УДК 621.391.83:004.94(045)

Ю.В. Куц, д-р техн. наук
О.В. Монченко

КРУГОВА МЕДІАНА В СТАТИСТИЧНОМУ АНАЛІЗІ КУТОВИХ ДАНИХ

НАУ, кафедра інформаційно-вимірювальних систем, e-mail: galena79@mail.ru

Запропоновано методику визначення вибіркової кругової медіани при статистичній обробці реалізацій випадкових кутів. Наведено приклад використання запропонованої методики на основі моделювання в системі MatLab.

Вступ

Однією з числових характеристик розподілів імовірностей є медіана. За визначенням [1] для дійсної випадкової величини $\xi(\omega)$, де ω – елемен-тарна подія з простору подій Ω, з функцією розподілу $F(x)$, $x \in R$, медіаною називається значення x_m, для якого

$$F(x_m) \le 0{,}5;$$
$$F(x_m + 0) \ge 0{,}5.$$

Довільна випадкова величина має принаймні одну медіану. Якщо $F(x)$ строго монотонна функція, то медіана єдина.

У математичній статистиці для оцінювання медіани розподілу за результатами незалежних спостережень $\{x_1,\dots x_n\}$ використовують вибіркову медіану [1; 2]. Ця порядкова статистика визначається як медіана відповідного варіаційного ряду $x_{(1)},\dots x_{(n)}$:

$$x_m = \mathbf{Me}(x_1,..x_n) = \begin{cases} x_{(k+1)}, & n = 2k+1, \\ \frac{1}{2}\left[x_{(k)} + x_{(k+1)}\right], & n = 2k, \end{cases}$$

де **Me** – оператор медіани; $x_{(k)}$ – k- й член варіаційного ряду, утвореного з послідовності $\{x_1,\dots x_n\}$.

Наприклад, для послідовності

$x = \{10, 78, 233, 4, 90, 342, 135, 197, 215\}$

маємо $n = 9$ і варіаційний ряд {4, 10, 78, 90, 135, 197, 215, 233, 342}. Отже, медіаною цієї послідовності є значення $x_{(5)} = 135$.

Медіанна фільтрація як ефективний метод нелінійної обробки сигналів з метою статистичного згладжування та пригнічення дії імпульсних завад відома понад тридцять років. Вона була запропонована Тьюки (J.W. Tukey) 1971 р. як метод аналізу часових рядів. Медіанна фільтрація реалізується під час руху ковзного вікна певної ширини (апертури) вздовж часового ряду шляхом заміни значення елемента ряду в центрі апертури медіаною тих значень ряду, що відбираються з ряду за допомогою апертури.

Якщо позначити оператор одновимірної медіанної фільтрації як $\mathbf{MF}_{2k+1}$, де $2k+1$ – ширина апертури фільтра, то реакція фільтра визначається співвідношенням

$$\{y_n\} = \mathbf{MF}_{2k+1}\{x_n\},\ n \in N,$$

де

$$y_n = \mathbf{Me}(x_{n-k},\dots x_n,\dots x_{n+k}).$$

Припускається, що послідовність $\{x_n\}$ є нескінченною.

Для послідовностей скінченного обсягу значень виникає питання про визначення медіани в примежових областях послідовності.

У цьому разі медіанним фільтром відбирається число елементів ряду, менше за ширину апертури, і медіана зазвичай обраховується з меншого за $2k+1$ обсягу елементів ряду.

Пізніше завдяки значній обчислювальній простоті і швидкодії цей метод фільтрації був застосований для обробки зображень та інших масивів інформації значного обсягу.

Порівняно з класичною лінійною фільтрацією нижніх частот медіанна фільтрація має низку суттєвих переваг:

– вона зберігає різкі перепади, у той час як лінійна низькочастотна фільтрація їх згладжує;

– вона дуже ефективна при згладжуванні імпульсного шуму.

