УДК 519.7

ПРОВЕРКА ГИПОТЕЗЫ О НЕЗАВИСИМОСТИ ДВУМЕРНЫХ СЛУЧАЙНЫХ ВЕЛИЧИН С ИСПОЛЬЗОВАНИЕМ НЕПАРАМЕТРИЧЕСКОГО АЛГОРИТМА РАСПОЗНАВАНИЯ ОБРАЗОВ

© **А. В. Лапко[1,2], В. А. Лапко[1,2]**

[1]*Институт вычислительного моделирования СО РАН,*
660036, г. Красноярск, Академгородок, 50, стр. 44
[2]*Сибирский 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 новая методика проверки гипотезы о независимости двумерных случайных величин. Рассматриваемая методика основывается на использовании непараметрического алгоритма распознавания образов, соответствующего критерию максимального правдоподобия. В отличие от традиционной постановки задачи априори отсутствует обучающая выборка. Исходная информация представляется статистическими данными, которые составляют значения двумерных случайных величин. Законы распределения случайных величин в классах оцениваются по исходным статистическим данным для условий их зависимости и независимости. При выборе оптимальных коэффициентов размытости непараметрических оценок плотностей вероятностей в качестве критерия используется максимум функций правдоподобия. В этих условиях вычисляются оценки вероятности ошибки распознавания образов в классах. По минимальному значению оценки вероятности ошибки распознавания образов принимается решение о независимости либо зависимости случайных величин. Эффективность разработанной методики подтверждается результатами вычислительных экспериментов при проверке гипотезы о независимости либо линейной зависимости двумерных случайных величин.

Ключевые слова: проверка гипотезы о независимости случайных величин, двумерные случайные величины, распознавание образов, ядерная оценка плотности вероятности, критерий максимального правдоподобия, доверительное оценивание вероятностей.

DOI: 10.15372/AUT20210205

Введение. Сведения о зависимости либо независимости случайных величин являются необходимым условием синтеза эффективных алгоритмов обработки информации и принятия решений. В [1–3] исследованы свойства непараметрической оценки плотности вероятности типа Розенблатта — Парзена независимых случайных величин. Установлено, что наличие априорных сведений о независимости случайных величин позволяет повысить аппроксимационные свойства непараметрической оценки их плотности вероятности по сравнению с ядерной статистикой для зависимых случайных величин. Данное преимущество возрастает с увеличением размерности случайных величин. Полученные результаты подтверждаются при исследовании асимптотических свойств непараметрической оценки уравнения разделяющей поверхности в двухальтернативной задаче распознавания образов [4].

Традиционная методика проверки гипотезы о независимости случайных величин основана на использовании универсального χ^2-критерия Пирсона. Однако его формирование содержит трудно формализуемый этап разбиения области значений случайных величин на многомерные интервалы [5]. Поэтому возникает задача разработки новой методики проверки гипотезы, обеспечивающей обход проблемы декомпозиции области значений случайных величин. Подобная задача решена при проверке гипотезы о тождественности законов распределения случайных величин на основе использования непараметрического алгоритма распознавания образов [6–8]. Показана возможность её замены задачей проверки гипотезы о равенстве ошибки распознавания образов определённому пороговому значению. Обучающая выборка при синтезе непараметрического алгоритма распознавания образов формируется по статистическим данным, характеризующим законы распределения сравниваемых случайных величин.

Цель представленной работы состоит в развитии предложенного подхода к задаче проверки гипотезы о независимости случайных величин с использованием непараметрического алгоритма распознавания образов.

Методика проверки гипотезы о независимости случайных величин. Пусть имеется выборка $V = (x^i,\ i = \overline{1,n}\,)$ объёма n, составленная из независимых наблюдений двумерной случайной величины $x = (x_1, x_2)$. Выборка V извлекается из генеральных совокупностей, характеризующихся плотностями вероятности $p(x_1)p(x_2)$ или $p(x_1, x_2)$. Необходимо по статистическим данным V проверить гипотезу

$$H_0\colon\quad p(x_1, x_2) \equiv p(x_1)p(x_2) \tag{1}$$

о независимости случайных величин x_1, x_2.

