УДК 006.91:53.089.68

В.С. Єременко, В.М. Мокійчук

Національний авіаційний університет, Київ

УНІВЕРСАЛЬНИЙ МЕТОД ІДЕНТИФІКАЦІЇ ЗАКОНУ РОЗПОДІЛУ ВИБІРКОВИХ ДАНИХ

У статті розглянуто метод ідентифікації закону розподілу даних вимірювань, який ґрунтується на функціональному перетворенні даних і приведення їх закону розподілу до рівномірного. Експериментально досліджено стійкість запропонованого методу до наявності у вибірці значень з надмірними похибками. Наведено приклади застосування запропонованого методу.

Ключові слова: *результат вимірювання, закони розподілу, надмірні похибки, критерії рівномірності, критерій Фроцині.*

Вступ

Проблема ідентифікації закону розподілу вибіркових даних виникає у метрологічних задачах при оцінюванні розширеної невизначеності за типом А та у задачах побудови вирішальних правил діагностики. Більшість метрологічних документів регламентує застосування спрощеного підходу, який ґрунтується на гауссівській моделі розподілу вибіркових даних. Тому регламентована достатня кількість статистичних критеріїв згоди емпіричного закону розподілу з законом Гаусса [2, 3]. На практиці закони розподілу даних вимірювань у багатьох випадках відрізняються від гауссівської моделі і її застосування може призвести до значних помилок як при оцінюванні розширеної невизначеності, так і при ухваленні діагностичних рішень. У таких випадках застосовують моделі стандартних законів розподілу (рівномірний, трикутний, експоненційний, тощо) [1], апроксимацію сімействами розподілів (Пірсона, Джонсона, Тьюкі) або апроксимацію рядами (Грамма-Шальє, Еджворта та ін.) [4]. Слід зауважити, що для зазначених моделей відсутні спеціалізовані критерії згоди, тому для вирішення задачі оцінювання згоди використовуються критерії Колмогорова-Смірнова або χ^2-Пірсона. Основним обмеженням до застосування цих критеріїв є необхідність наявності значного обсягу експериментальних даних. Отже постає задача обґрунтування та дослідження універсального метода ідентифікації закону розподілу вибіркових даних.

Метод ідентифікації закону розподілу вибіркових даних

Нехай маємо випадкову величину ξ із інтегральною функцією розподілу $F_\xi(x) \in [0...1]$ та величину η, яка пов'язана з величиною ξ функціональним перетворенням $\eta = f(\xi)$. Тоді інтегральна функція розподілу випадкової величини η визначається як

$$F_\eta(x) = F_\xi(f^{-1}(x)),$$

де $f^{-1}(\cdot)$ – обернена до $f(\cdot)$ функція.

Якщо за функцію $f(\cdot)$ взяти інтегральну функцію розподілу $F_\xi(x)$, то отримаємо

$$F_\eta(x) = F_\xi(F_\xi^{-1}(x)) = x,$$

яка відповідає функції рівномірного розподілу з областю значень [0…1], рис. 1. У випадку, якщо функція $f(\cdot)$ не відповідає функції розподілу вибіркових даних ξ, то закон розподілу випадкової величини η не буде рівномірним (рис. 2).

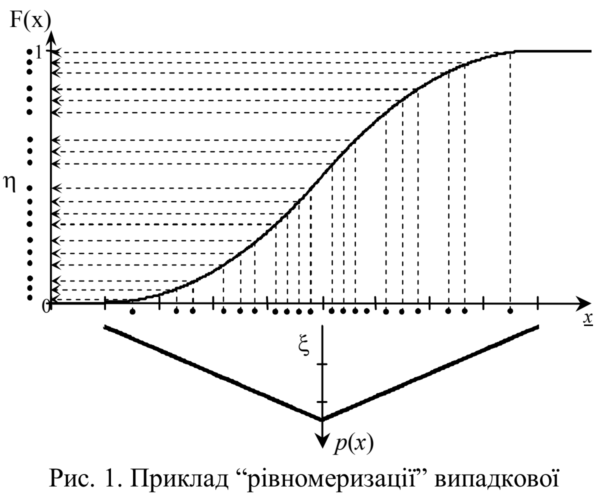

Рис. 1. Приклад "рівномеризації" випадкової величини ξ з трикутним законом розподілу

