

Метод дискретизации области значений многомерной случайной величины

А. В. Лапко^{1,2}, В. А. Лапко^{1,2}

¹Институт вычислительного моделирования Сибирского отделения РАН, Красноярск, Россия

²Сибирский государственный аэрокосмический университет им. акад. М. Ф. Решетнева, Красноярск, Россия, e-mail: lapko@icm.krasn.ru

Рассмотрен метод дискретизации области значений многомерной случайной величины. Исследована его зависимость от объема, размерности исходной информации и вида плотностей вероятностей. Проведено сравнение полученных результатов с правилом Скотта для многомерной случайной величины с нормальным законом распределения.

Ключевые слова: непараметрическая оценка плотности вероятности, дискретизация области значений случайных величин, выбор длины интервала, правило Скотта.

Проблема дискретизации области значений многомерных случайных величин возникает при оценивании их плотности вероятности, доверительного оценивания плотности вероятности и проверке статистических гипотез с использованием критерия Пирсона. В настоящее время разработан ряд методов дискретизации интервала значений одномерной случайной величины [1–3].

Для выбора количества интервалов дискретизации области значений многомерной случайной величины используются эвристические рекомендации. Количество многомерных интервалов дискретизации в работе [4] предлагается определять из условия, при котором минимальное число наблюдений случайной величины, попадающих в один интервал, соответствовало числам 3, 5, 10. Известны формулы дискретизации многомерной случайной величины, полученные на основе анализа асимптотических свойств гистограммы [5, 6]. Такие формулы наиболее часто применяются при дискретизации области значений случайных величин с нормальными законами распределения.

Впервые выбор оптимального количества интервалов дискретизации одномерной и двухмерной случайных величин из условия минимума среднего квадратического отклонения (СКО) ядерной оценки плотности вероятности предложен в исследованиях [7–9]. Авторы настоящей статьи обобщили полученные результаты и провели анализ эффективности формулы дискретизации области значений случайных величин на многомерных интервалах.

Формулы оптимальной дискретизации области значений случайных величин. В работах [10, 11] для восстановления плотности вероятности многомерной случайной величины $x = (x_v, v=1, k)$ в условиях выборок $V = (x'_1, \dots, x'_k, i=1, n)$ большого объема n предложена непараметрическая оценка

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

где c_v – коэффициент размытости ядерной функции $\Phi(u_v)$;

$\bar{P}^j$ – оценка вероятности принадлежности случайной величины x к j -му многомерному интервалу с координатами его центра $z = (z_1, \dots, z_k)$.

Синтез статистики (1) предполагает декомпозицию области значений случайной величины $x = (x_1, \dots, x_k)$ на N многомерных интервалов с центрами $z^j = (z_v^j, v=1, k)$. Далее необходимо определить соответствующие этим интервалам частоты $\bar{P}^j, j=1, N$. Тогда исходная выборка V преобразуется в данные $\bar{V} = (\bar{P}^j, z^j, j=1, N)$, по которым оценивается плотность вероятности $p(x_1, \dots, x_k)$ в виде статистики (1).

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

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

$$\int_{-\infty}^{+\infty} \Phi(u_v) du_v = 1, \quad \int_{-\infty}^{+\infty} u_v^2 \Phi(u_v) du_v = 1,$$

а коэффициенты размытости этих функций $c_v(N) \rightarrow 0, v=1, k$ при $N \rightarrow \infty$.

Заметим, что рассмотренная непараметрическая оценка плотности вероятности (1) относится к семейству нормированных функций. Аппроксимационные свойства $\bar{p}(x_1, \dots, x_k)$ зависят от объема n исходной информации, размерности k случайной величины $x = (x_1, \dots, x_k)$, параметров дискретизации области её значений и особенностей восстанавливаемой плотности вероятности. Причём зависимость асимптотического выражения СКО $W_2(N)$ непараметрической оценки плотности вероятности от количества N интервалов дискретизации имеет экстремальный характер. Данный факт обосновывает возможность нахождения оптимального количества интервалов дискретизации. Таким образом, процедуру выбора оптимального количества интервалов дискретизации области Δ значений одномерной случайной величины x с плотностью вероятности $p(x)$ можно записать следующим образом [7]:

$$N_1^* = \left(n \Delta \int_{-\infty}^{+\infty} p^2(x) dx \right)^{1/2} = \alpha_1 \sqrt{n}. \quad (2)$$

