

УДК 519.7 + 004.93

Doi: 10.31772/2712-8970-2025-26-1-48-59

Для цитирования: Шаруева А. В. Непараметрическая методика проверки гипотезы о независимости случайных величин и ее применение при анализе данных дистанционного зондирования // Сибирский аэрокосмический журнал. 2025. Т. 26, № 1. С. 48–59. Doi: 10.31772/2712-8970-2025-26-1-48-59.

For citation: Sharueva A. V. [Nonparametric method for testing the hypothesis of independence of random variables and its application in the analysis of remote sensing data]. *Siberian Aerospace Journal*. 2025, Vol. 26, No. 1, P. 48–59. Doi: 10.31772/2712-8970-2025-26-1-48-59.

Непараметрическая методика проверки гипотезы о независимости случайных величин и ее применение при анализе данных дистанционного зондирования

А. В. Шаруева

Сибирский государственный университет науки и технологий имени академика М. Ф. Решетнева
Российская Федерация, 660037, г. Красноярск, просп. им. газ. «Красноярский рабочий», 31
E-mail: anna-denisyuk@yandex.ru

Аннотация. Проверка гипотезы о независимости случайных величин является одним из основных этапов системного анализа статистических данных. На её результатах осуществляется синтез эффективных алгоритмов принятия решений. Традиционная методика проверки гипотезы о независимости случайных величин основана на использовании критерия Пирсона и содержит трудно формализуемый этап разбиения области значений случайных величин на многомерные интервалы. Предложена методика проверки гипотезы о независимости случайных величин, которая использует непараметрический алгоритм распознавания образов, соответствующий критерию максимального правдоподобия. Её применение позволяет обойти проблему декомпозиции области значений случайных величин на интервалы. Идея подхода состоит в формировании по исходным статистическим данным обучающей выборки для решения двухальтернативной задачи распознавания образов. Каждый класс определяется в предположении независимости либо зависимости случайных величин, что проявляется в различии их законов распределения в классах. В этих условиях появляется возможность замены исходной гипотезы на задачу проверки достоверности отличия вероятностей ошибок распознавания образов в классах. С использованием аппарата теории графов предлагаемая методика развита при формировании наборов независимых случайных величин. Полученные результаты обобщены при проверке гипотезы о независимости случайных величин для больших объёмов статистических данных на основе сжатия исходной информации. Это позволяет на порядки повысить вычислительную эффективность решаемой задачи. В статье обосновывается методика проверки гипотезы о независимости случайных величин, основанная на использовании непараметрического алгоритма распознавания образов в условиях больших объёмов статистических данных. Приводятся результаты сравнения методики с общепризнанным критерием согласия Пирсона при исследовании неоднозначных зависимостей между случайными величинами различной сложности. Эффективность предлагаемой методики подтверждается результатами применения при обработке информации дистанционного зондирования антропогенных территорий в окрестности города Красноярска.

Ключевые слова: проверка гипотезы о независимости случайных величин, ядерная оценка плотности вероятности, регрессионная оценка плотности вероятности, распознавание образов, критерий Пирсона, дистанционное зондирование.

Nonparametric method for testing the hypothesis of independence of random variables and its application in the analysis of remote sensing data

A. V. Sharueva

Reshetnev Siberian State University of Science and Technology
31, Krasnoyarskii rabochii prospekt, Krasnoyarsk, 660037, Russian Federation
E-mail: anna-denisyuk@yandex.ru

Abstract. Testing the hypothesis of independence of random variables is one of the main stages of system analysis of statistical data. Based on its results, a synthesis of effective decision-making algorithms is carried out. The traditional method of testing the hypothesis of independence of random variables is based on the use of the Pearson criterion, which contains a difficult to formalize stage of dividing the range of values of random variables into multidimensional intervals. A method for testing the hypothesis of independence of random variables is proposed, which uses a nonparametric pattern recognition algorithm corresponding to the maximum likelihood criterion. Its application makes it possible to circumvent the problem of decomposing the range of values of random variables into intervals. The idea of the approach is to form a training sample based on the initial statistical data to solve a two-alternative pattern recognition problem. Each class is defined under the assumption of independence or dependence of random variables, which is manifested in the difference in their distribution laws in the classes. Under these conditions, it becomes possible to replace the initial hypothesis with the task of checking the reliability of the difference in the probabilities of pattern recognition errors in classes. Using the apparatus of graph theory, the proposed method is developed in the formation of sets of independent random variables. The obtained results are generalized when testing the hypothesis of independence of random variables for large volumes of statistical data based on compression of the original information. This allows to increase the computational efficiency of the problem being solved. The article substantiates a method for testing the hypothesis of independence of random variables, based on the use of a nonparametric pattern recognition algorithm in conditions of large volumes of statistical data. The results of comparing the technique with the generally recognized Pearson consensus criterion in the study of ambiguous dependencies between random variables of varying complexity are presented. The effectiveness of the proposed method is confirmed by the results of its application in processing remote sensing information from anthropogenic territories in the vicinity of the city of Krasnoyarsk.

Keywords: testing the hypothesis of independence of random variables, kernel probability density estimation, regression probability density estimation, pattern recognition, Pearson criterion, remote sensing.

