УДК 519.24

В.А. Володарський, студент гр. ПА-91мп, А.А. Помилуйко, студент гр.ПА-91мп
КПІ ім. Ігоря Сікорського

ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ РОБАСТНИХ ПРОЦЕДУР

Анотація. В статті досліджується ефективність стійких до викидів ітераційних алгоритмів визначення за вибірковими даними параметрів генерального розподілу.

Ключові слова: викиди, відтворюваність, робастність, абсолютне медіанне відхилення, ітерація.

ВСТУП

На сучасному етапі, коли випробування можуть проводитися в різних лабораторіях і навіть в різних країнах, для атестації методики випробувань при заданих умовах, проводять спільний міжлабораторний експеримент, за результатами якого нормуються показники точності – зміщення та відтвореність [1]. Наявність у вибірці навіть невеликого числа спостережень, які різко виділяються і називаються промахами, здатне кардинально змінити результат статистичного дослідження. Промахи виявляють за допомогою відповідних критеріїв та вилучаються з вибірки [2].

Оскільки випробування однотипної продукції, виконані за єдиною методикою, можуть здійснюватися декількома лабораторіями в різних умовах, результати, природно, будуть дещо відрізнятися. Врахування впливових величин аналітичним шляхом практично неможливо. Єдиним підходом для вирішення цієї задачі є проведення міжлабораторних спільних випробувань. Міжлаборатний експеримент фактично є фізичною моделлю реалізації методики з залученням лабораторій, що мають близький професійний рівень і спеціалізуються в даному виді випробувань. При спільному експерименті існує припущення, що всі залучені лабораторії мають однакову повторюваність [1]. Але, з об'єктивних причин, це не завжди виконується.

При такому підході, при нормуванні показників точності методики випробувань, застосування статистичних критеріїв виключення викидів, як це робиться у вимірюванні, не можливо, бо це приведе до невірного визначення показників.

Для виключення впливу викидів при оцінюванні показників точності методики випробувань застосовуються робастні методи [3], які надають можливість використовувати всі наявні експериментальні дані.

ПОСТАНОВКА ЗАДАЧІ

Метою роботи є аналіз ефективності ітераційного робастного алгоритму, на підставі вибіркових даних, параметрів закону розподілу, які відповідають правильності та відтворюваності методики випробування.

ОСНОВНА ЧАСТИНА

Вибіркові дані, у загальному випадку, можуть мати деяку розбіжність з передбачуваним розподілом (особливо при малих обсягах) – містити деякі значення, що підпорядковуються іншому розподілу.

Вихідним є те, що деяка центральна частина розподілу експериментальних даних відповідає розподілу генеральної сукупності [4]. Тому при опрацюванні даних, які знаходяться в цій частині розподілу, доцільно використовувати метод найменших квадратів (МНК). Але метод найменших модулів (МНМ) є більш стійким до викидів, ніж МНК, тобто дає найкращий результат при найбільш несприятливому розподілі.

Виходячи з цього, при створенні робастних методів робиться «симбіоз» – для деякої центральної частини розподілу використовується МНК, а для іншої частини для зменшення впливу викидів, але зі збереженням наявних даних, застосовується МНМ. Граничне значення границі переходу від МНК до МНМ відповідає $\varphi = c\sigma$. Константа c регулює ступінь робастності і її значення залежить від ступеня «засмічення» [3]. Зазвичай обирають значення c = 1,5.

Найбільш стійким до викидів є інтервал, що знаходиться між вибірковими квартилями. У припущенні про можливий закон розподілу довжина інтервалу однозначно відповідає дисперсії цього розподілу. У якості початкової оцінки центру розподілу береться вибіркова медіана. Як початкова оцінка при переході від повного розподілу до усіченого, береться медіана абсолютних відхилень MAD (*median absolute deviation*).

$$MAD_n = med\{|x_i - M_n|\}. \tag{1}$$

де, $M_n = med\{x_i\}$, x_i – елемент вибірки, а індекс *n* відповідає числу елементів.

Первинна оцінка СКВ, яка є стійкою до викидів для вибірки з *n* елементів, знаходиться на підставі нормального інтерквартильного розмаху і складає $S^*_{(0+1)} = 1{,}483 \cdot MAD_n$. Константа 1,483 використовується для нормування при переході від «інтерквартильного» сегменту до генерального розподілу. Значення $S^*_{(0+1)}$ використовується при переході до робастної ітераційної процедури уточнення параметрів розподілу [4].

Перед початком цієї процедури визначається точка переходу від МНК до МНМ

$$\varphi_1 = 1{,}5 \cdot S^*_{(0+1)}. \tag{2}$$

та нижнього і верхнього граничних значень

$$x^*_{1\min} = x^*_0 - \varphi_1; \quad x^*_{1\max} = x^*_0 + \varphi_1. \tag{3}$$

Молодший елемент ранжируваного ряду вихідних даних порівнюють з $x^*_{1\min}$ а старший елемент порівнюють з $x^*_{1\max}$. Якщо молодший елемент менше нижнього граничного значення, то йому присвоюється значення $x^*_{1\min}$. У випадку, коли старший елемент буде більше верхнього граничного значення, то замість нього в ряд вводиться $x^*_{1\max}$. Всі ж інші елементи ряду залишають без зміни. Приходимо до модифікованого чисельного ряду, для якого обчислюється середнє $\bar{x}^*_1$, яке є уточненим значенням центру розподілу, та СКВ «усіченого» розподілу S^*_1 (так зване «старе»), яке використовується для обчислення уточненого на першому кроці, робастного СКВ («нового») $S^*_{(1+1)} = 1{,}134 \cdot S^*_1$.