Водночас існує значний клас задач, пов’язаних з обробкою кутових спостережень, в яких застосування медіанної фільтрації також може виявитися ефективним засобом статистичного згладжування.

До таких задач належать, наприклад, визначення фазової характеристики сигналів в умовах дії завад, підвищення завадостійкості передачі інформації в системах з фазовою маніпуляцією тощо.

Мета статті – дослідження властивостей вибіркової кругової медіани в задачах обробки результатів кутових спостережень на основі медіанної фільтрації.

Постановка завдання

Проводиться серія n незалежних вимірювань випадкового кута $\psi(\omega) \in [0,2\pi)$ зі щільністю розподілу ймовірностей $p(\theta)$, $\theta \in [0,2\pi)$.

Вимірювання проводиться за незмінних умов. Результатом вимірювань є вибірка значень кута
$\{\theta_i, i = \overline{1,n}\}$, $\theta_i \in [0,2\pi)$,
об'єму
$n = 2k + 1$, $k \in N$,
яка розглядається як реалізація випадкового вектора
$\boldsymbol{\psi}(\omega) = (\psi_1(\omega),...\psi_n(\omega))$
з n незалежними випадковими кутами
$\psi_i(\omega) \in [0,2\pi)$, $\omega \in \Omega$,
де ω – елементарна подія з простору подій Ω, $i = \overline{1,n}$.

Необхідно дослідити особливості визначення вибіркової кругової медіани під час виконання медіанної фільтрації.

Розв'язок поставленої задачі

Уведемо визначення. Круговою медіаною неперервного розподілу на колі $G(\theta)$, $\theta \in [0, 2\pi)$ випадкового кута $\psi(\omega)$ називатимемо те значення кута $\theta_m \in [0,2\pi)$, для якого значення $Q(\theta_m - 0{,}5\pi)$ максимальне та яке є одним із розв'язків рівняння

$$Q(\theta_m) = F(\theta_m + \pi) - F(\theta_m) - 0{,}5 = \int_{\theta_m}^{\theta_m+\pi} p(\theta)d\theta - 0{,}5 = 0, \tag{1}$$

де $F(\theta_m)$ – значення функції розподілу випадкового кута в точці θ_m; $p(\theta)$ – щільність розподілу ймовірностей випадкового кута $\psi(\omega)$ [3].

Як зразок розглянемо розподіл зі щільністю ймовірності Мізеса:

$$p_M(\theta) = \frac{1}{2\pi I_0(k)} \exp\{k\cos(\theta - \mu)\},\ |\mu| < \infty,$$

$$k > 0,\ \theta \in [0, 2\pi), \tag{2}$$

де k – параметр концентрації розподілу в околі μ;. μ – круговий середній напрямок випадкового кута.

За визначенням (1) медіана даного розподілу становить $\theta_m = 0{,}5$.

Спроба обчислити медіану за її визначенням для розподілів на прямій [2] дає помилкове значення $\theta_m' \approx 1{,}71$.

Розглянемо основні властивості кругової медіани.

Для одновершинних розподілів медіана завжди визначається однозначно.

Одновершинні неперервні розподіли на колі відрізняються тим, що в інтервалі $[0,2\pi)$ існують такі точки θ_1 і θ_2, що під час руху точки θ по одиничному колу від θ_2 до θ_1 в обох напрямках функція

$$p(\theta) = \frac{dF(\theta)}{d\theta}$$

є монотонно неспадною функцією.

У цьому разі точку θ_1 називають модою, а точку θ_2 – антимодою розподілу. Розглянута властивість сформульована і доведена у праці [3] у вигляді такої теореми.

Теорема. Якщо неперервний розподіл на колі зі щільністю розподілу ймовірності $p(\theta)$ одновершинний і

$p(\theta) \neq p(\theta + \pi)$ $\theta \in [0, 2\pi)$

для майже всіх θ, то кругова медіана в інтервалі $[0,2\pi)$ визначається однозначно.