Введение

Универсальным и общепризнанным критерием проверки гипотез о распределениях случайных величин, включая их независимость, является критерий Пирсона [1]. При его использовании необходимо решать задачи разбиения области значений случайных величин на многомерные интервалы и устанавливать закон распределения критерия, определяющего зависимости между вероятностными характеристиками случайных величин. В работах [2–4] предложен новый подход, позволяющий упростить проверку гипотезы о независимости случайных величин с применением непараметрического алгоритма распознавания образов ядерного типа, соответствующего критерию максимального правдоподобия. Идея подхода состоит в решении двухальтернативной задачи распознавания образов. Рассматриваемые классы определяются предположениями о зависимости и независимости случайных величин. На этой основе формируется обучающая выборка по исходным статистическим данным о наблюдениях случайных величин и решается задача распознавания образов. Соотношение между оценками вероятностей ошибок распознавания введённых классов подтверждает либо опровергает рассматриваемую гипотезу.

Цель данной работы состоит в обобщении и развитии непараметрического метода проверки гипотезы о независимости случайных величин для условий большого объёма статистических данных и его применении при анализе информации о дистанционном зондировании антропогенных территорий.

Методика проверки гипотезы о независимости случайных величин

Пусть имеется выборка $V = (x^i, i = \overline{1, n})$ объёма n , составленная из независимых наблюдений двухмерной случайной величины $x = (x_1, x_2)$. Предположим, что выборка V извлекается из генеральных совокупностей, характеризующихся плотностями вероятностей $p(x_1)p(x_2)$ или $p(x_1, x_2)$. Необходимо по статистическим данным V проверить гипотезу

$$H_0: p(x_1, x_2) \equiv p(x_1)p(x_2)$$

о независимости случайных величин x_1, x_2 .

Для проверки гипотезы H_0 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

$$m(x): \begin{cases} x \in \Omega_1, \text{ если } p(x_1, x_2) < p(x_1)p(x_2), \\ x \in \Omega_2, \text{ если } p(x_1, x_2) > p(x_1)p(x_2). \end{cases}$$

В отличие от традиционной постановки задачи распознавания образов при синтезе решающего правила $m(x)$ 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 восстановить плотности вероятностей $p(x_1, x_2), p(x_1)p(x_2)$, используя их непараметрические оценки типа Розенблатта – Парзена [5; 6],

$$\begin{aligned} \bar{p}(x_1, x_2) &= \frac{1}{n c_1 c_2} \sum_{i=1}^n \Phi\left(\frac{x_1 - x_1^i}{c_1}\right) \Phi\left(\frac{x_2 - x_2^i}{c_2}\right), \\ \bar{p}(x_1) \bar{p}(x_2) &= \frac{1}{n^2 c_1 c_2} \sum_{i=1}^n \sum_{j=1}^n \Phi\left(\frac{x_1 - x_1^i}{c_1}\right) \Phi\left(\frac{x_2 - x_2^j}{c_2}\right). \end{aligned}$$

В статистиках $\bar{p}(x_1, x_2), \bar{p}(x_1)\bar{p}(x_2)$ ядерные функции $\Phi(u_v)$ удовлетворяют условиям положительности, симметричности и нормированности.

Значения коэффициентов размытости $c_v, v = 1, 2$ ядерных функций убывают с ростом объёма n выборки статистических данных V . Тогда непараметрическое решающее правило классификации случайных величин $x = (x_1, x_2)$ запишется как

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

Оптимальные коэффициенты размытости ядерных функций решающего правила $\bar{m}(x)$ выбираются на основе анализа аппроксимационных свойств непараметрических оценок плотно-

стей вероятностей $\bar{p}(x_1, x_2)$, $\bar{p}(x_1)$, $\bar{p}(x_2)$ из условия минимума, соответствующих им оценок среднеквадратических отклонений от $p(x_1, x_2)$, $p(x_1)$, $p(x_2)$. Например, для $\bar{p}(x_1)$ подобным критерием является [7–11]

$$\int_{-\infty}^{\infty} \bar{p}^2(x_1) dx_1 - \frac{2}{n} \sum_{j=1}^n \bar{p}(x_1^j).$$

Определим оценки вероятностей ошибок распознавания образов $\bar{\rho}_1(\bar{c}(1))$, $\bar{\rho}_2(\bar{c}(2))$ решающим правилом $\bar{m}(x)$ 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 классу Ω_t ,

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

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

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

– «решение» алгоритма $\bar{m}(x)$ 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 $\bar{m}(x)$, исключается из процесса формирования статистик $\bar{p}(x_1, x_2)$, $\bar{p}(x_1)$, $\bar{p}(x_2)$.

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

Тогда 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.

При ограниченном объёме n выборки V возникает задача доверительного оценивания вероятностей ошибок распознавания образов. Для её решения используется традиционная методика доверительного оценивания вероятностей либо критерий Колмогорова – Смирнова.

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

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

Здесь β – вероятность (риск) отвергнуть гипотезу $\bar{H}_0: \rho_1 = \rho_2$. 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.