Аналитическая зависимость (2) по виду близка к формуле Гаеда [12] и совпадает с ней при оценивании плотности вероятности случайной величины с равномерным законом распределения. Установлено, что коэффициент α_1 определяется только видом плотности вероятности и не зависит от её параметров. Результаты исследования асимптотических свойств непараметрической оценки плотности вероятности (1) при $k=2$ положены в основу синтеза формулы выбора оптимального количества интервалов дискретизации области значений двумерной случайной величины, которая справедлива и для условий разных интервалов Δ_1, Δ_2 значений компонентов x_1, x_2 [9]:

$$N_2^* = \left(\frac{3}{4} n \Delta^2 \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p^2(x_1, x_2) dx_1 dx_2 \right)^{1/2} = \alpha_2 \sqrt{n}, \quad (3)$$

где $p(x_1, x_2)$ – плотность вероятности двумерной случайной величины $x=(x_1, x_2)$.

Количество N_2^* интервалов дискретизации каждого компонента $x_v, v=1, 2$ при $\Delta_1=\Delta_2=\Delta$ соответствует значению $N_2^* = \sqrt{N_2^*}$. Значения коэффициента α_2 определяются только видом плотности вероятности и не зависят от её параметров. Для независимых случайных величин x_1, x_2 с равномерным, линейным и нормальным законами распределения коэффициент α_2 принимает значения 0,866; 1,15; 1,46, соответственно.

Вывод формулы оптимальной дискретизации области значений многомерной случайной величины сопряжён с вычислительными проблемами формирования асимптотического выражения $W_2(N)$ СКО $\bar{p}(x_1, \dots, x_k)$ от $p(x_1, \dots, x_k)$, что не позволяет получить аналитически значимые результаты.

Однако анализ предложенных формул оптимальной дискретизации областей определения одномерной и двумерной плотностей вероятности создаёт основу их обобщения на многомерный случай. В соответствии с методом аналогий предложим формулу дискретизации многомерной случайной величины:

$$N_k^* = \left(\alpha(k) n \Delta^k \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} p^2(x_1, \dots, x_k) dx_1, \dots, dx_k \right)^{1/2} = \alpha_k \sqrt{n}. \quad (4)$$

Коэффициент $\alpha(k) \leq 1$ и его значения уменьшаются с ростом размерности k случайной величины. Проанализируем формулы (2), (3) и сделаем вывод, что $\alpha(k) = (2k-1)/k^2$. При $k=1, k=2$ значение $\alpha(1)=1, \alpha(2)=3/4$, что согласуется с формулами (2), (3). Зависимость коэффициента α_k от размерности k случайной величины $x=(x_1, \dots, x_k)$ и вида её плотности вероятности $p(x_1, \dots, x_k)$ представлена в табл. 1.

Из табл. 1 следует, что значения коэффициентов α_k определяются размерностью k случайной величины, видом плотности вероятности и не зависят от её параметров.

Анализ аппроксимационных свойств непараметрической оценки плотности вероятности. Проанализируем зависимости оптимального количества интервалов дискре-

Т а б л и ц а 1

Значения коэффициентов α_k в формуле дискретизации (4) в условиях плотности вероятности $p(x_1, \dots, x_k)$ случайной величины $(x_1, \dots, x_k)$ и её размерности k

Размерность k случайной величины x	Закон распределения случайной величины		
	Равномерный	Нормальный	Экспоненциальный
1	1	1,301	1,516
2	0,866	1,466	1,992
3	0,745	1,641	2,599
4	0,661	1,895	3,498
5	0,600	2,236	4,812
6	0,553	2,680	6,724
7	0,515	3,249	9,501
8	0,484	3,973	13,542
9	0,458	4,891	19,434
10	0,436	6,055	28,041

тизации N_k^* , вычисляемых по формуле (4), от объёма n статистических данных, их размерности $k=3$ для плотностей вероятности $p(x_1, x_2, x_3) = \prod_{v=1}^3 p(x_v)$ с равномерным $p_1(x_1, x_2, x_3)$, нормальным $p_2(x_1, x_2, x_3)$ и экспоненциальным $p_3(x_1, x_2, x_3)$ законами распределения. Значения компонентов x_v с равномерным законом распределения определены на интервале $[0; 6]$. Плотности вероятностей $p_2(x_1), p_2(x_2), p_2(x_3)$ характеризуются нулевым математическим ожиданием и СКО, равным 1. Параметры экспоненциальных законов распределения $p_3(x_1), p_3(x_2), p_3(x_3)$ подобраны из условия равенства вероятности принадлежности β случайных величин $x=(x_1, x_2, x_3)$ интервалу $[0; 6]$ и значению $\beta=0,997$. При анализе исследуемых закономерностей будем ориентироваться на количество интервалов дискретизации N_k^* без их предварительного округления.