Інша важлива властивість медіани стосується кругового середнього відхилення. У разі одновершинного розподілу кругове середнє відхилення досягає мінімуму в точці θ_m [3]. Кругове середнє відхилення $\delta(\nu)$ випадкового кута $\psi(\omega)$ відносно фіксованого кута ν визначається як математичне сподівання випадкового кута

$$\min\{\psi'(\omega),\ 2\pi - \psi'(\omega)\},$$

де $\psi'(\omega) = (\psi(\omega) - \nu)(\mathrm{mod}\, 2\pi)$, $x(\mathrm{mod}\, 2\pi)$ – операція обчислення лишку числа x за модулем 2π, тобто $\delta(\nu)$ визначається як величина

$$\delta(\nu) = \int_0^{\pi} \theta dF(\theta + \nu) + \int_{\pi}^{2\pi} (2\pi - \theta) dF(\theta + \nu).$$

Якщо розподіл випадкового кута симетричний і одновершинний, то круговий середній напрямок, медіана і мода збігаються.

Властивість кругової медіани мінімізувати кругове середнє відхилення, власне, і може бути покладена в основу застосування кругової медіанної фільтрації для статистичного згладжування результатів вимірювання випадкових кутів.

Під час оцінювання випадкових кутів за результатами отриманої вибірки $\{\theta_i, i = \overline{1,n}\}$, $\theta_i \in [0,2\pi)$ користуються оцінкою медіани у вигляді вибіркової кругової медіани [2]. Під вибірковою круговою медіаною розуміють кут, якому відповідає точка P на колі з такими властивостями:

– половина точок вибірки лежить з одного боку діаметра PQ;

– більшість точок вибірки ближче до P, ніж до Q.

На колі одиничного радіуса точками позначено значення певної вибірки кутів об'єму $n = 21$.

Діаметр PQ ділить коло на дві частини таким чином, що з кожного боку від діаметра знаходиться по 10 значень кутів (рис. 1).

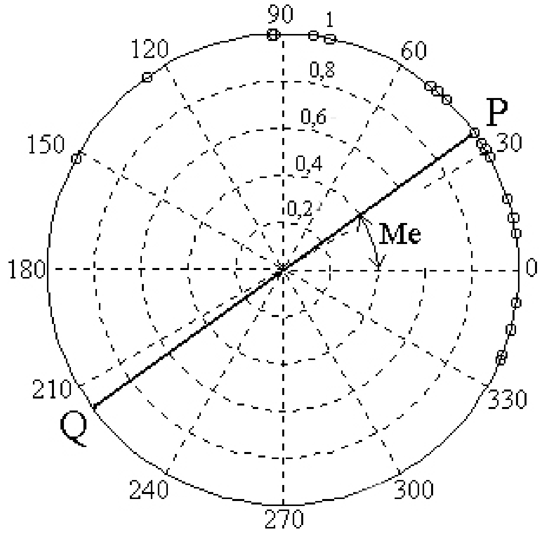

Рис.1. Вибіркова кругова медіана

Позначивши оператор одновимірної медіанної фільтрації на колі як $\mathbf{MFC}_{2k+1}$, результат на виході медіанного фільтра можна подати як

$\{\varphi_n\} = \mathbf{MFC}_{2k+1}\{\theta_n\},\ n \in N$,

де

$\varphi_n = \mathbf{Mec}(\theta_{n-k},...\theta_n,...\theta_{n+k})$,

де **Mec** – оператор кругової вибіркової медіани, який реалізує обробку даних відповідно до визначення.