Формирование наборов независимых случайных величин

Имеется выборка наблюдений $V = (x_v^i, v = \overline{1, k}, i = \overline{1, n})$ объёма n , составленная из статистически независимых наблюдений компонент многомерной случайной величины $x = (x_v, v = \overline{1, k})$. Вид плотности вероятности $p(x)$ априори неизвестен. Необходимо по статистическим данным V , используя предложенный выше критерий проверки гипотез [13–16]

$$H_{vj}: p(x_v, x_j) \equiv p(x_v) p(x_j)$$

для компонент $x_v, v = \overline{1, k}, x_j, j = \overline{1, k}, v > j$, сформировать наборы независимых случайных величин $x(t) = (x_v, v \in I_t), t = \overline{1, m}$. Количество m наборов компонент случайной величины x неизвестно, а I_t – множество номеров компонент, составляющих набор $x(t)$.

Предлагаемая методика основана на выполнении следующих действий:

1. В соответствии с приведёнными выше рекомендациями проверить гипотезы H_{vj} для каждой пары компонент (x_v, x_j) многомерной случайной величины $x = (x_v, v = \overline{1, k})$. Количество таких пар соответствует значению $k(k-1)/2$.

2. По результатам этапа 1 построить информационный граф $G(X, A)$, где X – множество его вершин, соответствующих компонентам случайной величины x , а A – множество ребер. Между двумя вершинами x_v, x_j имеется ребро, если выполняется гипотеза H_{vj} , т. е. компоненты x_v, x_j являются независимыми.

3. Провести анализ информационного графа $G(X, A)$ и определить его полные подграфы $G(X_t, A_t), t = \overline{1, m}$. Каждая пара вершин подграфа $G(X_t, A_t)$ имеет ребро, если компоненты случайной величины x являются независимыми. Обнаружить полные подграфы с использованием алгоритмов разрезания исходного графа, которые основаны на анализе его матрицы смежности. Компоненты $x_v, v \in I_t$, соответствующие вершинам полного подграфа $G(X_t, A_t)$, образуют набор независимых случайных величин.

Модификация методики проверки гипотезы о независимости случайных величин в условиях больших объёмов статистических данных

При больших объёмах n статистических данных $V = (x_1^i, x_2^i, i = \overline{1, n})$ в предложенной методике используются регрессионные оценки плотностей вероятностей $\bar{p}(x_1, x_2), \bar{p}(x_1), \bar{p}(x_2)$. Эти оценки основаны на сжатии исходной информации, например, $V_1 = (x_1^i, i = \overline{1, n})$ в массив данных $\bar{V}_1 = (\bar{p}_1^j, z^j, j = \overline{1, N})$ путём декомпозиции области значений x_1 на N интервалов. Здесь z^j – центры интервалов дискретизации значений x_1 , а $\bar{p}_1^j = \bar{P}_1^j / \Delta$ – оценка плотности вероятности в j -м интервале; Δ – длина интервала дискретизации; $\bar{P}_1^j$ – частота встречаемости значений x_1^i из выборки V_1 в интервале под номером j . Тогда регрессионная оценка плотности вероятности $p(x_1)$ по данным $\bar{V}_1$ имеет вид [17; 18]

$$\bar{p}(x_1) = \frac{1}{c_1} \sum_{j=1}^N \bar{P}_1^j \Phi\left(\frac{x_1 - z^j}{c_1}\right).$$

Предлагаемый подход позволяет на порядки сократить объём n исходной статистической информации при оценивании плотностей вероятностей. Особенность статистики типа $\bar{p}(x_1)$ позволяет значительно упростить выбор коэффициентов размытости c ядерных функций в статистике $\bar{p}(x_1)$ из условия минимума критерия

$$\frac{1}{N} \sum_{i=1}^N (\bar{p}_1^i - \bar{p}(x_1^i)).$$

По аналогии осуществляется оценивание плотностей вероятностей $p(x_2)$, $p(x_1, x_2)$. Регрессионные оценки плотностей вероятностей используются при проверке гипотезы о независимости случайных величин в соответствии с предложенной методикой.

Анализ результатов вычислительного эксперимента

Проведено сравнение эффективности предлагаемой методики проверки гипотезы о независимости двумерных случайных величин и критерия Пирсона в условиях неоднозначных зависимостей при различных объёмах статистических данных [19–21]. Датчики случайных величин x_1 , x_2 формировались на основе равномерного закона распределения x_1 , который использовался при вычислении значений x_2 в виде нелинейных преобразований x_1 . При этом на значения x_2 накладывались помехи с нормальным законом распределения, который имеет нулевое математическое ожидание и среднеквадратическое отклонение σ . Пример значений случайных величин x_1 и x_2 приведён на рис. 1.

