

УДК 006.91:519.2

С.В. Шенгур

Національний авіаційний університет, Київ

ОПРАЦЮВАННЯ НАДМАЛИХ ВИБІРОК ВИПАДКОВИХ КУТІВ

Запропоновано методику опрацювання вибірок обмеженого обсягу, одержаних за результатами кутових спостережень. Наведено результати експериментальних досліджень.

Ключові слова: випадковий кут, вибіркові кругові характеристики, невизначеність, бутстреп.

Вступ

Постановка проблеми. Результати вимірювання, у відповідності до вимог національних та міжнародних стандартів [1, 2], повинні супроводжуватись показниками точності. В якості таких показників найчастіше використовують розширену невизначеність U або довірчий інтервал Δ_d .

Довірчим інтервалом результатів вимірювання випадкового кута є довжина найменшої дуги кола одиничного радіуса, що включає найбільш достовірні значення цієї статистики [3].

Невизначеність вимірювання (англ. measurement uncertainty) – параметр, що пов'язаний з результатом вимірювання та характеризує розсіяння значень, які обґрунтовано могли б бути приписані вимірюваній величині [2]. Оскільки на практиці вимірюваній величині приписуються значення, отримані в результаті вимірювання, то вказаний показник характеризує розсіяння результатів вимірювання.

Розподіл результату вимірювання випадкової величини, ґрунтуючись на результатах попередніх вимірювань, як правило приймають близьким до гауссівського [3]. За наявності достатньої статистики виконують перевірку на гауссовість за одним з відомих критеріїв – Колмогорова, χ^2 -квадрат.

Проте, такі критерії передбачають побудову за даними спостереження гістограми, а отже, мають обмеження застосування до вибірок менше 35 значень. Та навіть за такого обсягу даних дають нестабільний результат. На практиці не рідко трапляються випадки, коли вибірка, одержана за даними спостереження, не перевищує 10 значень. Це зумовлено, передусім, високою вартістю експерименту або неможливістю його повторного проведення. Загальнозначеного критерію надмалої вибірки немає. Будемо користуватися цим терміном, коли $n \leq 9$ [4]. Опрацювання надмалих вибірок вимагає специфічного підходу, за винятком ситуацій, коли закон розподілу випадкового кута є відомим.

Традиційний метод розрахунку та представлення результату вимірювання випадкових кутових даних передбачає отримання за результатами вимірювання вибірки кутів певного обсягу, визначення

вибіркових кругових середнього та стандартного відхилення, оцінки невизначеності вимірювання як симетричного відносно вибіркового кругового середнього інтервалу значень кутів, величина якого формується як подвоєний добуток вибіркового кругового стандартного відхилення та фактор покриття для заданого рівня довіри та відомого розподілу ймовірності випадкової кутової величини. Застосування традиційного методу до надмалих вибірок, закон розподілу яких апріорно не відомий, є необґрунтованим та може призвести до грубих похибок.

Аналіз досліджень і публікацій. У роботі [5] наведено та експериментально апробовано методику оцінювання основних вибіркових кругових характеристик – середнього, моди, медіани, за вибірками обсягом до 100 значень.

Мета роботи – розробити та експериментально дослідити методику оцінювання точкових та інтервальних характеристик випадкових кутових величин за вибірками надмалої обсягу та апріорно невідомого розподілу.

Вирішення задачі

Задачу оцінювання результатів вимірювання випадкових кутів за вибірками надмалої обсягу запропоновано вирішувати на базі бутстреп-методу [4, 5]. Бутстреп-методи статистичного опрацювання даних відомі з 70-х років минулого століття. Їх перевага полягає у збільшенні статистики без збільшення вихідних даних.

Методика оцінювання довірчого інтервалу для кругового середнього значення вибірки ВК

$$\Theta = \{\theta_1, \dots, \theta_j, \dots, \theta_n\}$$

обсягу $n = 3 \dots 9$ має наступний вигляд.

1. З вхідної вибірки ВК

$$\Theta = \{\theta_1, \dots, \theta_j, \dots, \theta_n\}$$

формується нову вибірку

$$\Theta^{(1)} = \{\theta'_1, \dots, \theta'_j, \dots, \theta'_n\}$$

обсягом n шляхом перестановки з поверненням.

