УДК 681.142

О. Ю. Года, к.т.н., ст. викладач, проф., Н.А. Яремчук, к.т.н., доц.
КПІ ім. Ігоря Сікорського

СПОСОБИ ОТРИМАННЯ СТАНДАРТНОЇ НЕВИЗНАЧЕНОСТІ ВИБІРКОВОЇ МЕДІАНИ

Анотація. В статті наведено результати аналізу способів розрахунку стандартної невизначеності вибіркової медіани як характеристики центру вибірки при багаторазових вимірюваннях ординальних властивостей. Стандартна невизначеність розраховується за розширеною невизначеністю, яку знаходять за номерами порядкових статистик, що відповідають певному рівню довіри. При розрахунках використовуються вкладений інтервал нечіткої множини і рівневі множини.

Ключові слова: вибіркова медіана, стандартна невизначеність.

ВСТУП І КОРОТКИЙ ОГЛЯД ЛІТЕРАТУРНИХ ДЖЕРЕЛ

При опрацюванні багаторазових вимірювань обирають ефективну оцінку результату вимірювання згідно розподілу вибірки, і цією оцінкою для класу експоненціальних розподілів є вибіркова медіана, що визначається за впорядкованою вибіркою об'єму n.

Крім того медіана є робастною оцінкою, тобто стійкою до аномальних результатів і викидів, і її обирають, якщо бажано використати такі оцінки, що мало залежать від порушень прийнятих моделей вхідних сигналів. Оскільки медіана є непараметричною оцінкою, її застосовують в якості результату при багаторазових вимірюваннях ординальних величин, при опрацюванні яких не використовуються алгебраїчні операції.

За літературними джерелами існують наступні способи для розрахунку розширеної невизначеності медіани. Так при $n \geq 10$, за [1], довірчий інтервал, що характеризує точність визначення медіани за довірчою ймовірністю P, становить (x_U, x_V), де U – найбільше ціле число, що менше за $(n+1-z_p\sqrt{n})/2$, V – найменше ціле число, що більше за $(n+1+z_p\sqrt{n})/2$, де z_p - квантиль стандартного гаусівського розподілу.

За умов нормальності розподілу вибірки і при великому об'ємі вибірки n дисперсію медіани можна обчислити за співвідношенням [2]:

$$D(med) = \pi\sigma^2/(2n), \qquad (1)$$

де σ^2 – дисперсія середнього арифметичного вибірки.

Отже за умов нормальності розподілу оцінку дисперсії можна отримати лише за параметричною оцінкою дисперсії середнього арифметичного. Тому формула (1) непридатна для визначення дисперсії медіани по-перше при невідомому розподілі вибірки, а по-друге – при опрацюванні результатів вимірювання ординальних величин, де використовуються оцінки, засновані на порядкових статистиках.

Для розрахунку розширеної невизначеності медіани при об'ємі вибірки $n<10$ може бути використана неповна β–функція I_γ [2], за якою співвідношення для вільного від розподілу толерантного інтервалу (x_r, x_s) включає п'ять величин: γ - мінімальна доля $F(x)$, що входить до покривного інтервалу; β - покривна ймовірність; n - об'єм вибірки; r,s - положення

порядкових статистик в вибірці, які обирають симетрично, так що $s = n - r + 1$.

$$P\{F(x_s) - F(x_r) \geq \gamma\} = 1 - I_\gamma(s - r, n - s + r + 1) = \beta. \qquad (2)$$

Якщо $r = 1$, так що використовуються крайні члени вибірки, (2) приймає вигляд: $I_\gamma(n-1;2) = 1 - \beta$, що визначає ймовірність β, за якою розмах вибірки з n результатів вимірювань покриває долю γ розподілу, з якого отримана вибірка. Таблиці для розрахунку розширеної невизначеності медіани з використанням (2) для n від 2 до 10 наведено в роботі [3].

Але, якщо статистична похибка медіани об'єднується з іншими складовими невизначеності, то для отримання сумарної (комбінованої) стандартної невизначеності необхідно розробити перехід від розширеної до стандартної невизначеності медіани. Вирішенню цієї проблеми присвячено аналіз, наведений у наступному розділі.

АНАЛІЗ ПІДХОДІВ, ЩО ЗАСТОСОВУЮТЬСЯ ПРИ ОБЧИСЛЕННІ СТАНДАРТНОЇ НЕВИЗНАЧЕНОСТІ МЕДІАНИ ЗА ЇЇ РОЗШИРЕНОЮ НЕВИЗНАЧЕНІСТЮ

Перший підхід полягає в використанні вкладеного інтервалу нечіткої змінної [4] за яким:

$$\sigma(med) = \Delta_p / \sqrt{-2\ln(1-P)}, \qquad (3)$$

де Δ_p – границя вкладеного інтервалу, ймовірність знаходження в якому P.