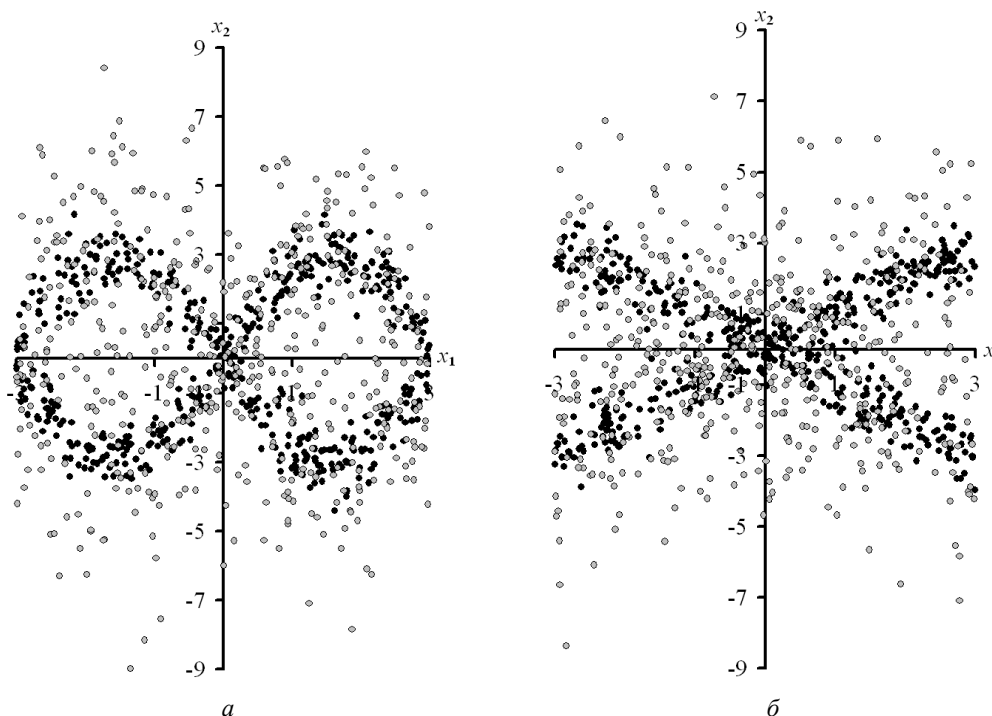


Рис. 1. Значения случайных величин x_1 , x_2 из выборки исходных статистических данных V при $n = 500$ и $\sigma = 0,5$ (темные точки), а при $\sigma = 2$ (серые точки) при использовании зависимостей различной сложности

Fig. 1. Values x_1 , x_2 of random variables from a sample of initial statistical data V at $n = 500$ and $\sigma = 0.5$ (dark dots), and at $\sigma = 2$ (gray dots) when using dependencies of varying complexity

При проверке гипотезы о независимости компонент двухмерной случайной величины на основе критерия Пирсона используются результаты оптимального выбора количества интервалов дискретизации [22–24]

$$N^* = \left(\frac{3}{4} \Delta_1 \Delta_2 \|p(x_1, x_2)\|^2 n \right)^{1/2}.$$

Значение $\|p(x_1, x_2)\|^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p^2(x_1, x_2) dx_1 dx_2$, а Δ_v – длина интервала значений случайной

величины x_v , $v = 1, 2$. Традиционным формулам дискретизации области значений случайных величин посвящены работы [25–27].

По результатам вычислительного эксперимента предлагаемая методика и критерий Пирсона при анализе неоднозначных зависимостей между случайными величинами в условиях относительно малых объёмов статистических данных и средних квадратических отклонений σ помех сопоставимы и безошибочно определяют зависимость случайных величин. Данный вывод не соблюдается при зависимости между случайными величинами (рис. 1, а), когда критерий Пирсона не устанавливает зависимость в условиях $n = 100$ и $\sigma \in [0.5; 2]$. С увеличением σ эффективность сравниваемых критериев снижается. Этот факт объясняется особенностями неоднозначных зависимостей и большими значениями σ , когда область определения случайных величин скрывает искомую зависимость. С увеличением объёма n исходных данных эффективность сравниваемых критериев проверки гипотезы о независимости случайных величин повышается. Этот вывод является ожидаемым, так как с ростом n повышаются асимптотические свойства непараметрических оценок плотностей вероятностей и частот встречаемости случайных величин в их двухмерных интервалах. Преимущество предлагаемой методики проверки гипотезы о независимости случайных величин наблюдается при малых значениях σ , ограниченных и больших n . При больших n и σ часто обнаруживается преимущество критерия Пирсона, если используется процедура оптимальной дискретизации области значений двухмерной случайной величины [22].

Применение предлагаемой методики при анализе данных дистанционного зондирования

Разработанная методика апробирована при анализе данных дистанционного зондирования [2; 28]. Объектом исследования являются антропогенные территории (карьер, пригородная застройка) в окрестности города Красноярск. Исходная информация формировалась по фрагментам съёмки спутника Sentinel-2 на 26.08.2021 (рис. 2). Использовались спектральные каналы x_j , $j = \overline{1, 9}$, которые характеризуются длинами волн (нанометры): x_1 – (458–523), x_2 – (543–578), x_3 – (650–680), x_4 – (698–713), x_5 – (733–748), x_6 – (773–793), x_7 – (785–899), x_8 – (1565–1655), x_9 – (2100–2280).

Предлагаемая методика позволяет формировать пары независимых и зависимых случайных величин, изменяя соотношение между их параметрами. Применение методики позволило обнаружить 31 и 29 пар спектральных признаков с сильной линейной зависимостью соответственно для объектов «карьер», «пригородная застройка». Полученные результаты представлены на рис. 3.