2. Для сформованої вибірки

$$\Theta^{(1)} = \{\theta'_1, \dots, \theta'_j, \dots, \theta'_n\}$$

обчислюють значення оцінки кругового середнього $\bar{\theta}^{(1)}$ за:

$$\bar{\theta} = \left\{ \arctg \frac{S}{C} + \frac{\pi}{2} \{ 2 - (\text{sign}S)[1 + \text{sign}C] \} \right\}, \quad (1)$$

де

$$C = \frac{1}{n} \sum_{j=1}^n \cos \theta_j, \quad S = \frac{1}{n} \sum_{j=1}^n \sin \theta_j. \quad (2)$$

3. З вибірки $\Theta^{(1)} = \{\theta'_1, \dots, \theta'_j, \dots, \theta'_n\}$ формують нову вибірку ВК $\Theta^{(11)} = \{\theta''_1, \dots, \theta''_j, \dots, \theta''_n\}$ обсягом n шляхом перестановки з поверненням.

4. Для вибірки $\Theta^{(11)} = \{\theta''_1, \dots, \theta''_j, \dots, \theta''_n\}$ обчислюють значення оцінки кругового середнього $\bar{\theta}^{(11)}$.

5. Пункти 3 та 4 повторюють B разів (нехай $B=200$) та формують ряд $\bar{\theta}^{(11)}, \dots, \bar{\theta}^{(1b)}, \dots, \bar{\theta}^{(1B)}$.

6. Обчислюють значення ймовірнісних складових p та q :

$$p_0 = 0, \quad p_i = p_{i-1} + (1/B) I \left[\bar{\theta} < \bar{\theta}^{(1i)} \right], \quad i = 1, \dots, B \quad (3)$$

та

$$q_0 = 1, \quad q_i = q_{i-1} - (1/B) I \left[\bar{\theta} > \bar{\theta}^{(1i)} \right], \quad i = 1, \dots, B, \quad (4)$$

де $I[A] = 1$ у випадку, коли подія A істинна, та $I[A] = 0$ – коли неістинна.

7. Пункти 1-6 повторюють B разів та формують “бутстреп”-розподіл $\bar{\theta}^{(1)}, \dots, \bar{\theta}^{(b)}, \dots, \bar{\theta}^{(B)}$ та ряди P_1, \dots, P_B та Q_1, \dots, Q_B .

8. Значення “бутстреп”-розподілу розташовують у порядку зростання $\bar{\theta}^{(1)} \leq \dots \leq \bar{\theta}^{(B)}$. Для $100(P_{\hat{\alpha}\hat{\alpha}})\%$ довірчого інтервалу для невідомого кругового середнього значення знаходять цілочисельне i таке, що відноситься до максимального p_i і менше або дорівнює $\frac{1}{2}(1 - P_{\hat{\alpha}\hat{\alpha}})$ та цілочисельне k таке, що відноситься до максимального q_k і менше або дорівнює $\frac{1}{2}(1 - P_{\hat{\alpha}\hat{\alpha}})$. Тоді межі $100(P_{\hat{\alpha}\hat{\alpha}})\%$ довірчого інтервалу представляють як:

$$\left(\bar{\theta}^{(i+1)}, \bar{\theta}^{(k-1)} \right). \quad (5)$$

9. Значення складових розширеної невизначеності кругового середнього U_{mean}^- та U_{mean}^+ обчислюють як довжини дуг, що охоплюють вибірки $\bar{\theta}^{(1+1)}, \dots, \bar{\theta}^{(b)}, \dots, \bar{\theta}$ та $\bar{\theta}, \dots, \bar{\theta}^{(b)}, \dots, \bar{\theta}^{(B-1)}$. Значення довжин дуг обчислюють за формулою вибіркового кругового розмаху:

$$T_j = \theta_{j+1} - \theta_j, \quad j = 1, \dots, n-1, \quad T_n = 2\pi - \theta_n + \theta_1, \\ W = 2\pi - \max \{ T_1, \dots, T_n \}. \quad (6)$$

10. Формують результат вимірювання як:

$$\left((\bar{\theta} - U_{\text{mean}}^-) \bmod 2\pi, (\bar{\theta} + U_{\text{mean}}^+) \bmod 2\pi \right), P_{\hat{\alpha}\hat{\alpha}}. \quad (7)$$

Для проведення комп'ютерних експериментів з оцінювання вибіркового кругового середнього напряму та його показників точності за надмалими вибірками випадкових кутових величин застосовано методику [5]:

1. Формування симетричної (асиметричної) вибірки обсягом $N=100000$ значень випадкових кутів з діапазону $[0, 2\pi)$.

2. Формування з вибірки п.1 вибірок обсягу $n = 3 \div 9$ значень.

3. Обчислення усереднених значень вибіркового кругового середнього за результатами 1000 експериментів для кожного обсягу вибірки ВК з використанням традиційного методу та методу бутстреп.