У практиці обробки результатів кутових вимірювань часто мають справу з групованими даними. Такі дані отримують тоді, коли все коло розбивають на m рівних клас-інтервалів величиною $\frac{2\pi}{m}$ і спостерігають частоти V_i, $i \in [1, m]$ попадання кутів в певні клас-інтервали. У цьому разі доцільно обирати об'єм вибірки парним, тобто $n = 2k,\ k \in N$. Якщо виконується умова $V_i < \frac{n}{2} < V_{i+1}$, то, застосовуючи інтерполяцію, можна визначити вибіркову кругову медіану за групованими даними так:

$$\mathbf{Me}_{\text{Г}}\{\theta_1,..\theta_n\} = \left[\frac{2\pi}{m}\left(i + \frac{0.5n - V_i}{V_{i+1} - V_i}\right) + \pi k\right](\mathrm{mod}2\pi), \quad (3)$$

де $k = 0$ в разі концентрації значень кутів в околі точки Q_i; $k = -1$ в разі концентрації значень кутів в околі точки P_i.

Моделювання задачі визначення вибіркової кругової медіани під час статистичного згладжування результатів кутових вимірювань

Приклад 1. Нехай необхідно знайти медіану для такої вибірки випадкового кута 1,777; 0,016; 2,186; 0,506; 0,751; 0,338; 0,072; 0,890; 1,202; 0,421; 4,221 рад об'єму n=11.

З рис. 2 видно, що для θ_m=0,506 половина значень вибірки розміщується праворуч, а друга половина – ліворуч діаметра, який проходить через медіану.

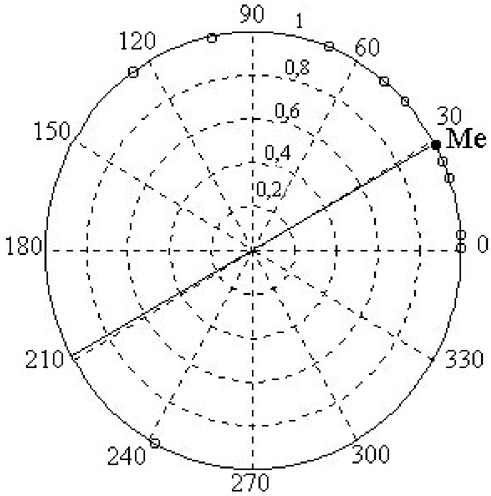

Рис. 2. Результати побудови значень вибірки та вибіркової медіани на колі

Наступний приклад ілюструє методику розрахунку вибіркової медіани за групованими даними.

Приклад 2. Обчислимо медіану за групованими даними в два етапи. На першому підготовчому етапі сформуємо вибірку випадкового кута об'єму $n = 50$ з розподілом (2) і параметрами $k = 1$, $\mu = 0{,}5$ рад (~28,6°) за методикою, викладеною в праці [4]. На другому етапі безпосередньо обчислимо значення медіани.

За результатами моделювання даних отримано таку вибірку об'єму n=50: 27, 75, 14, 84, 331, 70, 44, 62, 282, 192, 347, 321, 89, 169, 5, 113, 85, 330, 328, 90, 304, 199, 349, 22, 96, 0, 40, 57, 338, 55, 44, 79, 324, 32, 101, 84, 57, 67, 335, 292, 335, 86, 309, 23, 7, 66, 66, 93, 353, 60.

Утворимо з цієї вибірки варіаційний ряд:

0, 5, 7, 14, 22, 23, 27, 32, 40, 44, 44, 55, 57, 57, 60, 62, 66, 66, 67, 70, 75, 79, 84, 84, 85, 86, 89, 90, 93, 96, 101, 113, 169, 192, 199, 282, 292, 304, 309, 321, 324, 328, 330, 331, 335, 335, 338, 347, 349, 353.

Неважко переконатися в тому, що за не групованими даними значення медіани становить $\theta_m = 44^{\circ}$.

Згрупуємо значення варіаційного ряду. З цією метою поділимо коло, наприклад, на дванадцять клас-інтервалів по 30°. Розрахунок медіани виконаємо табличним способом [3], який спрощує допоміжні розрахунки. Результати розрахунків для цього прикладу наведено в таблиці.