Дополнительно обнаружены нелинейные зависимости между спектральными признаками для объекта «карьер»

$$(x_1, x_9), (x_1, x_8), (x_1, x_7), (x_1, x_5), (x_1, x_4)$$

и объекта «пригородная застройка»

$$(x_7, x_9), (x_4, x_9), (x_3, x_9), (x_2, x_9), (x_1, x_9), (x_1, x_8), (x_1, x_7).$$

Полученные результаты являются достоверными для всех пар спектральных признаков, так как соблюдается условие $|\bar{\rho}_1 - \bar{\rho}_2| > D_\beta$ при $D_\beta = 0,029$ и риске $\beta = 0,025$ отвергнуть гипотезу H_0 равенства значений ρ_1, ρ_2 .

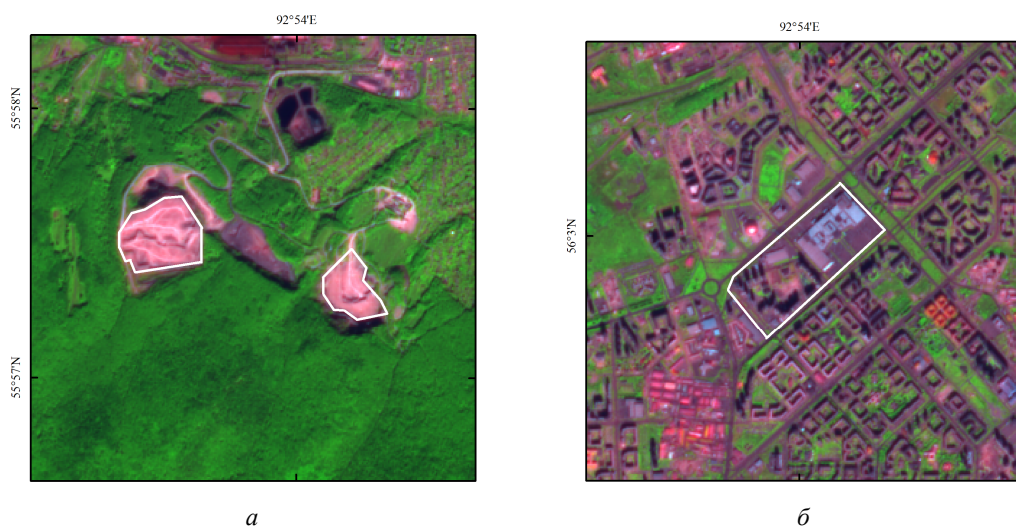


Рис. 2. Фрагменты спутниковой съемки Sentinel-2. Антропогенные территории:
а – карьер; б – пригородная застройка

Fig. 2. Fragments of Sentinel-2 satellite imagery. Anthropogenic territories:
а – quarry; б – suburban development

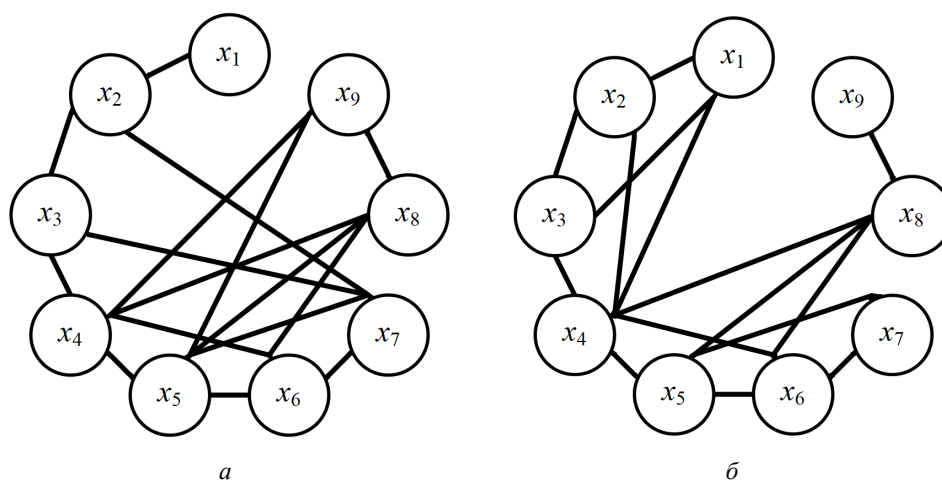


Рис. 3. Иллюстрация сильной линейной зависимости между парами спектральных признаков (x_i, x_j) , характеризующихся оценками коэффициентов корреляции больше 0,9:
а – карьер; б – пригородная застройка

Fig. 3. Illustration of a strong linear relationship between pairs of spectral features (x_i, x_j) characterized by correlation coefficient estimates greater than 0,9:
а – quarry; б – suburban development

Рассмотрена задача обнаружения антропогенных территорий по спектральным данным. Ошибка их распознавания в пространстве спектральных признаков $x = (x_j, j = \overline{1, 9})$ по обучающей выборке $V = (x^i, \sigma(i), i = \overline{1, n})$ равна 0,012, где $n = n_1 + n_2$, $n_1 = 3377$ («карьер», $\sigma(i) = 1$),

$n_2 = 5049$ («пригородная застройка», $\sigma(i) = 2$). При исключении из обучающей выборки, например, спектральных признаков (x_4, x_5) , (x_5, x_6) , (x_4, x_5, x_6) , оценки ошибок распознавания образов соответствуют значениям 0,011; 0,01; 0,008. Полученное снижение ошибок распознавания образов не является достоверным по сравнению с оценкой ошибки в пространстве признаков x_j , $j = \overline{1, 9}$. Однако полученный результат обосновывает возможность сокращения спектральных признаков при синтезе алгоритмов принятия решений и упрощения их оптимизации.