4. Обчислення значень границь довірчих інтервалів для кругового середнього напряму з використанням методів бутстреп, Чебишева та традиційного методу для гауссівських розподілів.

5. Проведення порівняльного аналізу методів (рис. 3).

Для кожного обсягу вибірки з діапазону $n = 3 \div 9$ значень експеримент повторено 1000 разів для перевірки точності (правильності та прецизійності) оцінок, як основних показників якості результату вимірювання за ДСТУ ГОСТ ИСО 5725-1:2005.

Правильність оцінювалась як відхилення усередненого значення одержаної оцінки від істинного $\hat{\theta}_{\text{нб}}$ (умовно істинного – у випадку складених законів розподілу генеральної сукупності) значення оцінки кругової характеристики, що процентному співвідношенні має вигляд:

$$\delta_{\hat{\Delta}} = \frac{\hat{\theta}_{\text{нб}} - \hat{\theta}_{\hat{\Delta}}}{\hat{\theta}_{\text{нб}}} \cdot 100\% \quad \text{та} \quad \delta_{\hat{\Delta}} = \frac{\hat{\theta}_{\text{нб}} - \hat{\theta}_{\hat{\Delta}}}{\hat{\theta}_{\text{нб}}} \cdot 100\%. \quad (8)$$

Прецизійність оцінювалась порівнянням усереднених значень стандартних відхилень $\hat{\sigma}_{\hat{\Delta}}$ та $\hat{\sigma}_{\text{нб}}$ оцінок вибірових кругових характеристик.

Правильність оцінок вибіркового кругового середнього значення для двох досліджуваних методів однакова для симетричної та асиметричної сукупно-

стей на усіх значеннях обсягу вибірки з діапазону $3 \leq n \leq 9$. Алгоритм формування результату вимірювання наведено на рис. 1. Результати досліджень стандартних відхилень оцінок зведені в табл. 1.

Таблиця 1

Значення стандартних відхилень оцінки вибіркового кругового середнього та границь його розсіяння

Обсяг вибірки n	95% границі розсіяння оцінки		Стандартне відхилення	
	Традиційний метод	Бутстреп метод	$\hat{\sigma}_{\bar{\theta}}$	$\hat{\sigma}_{\Delta}$
Симетричний розподіл				
3	1,27; 2,18	1,27; 2,19	0,22	0,22
4	1,27; 2,22	1,27; 2,22	0,2	0,2
5	1,26; 2,19	1,27; 2,19	0,19	0,19
6	1,31; 2,03	1,30; 2,03	0,16	0,16
7	1,34; 2,22	1,34; 2,22	0,17	0,16
8	1,39; 2,04	1,39; 2,03	0,14	0,14
9	1,42; 1,94	1,42; 1,94	0,13	0,13
Асиметричний розподіл				
3	1,17; 2,64	1,17; 2,64	0,22	0,22
4	1,23; 2,65	1,23; 2,66	0,19	0,19
5	1,21; 2,35	1,20; 2,35	0,18	0,18
6	1,29; 2,39	1,29; 2,40	0,18	0,18
7	1,3; 2,26	1,29; 2,23	0,16	0,16
8	1,37; 2,29	1,37; 2,30	0,15	0,15
9	1,37; 2,16	1,36; 2,16	0,14	0,14

Результати оцінювання вибіркового кругового середнього напряму традиційним та бутстреп методами в межах похибки статистичного експерименту не відрізняються як у випадку симетрично розподіленої генеральної сукупності, так і асиметрично розподіленої. Розсіяння значень ВКС, обчислені двома методами, однакові.

Результати дослідження границь довірчого інтервалу для ВСК відображено на графіку, наведеному на рис. 2. Обчислення довірчого інтервалу проведено з використанням методів: традиційного (крива 1), бутстреп (крива 2) та Чебишева (крива 3). Дослідження проведено з симетричною (рис. 2, а) та асиметричною (рис. 2, б) генеральними сукупностями.