Константа 1,134 дозволяє перерахувати СКВ, обчислене для «усіченого» розподілу до генерального [5].

Другий крок ітераційної процедури починаємо з обчислення граничних значень $x^*_{2\min}$ та $x^*_{2\max}$. Для цього застосовуються співвідношення (3), в яких використовується $\bar{x}^*_1$.

Ітераційна процедура продовжується, поки розходження між параметрами розподілу на поточному і попередньому кроці не стане менше заданого значення.

Для прикладу реалізації робастного ітераційного алгоритму скористаємося ранжируваними даними, наведеними у [1].

Дані були отримані при проведені міжлабораторного спільного експерименту при залучені n = 9 лабораторій (підкресленням позначено найменший та найбільший лабораторні результати):

24,140 20,155 19,500 20,300 20,705 17,570 20,100 20,940 21,185

В таблиці 1 представлені дані для j-ого кроку ітерації.

Таблиця 1. Покрокове визначення середнього значення та СКВ

Номер ітерації j	*0*	*1*	*2*	*3*	*4*
φ_j		1,424	1,478	1,514	1,539
$x^*_j-\varphi_j$ $x^*_j+\varphi_j$		18,876 21,724	18,909 21,865	18,893 21,921	18,872 21,950
$x^*_{1(j)}$	17,570	18,876	18,909	18,893	18,872
$x^*_{2(j)}$	19,500	19,500	19,500	19,500	19,500
$x^*_{3(j)}$	20,100	20,100	20,100	20,100	20,100
$x^*_{4(j)}$	20,155	20,155	20,155	20,155	20,155
$x^*_{5(j)}$	20,300	20,300	20,300	20,300	20,300
$x^*_{6(j)}$	20,705	20,705	20,705	20,705	20,705
$x^*_{7(j)}$	20,940	20,940	20,940	20,940	20,940
$x^*_{8(j)}$	21,185	21,185	21,185	21,185	21,185
$x^*_{9(j)}$	24,140	21,724	21,865	21,921	21,950
Середнє $\bar{x}^*_j$	20,511	20,387	20,407	20,411	20,412
Стандартне відхилення $\bar{S}^*_j$	1,727	0,869	0,890	0,905	0,916
Нове $\bar{x}^*_{j+1}$	20,300	20,387	20,407	20,411	20,412

Нове $\bar{S}^*_{(j+1)}$	0.949	0,985	1,009	1,026	1,039

Аналіз результатів, які були отримані при моделюванні, дозволяє зробити висновок, що значення граничних елементів вибірки практично не впливають на робастне оцінювання параметрів закону розподілу. Ефективність алгоритму пояснюється тим, що абсолютне медіанне відхилення визначається по відношенню до медіани $x^*_{5(0)} = 20{,}300$ вибірки.

Також проаналізовано випадок, коли впорядкований числовий ряд включає два близько розташованих елементи, які є викидами. При цьому має місце, так званий, замаскований ефект викидів. Застосування критерія Тітьена-Мура дозволяє запобігти «маскувальному ефектові» і виключити викиди. Але знайдені оцінки параметрів розподілу відрізняються від істинних значень.

Для оцінювання ефективності алгоритму при наявності викидів з «маскувальним ефектом» у стовпчику j = 0 таблиці 1, проведена модифікація $x^*_{2(0)} = 18{,}250$. Проведений аналіз показав, що ітераційний алгоритм «працює» і, в даному випадку, він залишається стійким навіть при наявності замаскованих викидів.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] Точність (правильність і прецизійність) методів та результатів вимірювання. Частина 2. Основний метод визначення повторюваності та відтворюваності стандартного методу вимірювань: (ISO/IEC 5725-2:1994, IDT) ДСТУ ИСО 5725-2:2005. – [Чинний від 2006.07.01]. – К.: Держспоживстандарт України, 2006. – 48 с. – (Національний стандарт України).

[2] Ціленко, В.Д. Невизначеність вимірювання: монографія / В.Д. Ціленко, Н.А. Яремчук. – К.: ІВЦ «Політехніка», 2002. – 176 с.

[3] Sarhan Ahmed E., Greenberg Bernard G.: Contributions to order statistics (Вклады в порядковые статистики). – John Wiley & Sons, 1962, pp. 482.

[4] Хьюбер, П. Робастность в статистике: монографія / П. Хьюбер ; пер. з англ. И. Моховой, В. Хохлова – М. : Мир, 1984, – 304с.

[5] Odporna ocean dokładności metod pomiarowych / E. Volodarsky, L. Kosheva, Z. Warsza // Pomjari, awtomatica, control. – 2012. – №4. – P. 396-401.

Наук. керівник – к.т.н., доц. Добролюбова М.В.