Заключение

Методика проверки гипотезы о независимости пар случайных величин, основанная на использовании непараметрического алгоритма распознавания образов, позволяет обойти проблему дискретизации области значений случайных величин на многомерные интервалы. Эта проблема свойственна общепризнанному критерию Пирсона. Определены условия компетентности предлагаемого метода и критерия Пирсона при анализе однозначных и неоднозначных зависимостей между случайными величинами. С использованием аппарата теории графов предлагаемая методика развита при формировании наборов независимых случайных величин. Полученные результаты обобщены при проверке гипотезы о независимости случайных величин для больших объёмов статистических данных на основе сжатия исходной информации, что позволяет на порядки повысить вычислительную эффективность решаемых задач. Эффективность предложенной методики подтверждена при анализе данных дистанционного зондирования антропогенных территорий и оценивании их состояний. При наличии набора спектральных признаков, характеризующихся сильной линейной зависимостью между его парами, появляется возможность сокращения количества спектральных признаков при распознавании антропогенных территорий с уменьшением оценки вероятности ошибки их распознавания.

Библиографические ссылки

1. Академик Пугачёв Владимир Семёнович: к столетию со дня рождения / под ред. И. Н. Синицына. М. : Торус Пресс, 2011. 376 с.
2. Шаруева А. В., Лапко А. В., Лапко В. А. Непараметрические методы проверки гипотез о распределениях случайных величин при анализе данных дистанционного зондирования. Новосибирск : СО РАН, 2024. 189 с.
3. Лапко А. В., Лапко В. А. Проверка гипотезы о независимости двумерных случайных величин с использованием непараметрического алгоритма распознавания образов // Автометрия. 2021. Т. 57, № 2. С. 41–48. DOI: 10.15372/AUT20210205.
4. Лапко А. В., Лапко В. А., Бахтина А. В. Исследование методики проверки гипотезы о независимости двумерных случайных величин с использованием непараметрического классификатора // Автометрия. 2021. Т. 57, № 6. С. 90–100. DOI: 10.15372/AUT20210610.
5. Parzen E. On estimation of a probability density function and mode // Annals of Mathematical Statistics. 1962. Vol. 33, No. 3. P. 1065–1076. DOI: 10.1214/aoms/1177704472.
6. Епанечников В. А. Непараметрическая оценка многомерной плотности вероятности // Теория вероятности и ее применения. 1969. Т. 14, №1. С. 156–161.
7. Лапко А. В., Лапко В. А. Анализ методов оптимизации непараметрической оценки плотности вероятности по коэффициенту размытости ядерных функций // Измерительная техника. 2017. № 6. С. 3–8.
8. Лапко А. В., Лапко В. А. Ядерные оценки плотности вероятности и их применение. Красноярск : СибГУ им. М. Ф. Решетнёва, 2021. 308 с.
9. Rudemo M. Empirical choice of histogram and kernel density estimators // Scandinavian Journal of Statistics. 1982. No. 9. P. 65–78.