Отже, застосувавши до величини η критерій згоди для рівномірного закону (для інтервалу [0…1]) можна зробити висновок про відповідність обраної функції $f(\cdot)$ та закону розподілу випадкової величини ξ. На рис. 2 наведено приклад "рівномеризації" випадкової величини ξ з законом розподілу Гаусса (*а*). Щільність розподілу величини η отримана при застосуванні як функції $f(\cdot)$ функції Гаусса зображена на рис. 2. *б*, при застосуванні як функції $f(\cdot)$ функції Сімпсона (трикутного закону) отримали щільність розподілу зображену на рис. 2. *в*.

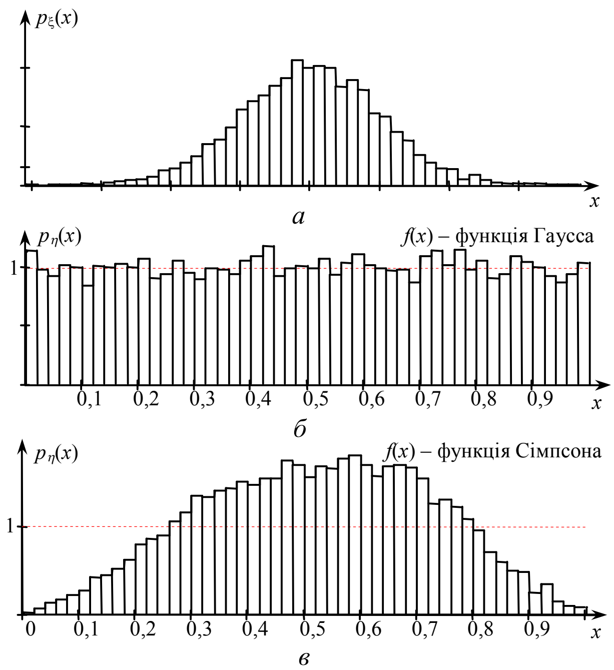

Рис. 2. Приклад "рівномеризації" випадкової величини ξ з законом розподілу Гаусса

Отже, послідовність застосування методу наступна.

1. Використовуючи апріорну інформацію, обираємо модель закону розподілу експериментальних даних $\xi_1 ... \xi_n$ – $F_\xi(x)$.

2. Оцінюємо параметри моделі $F_\xi(x)$.

3. Виконуємо «рівномеризацію» даних шляхом функціонального перетворення $\eta_i = F_\xi(\xi_i), i = \overline{1,n}$.

4. Застосовуємо критерій згоди для рівномірного закону, і робимо висновок: якщо закон розподілу $F_\eta(x)$ визнано рівномірним, то обрана модель $F_\xi(x)$ відповідає закону розподілу даних вимірювань ξ. У протилежному випадку модель є неадекватною.

Відома значна кількість статистичних критеріїв перевірки рівномірності розподілу: Кімбела, Морана, Шермана, Фроцині, тощо. Ці критерії відрізняються потужністю, наявністю табличних граничних значень або їх апроксимуючих рівнянь, а також складністю обчислень. Для порівняльного аналізу був обраний критерій Фроцині.

Результати досліджень. Розглянемо чутливість запропонованого методу до наявності аномальних значень у досліджуваних вибірках. Передбачається, що за рахунок використання нелінійних перетворень вихідних даних, запропонований метод буде менш чутливим до викидів (надмірних похибок).