Аналізуючи графік, можна зробити висновок про те, що “бутстреп” метод враховує асиметричність генеральної сукупності, до якої належить вибірка ВК та забезпечує найменший довірчий інтервал навколо середнього напряму порівняно з іншими методами. На вибірках обсягом 3...7 значень метод Чебишева забезпечив кращий результат порівняно з традиційним. Враховуючи цю залежність, можна рекомендувати до застосування метод Чебишева для опрацювання надмалих вибірок ВК обсягом до 7 значень у

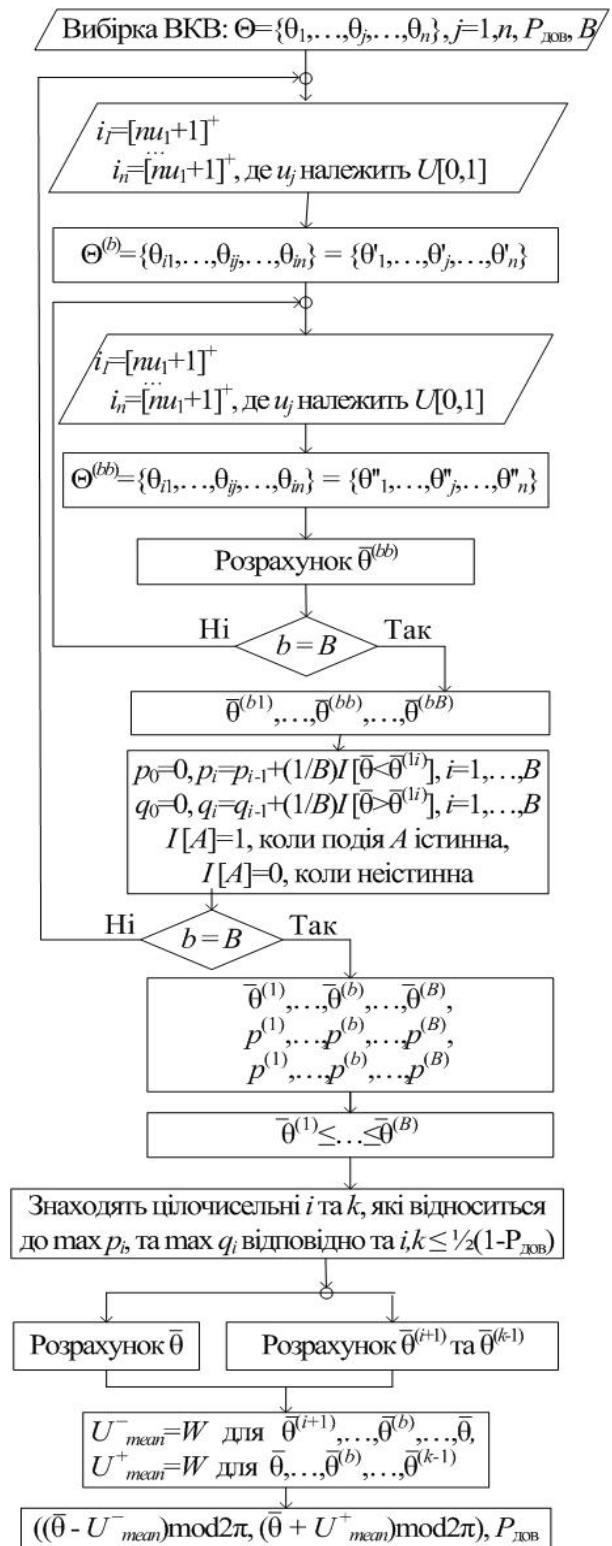


Рис. 1. Алгоритм формування результату вимірювання

випадках, коли сформульовані вимоги до обмеження затрат продуктивності обчислювальних ресурсів. Графік процентного співвідношення дов. Враховуючи цю залежність, можна рекомендувати до застосування метод Чебишева для опрацювання надмалих вибірок ВК обсягом до 7 значень у випадках, коли сформульовані вимоги до обмеження затрат продук-

тивності обчислювальних ресурсів. Графік процентного співвідношення довжини жини довірчого інтервалу, обчисленого “бутстреп” методом порівняно з традиційними (крива 1) та методом Чебишева (крива 2) наведено на рис. 3.

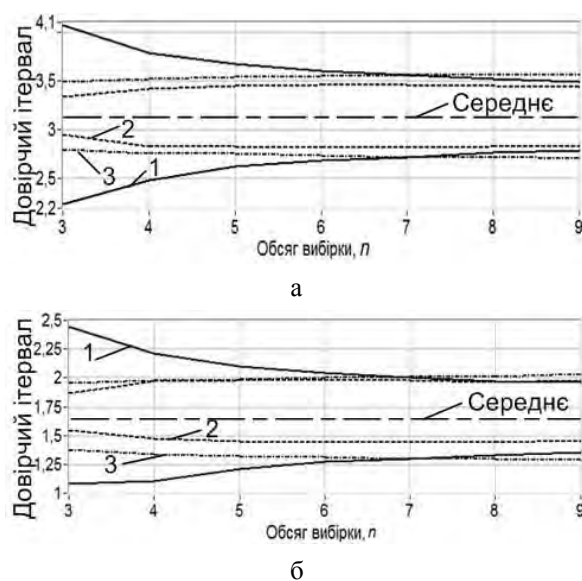


Рис. 2. Графіки дослідження границь довірчого інтервалу для ВКС за надмалими вибірками ВК; генеральна сукупність належить до: а – симетричного розподілу; б – асиметричного розподілу