Увеличение объёма статистических данных сопровождается ростом оптимального количества интервалов дискретизации области определения плотности вероятности $p(x_1, x_2, x_3)$ пропорционально значению $n^{1/2}$.

Для экспоненциального закона распределения увеличивается количество интервалов дискретизации по сравнению с нормальным и равномерным законами. Данный факт объясняется неравномерным распределением статистических данных в области определения плотности вероятности $p(x_1, x_2, x_3)$ и усложнением условий асимптотической сходимости её непараметрической оценки (1) [10, 11]. Отношение количества многомерных интервалов дискретизации области определения плотностей вероятности, вычисляемых по формуле (4), не зависит от объёма n статистических данных, а определяется видом плотности вероятности и размерностью случайных величин. Например, отношение $N_3^*(2)/N_3^*(1)$ количества интервалов дискретизации $N_3^*(2), N_3^*(1)$, соответствующих плотностям

вероятности $p_2(x_1, x_2, x_3)$, $p_1(x_1, x_2, x_3)$ случайных величин $x=(x_1, x_2, x_3)$ равно 2,2, а $N_3^*(3)/N_3^*(2) = 1,58$.

С ростом размерности k усложняются условия восстановления плотностей вероятности, что приводит к увеличению рассматриваемых отношений. При $k=4$ отношение $N_4^*(2)/N_4^*(1)$ соответствует значению 2,86, а $N_4^*(3)/N_4^*(2) = 1,84$. Здесь $N_4^*(2)$ – количество оптимальных интервалов дискретизации области значений случайных величин $x=(x_1, \dots, x_4)$, вычисляемых по формуле (4), с плотностью вероятности $p_2(x_1, x_2, x_3)$. Отмеченные закономерности обусловливаются видом сравниваемых плотностей вероятности и не зависят от их параметров. Обнаруженные закономерности объясняются тем, что коэффициент α_k в формуле (4) определяется только видом плотности вероятности и не зависит от её параметров.

При увеличении значений N_k^* с ростом объёма n статистических данных естественно ожидать уменьшение длины $\bar{\Delta} = \Delta / (N_k^*)^{1/k}$ интервалов дискретизации каждого компонента x_1, x_2, x_3 . С усложнением закона распределения случайных величин длины интервалов $\bar{\Delta}$ уменьшаются. Например, отношение $\bar{\Delta}(2)/\bar{\Delta}(1)$, соответствующее плотностям вероятности $p_2(x_1, x_2, x_3)$, $p_1(x_1, x_2, x_3)$ при $k=3$, равно 0,768, а

Т а б л и ц а 2

Количество интервалов дискретизации $\bar{N}_3^*$, определяемых по предложенной методике с использованием формулы (4) при $k=3$, и соответствующие им значения $\bar{\Delta}$ для каждого компонента x_v , $v=1,3$, случайной величины x

Объём n исходных данных	Закон распределения случайной величины					
	Равномерный $p_1(x_1, x_2, x_3)$		Нормальный $p_2(x_1, x_2, x_3)$		Экспоненциальный $p_3(x_1, x_2, x_3)$	
	$\bar{N}_3^*$	$\bar{\Delta}$	$\bar{N}_3^*$	$\bar{\Delta}$	$\bar{N}_3^*$	$\bar{\Delta}$
500	27	2	64	1,5	64	1,5
900	27	2	64	1,5	125	1,2
1300	27	2	64	1,5	125	1,2
5000	64	1,5	125	1,2	216	1
10^4	125	1,2	216	1	343	0,857
10^5	343	0,857	729	0,667	1000	0,6

Т а б л и ц а 3

Количество интервалов дискретизации $\bar{N}_3^*$, определяемых по формуле Скотта, и соответствующие им значения $\bar{\Delta}$ для каждого компонента x_v , $v=1,3$, случайной величины с нормальным законом распределения

Параметр	Объём n исходных данных					
	500	900	1300	5000	10^4	10^5
$\bar{N}_3^*$	216	343	343	729	1331	4913
$\bar{\Delta}$	1	0,857	0,857	0,667	0,545	0,353

$\bar{\Delta}(3)/\bar{\Delta}(2) = 0,858$. Эти отношения не зависят от объёма n статистических данных. Отмеченные тенденции сохраняются и при увеличении размерности случайных величин.