Для проверки гипотезы H_0 (1) будем решать двухальтернативную задачу распознавания образов. Под классами Ω_1, Ω_2 понимаются области определения плотностей вероятностей $p(x_1)p(x_2)$, $p(x_1, x_2)$. В этих условиях байесовское решающее правило, соответствующее критерию максимального правдоподобия, имеет вид

$$m(x)\colon \begin{cases} x \in \Omega_1, & \text{если } p(x_1, x_2) < p(x_1)p(x_2); \\ x \in \Omega_2, & \text{если } p(x_1, x_2) > p(x_1)p(x_2). \end{cases} \tag{2}$$

В отличие от традиционной постановки задачи распознавания образов при синтезе решающего правила (2) априори отсутствует обучающая выборка, содержащая сведения о принадлежности элементов выборки V к тому или иному классу. Эти сведения должны обнаруживаться в процессе реализации методики проверки гипотезы H_0, которая основана на выполнении следующих действий.

По выборке V восстановим плотности вероятностей $p(x_1, x_2)$, $p(x_1)p(x_2)$, используя их непараметрические оценки типа Розенблатта — Парзена [9, 10]:

$$\bar{p}(x_1, x_2) = \frac{1}{nc_1c_2} \sum_{i=1}^{n} \Phi\Big(\frac{x_1 - x_1^i}{c_1}\Big) \Phi\Big(\frac{x_2 - x_2^i}{c_2}\Big), \tag{3}$$

$$\bar{p}(x_1)\bar{p}(x_2) = \frac{1}{n^2c_1c_2} \sum_{i=1}^{n} \sum_{j=1}^{n} \Phi\Big(\frac{x_1 - x_1^i}{c_1}\Big) \Phi\Big(\frac{x_2 - x_2^j}{c_2}\Big). \tag{4}$$

В статистиках (3), (4) ядерные функции $\Phi(u_v)$ удовлетворяют следующим условиям:

$$\Phi(u_v) = \Phi(-u_v), \qquad 0 \leqslant \Phi(u_v) < \infty, \qquad \int\limits_{-\infty}^{+\infty} \Phi(u_v)\, du_v = 1,$$

$$\int\limits_{-\infty}^{+\infty} u^m \Phi(u_v)\, du_v < \infty, \qquad 0 \leqslant m < \infty, \quad v = 1, 2.$$

Значения коэффициентов размытости c_v ядерных функций убывают с ростом объёма n выборки статистических данных V. С учётом выражений (2)–(4) непараметрическое решающее правило классификации случайных величин $x = (x_1, x_2)$ запишется как

$$\bar{m}(x)\colon \begin{cases} x \in \Omega_1, & \text{если } \bar{p}(x_1, x_2) < \bar{p}(x_1)\bar{p}(x_2); \\ x \in \Omega_2, & \text{если } \bar{p}(x_1, x_2) > \bar{p}(x_1)\bar{p}(x_2). \end{cases} \tag{5}$$

В условиях подобной неопределённости будем выбирать оптимальные коэффициенты размытости ядерных функций решающего правила (5) на основе анализа аппроксимационных свойств непараметрических оценок $\bar{p}(x_1, x_2)$, $\bar{p}(x_1)\bar{p}(x_2)$ плотностей вероятностей $p(x_1, x_2)$, $p(x_1)p(x_2)$. Для выбора оптимальных коэффициентов размытости непараметрической оценки плотности вероятности $p(x_1, x_2)$ в качестве критерия используется, например, максимум функции правдоподобия [11, 12]

$$L(c_1, c_2) = \prod_{j=1}^{n} \bar{p}(x_1^j, x_2^j), \quad \bar{p}(x_1^j, x_2^j) = \frac{1}{(n-1)c_1 c_2} \sum_{i=1,\ i \neq j}^{n} \Phi\Big(\frac{x_1^j - x_1^i}{c_1}\Big)\Phi\Big(\frac{x_2^j - x_2^i}{c_2}\Big). \tag{6}$$

По аналогии с выражением (6) нетрудно определить критерий выбора оптимальных коэффициентов размытости статистики $\bar{p}(x_1)\bar{p}(x_2)$ (4).

