УДК 519.2

С. В. Заболотній
Черкаський державний технологічний університет,
кафедра радіотехніки

МЕТОД ПЕРЕВІРКИ СТАТИСТИЧНОЇ ГІПОТЕЗИ НА ОСНОВІ РОЗКЛАДУ ВИПАДКОВИХ ВЕЛИЧИН У ПРОСТОРІ З ПОРІДНИМ ЕЛЕМЕНТОМ

Викладено новий метод перевірки статистичних гіпотез, який ґрунтується на апараті розкладу випадкових величин у просторі з порідним елементом (просторі Кунченко). Здійснено синтез вирішного правила та отримано аналітичні вирази, які дають змогу розраховувати ймовірності похибок прийняття рішень.

The new method of verification of statistical hypotheses, which is based on the apparatus of decomposition of random values in space with a generating element is expounded in work (space of Kunchenko). The synthesis of solution is carried out governed and analytical expressions which allow to calculate probabilities of errors of making a decision are got.

Вступ. Теорія перевірки статистичних гіпотез, основи якої заклали Пірсон та Нейман майже століття тому, є основним математичним базисом, що застосовується для синтезу методів та алгоритмів, призначених для розв'язання задач виявлення і розпізнавання сигналів та образів. Традиційно ця проблема є актуальною для таких високотехнологічних галузей, як автоматичне керування, радіотехніка та телекомунікація. Не менш важливою прикладною сферою застосування цього математичного апарату є розроблення методик статистичного аналізу економічних, біомедичних та соціальних даних.

Нині існує значний доробок у цьому напрямі та велика кількість різноманітних методів перевірки статистичних гіпотез. Аналіз ситуації вказує на те, що найпоширенішим є так званий байєсовий підхід, відповідно до якого синтез статистичних критеріїв узгодження та правил прийняття рішень ґрунтується на формуванні відношення правдоподібності. Такий підхід потребує повного опису статистичних даних у вигляді щільності розподілу ймовірностей та в загальному випадку вимагає достатньо складних аналітичних викладок, необхідних для визначення якісних характеристик вирішних правил, які часто вдається отримати лише для асимптотичного випадку, коли обсяг вибірки прямує до нескінченності. Розв'язання подібних задач значно спрощується, якщо статистичні дані можуть бути адекватно описані гауссовим розподілом. Використання подібної моделі дає змогу синтезувати доволі прості та ефективні лінійні вирішні правила, які, залежно від конкретної ситуації, можна легко оптимізувати за певним імовірнісним критерієм (Байєса, Неймана–Пірсона, мінімаксним тощо) [1].

Одним із альтернативних способів побудови вирішних правил є застосування математичного апарату стохастичних поліномів. Зокрема, в роботі [2] застосовується спосіб подання логарифма відношення правдоподібності у вигляді рядів, базисними функціями яких є деякі нелінійні (наприклад, степеневі) перетворення вибіркових даних, а коефіцієнти ряду знаходять за умови забезпечення екстремуму певного якісного критерію прийняття рішень. Результати досліджень із застосування цього підходу для виявлення сигналів свідчать про доцільність його використання саме при негауссовому характері статистичних даних. Характерною рисою цих алгоритмів є те, що вони ґрунтуються на неповному ймовірнісному описі у вигляді скінченної послідовності моментних або кумулянтних функцій.

У цій роботі пропонується метод, який також ґрунтується на використанні стохастичних поліномів, але за методикою побудови вирішних правил істотно відрізняється від вказаного вище. В його основі лежить апарат [3] розкладу у просторі з порідним елементом (просторі Кунченко), результатом застосування якого є представлення випадкових величин у вигляді стохастичних рядів, базисними функціями яких є деякі нелінійні перетворення від випадкової величини, що розкладається [4]. Принципи застосування цього математичного апарату для розв'язання задач перевірки статистичних гіпотез були вперше сформульовані в тезисних роботах [5, 6], розвинуті в [7] та застосовані для побудови процедур виявлення та розпізнавання сигналів і образів в [8–10]. Проте в цих роботах залишалось відкритим питання аналітичного визначення границь прийняття рішень та розрахунку ймовірностей помилок вирішних правил.