Експериментальні дослідження методу проводилися за допомогою імітаційного моделювання, в основу якого покладено метод статистичних випробувань Монте-Карло. В якості тестових генерувалося 10000 вибірок по 34 значення випадкової величини із генеральної сукупності з гауссівським законом розподілу. Таким же чином генерувалися вибірки з викидами, значення яких дорівнювало 5σ. Для перевірки вибірки на нормальність використовувався RS-критерій [5]. Дослідження показали, що для вибірок без аномальних значень ймовірність прийняття правильного рішення склала 95%, а для вибірок з аномальними значеннями – 10%. Емпіричні розподіли статистик критерію для розглянутих випадків представлені на рис. 3.

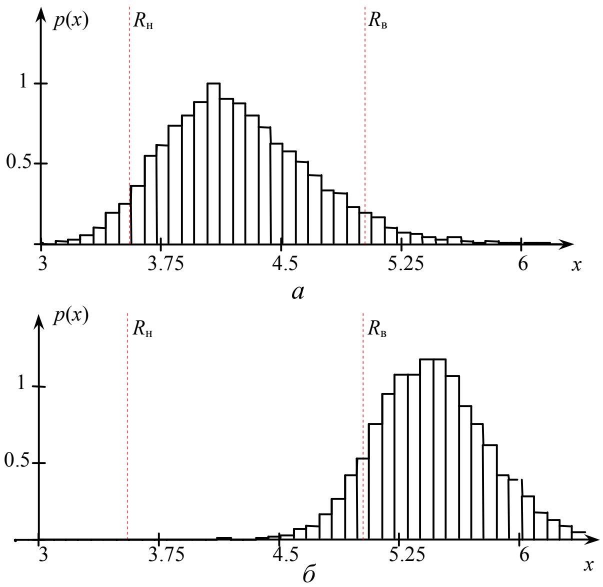

Рис. 3. Розподіл RS-статистики без (а) та з аномальним значенням (б) у вибірці обсягом 34 значення

Для цих самих згенерованих вибірок проводилася перевірка на нормальність за допомогою запропонованого методу "рівномеризації" із подальшим застосуванням критерію згоди Фроцині [5]. Статистика критерію має вид:

$$B = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left| U_i - \frac{i - 0.5}{n} \right|,$$

де U_i – варіаційний ряд побудований за аналізованою вибіркою.

За результатами проведеного моделювання ймовірність прийняття правильного рішення для вибірок без аномальних значень склала 95%, а з аномальними значеннями 53%. Розподіли статистик Фроцині для обох випадків наведені на рис. 4. Порівнюючи результати наведені на рисунках можна зробити висновок про більшу стійкість запропонованого методу до наявності аномальних значень, оскільки статистика Фроцині має менше зміщення відносно граничного значення.

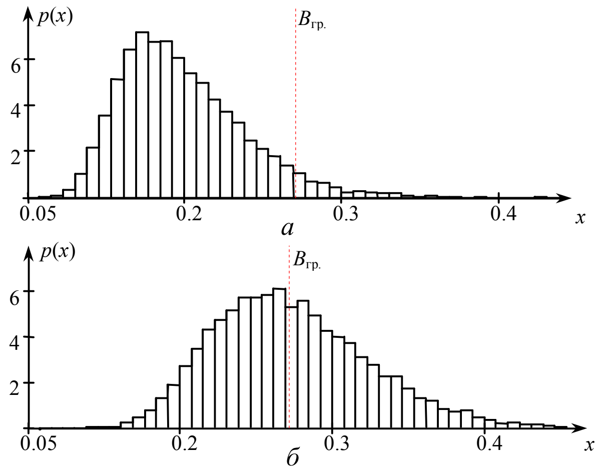

Рис. 4. Розподіл статистики Фроцині без (а) та з аномальним значенням (б) у вибірці обсягом 34 значення

В табл. 1 для порівняння наведені оцінки ймовірності прийняття рішення про гауссовість закону розподілу даних (вибірки по 34 значення) в залежності від величини аномального значення для обох випадків – без перетворення і з застосуванням запропонованого методу.