Для застосування формули (3) попередньо визначаються ймовірності знаходження медіани в границях, встановлених за номерами порядкових статистик (2). Так, наприклад, для $n = 7$ з застосуванням таблиць [3] отримуємо:

$$P(x_1 \leq med \leq x_7) = 0.98,\ P(x_2 \leq med \leq x_6) = 0.86,\ P(x_3 \leq med \leq x_5) = 0.42.$$

Якщо використати, наприклад, ці інтервали для ранжованого ряду $x = \{11,\ 12,\ 13,\ 14,\ 15,\ 16,\ 17\}$ з медіаною $med = 14$, отримаємо $\sigma_1(med) = 1.03$; $\sigma_2(med) = 1.01$; $\sigma_3(med) = 0.96$. Отримані результати свідчать про хорошу збіжність значень СКВ.

Але якщо в вибірці мають місце значення, що повторюються, може виникнути неоднозначність при обчисленні СКВ. Тобто може створитись ситуація, що різним номерам порядкових статистик відповідають однакові значення, а однаковим інтервалам відповідають різні ймовірності.

Тому другий підхід полягає в поданні довірчих інтервалів вибіркової медіани як множини α–рівнів, де $\alpha = 1 - P$ [5] і в горизонтальній побудові нечіткої множини [6,7], що представляє функцію приналежності з використанням об'єднання множин α–рівнів. Тобто вирішується задача відновлення функцій приналежності нечіткої множини за множиною α–рівнів.

Записуємо усі множини α–рівня, а потім об'єднуємо відповідні їм нечіткі множини. Вибірка з повтореннями: $x = \{11,\ 11,\ 12,\ 13,\ 13,\ 14,\ 14\}$

$$\alpha_1 A_{\alpha_1 = 0,02} = \{0.02|11;\ 0.02|11;\ 0.02|12;\ 0.02|13;\ 0.02|14;\ 0.02|14;\ \},$$

$$\alpha_2 A_{\alpha_2 = 0,14} = \{0.14|11;\ 0.14|12;\ 0.14|13;\ 0.14|13;\ 0.14|14\},$$

$\alpha_3 A_{\alpha_3=0,58} = \{0.58|12;\ 0.58|13;\ 0.58|13\}$.

Об'єднуємо результати з використанням оператора логічна сума:
$(0,02 \cup 0,14)|11;\ (0,02 \cup 0,14 \cup 0.58)|12;\ (0,02 \cup 0,14 \cup 0.58)|13;\ (0,02 \cup 0,14)|14 =$
$= 0.14|11;\ 0.58|12;\ 0.58|13;\ 0.14|14.$

З урахуванням правила розбиття одиниці отримуємо :
$0.097|11;\ 0.403|12;\ 0.403|13;\ 0.097|14.$

За отриманою функцією приналежності визначаємо вкладений інтервал і стандартну невизначеність.

ВИСНОВОК

При багаторазових вимірюваннях ординальних властивостей в якості центру вибірки використовують вибіркову медіану. Розширену невизначеність медіани можна знайти за номерами порядкових статистик, що відповідають певному рівню довіри. Але, якщо похибку медіани об'єднують з іншими складовими невизначеності (наприклад, інструментальними), то виникає необхідність в розрахунку стандартної невизначеності медіани.

Для цієї мети в роботі обрано спосіб з використанням вкладеного інтервалу. Але при використанні цього способу може виниками неоднозначність в виборі інтервалу за певної ймовірності при наявності в вибірці результатів вимірювання, що повторюються. У цьому випадку задачу можна вирішити з використанням горизонтальної побудови функції приналежності нечіткої множини за множинами α -рівня.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Яремчук Н.А. Анализ способов получения стандартной неопределенности выборочной медианы Грановський В.А., Методы обработки экспериментальных данных при измерениях / В.А. Грановський, Т.Н. Сирая – Л.: Энергоиздат, 1990. – 288с.

2. Кендалл М. Статистические выводы и связи / М. Кендалл, А. Стьюарт Редакция физико-математической литературы изд-ва «Наука», 1973.- 899с.

3. Яремчук Н.А. Оцінювання комплексного показника якості складного об'єкта з характеристикою невизначеності / Н.А. Яремчук, О.Ю. Года, В.В. Проскін // Український метрологічний журнал. -2018.-№2. – с.9-17

4. Солопченко Г.Н. Представление измеряемых величин и погрешностей измерений как нечетких переменных / Г.Н. Солопченко // Измерительная техника.- 2007. - №2. - С. 3-7.

5. Яремчук Н.А. Анализ способов получения стандартной неопределенности выборочной медианы / Н.А. Яремчук, О.Ю. Года, А.Ю. Скрипий // Метрологія та вимірювальна техніка (Метрологія – 2018): матеріали XI міжнародної науково-технічної конференції, 9-11 жовтня. – Харків, 2018. – с. 45-46.

6. Пегат А. Нечеткое моделирование и управление / А. Пегат : пер. С англ. –М.: Бином. Лаборатория знаний . – 2009. – 798с.

7. Конышева Л.К. Основы теории нечетких множеств: Учебное пособие / Л.К. Конышева, Д.М. Назаров . – Спб.: Питер.- 2011. – 192с.

Наук. керівник – к.т.н., доц., проф. Яремчук Н.А.