10. Bowman A. W. A comparative study of some kernel-based non-parametric density estimators // *Journal of Statistical Computation and Simulation*. 1982. Vol. 21. P. 313–327.
11. Hall P. Large-sample optimality of least squares cross-validation in density estimation // *Annals of Statistics*. 1983. Vol. 11. P. 1156–1174.
12. Шараханэ А. С., Железнов И. Г., Ивницкий В. А. Сложные системы. М. : Высш. шк., 1977. 248 с.
13. Лапко А. В., Лапко В. А., Бахтина А. В. Формирование наборов независимых компонент многомерной случайной величины на основе непараметрического алгоритма распознавания образов // *Измерительная техника*. 2021. № 9. С. 3–9. DOI: 10.32446/0368-1025it.2021-9-3-9.
14. Непараметрический алгоритм распознавания образов в задаче проверки гипотезы о независимости случайных величин / И. В. Зеньков, А. В. Лапко, В. А. Лапко и др. // *Компьютерная оптика*. 2021. Т. 45, № 5. С. 767–772. DOI: 10.18287/2412-6179-CO-871.
15. Методика последовательного формирования набора компонент многомерной случайной величины с использованием непараметрического алгоритма распознавания образов / И. В. Зеньков, А. В. Лапко, В. А. Лапко и др. // *Компьютерная оптика*. 2021. Т. 45. № 6. С. 926–933. DOI: 10.18287/2412-6179-CO-902.
16. Лапко А. В., Лапко В. А., Шаруева А. В. Непараметрический алгоритм распознавания образов в задаче формирования наборов независимых случайных величин // *Информатика и системы управления*. 2024. Т. 79, №1. С. 81–90. DOI: 10.22250/18142400_2024_79_1_81.
17. Лапко А. В., Лапко В. А. Регрессионная оценка плотности вероятности и её свойства // *Системы управления и информационные технологии*. 2012. № 3-1 (49). С. 152–156.
18. Лапко А. В., Лапко В. А. Регрессионная оценка многомерной плотности вероятности и её свойства // *Автометрия*. 2014. Т. 50, № 2. С. 50–56.
19. Лапко А. В., Лапко В. А., Бахтина А. В. Применение непараметрического алгоритма распознавания образов в задаче проверки гипотезы о независимости переменных неоднозначных функций // *Измерительная техника*. 2022. № 1. С. 17–22. DOI: 10.32446/0368-1025it.2022-01-17-22.
20. Лапко А. В., Лапко В. А., Бахтина А. В. Сравнение методики проверки гипотезы о независимости двумерных случайных величин, основанной на непараметрическом классификаторе // *Искусственный интеллект и принятие решений*. 2022. № 1. С. 45–56. DOI: 10.14357/20718594220105.
21. Лапко А. В., Лапко В. А., Бахтина А. В. Сравнение методик проверки гипотезы о независимости случайных величин, основанных на непараметрическом классификаторе и критерии Пирсона // *Автометрия*. 2023. Т. 59, № 5. С. 36–46.
22. Лапко А. В., Лапко В. А. Выбор оптимального количества интервалов дискретизации области значений двумерной случайной величины // *Измерительная техника*. 2016. № 2. С. 14–17.
23. Лапко А. В., Лапко В. А. Метод дискретизации области значений многомерной случайной величины // *Измерительная техника*. 2019. № 1. С. 16–20. DOI: 10.32446/0368-1025it.2019-1-16-20.
24. Лапко А. В., Лапко В. А. Оценивание параметров формулы оптимальной дискретизации области значений двумерной случайной величины // *Измерительная техника*. 2018. № 5. С. 9–13.
25. Sturges H. A. The choice of a class interval // *Journal of the American Statistical Association*. 1926. Vol. 21, No. 153. P. 65–66.
26. Heinhold I., Gaede K. *Ingenieur statistik*. München: Springer Verlag, 1964. 352 p.
27. Scott D.W. *Multivariate Density Estimation: Theory, Practice, and Visualization*. New Jersey: John Wiley & Sons, 2015. 384 p.
28. Лапко А. В., Лапко В. А., Бахтина А. В. Применение непараметрической методики проверки гипотезы о независимости случайных величин в условиях большого объёма статистических данных // *Измерительная техника*. 2023. № 10. С. 17–24. DOI: 10.32446/0368-1025it.2023-10-17-24.

References

1. Sinitsyna I. N. *Akademik Pugachev Vladimir Semenovich: k stoletiyu so dnya rozhdeniya* [Academician Vladimir Semenovich Pugachev: on the centenary of his birth]. Moscow, Torus Press Publ., 2011, 376 p.
2. Sharueva A. V., Lapko A. V., Lapko V. A. *Neparametricheskiye metody proverki gipotez o raspredeleniyakh sluchaynykh velichin pri analize dannykh distantsionnogo zondirovaniya* [Nonparametric methods for testing hypotheses about distributions of random variables in the analysis of remote sensing data]. Novosibirsk, SO RAN Publ., 2024, 189 p.
3. Lapko A. V., Lapko V. A. Testing the Hypothesis of the Independence of Two-Dimensional Random Variables Using a Nonparametric Algorithm for Pattern Recognition. *Optoelectronics, Instrumentation and Data Processing*. 2021, Vol. 57, No. 2, P. 149–155.
4. Lapko A. V., Lapko V. A., Bakhtina A. V. Study of the Method for Verification of the Hypothesis on Independence of Two-Dimensional Random Quantities Using a Nonparametric Classifier. *Optoelectronics, Instrumentation and Data Processing*. 2022, Vol. 57, No. 6, P. 639–648.
5. Parzen E. On estimation of a probability density function and mode. *Annals of Mathematical Statistics*. 1962, Vol. 33, No. 3, P. 1065–1076.
6. Epanechnikov V. A. [Non-parametric estimation of a multivariate probability density]. *Theory of Probability & Its Applications*. 1969, Vol. 14, No. 1, P. 156–161 (In Russ.).
7. Lapko A. V., Lapko V. A. Analysis of optimization methods for nonparametric estimation of the probability density with respect to the blur factor of kernel functions. *Measurement Techniques*. 2017, Vol. 60, No. 6, P. 515–522.
8. Lapko A. V., Lapko V. A. *Yadernye otsenki plotnosti veroyatnosti i ikh primeneniye* [Kernel probability density estimates and their applications]. Krasnoyarsk, SibGU im. M.F. Reshetnev Publ., 2021, 208 p.
9. Rudemo M. Empirical choice of histogram and kernel density estimators. *Scandinavian Journal of Statistics*. 1982, No. 9, P. 65–78.
10. Bowman A. W. A comparative study of some kernel-based non-parametric density estimators. *Journal of Statistical Computation and Simulation*. 1982, Vol. 21, P. 313–327.
11. Hall P. Large-sample optimality of least squares cross-validation in density estimation. *Annals of Statistics*. 1983, Vol. 11, P. 1156–1174.
12. Sharakhane, A. S., Zheleznov I. G., Ivnitiskii V. A. *Slozhnye sistemy* [Complex systems]. Moscow, Vysshaya Shkola Publ., 1977, 248 p.
13. Lapko A. V., Lapko V. A., Bakhtina A. V. Formation of Sets of Independent Components of a Multidimensional Random Variable Based on a Nonparametric Pattern Recognition Algorithm. *Measurement Techniques*. 2021, Vol. 64, No. 9, P. 689–696.
14. Zenkov I. V., Lapko A. V., Lapko V. A., Kiryushina E. V., Vokin V. N. Nonparametric pattern recognition algorithm for testing a hypothesis of the independence of random variables. *Computer Optics*. 2021, Vol. 45, No 5, P. 767–772.
15. Zenkov I. V., Lapko A. V., Lapko V. A., Kiryushina E. V., Vokin V. N., Bakhtina A. V. A method of sequentially generating a set of components of a multidimensional random variable using a nonparametric pattern recognition algorithm. *Computer Optics*. 2021, Vol. 45, No. 6, P. 926–933.
16. Lapko A. V., Lapko V. A., Sharueva A. V. *Neparametricheskii algoritim raspoznavaniya obrazov v zadache formirovaniya naborov nezavisimyykh sluchainyykh velichin* [A nonparametric pattern recognition algorithm in the problem of forming sets of independent random variables]. *Informatika i sistemy upravleniya*. 2024, Vol. 79, No. 1, P. 81–90 (In Russ.).
17. Lapko A. V., Lapko V. A. Regressionnaya otsenka plotnosti veroyatnosti i ee svoistva [Regression estimation of probability density and its properties]. *Sistemy upravleniya i informatsionnye tekhnologii*. 2012, No. 3-1 (49), P. 152–156 (In Russ.).
18. Lapko A. V., Lapko V. A. Regression estimate of the multidimensional probability density and its properties. *Optoelectronics, Instrumentation and Data Processing*. 2014, Vol. 50, No. 2, P. 148–153.