Розрахунок медіани варіаційного ряду за групованими даними

Права межа клас-інтервалів V_i, град	Частота θ_i	Накопичена частота	Накопичена частота кутів, менших θ_i-180°	Частота в інтервалі (θ_i-180°, θ_i)	Інтервал (θ_i-180°, θ_i), град
30	7	7	–	–	–
60	8	15	–	–	–
90	13	28	–	–	–
120	4	32	–	–	–
150	0	32	–	–	–
180	1	33	0	33	0–180
210	2	35	7	28	30–210
240	0	35	15	20	60–240
270	0	35	28	7	90–270
300	2	37	32	5	120–300
330	6	43	32	11	150–330
360	7	50	33	17	180–360

Зокрема, у першій колонці таблиці наведено значення правої межі клас-інтервалів у градусах, у другій – частоти V_i, $i=\overline{1{,}2}$ попадання кутів у відповідний клас-інтервал, у третій – накопичені частоти, тобто значення $\sum_{j=1}^{i} V_j$ від першого до i-го клас-інтервалу, у четвертій – накопичена частота кутів менших θ_i-180° (повторені дані першої половини колонки 3), у п'ятій – частота в інтервалі величиною 180° та різними початковими кутами, тобто в інтервалах (θ_i–180°, θ_i), у шостій – значення інтервалів (θ_i–180°, θ_i).

Частоти кутів в інтервалах (θ_i –180°, θ_i) отримано шляхом віднімання значень четвертої колонки від значень третьої (п'ята колонка таблиці).

З таблиці видно, що половина значень всіх кутів (тобто 25 значень кутів) розміщена між частотами сьомого (i = 7) і восьмого інтервалів. Менша частина кутів знаходиться в інтервалі (120°, 300°).

З урахуванням цього та згідно з виразом (3) маємо значення вибіркової кругової медіани

$$\theta_{г} = \left[30\left(7-\frac{25-28}{20-28}\right)-180\right]\left(\mathrm{mod}\,360\right) \approx 41{,}2^{\circ}.$$

Отже, $\theta_{Г}$ наближається до медіани θ_m. Отриманий результат лежить у межах статистичної похибки оцінювання.

Висновки

1. У разі статистичної обробки результатів вимірювання випадкових кутів з одномодальними законами розподілів імовірностей вибіркова кругова медіана є важливою статистикою, яка вказує на ту точку на колі, в околі якої спостерігається найбільша концентрація вибіркових значень кутів.
2. Наведено визначення та основні властивості вибіркової кругової медіани, розглянуто особливості її обчислення порівняно з вибірковою медіаною випадкових величин, розподілених на прямій.
3. Розглянуто приклади визначення вибіркової кругової медіани, які демонструють методику її обчислення і засвідчують правильність отриманих формул.
4. Наведені дані можуть бути застосовані під час розробки та реалізації медіанної фільтрації результатів аналізу випадкових кутів.

Література

1. *Математическая* энциклопедия. Т. 3/ Гл. ред. И.М. Виноградов. – М.: Сов. энцикл., 1982. – 1184 с.
2. *Теорія* ймовірностей, випадкові процеси та математична статистика / В.П. Бабак, Б.Г. Марченко, М.Є. Фриз. – К.: Техніка, 2004. – 288 с.
3. *Мардиа К.* Статистический анализ угловых наблюдений. – М.: Наука. Гл. ред. физ.-мат. лит-ры изд-ва, 1979. – 240 с.
4. *Куц Ю.В., Куц Н.Е.* Моделирование случайных величин, распределенных на окружности // Зб. наук. пр. – К.: Ін–т проблем моделювання в енергетиці, 2003. – Вип. 21. – С. 101–106.

Стаття надійшла до редакції 15.11.05.

Предложена методика определения выборочной круговой медианы при статистической обработке реализаций случайных углов. Приведен пример использования предложенной методики на основе моделирования в системе MatLab.

Definition a round mediane in statistic processing of the random angles is offered. The realization of the offered technique is considered on an example of the use model in MatLab.