Отметим, что выбор оптимальных коэффициентов размытости непараметрических оценок плотностей вероятностей $\bar{p}(x_1, x_2)$, $\bar{p}(x_1)\bar{p}(x_2)$ может осуществляться из условия минимума статистических оценок средних квадратических отклонений $\bar{p}(x_1, x_2)$, $\bar{p}(x_1)\bar{p}(x_2)$ от $p(x_1, x_2)$, $p(x_1)p(x_2)$ соответственно [13–19].

Оптимизацию непараметрического решающего правила (5) по коэффициентам размытости ядерных функций c_1, c_2 можно упростить, если положить в статистиках (3), (4) значения $c_v = c\bar{\sigma}_v$, $v = 1, 2$. Здесь $\bar{\sigma}_v$ — оценка среднего квадратического отклонения случайной величины x_v в выборке V. Данное утверждение является очевидным, так как большей длине интервала значений x_v соответствует больший коэффициент размытости c_v ядерной функции $\Phi(u_v)$, $v = 1, 2$. Подобный подход использовался при построении «быстрых» процедур оптимизации непараметрических оценок плотности вероятности ядерного типа [20–23].

Значения оценок средних квадратических отклонений $\bar{\sigma}_v$ определяются по статистическим данным выборки V:

$$\bar{\sigma}_v = \Big(\frac{1}{n-1} \sum_{i=1}^{n} (x_v^i - \bar{x}_v)^2\Big)^{1/2}, \qquad v = 1, 2.$$

Здесь $\bar{x}_v$ — среднее значение случайной величины x_v, которое вычисляется по выборке V.

Поэтому появляется возможность оптимизацию непараметрического алгоритма распознавания образов (5) проводить лишь по одному параметру c коэффициентов размытости ядерных функций.

Определим оценки вероятностей ошибок распознавания образов $\bar{\rho}_1(\bar{c}_1, \bar{c}_2)$, $\bar{\rho}_2(\bar{c}_1, \bar{c}_2)$ решающим правилом (5) по исходным статистическим данным V при оптимальных коэффициентах размытости $\bar{c}(1) = (\bar{c}_1(1), \bar{c}_2(1))$, $\bar{c}(2) = (\bar{c}_1(2), \bar{c}_2(2))$ ядерных функций статистик $\bar{p}(x_1)\bar{p}(x_2)$, $\bar{p}(x_1, x_2)$ соответственно.

Значения $\bar{\rho}_t(\bar{c}(1), \bar{c}(2))$ вычисляются в режиме «скользящего экзамена» по выборке V в предположении, что её элементы принадлежат к классу Ω_t:

$$\bar{\rho}_t(\bar{c}(1), \bar{c}(2)) = \frac{1}{n} \sum_{j=1}^{n} 1(\delta(j), \bar{\delta}(j)), \qquad t = 1, 2,$$

где $\delta(j) = t$ — указания типа $x^j = (x_1^j, x_2^j) \in \Omega_t$;

$$\bar{\delta}(j) = \begin{cases} t, & \text{если } x^j \in \Omega_t; \\ 0, & \text{если } x^j \notin \Omega_t, \end{cases}$$

— решение алгоритма (5) о принадлежности ситуации x^j к одному из классов Ω_t, $t = 1, 2$.

При вычислении $\bar{\rho}_t(\bar{c}(1), \bar{c}(2))$ в соответствии с методикой «скользящего экзамена» ситуация $x^j = (x_1^j, x_2^j)$ из выборки V, которая подаётся на контроль в алгоритм (5), исключается из процесса формирования статистик (3), (4).

Индикаторная функция определяется выражением

$$1(\delta(j), \bar{\delta}(j)) = \begin{cases} 0, & \text{если } \delta(j) = \bar{\delta}(j); \\ 1, & \text{если } \delta(j) \neq \bar{\delta}(j). \end{cases}$$

Обозначим через $\bar{\bar{\rho}}_t$ минимальное значение оценки вероятности ошибки распознавания образов в предположении, что элементы выборки V принадлежат к классу Ω_t, $t = 1, 2$. Сравним значения $\bar{\bar{\rho}}_1$, $\bar{\bar{\rho}}_2$.