19. Lapko A. V., Lapko V. A., Bakhtina A. V. Application of a nonparametric pattern recognition algorithm to the problem of testing the hypothesis of the independence of variables of multi-valued functions. *Measurement Techniques*. 2022, Vol. 65, No. 1, P. 17–23.
20. Lapko A. V., Lapko V. A., Bakhtina A. V. Comparison of the Methodology for Hypothesis Testing of the Independence of Two-Dimensional Random Variables Based on a Nonparametric Classifier. *Scientific and Technical Information Processing*. 2023, Vol. 50, No. 6, P. 572–581.
21. Lapko A. V., Lapko V. A., Bakhtina A. V. Comparison of Methods for Testing the Hypothesis of Independence of Random Variables Based on a Nonparametric Classifier and Pearson's Chi-Squared Test. *Optoelectronics, Instrumentation and Data Processing*, 2023, Vol. 59, No. 5, P. 551–560.
22. Lapko A. V., Lapko V. A. Selection of the Optimal Number of Intervals Sampling the Region of Values of a Two-Dimensional Random Variable. *Measurement Techniques*. 2016, Vol. 59, No. 2, P. 122–126.
23. Lapko A. V., Lapko V. A. Discretization method for the range of values of a multi-dimensional random variable. *Measurement Techniques*. 2019, Vol. 62, No. 1, P. 16–22.
24. Lapko A. V., Lapko V. A. Estimation of parameters of the formula for optimal discretization of the range of values of a two-dimensional random variable. *Measurement Techniques*. 2018, Vol. 61, No. 5, P. 427–433.
25. Sturges H. A. The choice of a class interval. *Journal of the American Statistical Association*, 1926, Vol. 21, P. 65–66.
26. Heinhold I., Gaede K. W. *Ingenieur statistic*. München, Springer Verlag, 1964, 352 p.
27. Scott D. W. *Multivariate Density Estimation: Theory, Practice, and Visualization*. New Jersey, John Wiley & Sons, 2015, 384 p.
28. Lapko A. V., Lapko V. A., Bakhtina A. V. Application of a nonparametric procedure for testing the hypothesis about the independence of random variables given a large amount of statistical data. *Measurement Techniques*, 2024, Vol. 66, P. 744–754.

© Шаруева А. В., 2025

Шаруева Анна Владимировна – заведующий лабораторией дистанционного зондирования, ассистент; Сибирский государственный университет науки и технологий имени академика М. Ф. Решетнева. E-mail: anna-denisyuk@yandex.ru. <https://orcid.org/0009-0003-4255-4554>.

Sharueva Anna Vladimirovna – head of the remote sensing laboratory, assistant; Reshetnev Siberian State University of Science and Technology. E-mail: anna-denisyuk@yandex.ru. <https://orcid.org/0009-0003-4255-4554>.

Статья поступила в редакцию 27.12.2024; принята к публикации 13.01.2025; опубликована 11.04.2025
The article was submitted 27.12.2024; accepted for publication 13.01.2025; published 11.04.2025

Статья доступна по лицензии Creative Commons Attribution 4.0
The article can be used under the Creative Commons Attribution 4.0 License