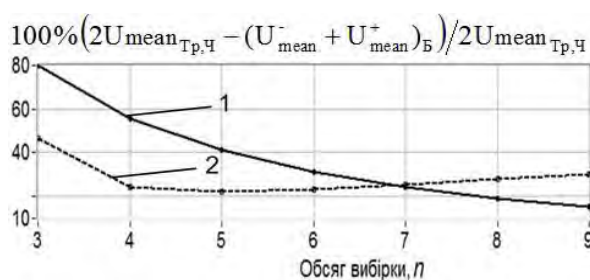


Рис. 3. Графік дослідження процентного співвідношення довжин довірчого інтервалу, обчислених досліджуваними методами

З графіку видно, що застосування методу “бутстреп” під час оцінювання довірчого інтервалу для кругового середнього напрямку забезпечує зменшення його значення на 12-80% порівняно з традиційним та на 20-47% порівняно з методом Чебишева в інтервалі значень обсягу вибірки $n \in (3, 9)$.

Висновки

У роботі запропоновано та експериментально досліджено методику опрацювання надмалих вибірок випадкових кутів. Досягнуто підвищення точності інтервального оцінювання вибіркового кругового середнього, обчисленого за надмалими вибірками (3-9 значень).

Підвищення точності, одержане запропонованим методом, виражене довжиною довірчого інтервалу та становить: відносно традиційного методу 12-80%; відносно методу Чебишева – 20-47%.

Список літератури

1. ДСТУ 2681-94 Метрологія. Терміни та визначення: Чин. з 1995-01-01. – К.: Держстандарт України, 1994. – 68 с.
2. Guide to the expression of uncertainty in measurement (GUM): First edition. – ISO, 1993. – 101 p.
3. Куц Ю.В. Статистична фазометрія / Ю.В. Куц, Л.М. Щербак – В.: Тернопіль, 2009. – 383с.
4. Fisher N.I. Statistical analysis of circular data. / N.I. Fisher. – Cambridge: Cambridge University Press, 2000. – 277 p.
5. Шенгур С.В. Опрацювання вибірок випадкових кутів з апріорно невідомого розподілу / С.В. Шенгур // Системи обробки інформації. – Х.: ХВПС, 2013. – Вип. 8 (115). – С. 130-135.
6. Куц Ю.В. Подання результату кутових вимірювань в концепції невизначеності / Ю.В. Куц, С.В. Шенгур // Системи обробки інформації. – Х.: ХВПС, 2013. – Вип. 3 (110). – С. 97-100.
7. Kuts Y.V. Circular data point and interval evaluation / Y.V. Kuts, S.V. Shengur, A.V. Dergunov // «Aviation in the XXI-st century 2012»: proc. VI World congress: – К.:NAU, 2012. – P. 1.6.5 – 1.6.7.
8. Mardia K.V. Statistics of directional data / K.V. Mardia, P.E. Jupp. – London: Academic Press Inc., 1972. – 415 p.
9. Kuts Y.V. Circular measurement data modeling and statistical processing in LabView / Y.V. Kuts, S.V. Shengur, L.M. Shcherbak // «MRRS-2011»: proceedings III International symposium: – К.:NAU, – P. 317-320.
10. Jammalamadaka S. Rao. Topics in circular statistics / S. Rao Jammalamadaka, A. SenGupta. – Singapore: World Scientific Publishing Co. Pte. Ltd., 2001 – 322 p.
11. Efron B. An introduction to the bootstrap / B. Efron, R. Tibshirani. – Boca Raton, FL: Chapman&Hall/CRC, 1993. – 436 p.

Надійшла до редколегії 27.02.2014

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

ОБРАБОТКА СВЕРХМАЛЫХ ВЫБОРОК СЛУЧАЙНЫХ УГЛОВ

С.В. Шенгур

Предложена методика обработки выборок ограниченного объема, полученных по результатам угловых наблюдений. Приведены результаты экспериментальных исследований.

Ключевые слова: случайный угол, выборочные круговые характеристики, неопределенность, бутстреп.

PROCESSING OF VERY SMALL RANDOM CIRCLE SAMPLES

S.V. Shengur

The technique of very small random circle samples processing is proposed. Experimental researches are shown.

Keywords: Random circle, sample circular characteristics, uncertainty, bootstrap.