Гипотеза H_0 справедлива, если $\bar{\bar{\rho}}_1 < \bar{\bar{\rho}}_2$. В противном случае при $\bar{\bar{\rho}}_2 < \bar{\bar{\rho}}_1$ случайные величины x_1 и x_2 являются зависимыми.

Естественно, что при ограниченных объёмах n выборки V возникает задача доверительного оценивания вероятностей ошибок распознавания образов. Для её решения может применяться традиционная методика доверительного оценивания вероятностей [5] либо критерий Колмогорова — Смирнова [24].

Например, при использовании критерия Колмогорова — Смирнова отклонение $\bar{D}_{12} = |\bar{\bar{\rho}}_1 - \bar{\bar{\rho}}_2|$ сравнивается с пороговым значением

$$D_\beta = \sqrt{-\ln(\beta/2)/n}\,.$$

Здесь β — вероятность (риск) отвергнуть гипотезу $\bar{H}_0$: $\rho_1(c_1, c_2) = \rho_2(c_1, c_2)$. Если выполняется соотношение $\bar{D}_{12} < D_\beta$, то гипотеза $\bar{H}_0$ справедлива и риск её отвергнуть не превышает значения β. При $\bar{D}_{12} > D_\beta$ гипотеза $\bar{H}_0$ отвергается.

Анализ результатов вычислительного эксперимента. Исследуем зависимость эффективности предлагаемой методики проверки гипотезы о независимости двумерных случайных величин от объёма исходных статистических данных. Будем считать, что случайные величины x_1 и x_2 имеют законы распределения Гаусса. При формировании значений x_1, x_2 в выборке V используются датчики случайных величин

$$x_1^i = M(x_1) + \sigma_1\Big(\sum_{j=1}^{12} \varepsilon_1^j - 6\Big), \qquad x_2^i = x_1^i + \sigma_2\Big(\sum_{j=1}^{12} \varepsilon_2^j - 6\Big), \qquad i = \overline{1, n},$$

где $M(x_1)$ — математическое ожидание случайной величины x_1; σ_1, σ_2 — средние квадратические отклонения x_1 и x_2; ε_1, ε_2 — случайные величины с равномерными законами распределения на интервале $[0; 1]$.

При выборе значений σ_1, σ_2 изменяется показатель зависимости (коэффициент корреляции) между случайными величинами x_1, x_2, который вычисляется по полученным статистическим данным V. Значения $\bar{\bar{\rho}}_1^t$, $\bar{\bar{\rho}}_2^t$ при конкретном объёме $n(t) = n$ выборки V определяются 50 раз. Полученные данные $\bar{\bar{\rho}}_1^t$, $\bar{\bar{\rho}}_2^t$, $t = \overline{1, 50}$ усредняются, а их результаты обозначаются $\tilde{\rho}_1$, $\tilde{\rho}_2$ и представлены на рис. 1.

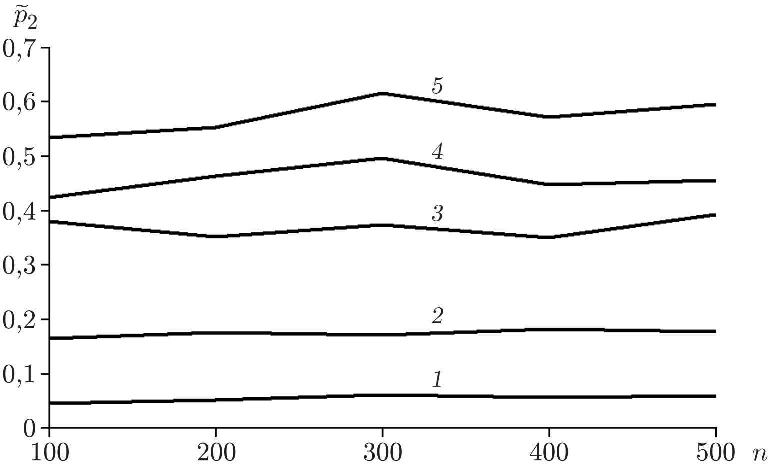

Рис. 1. Зависимость усреднённых оценок вероятностей ошибок принадлежности элементов V к независимым случайным величинам x_1, x_2 от объёма выборки n (кривые *1*—*5* соответствуют значениям коэффициентов корреляции $r = 0{,}9$; 0,7; 0,45; 0,33; 0)