Метою роботи є розроблення із застосуванням апарату розкладу випадкових величин в просторі з порідним елементом (просторі Кунченко) нового методу перевірки статистичної гіпотези.

Постановка задачі. Нехай $\vec{X} = \{x_1, x_2, \ldots, x_N\}$ – N -вимірний вектор, що містить статистично незалежні однаково розподілені вибіркові значення. Основною гіпотезою H_0 вважається ситуація, коли $\vec{X}$ є реалізацією випадкової величини x з одновимірним законом розподілу ймовірностей $P_x(x)$. Розподіл складових вектора $\vec{X}$ за будь-яким іншим законом вважається альтернативою H_1.

Необхідно із застосуванням апарату розкладу випадкових величин у стохастичні ряди синтезувати та проаналізувати характеристики вирішного правила, яке дає змогу на основі опрацювання вибірки $\vec{X}$ підтвердити (або спростувати) реалізацію гіпотези H_0.

Результати дослідження.

1. Принципи розкладу випадкових величин у стохастичні ряди

Відповідно до теорії розкладу в просторі з порідним елементом [3] деяку випадкову величину x за певних умов можна подати у вигляді стохастичного функціонального ряду

$$x = k_0 + \sum_{i=1}^{\infty} k_i j_i(x), \tag{1}$$

де рівність в (1) трактується як рівність у середньоквадратичному, тобто

$$E\left\{x - \left[k_0 + \sum_{i=1}^{\infty} k_i j_i(x)\right]\right\}^2 = 0.$$

Особливістю такого розкладу є те, що його базисні функції $j_i(\cdot)$ являють собою у певний спосіб впорядковані нелінійні перетворення від випадкової величини, яка розкладається, і їх треба вибирати такими, щоб існували їх математичні сподівання

$$E\{j_i(\cdot)\} = \Psi_i < \infty.$$

Відомо, що практичне застосування будь-якого розкладу для розв'язання прикладних задач вимагає обмеження кількості членів ряду. В нашій ситуації таке обмеження призводить до виникнення похибки розкладу

$$h_S = x - x_S = x - \left[k_0 + \sum_{i=1}^{S} k_i j_i(x)\right], \tag{2}$$

тобто різниці між випадковою величиною x (яку називають порідною) та її наближеним поданням у вигляді стохастичного полінома x_S порядку S. Очевидно також, що похибка h_S є теж випадковою величиною.

У роботі [4] показано, що для мінімізації середньоквадратичної похибки розкладу, що еквівалентно мінімізації дисперсії випадкової величини h_S, необхідно, щоб коефіцієнти k_i, $i=\overline{1,S}$ знаходились як розв'язок системи лінійних алгебраїчних рівнянь

$$\sum_{i=1}^{S} k_i F_{i,j} = B_j, \quad j=\overline{1,S}, \tag{3}$$

де $F_{i,j} = E\{[j_i(x)-\Psi_i][j_j(x)-\Psi_j]\}$, $B_i = E\{[x-\Psi_0][j_i(x)-\Psi_i]\}$, $\Psi_0 = E\{x\}$.

Крім того, значення коефіцієнта k_0 повинно забезпечувати умову центрування похибки розкладу (рівність нулю математичного сподівання випадкової величини h_S), тобто

$$k_0 = \Psi_0 - \sum_{i=1}^{S} k_i \Psi_i . \tag{4}$$

Набір коефіцієнтів розкладу, знайдений з (3) та (4), називають оптимальним, адже він реалізує узгодженість між порідною випадковою величиною x та її представленням у вигляді стохастичного полінома x_S, забезпечуючи мінімум середньоквадратичної похибки розкладу за визначеного степеня полінома S.

Формування вирішних функцій із застосуванням стохастичних поліномів. Відомо, що одним із ключових моментів будь-якого методу перевірки статистичних гіпотез є спосіб формування статистичного критерію (вирішного правила), який передбачає вибір відповідного функціонального перетворення $f(\vec{X})$ для відображення N-вимірного вектору $\vec{X}$ в одну точку, тобто єдине числове значення, порівняння якого з певною граничною величиною (величинами) дає можливість приймати рішення про реалізацію відповідної гіпотези. Вибір граничних значень, які розмежовують область прийняття та відхилення рішень, залежить від вибраного критерію якості, який є функцією ймовірностей прийняття помилкових рішень та ризиків, з якими вони пов'язані. Теоретичний розрахунок границь вимагає визначення закону розподілу статистики $f(\vec{X})$, яка через обмеженість обсягу вибірки N є теж випадковою величиною.