Если выполняется соотношение $\bar{\Delta} > \Delta/2$, то в соответствующих ему условиях (n, k) значения компонента случайной величины $x=(x_1, \dots, x_k)$ нельзя разбить на количество интервалов более 1. Подобные ситуации встречаются, например, при декомпозиции области определения нормального закона распределения в условиях: $n=500$ при $k>5$; $1500>n>500$ при $k>6$; $10^4>n>1500$ при $k>8$. Декомпозицию области значений случайных величин с равномерным законом распределения невозможно осуществить в следующих условиях: $n=500$ и $k>3$; $1500>n>500$ и $k>4$; $10^4>n>1500$ и $k>5$. Полученные результаты согласуются с утверждением, что для эффективного оценивания плотности вероятности многомерной случайной величины необходимы тысячи её наблюдений [4].

Определим методику расчёта количества интервалов дискретизации области значений многомерных случайных величин. С помощью формулы (4) будем находить значения N_k^* и $\bar{N}_k^* = \lfloor (N_k^*)^{1/k} \rfloor + 1$, где $\lfloor (N_k^*)^{1/k} \rfloor$ – целая часть этого числа. Тогда длина интервала дискретизации случайных величин x_1, x_2, x_3 определяется значением $\bar{\Delta} = \Delta / \bar{N}_k^*$, а общее количество многомерных интервалов дискретизации $\bar{N}_k^* = (\bar{N}_k^*)^k$. Здесь Δ – интервал изменения значений каждой случайной величины x_1, x_2, x_3 .

Полученные ранее результаты, характеризующие зависимости N_k^* от объёма n исходных данных, их размерности и вида плотности вероятности согласуются со свойствами количества интервалов дискретизации $\bar{N}_k^*$ в условиях их целых значений (табл. 2).

Сравним количество интервалов дискретизации, полученных с использованием предложенных рекомендаций и формулы Скотта [5]. Для независимых случайных величин с нормальным законом распределения и принятых условий вычислительного эксперимента формулу Скотта можно представить в виде

$$\bar{N}_k^* = \left(\left\lfloor \frac{\Delta}{\bar{\Delta}} \right\rfloor + 1 \right)^k, \quad \bar{\Delta} = 3,5\sigma n^{-1/(2+k)}, \quad (5)$$

где σ – СКО случайных величин $x_1, \dots, x_k$.

Значения $\bar{N}_k^*$, $\bar{\Delta}$ для $k=3$ приведены в табл. 3.

С ростом объёма n статистических данных количество $\bar{N}_k^*$ интервалов дискретизации возрастает. Например, при дискретизации области значений случайных величин с нормальным законом распределения $p_2(x_1, x_2, x_3)$ значения $N_3^*(2)=64$ при $n=900$ и $N_3^*(2)=125$ при $n=5000$. Для экспоненциального закона распределения количество интервалов дискретизации $N_3^*(3)=64$ при $n=900$ и $N_3^*(3)=216$ при $n=5000$.

Заметим, что если применить формулу Скотта, то количество интервалов дискретизации области значений случайных величин значительно превысит значение $\bar{N}_3^*$, найденное в соответствии с рекомендациями данной статьи. Количество интервалов дискретизации, найденное по правилу Скотта (5), может в 3–7 раз превышать значения $\bar{N}_3^*$, которые вычислены по формуле (4). Данный факт особенно проявляется при больших объёмах статистических данных.

