученные результаты создают основу их развития для решения задачи быстрого выбора коэффициентов размытости многомерных ядерных оценок плотностей вероятностей, включая наборы зависимых случайных величин.

СПИСОК ЛИТЕРАТУРЫ

1. Лапко А. В., Лапко В. А. Непараметрические алгоритмы оценивания состояний природных объектов // Автометрия. 2018. **54**, № 5. С. 33–39. DOI: 10.15372/AUT20180504.
2. Зеньков И. В., Лапко А. В., Лапко В. А. и др. Непараметрический алгоритм автоматической классификации многомерных статистических данных большого объёма и его применение // Компьютерная оптика. 2021. **45**, № 2. С. 253–260. DOI: 10.18287/2412-6179-CO-801.
3. Зеньков И. В., Лапко А. В., Лапко В. А. и др. Непараметрический алгоритм распознавания образов в задаче проверки гипотезы о независимости случайных величин // Компьютерная оптика. 2021. **45**, № 5. С. 767–772. DOI: 10.18287/2412-6179-CO-871.
4. Лапко А. В., Лапко В. А., Бахтина А. В. Исследование методики проверки гипотезы о независимости двухмерных случайных величин с использованием непараметрического классификатора // Автометрия. 2021. **57**, № 6. С. 90–100. DOI: 10.15372/AUT20210610.
5. Rudemo M. Empirical choice of histogram and kernel density estimators // Scand. Journ. Statist. 1982. **9**, N 2. P. 65–78.
6. Hall P. Large-sample optimality of least squares cross-validation in density estimation // Ann. Statist. 1983. **11**, N 4. P. 1156–1174.
7. Bowman A. W. A comparative study of some kernel-based non-parametric density estimators // Journ. Statist. Comput. and Simulation. 1982. **21**. P. 313–327. DOI: 10.1080/00949658508810822.

8. **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.
9. **Варжапетян А. Г., Михайлова Е. Ю.** Методы выбора определяющих характеристик непараметрических алгоритмов идентификации моделей надёжности сложных систем по эксплуатационным данным // Вопросы кибернетики. 1982. Вып. 94. С. 77–87.
10. **Silverman B. W.** Density Estimation for Statistics and Data Analysis. London: Chapman & Hall, 1986. 175 p.
11. **Sheather S., Jones M.** A reliable data-based bandwidth selection method for density estimation // Journ. Royal Statist. Soc. Ser. B. 1991. **53**, N 3. P. 683–690. DOI: 10.1111/j.2517-6161.1991.tb01857.x.
12. **Sheather S. J.** Density estimation // Statist. Sci. 2004. **19**, N 4. P. 588–597. DOI: 10.1214/088342304000000297.
13. **Terrell G. R., Scott D. W.** Oversmoothed nonparametric density estimates // Journ. Amer. Statist. Association. 1985. **80**, N 389. P. 209–214. DOI: 10.2307/2288074.
14. **Jones M. C., Marron J. S., Sheather S. J.** A brief survey of bandwidth selection for density estimation // Journ. Amer. Statist. Association. 1996. **91**, N 433. P. 401–407. DOI: 10.2307/2291420.
15. **Лапко А. В., Лапко В. А.** Модифицированный алгоритм быстрого определения коэффициента размытости ядерной оценки плотности вероятности // Автометрия. 2020. **56**, № 6. С. 11–18. DOI: 10.15372/AUT20200602.
16. **Лапко А. В., Лапко В. А.** Ядерные оценки плотности вероятности и их применение. Красноярск: СибГУ им. М. Ф. Решетнёва, 2021. 308 с.
17. **Лапко А. В., Лапко В. А.** Оценивание нелинейного функционала от плотности вероятности при оптимизации непараметрических решающих функций // Измерительная техника. 2021. № 1. С. 14–20. DOI: 10.32446/0368-1025it.2021-1-14-20.
18. **Епанечников В. А.** Непараметрическая оценка многомерной плотности вероятности // Теория вероятности и её применения. 1969. **14**, № 1. С. 156–161.

Поступила в редакцию 08.02.2022

После доработки 21.02.2022

Принята к публикации 27.02.2022