Результаты вычислительных экспериментов подтверждают эффективность предлагаемой методики. При коэффициенте корреляции $r \geqslant 0{,}35$ применение рассматриваемой методики позволяет исключить ошибки отнесения исходных статистических данных V к независимым случайным величинам в 50 вычислительных экспериментах. Обозначим эту оценку вероятности подтверждения гипотезы H_0 через $\bar{P}_1 = 0$. Если $r = 0$, значения $\bar{P}_1 = 0{,}6$; 0,62; 0,8 при объёмах статистических данных $n = 100$; 200; 500 соответственно.

Проведём анализ значений $\tilde{\rho}_1$, $\tilde{\rho}_2$, определяющих критерий проверки гипотезы H_0 в вычислительных экспериментах. С ростом коэффициента корреляции r наблюдается уменьшение усреднённой оценки вероятности ошибки $\tilde{\rho}_2$ принадлежности элементов выборки V к классу Ω_2 значений зависимых случайных величин. Например, при увеличении r в интервале $[0{,}45; 0{,}9]$ оценка вероятности ошибки распознавания образов $\tilde{\rho}_2$ уменьшается от значения 0,37 до 0,055. Данный факт объясняется уменьшением области пересечения классов Ω_1, Ω_2 и, как следствие, увеличением значений ядерной оценки плотности вероятности $\bar{p}(x_1, x_2)$ по сравнению с $\bar{p}(x_1)\bar{p}(x_2)$ в непараметрическом решающем правиле (5), что приводит к уменьшению $\tilde{\rho}_2$. При уменьшении r в интервале $[0, 33; 0]$ оценки вероятностей ошибки распознавания образов $\tilde{\rho}_2$ увеличиваются от значения 0,45 до 0,53. В этих условиях наблюдается тенденция сближения области пересечения классов Ω_1, Ω_2 и значений $\tilde{\rho}_2$, $\tilde{\rho}_1$ как критерий тождественности законов распределения $p(x_1, x_2)$, $p(x_1)p(x_2)$ сравниваемых случайных величин.

Обнаружена устойчивость предлагаемой методики к объёму исходных статистических данных при конкретных значениях коэффициента корреляции, которая проявляется в близких значениях $\tilde{\rho}_2$ при $n \in [100; 500]$. Например, при $r = 0{,}9$ значения $\tilde{\rho}_2 \in [0{,}044; 0{,}06]$, а в условиях $r = 0{,}45$ значения $\tilde{\rho}_2 \in [0{,}35; 0{,}39]$ (см. рис. 1). Отмеченная закономерность ослабевает с уменьшением значений r. Данный вывод подтверждается значениями $\tilde{\rho}_2$ в интервале $[0{,}423; 0{,}496]$ при $r = 0{,}33$.

Приведённые утверждения являются достоверными, что проверяется с использованием критерия Колмогорова — Смирнова при риске $\beta = 0{,}05$ отвергнуть проверяемую гипотезу. Сомнительные решения проявляются при значениях r, близких к нулю. В этих условиях предлагаемая методика обеспечивает достоверное решение при $n \geqslant 300$. Например, при $n = 300$, 400, 500 значения $\bar{D}_{12} = 0{,}23$; 0,141; 0,189, которые превышают пороговые значения $D_\beta = 0{,}111$; 0,096; 0,086. Полученные результаты подтверждают гипотезу о независимости случайных величин.