Рассмотрим интервалы дискретизации в области определения независимых случайных величин $x=(x_1, x_2, x_3)$ по нормальным законам распределения с минимальными и средними вероятностями принадлежности им значений x . К таким элементам отнесём, например, следующие многомерные интервалы:

$$\Delta(1)=(x_v \in [3\sigma - \bar{\Delta}, \infty), v=\overline{1,3});$$

$$\Delta(2)=(x_1 \in [-\bar{\Delta}/2, \bar{\Delta}/2]; x_v \in [3\sigma - \bar{\Delta}, \infty), v=\overline{2,3});$$

$$\Delta(3)=(x_v \in [-\bar{\Delta}/2, \bar{\Delta}/2], v=\overline{1,2}; x_3 \in [3\sigma - \bar{\Delta}, \infty)).$$

Пусть вероятность принадлежности трёхмерному интервалу $x \in \Delta(j)$ равна $P^t(j)$, $j=\overline{1,3}$, при использовании формулы дискретизации (4). По аналогии будем считать, что $P^2(j)$ – это вероятность принадлежности случайной величины x трёхмерному интервалу $x \in \Delta(j)$, $j=\overline{1,3}$, при применении правила Скотта. Обозначим через $N^t(j)$ – количество значений случайной величины x , попадающих в многомерный интервал $\Delta(j)$, $j=\overline{1,3}$, методами дискретизации с использованием формулы (4) ($t=1$) и правила Скотта ($t=2$). Значения $N^t(j)$ будем определять в соответствии с выражением

$$N^t(j) = P^t(j)n, j=\overline{1,3}, t=\overline{1,2}.$$

Для удобного анализа результатов вычислительного эксперимента значения $N^t(j)$ в табл. 4 не округлены. Верхние элементы таблицы соответствуют вероятностям $P^t(j)$ принадлежности значений случайных величин интервалам, а нижние представляют собой рассчитанное количество $N^t(j)$ случайных величин, принадлежащих $\Delta(j)$, $j=\overline{1,2}$. Результаты, представленные в табл. 4, подтверждают завышенные значения количества интервалов дискретизации области значений случайных величин, которые получены на основе правила Скотта. Их количество противоречит эвристическим рекомендациям, часто используемым в практических расчётах. Предложенный авторами подход дискретизации имеет преимущество по сравнению с правилом Скотта.

Обсуждение результатов. Формулы дискретизации области значений одномерной и двумерной случайных величин допускают их обобщение на многомерный вариант $x=(x_1, x_2, x_3)$. Количество интервалов дискретизации в многомерном случае пропорционально коэффициенту α_k и квадратному корню от объёма n исходных статистических данных. Значения коэффициента α_k обуславливаются размерностью k случайной величины, видом плотности вероятности и не зависят от её параметров. Увеличение объёма статистических данных сопровождается ростом количества интервалов дискретизации области определения плотности вероятности и пропорционально значению $n^{1/2}$. Отношение количества многомерных интервалов дискретизации области определения различных плотностей вероятности не зависит от объёма статистических данных, а определяется только видом плотности вероятности. Применение правила Скотта для многомерных

Зависимость значений $P^t(j)$, $N^t(j)$, $t=\overline{1,2}$, $j=\overline{1,2}$ от объёма n исходных статистических данных

Параметр	Объём n исходных данных					
	500	900	1300	5000	10^4	10^5
$P^1(1)$	$3 \cdot 10^{-4}$	$3 \cdot 10^{-4}$	$3 \cdot 10^{-4}$	$4,6 \cdot 10^{-5}$	$1,2 \cdot 10^{-5}$	$9,5 \cdot 10^{-7}$
$N^1(1)$	0,149	0,268	0,388	0,232	0,118	0,095
$P^1(2)$	$2,4 \cdot 10^{-3}$	$2,4 \cdot 10^{-3}$	$2,4 \cdot 10^{-3}$	$5,8 \cdot 10^{-4}$	$2 \cdot 10^{-4}$	$2,5 \cdot 10^{-5}$
$N^1(2)$	1,22	2,2	3,172	2,9	1,982	2,5
$P^1(3)$	0,02	0,02	0,02	$7,3 \cdot 10^{-3}$	$3,3 \cdot 10^{-3}$	$6,7 \cdot 10^{-4}$
$N^1(3)$	10	18	26	36,6	33,3	67
$P^2(1)$	$1,2 \cdot 10^{-5}$	$4 \cdot 10^{-6}$	$4 \cdot 10^{-6}$	$9,5 \cdot 10^{-7}$	$3,5 \cdot 10^{-7}$	$6,7 \cdot 10^{-8}$
$N^2(1)$	$5,6 \cdot 10^{-3}$	$3,7 \cdot 10^{-3}$	$5,4 \cdot 10^{-3}$	$4,7 \cdot 10^{-3}$	$3,5 \cdot 10^{-3}$	$6,7 \cdot 10^{-3}$
$P^2(2)$	$2 \cdot 10^{-4}$	$8,5 \cdot 10^{-5}$	$8,5 \cdot 10^{-5}$	$2,5 \cdot 10^{-5}$	$1,07 \cdot 10^{-5}$	$2,3 \cdot 10^{-6}$
$N^2(2)$	0,1	0,077	0,111	0,126	0,107	0,231
$P^2(3)$	$3,3 \cdot 10^{-3}$	$1,8 \cdot 10^{-3}$	$1,8 \cdot 10^{-3}$	$6,7 \cdot 10^{-4}$	$3,2 \cdot 10^{-4}$	$8 \cdot 10^{-5}$
$N^2(3)$	1,7	1,6	2,3	3,35	3,2	8