Принципи побудови вирішних правил прийняття рішень, що пропонуються в цій роботі, ґрунтуються на властивостях похибки представлення випадкових величин у вигляді стохастичних функціональних поліномів. Ця властивість полягає у тому, що при оптимальному виборі коефіцієнтів розкладу забезпечується мінімізація (для відповідного степеня полінома S) дисперсії похибки (випадкової величини h_S), яка є меншою за дисперсію порідної випадкової величини x і зі зростанням степеня полінома прямує до нуля [4].

Отже, на основі виразу (2) та з урахуванням (4) можемо сформувати статистику

$$f(\vec{X}) = \frac{1}{N}\sum_{v=1}^{N}\sum_{i=0}^{S} k_i [j_i(x_v) - \Psi_i], \tag{5}$$

де в (5) для узагальнення введені такі позначення $j_0(x_v) = x_v$; $k_0 = -1$.

Фактично статистика виду (5) являє собою вибіркове середнє (лінійну оцінку математичного сподівання) випадкової величини h_S. З іншого боку, цю статистику можна трактувати як середньозважену (на величину оптимальних коефіцієнтів) різницю між теоретичними (очікуваними) та експериментальними (вибірковими) значеннями математичних сподівань базисних функцій стохастичного полінома.

Характеристики поліноміальних вирішних функцій. У роботі [4] показано, що середні значення стохастичних поліномів виду (5) асимптотично розподілені за гауссовим законом, що є безпосереднім наслідком центральної граничної теореми. Як відомо, параметрами гауссового

розподілу є математичне сподівання E та дисперсія D, величина яких істотно залежить від того, чи реалізується очікувана основна гіпотеза H_0 (для якої оптимізовані коефіцієнти розкладу $k_i^{(0)}$) чи її альтернатива H_1.

Можна показати, що в оптимальному (при реалізації гіпотези H_0) випадку математичне сподівання статистики (5) дорівнює нулю, тобто $E_{(0)} = 0$, а дисперсія визначається за співвідношенням

$$D_{(0)} = \frac{s_x^2 - J_S}{N}, \tag{6}$$

де величину $J_S = \sum_{i=1}^{S} k_i B_i$ називають інформкуною, значення якої характеризує ступінь зменшення дисперсії похибки розкладу порівняно з дисперсією порідної величини [4].

Враховуючи той факт, що зі зростанням кількості членів ряду величина інформкуни монотонно зростає від нуля і прямує до дисперсії порідної випадкової величини $\lim_{S\to\infty} J_S = s_x^2$, з'являється додатковий важіль для зменшення дисперсії статистики (5) навіть за фіксованого обсягу вибірки. Очевидною платою за це є ускладнення розрахунків та додаткова апріорна інформація про очікувану гіпотезу у вигляді математичних сподівань базисних функцій Ψ_i, які необхідні для розрахунку оптимальних коефіцієнтів розкладу і формування статистики (5).

Вираз для дисперсії (6) із урахуванням позначень в (3) можна записати у загальнішому вигляді

$$D_{(0)} = \frac{1}{N}\left[B_0^{(00)} - \sum_{i=1}^{S} k_i^{(0)} B_i^{(00)} \right].$$

В альтернативній ситуації H_1, коли вибірка розподілена за деяким іншим, відмінним від очікуваного розподілом, математичне сподівання та дисперсія статистики (5) визначаються за такими виразами

$$E_{(1)} = \sum_{i=0}^{S} k_i^{(0)} \left[\Psi_i^{(1)} - \Psi_i^{(0)} \right];\; D_{(1)} = \frac{1}{N}\left[B_0^{(10)} - 2\sum_{i=1}^{S} k_i^{(0)} B_i^{(10)} + \sum_{i=1}^{S}\sum_{j=1}^{S} k_i^{(0)} k_j^{(0)} F_{i,j}^{(10)} \right] - \left[E_{(1)} \right]^2,$$

де індекси в дужках визначають тип розподілу (реальний та очікуваний), для якого повинні розраховуватися відповідні математичні сподівання базисних функцій.