Результаты вычислительных экспериментов сравнивались с доверительными границами коэффициента корреляции

$$\operatorname{th}\left(\frac{1}{2}\ln\frac{1+\bar{r}}{1-\bar{r}}-\frac{\varepsilon_\alpha}{\sqrt{n-3}}\right)<r<\operatorname{th}\left(\frac{1}{2}\ln\frac{1+\bar{r}}{1-\bar{r}}+\frac{\varepsilon_\alpha}{\sqrt{n-3}}\right),$$

где ε_α определяется соотношением $2F(\varepsilon_\alpha)=\alpha$; $\bar{r}$ — оценка коэффициента корреляции. Здесь $F(\varepsilon_\alpha)$ — функция Лапласа, а α — коэффициент доверия; $\operatorname{th}(\cdot)$ — гиперболический тангенс. Для этих условий при $\bar{r}=0$, $\alpha=0{,}95$ и $\varepsilon_\alpha=1{,}96$ доверительные границы для коэффициента корреляции определяются интервалами $r\in(\pm0{,}196)$, $(\pm0{,}139)$, $(\pm0{,}113)$, $(\pm0{,}098)$, $(\pm0{,}088)$, которые соответствуют объёмам статистических данных $n=100$, 200, 300, 400, 500.

Сравним эффективность предлагаемой методики с подходом, который использует коэффициент корреляции в качестве критерия линейной зависимости между случайными величинами. Для этого определим оценки вероятностей решений относительно гипотезы H_0, принимаемых в соответствии с предлагаемой методикой, в условиях значения коэффициента корреляции $r=0{,}196$. Отметим, что это значение r соответствует его доверительной границе при $n=100$ и $\alpha=0{,}95$. В этих условиях оценка вероятности подтверждения гипотезы H_0 соответствует значению $\bar{P}_1=0{,}4$, а её опровержение $\bar{P}_2=0{,}6$ в 50 вычислительных экспериментах. По значениям $\bar{P}_1$, $\bar{P}_2$ предлагаемая методика более чувствительна к изменению показателя r линейной зависимости между случайными величинами x_1, x_2. Полученные результаты согласуются с традиционным подходом проверки гипотезы о линейной зависимости случайных величин. Однако представленная методика распространяется на условия нелинейной зависимости между случайными величинами.

Заключение. Предложенная в данной работе методика проверки гипотезы о независимости случайных величин обеспечивает обход проблемы декомпозиции области значений случайных величин на многомерные интервалы, которая свойственна критерию Пирсона. Для решения данной проблемы используется непараметрический алгоритм распознавания образов, соответствующий критерию максимального правдоподобия. Оптимизация ядерных оценок плотностей вероятностей по коэффициентам размытости осуществляется из условия максимума функции правдоподобия. В предположении независимости либо зависимости случайных величин в исходных статистических данных определяются оценки вероятностей ошибок распознавания образов. По их минимальному значению принимается решение о независимости либо зависимости случайных величин.

Эффективность предлагаемой методики подтверждается результатами вычислительных экспериментов при проверке гипотезы о независимости двумерной случайной величины, компоненты которой характеризуются нормальными законами распределения. Установлено, что при коэффициенте корреляции $r\geqslant0{,}35$ между случайными величинами предлагаемая методика безошибочно отвергает исходную гипотезу при объёме исходных статистических данных от 100 до 500. При независимых случайных величинах в условиях, когда коэффициент корреляции равен нулю, исходная гипотеза подтверждается оценками вероятности 0,6; 0,62; 0,8 при объёме статистических данных n=100, 200, 500 соответственно. Наблюдается устойчивость значений используемого критерия проверки рассматриваемой гипотезы к изменению объёма статистических данных при конкретных условиях эксперимента.

Перспективными исследованиями в данном направлении являются применение предлагаемой методики при проверке гипотезы о нелинейной зависимости между случайными величинами и формирование набора независимых случайных величин, что позволит упростить задачу синтеза эффективных алгоритмов обработки информации.

Финансирование. Исследование выполнено при финансовой поддержке Российского фонда фундаментальных исследований, Правительства Красноярского края и Красноярского краевого фонда науки (проект № 20-41-240001).