независимых случайных величин с нормальным законом распределения приводит к завышенному количеству интервалов дискретизации, которые не согласуются с предложенными авторами данной статьи рекомендациями и эвристическими методами дискретизации. Количество интервалов дискретизации, определяемое на основе правила Скотта, не обеспечивает восстановление плотности вероятности в области её малых значений.

Перспективное направление развития рассмотренной методики дискретизации многомерной случайной величины предполагает оценивание функционала

$$\Delta^k \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} p^2(x_1, \dots, x_k) dx_1, \dots, dx_k.$$

Его значения определяются только видом плотности вероятности и не зависят от её параметров.

Заключение. Полученные результаты имеют важное значение при восстановлении плотности вероятности многомерных случайных величин в условиях статистических данных большого объёма, её доверительного оценивания и проверке гипотез о распределениях случайных величин с использованием критерия Пирсона. Исследования выполнены при поддержке РФФИ в рамках гранта № 18-01-00251.

Литература

1. **Sturges H. A.** The choice of a class interval // J. American Statistical Association. 1926. V. 21. P. 65–66. DOI: 10.1080/01621459.1926.10502161.

2. **Scott D. W.** Multivariate Density Estimation: Theory, Practice, and Visualization. N. Y.: Wiley, 1992.
3. **Hacine Gharbi A., Ravier P., Harba R., Mohamadi T.** Low bias histogram based estimation of mutual information for feature selection // Pattern Recognition Letters. 2012. V. 33. No. 10. P. 1302–1308. DOI: 10.1016/j.patrec.2012.02.022.
4. **Пугачёв В. С.** Теория вероятностей и математическая статистика: Учеб. пособие. М.: ФИЗМАТЛИТ, 2002.
5. **Scott D. W.** Multivariate Density Estimation: Theory, Practice, and Visualization. 2nd Edition. NJ: John Wiley & Sons, 2015.
6. **Devroye L., Lugosi G.** Bin width selection in multivariate histograms by the combinatorial method // Test. 2004. V. 13. No. 1. P. 129–145. DOI: 10.1007/BF02603004.
7. **Лапко А. В., Лапко В. А.** Оптимальный выбор количества интервалов дискретизации области изменения одномерной случайной величины при оценивании плотности вероятности // Измерительная техника. 2013. № 7. С. 24–27.
8. **Лапко А. В., Лапко В. А.** Выбор оптимального количества интервалов дискретизации области значений двумерной случайной величины // Измерительная техника. 2016. № 2. С. 14–17.
9. **Лапко А. В., Лапко В. А.** Сравнение эффективности методов дискретизации области значений зависимых случайных величин при синтезе непараметрической оценки двумерной плотности вероятности // Измерительная техника. 2017. № 4. С. 15–18.
10. **Lapko A. V., Lapko V. A.** Regression Estimate of the Multidimensional Probability Density and Its Properties // Optoelectronics, Instrumentation and Data Processing. 2014. V. 50. No. 2. P. 148–153. DOI: 10.3103/S875669901402006X.
11. **Лапко А. В., Лапко В. А.** Регрессионная оценка многомерной плотности вероятности и её свойства // Автоматизация. 2014. Т. 50. № 2. С. 50–56.
12. **Heinhold I., Gaede K.** Ingenieur statistic – München: Wien, Springer Verlag, 1964.

Method of sampling the range of values of a multidimensional random variable

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

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

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

The method of sampling the range of values of a multidimensional random variable is considered. We study its dependence on the volume, the dimension of the apriori information, and the form of the probability densities. The results are compared with the Scott's rule for a multidimensional random variable for normal distribution law.

Key words: *multivariate histograms, probability density, sampling the range of values of a random variable, bin width selection, Scott's rule.*