Безпосереднє застосування статистики виду (5), яка має симетричний відносно нуля розподіл, вимагає застосування двосторонніх критеріїв прийняття рішень, що в деяких практичних випадках є незручним. Для переходу до однобічного правостороннього критерію необхідно модифікувати цю статистику, наприклад, визначаючи її модуль або зводячи до квадрата. Застосувавши другий спосіб, отримуємо статистику виду

$$g\left(\mathbf{X}\right) = \left[f\left(\mathbf{X}\right) \right]^2. \tag{7}$$

Отже, використовуючи статистику (7), загальне правило прийняття рішень можна подати у вигляді

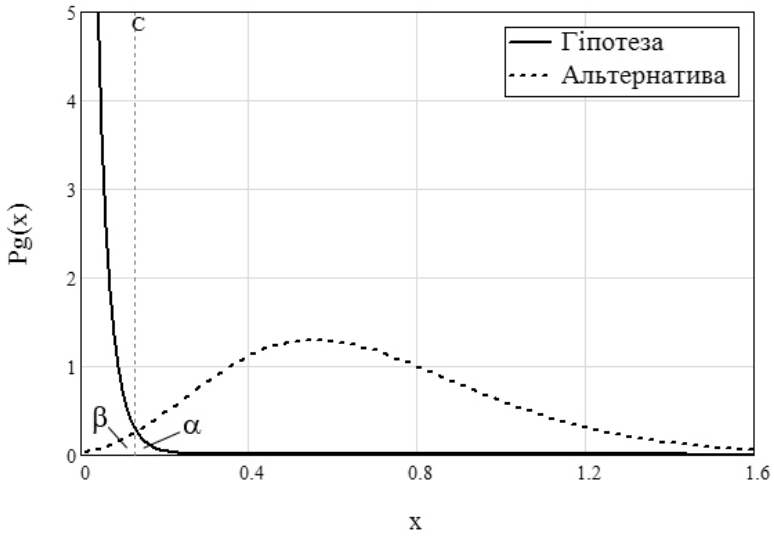
$$\left[\frac{1}{N} \sum_{v=1}^{N} \sum_{i=0}^{S} k_i \left[j_i(x_v) - \Psi_i \right] \right]^2 \underset{H_1}{\overset{H_0}{<>}} C \tag{8}$$

Очевидно, що подібне нелінійне перетворення призведе до трансформації розподілу результуючої квадратичної статистики $g\left(\mathbf{X}\right)$, який набуде такого вигляду

$$P_g(x) = \frac{1 + e^{-\frac{E^2}{D}}}{\sqrt{8pDx}} e^{-\frac{\left(\sqrt{x}-E\right)^2}{2D}}, \qquad (9)$$

де параметри E та D – математичне сподівання та дисперсія відповідного гауссового розподілу лінійної статистики $f\left(\vec{X}\right)$.

На рисунку наведено приклади графічних залежностей розподілу виду (9), які побудовані для випадків реалізації гіпотези та альтернативи із застосуванням як базисних функцій стохастичних поліномів степеневих перетворень вибіркових значень.



Щільності розподілу ймовірностей вирішної функції при гіпотезі та альтернативі

Аналітичне визначення розподілу вирішної функції та його параметрів при гіпотезі та альтернативі дає змогу оптимізувати правила прийняття рішень за допомогою вибору граничних значень порогу C відповідно до певного якісного критерію (Неймана–Пірсона, середнього ризику тощо) та визначати ймовірності помилок 1-го a та 2-го роду b .

Висновки. Запропоновано принципово новий метод перевірки статистичних гіпотез, який характеризується тим, що вирішна функція являє собою стохастичний поліном, коефіцієнти якого забезпечують мінімізацію її дисперсії за рахунок збільшення кількості членів ряду, що дає змогу зменшувати критичну область прийняття гіпотези навіть за фіксованого обсягу вибіркових значень. Отримані аналітичні вирази, які описують розподіл вирішного правила при гіпотезі та альтернативі, що дає змогу визначати межі критичних областей та ймовірності помилок як для загальнішого двостороннього критерію, так і для модифікованого однобічного.