СПИСОК ЛИТЕРАТУРЫ

1. **Лапко А. В., Лапко В. А.** Свойства непараметрической оценки многомерной плотности вероятности независимых случайных величин // Информатика и системы управления. 2012. **31**, № 1. С. 166–174.
2. **Лапко А. В., Лапко В. А.** Непараметрическая оценка плотности вероятности независимых случайных величин // Информатика и системы управления. 2011. **29**, № 3. С. 118–124.
3. **Лапко А. В., Лапко В. А.** Влияние априорной информации о независимости многомерных случайных величин на свойства их непараметрической оценки плотности вероятности // Системы управления и информационные технологии. 2012. **48**, № 2.1. С. 164–167.
4. **Лапко А. В., Лапко В. А.** Свойства непараметрической решающей функции при наличии априорных сведений о независимости признаков классифицируемых объектов // Автометрия. 2012. **48**, № 4. С. 112–119.
5. **Пугачёв В. С.** Теория вероятностей и математическая статистика: Учеб. пособие. М: Физматлит, 2002. 496 с.
6. **Лапко А. В., Лапко В. А.** Непараметрические алгоритмы распознавания образов в задаче проверки статистической гипотезы о тождественности двух законов распределения случайных величин // Автометрия. 2010. **46**, № 6. С. 47–53.
7. **Лапко А. В., Лапко В. А.** Сравнение эмпирической и теоретической функций распределения случайной величины на основе непараметрического классификатора // Автометрия. 2012. **48**, № 1. С. 45–49.
8. **Лапко А. В., Лапко В. А.** Методика проверки гипотез о распределениях многомерных спектральных данных с использованием непараметрического алгоритма распознавания образов // Компьютерная оптика. 2019. **43**, № 2. С. 238–244. DOI: 10.18287/2412-6179-2019-43-2-238-244.
9. **Parzen E.** On estimation of a probability density function and mode // Ann. Math. Statist. 1962. **33**, N 3. P. 1065–1076.
10. **Епанечников В. А.** Непараметрическая оценка многомерной плотности вероятности // Теория вероятности и ее применения. 1969. **14**, N 1. С. 156–161.
11. **Duin R. P. W.** On the choice of smoothing parameters for parzen estimators of probability density functions // IEEE Trans. Comp. 1976. **C-25**, Iss. 11. P. 1175–1179.
12. **Botev Z. I., Kroese D. P.** Non-asymptotic bandwidth selection for density estimation of discrete data // Methodology and Comput. Appl. Probability. 2008. **10**, N 3. P. 435–451.
13. **Rudemo M.** Empirical choice of histogram and kernel density estimators // Scandinavian Journ. Statist. 1982. N 9. P. 65–78.
14. **Bowman A. W.** A comparative study of some kernel-based non-parametric density estimators // Journ. Statist. Comput. Simulation. 1982. **21**. P. 313–327.
15. **Hall P.** Large-sample optimality of least squares cross-validation in density estimation // Ann. Statist. 1983. **11**, Iss. 4. P. 1156–1174.
16. **Jiang M., Provost S. B.** A hybrid bandwidth selection methodology for kernel density estimation // Journ. Statist. Comput. Simulation. 2014. **84**, № 3. P. 614–627.
17. **Dutta S.** Cross-validation revisited // Communications in Statistics — Simulation and Computation. 2016. **45**, N 2. P. 472–490.

18. **Heidenreich N.-B., Schindler A., Sperlich S.** Bandwidth selection for kernel density estimation: A review of fully automatic selectors // AStA Advances in Statistical Analysis. 2013. **97**, N 4. P. 403–433.
19. **Li Q., Racine J. S.** Nonparametric Econometrics: Theory and Practice. Princeton: Princeton University Press, 2007. 768 p.
20. **Лапко А. В., Лапко В. А.** Методика быстрого выбора коэффициентов размытости в непараметрическом классификаторе, соответствующем критерию максимума апостериорной вероятности // Автометрия. 2019. **55**, № 6. С. 76–86. DOI: 10.15372/AUT20190610.
21. **Scott D. W.** Multivariate Density Estimation: Theory, Practice, and Visualization. New Jersey: John Wiley & Sons, 2015. 384 p.
22. **Sheather S. J.** Density estimation // Statist. Science. 2004. **19**, N 4. P. 588–597. DOI: 10.1214/088342304000000297.
23. **Silverman B. W.** Density Estimation for Statistics and Data Analysis. London: Chapman and Hall, 1986. 175 p.
24. **Шаракшанэ А. С., Железнов И. Г., Ивницкий В. А.** Сложные системы. М.: Высш. шк., 1977. 248 с.

Поступила в редакцию 17.12.2020
После доработки 04.03.2021
Принята к публикации 04.03.2021