

УДК 519.21

КОМП'ЮТЕРНА ПРОГРАМА ДЛЯ ВІДПОВІДНОСТІ ВИБІРКОВИХ ДАНИХ ВЕКТОРНОМУ НОРМАЛЬНОМУ РОЗПОДІЛУ ІМОВІРНОСТЕЙ МЕТОДОМ ПЕРЕХОДУ ДО СУКУПНОСТІ СКАЛЯРНИХ ВИБІРОК (ПССВ)

С.М. Краснитський, доктор фіз. - мат. наук, професор
Київський національний університет технологій та дизайну
А.О. Кібітов, студент

Київський національний університет технологій та дизайну

Ключові слова: багатовимірний нормальний розподіл, векторна вибірка, коваріаційна матриця, гіпотеза про нормальність, метод ПССВ.

Постановка задачі. Нехай маємо вибірку векторних даних $X = (x_1, x_2, \dots, x_n)$, де $x_j = (x_{1j}, \dots, x_{mj})$, де «штрих» — знак транспонування (так що кожне x_j — вектор-стовпець з m дійсними координатами). Гіпотеза H_0 : теоретичним розподілом даної вибірки m -вимірний нормальний розподіл. Іншими словами, наявні вибіркові значення є реалізаціями m -вимірного випадкового вектора), що має m -вимірний нормальний розподіл [2].

Відомі результати. З числа досліджень щодо можливостей перевірки H_0 відзначимо роботи [3 – 5]. Можна назвати застосовані в цих роботах методи «прямими», оскільки переважна більшість застосованих в них алгоритмів заснована на безпосередній оцінці щільності багатовимірних розподілів імовірностей наявних вибіркових даних.

Серед зазначених джерел виділимо роботу [5], в якій перевірку H_0 пропонується виконувати на основі деяких можливостей теорії розпізнавання образів. Вихідним зразком береться множина точок, утворених реалізаціями нормально розподіленого випадкового вектора. Дана множина об'єднується з багатовимірними точками — реалізаціями досліджуваної вибірки. На точках вказаного об'єднання будується деякий граф, в якому певним чином виділяється група ребер, кількість яких і приймається в якості статистики критерію.

Методи перевірки гіпотези H_0 , про які мова йшла вище, спираються на коректні і зрозумілі дії, що використовують певні властивості багатовимірного нормального розподілу. З іншого боку, вказані методи базуються на оцінках за вибірковими даними коваріаційних матриць розподілів і обернених до таких матриць. Очевидно, при великих розмірностях даних це не є зручною обставиною. Метод перевірки гіпотези H_0 , котрий пропонується у даній роботі, не використовує зазначених операцій. Основна ідея зазначеного методу полягає у заміні роботи з характеристиками векторної вибірки дослідженням характеристик сукупності деяких скалярних вибірок. Ці вибірки є лінійними перетвореннями вихідної векторної вибірки. Згідно з теоремою 1.3.8,

перевірку нормальності вихідної вибірки можна замінити перевіркою нормальності *всіх* таких перетворень. Практично мова може йти про перевірку нормальності *як завгодно великої кількості* таких перетворень. Зауважимо, що більш рання редакція зазначеного методу була застосована в роботі [1].

Вихідні (початкові) дані методу ПССВ:

Об'єм вибірки, розмірність вибірових елементів, дійсна векторна вибірка X , представлена у формі матриця даних $X = X_{m \times n}$, кількість ітерацій k , параметр ітерації c — дійсний m -вимірний вектор, залежний від номеру ітерації l , рівень значущості ε .

Окрема ітерація (c -ітерація, точніше, $c(l)$ - ітерація)) полягає в наступному перетворенні вихідної матриці даних X :

$$c(X) = y, \text{ де } y = (y_1, y_2, \dots, y_n), y_j = \sum_{i=1}^m c_i x_{ij}, j = 1, \dots, n.$$

Таким чином, зазначене перетворення переводить векторну вибірку X в скалярну вибірку y . До одержаної вибірки застосовується один з відомих методів перевірки на нормальність вибірки скалярних величин. Дана програма розрахована на застосування певного варіанту відомого методу χ^2 (хі-квадрат) [6,7].

Алгоритм даної роботи (аббревіатура — ПССВ) не пов'язаний з необхідністю обернення оціночних коваріаційних матриць великої розмірності. Це надало можливість розробити відповідну комп'ютерну програму, котра може працювати з даними великих об'ємів і розмірностей.

Список використаних джерел

1. Заржицький Є.В., Краснитський М.С. Комп'ютерна програма для перевірки нормальності закону розподілу векторної вибірки // Комплексна дипломна робота на здобуття ступеня магістра комп'ютерних наук— Київ, КНУТД, 2005.
2. Краснитський С.М., Щербань В.Ю. Краснитський М.С., Заржицький Є.В. та ін. Векторні випадкові величини і випадкові процеси. — К.: Бумсервіс, 2008. — 191 с.
3. Smith, S. P.; Jain, A. K. (1988). "A test to determine the multivariate normality of a data set". *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 10 (5): 757.
4. Cox, D.R., Small, N. J. H. _Testing multivariate normality // *Biometrika*, Volume 65, Issue 2, 1978, pp. 263–272.
5. Smith, S. P.; Jain, A. K. (1988). "A test to determine the multivariate normality of a data set". *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 10 (5): 757.
6. Johnson, R.A., Wichern. D.W. Applied Multivariate Statistical Analysis (2017). Washington: Pearson, 773.
7. Gentle, J. E. (2009). Computational Statistics. *Statistics and Computing*. New York: Springer. pp. 315–316.