Необхідно зазначити, що цей метод можна модифікувати для складніших випадків неоднаково розподілених та статистично залежних вибіркових даних, що і є завданням подальших досліджень.

1. Леман Э. Проверка статистических гипотез. – М.: Наука, 1979. – 408 с. 2. Кунченко Ю.П., Мартыненко С.С., Палагин В.В. Разработка нелинейных обнаружителей сигналов при негауссовских помехах, оптимальных по дисперсионным критериям // Труды УкрТелеКом-95. Одесса. – 1995. – С.440–443. 3. Кунченко Ю.П. Полиномы приближения в пространстве с порождающим элементом. – К.: Наук. думка, 2003. – 243 с. 4. Кунченко Ю.П. Стохастические полиномы. – К.: Наук. думка, 2006. – 275 с. 5. Кунченко Ю.П., Заболотный С.В. Применение метода неортогонального разложения случайных величин и процессов для их распознавания // Праці 3-ї української конференції з автоматичного керування "Автоматика-96". – Т.1. –

Севастополь, 1996, С.193-194. 6. Заболотній С.В. Нелінійні перетворювання випадкових величин як перспективний аспект у побудові нових методів перевірки простих статистичних гіпотез // Праці 4-ї української конференції з автоматичного керування "Автоматика-97". – Т.2. – Черкаси, 1997. 7. Заболотній С.В., Коваль В.В. Застосування неортогонального розкладання для перевірки статистичних гіпотез. Матеріали 3-ї Всеукраїнської конференції молодих науковців "ІТОНТ-2002". – Черкаси. 2002. – С.233–234. 8. Заболотній С.В. Розпізнавання випадкових сигналів з дискретним часом на основі неортогонального розкладання випадкових величин // Праці Луганського відділення Міжнародної академії інформатизації № 1 (8). – Луганськ. – 2004. – С. 39–41. 9. Заболотній С.В., Коваль В.В., Салипа С.В. Виявлення відеосигналів із застосуванням нелінійних дискретних фільтрів з постійними коефіцієнтами // Електроніка та системи управління. – 2008. – № 3. – С.77–83. 10. Заболотній С.В. Статистичне розпізнавання образів на основі розкладу в просторі з порідним елементом // Вісник Нац. ун-ту "Львівська політехніка" "Комп'ютерні науки та інформаційні технології" – 2009. – № 638. – С.118–123.

УДК 004.35

Я.Р. Совин, Я.В. Решетар, В.В. Хома
Національний університет "Львівська політехніка",
кафедра захисту інформації

ЕЛЕКТРОННИЙ БЕЗПРОВІДНИЙ ІДЕНТИФІКАТОР ДЛЯ ДОСТУПУ ДО ПК

Наведено програмно-апаратні засоби, використані під час побудови електронного безпровідного ідентифікатора доступу до персонального комп'ютера на базі технології Bluetooth.

Hardware and software tools are resulted used for the construction of electronic wireless-token on the base of Bluetooth technology for access to the personal computer.

Вступ. З розвитком комп'ютерних технологій, проникненням їх у всі сфери життя і діяльності людини зростає необхідність в гарантуванні безпеки інформації. Це потребує розв'язання таких задач, як ідентифікація/автентифікація особи, управління доступом, захист інформації та трафіку. Засобами електронної ідентифікації можуть бути RFID-мітки, proximity-карти, брелоки TouchMemory, магнітні картки, смарт-карти, USB-ключі.

Головною тенденцією розвитку систем електронної ідентифікації є наділення програмно-апаратних засобів максимальною кількістю функцій зі збирання, оброблення інформації та прийняття рішень. Це дає змогу зменшити вплив "людського фактора" та звільнити користувача від виконання рутинних операцій.

Аналіз останніх досліджень та постановка задачі. На цей час серед засобів доступу до ПК найбільші функціональні можливості та найвищий рівень захисту мають електронні ідентифікатори у вигляді смарт-карт та USB-ключів або токенів [3, 9].

Смарт-карта – це пластикова карта з вбудованим мікроконтролером, який виконує функції розмежування доступу до інформації, що зберігається в пам'яті, обробки й обміну даними, а також