Таблиця 1

Оцінки ймовірності прийняття рішення

	Величина аномального значення				
Критерій	**3σ**	**3.5σ**	**4σ**	**4.5σ**	**5σ**
RS	0.84	0.66	0.44	0.24	0.11
Фроцині	0.92	0.88	0.81	0.70	0.53

Як видно з наведених досліджень, метод ідентифікації закону розподілу заснований на «рівномеризації» даних при дослідженні малих вибірок є менш чутливим до надмірних похибок. Ця властивість дозволяє в деяких випадках відмовитись від цензурування даних, яке на вибірках малого обсягу призводить до суттєвої втрати вимірювальної інформації.

Висновки

В статті запропоновано метод ідентифікації закону розподілу даних вимірювань, який ґрунтується на «рівномеризації» вибіркових даних. Він дозволяє досліджувати вибірки малого обсягу, для яких неможливо застосовувати методи Колмогорова-Смірнова та χ^2-Пірсона.

Проведені експериментальні дослідження показали стійкість результату ідентифікації до наявності у вибіркових даних надмірних похибок.

Перевагою метода є можливість застосування для ідентифікації будь-якого закону розподілу тільки одного статистичного критерію згоди для рівномірного закону, який можна обрати за необхідним критерієм ефективності.

Звичайно, розглянутий метод потребує подальшого дослідження щодо впливу точності оцінок параметрів моделі $f(\cdot)$.

Список літератури

1. Uncertainty of measurement. Part 3: Guide to the expression of uncertainty in measurement (GUM:1995) : ISO/IEC Guide 98–3:2008. – [Accepted 2008-09-30]. – Geneva : ISO, 2008. – 120 p. – (International standard).

2. Рекомендации по стандартизации. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Ч. I. Критерии типа хи-квадрат: Р 50.1.033-2001. – [введ. 2002-07-01]. – М.: ИПК Издательство стандартов, 2002 – 91 с.

3. Рекомендации по стандартизации. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Ч. II. Непараметрические критерии: Р 50.1.037-2002. – [введ. 2002-07-01]. – М.: ИПК Издательство стандартов, 2002 – 139 с.

4. Хан Г. Статистические модели в инженерных задачах / Г.Хан, С.Шапиро. – М.: Мир, 1969. – 395 с.

5. Кобзарь А. И. Прикладная математическая статистика. Для инженеров и научных работников. / А.И. Кобзар. – М.: ФИЗМАТЛИТ, 2006. – 816 с.

6. Еременко В.С. Оценка однородности выборок малого объема / В.С. Єременко, Ю.В. Куц, В.М. Мокійчук // Системи обробки інформації. – 2006. –№7(56) – С. 26–29.

Надійшла до редколегії 6.02.2013

Рецензент: д-р техн. наук, проф. Л.М. Щербак, Національний авіаційний університет, Київ.

УНИВЕРСАЛЬНЫЙ МЕТОД ИДЕНТИФИКАЦИИ ЗАКОНА РАСПРЕДЕЛЕНИЯ ВЫБОРОЧНЫХ ДАННЫХ

В.С. Еременко, В.М. Мокийчук

В статье рассмотрен метод идентификации закона распределения, основанном на функциональном преобразовании данных и приведения их закона распределения к равномерному. Экспериментально исследована устойчивость предложенного метода к наличию в выборке значений с грубыми погрешностями. Приведены примеры использования предложенного метода.

Ключевые слова: *результат измерения, законы распределения, грубые погрешности, критерии равномерности, критерий Фроцини.*

THE UNIVERSAL METHOD OF SAMPLED DATA DISTRIBUTION LAW IDENTIFICATION

V.S. Yeremenko, V.M. Mokiychuk

The method to identify the distribution low of measurement data is examined. Method is based on data functional transformation and reduction the form of distribution law to the uniform one. The robustness of given method against the gross errors was experimentally tested. The examples of method application were set.

Keywords:*: measurement result, distribution laws, gross errors, tests for uniformity, Frozini’s test.*