

*І. МОЩЕНКО, канд. техн. наук, О. НІКІТЕНКО, канд. техн. наук,
Ю. КОЗЛОВ, канд. техн. наук, Ю. ЖАРКО, канд. техн. наук*

ОСОБЛИВОСТІ СТАТИСТИЧНОЇ ОБРОБКИ ДАНИХ ЗАСОБАМИ СИСТЕМ КОМП'ЮТЕРНОЇ МАТЕМАТИКИ

Вступ

Під час досліджень та оцінювання параметрів фізичних коливальних систем виникають певні питання, які, на перший погляд, не мають явних точок контакту одна з одною:

1. Моделювання поведінки таких систем дозволяє теоретично визначити взаємний вплив одних параметрів системи на інші;
2. Механізми визначення фізичних явищ та параметрів системних конструктивних вузлів, які мають вплив на характеристики цих систем;
3. Визначення параметрів, які впливають на поведінку системи загалом або окремих її вузлів;
4. Визначення частотних характеристик систем зі схрещеними полями;
5. Створення системи оцінювання вибраних параметрів;
6. Вибір методів дослідження частотних параметрів;
7. Вимірювання параметрів частоти (збирання даних);
8. Вибір методів обробки зібраних даних;
9. Обробка зібраних статистичних даних.

З іншого боку, основними розділами теоретичної метрології традиційно є: основа забезпечення єдності вимірювань (включаючи стандартизацію одиниць фізичних величин, відтворення та методи передачі їх розміру); теорія помилок (невизначеностей) та методи оцінювання результатів вимірювань; методи та засоби виміральної техніки.

Зараз методи статистичної обробки використовують не тільки у виробництві, але також у плануванні, розробці системи маркетингу, матеріально-технічного постачання тощо. Особливу увагу приділяють якості планування, дизайну, виробництва продукції, але про якість обробки результатів вимірювань майже не згадують. Таким чином, передбачається, що точність обробки результатів є абсолютною.

Будь-яке вимірювання частоти або періоду (часу) закінчується обробкою отриманих результатів. Сьогодні обробку експериментальних даних часто здійснюють за допомогою комп'ютера. Вважається, що комп'ютерні розрахунки мають абсолютну точність, але це не відповідає дійсності. З іншого боку, статистичні розрахунки без застосування комп'ютера є працеємкими і вимагають використання багатьох таблиць стандартних розподілів кватилів. Спеціальні статистичні пакети вимагають від фахівців високого рівня знань в галузі математичної статистики.

Таким чином, мета цієї роботи – порівняння результатів статистичних обчислень, отриманих під час дослідження коливальних станів електровакуумних приладів зі схрещеними полями за допомогою різних математичних пакетів з результатами, що розраховані за теоретичними формулами.

Матеріали і методи

Електровакуумні прилади зі схрещеними полями мають високий рівень шуму. Нелінійна взаємодія електронного пучка та електромагнітних хвиль призводить до збудження хаотичних або комбінованих коливань. Було проведено дослідження спектру низько- і високочастотних режимів для доведення хаотичної поведінки коливань, що збуджуються в електроваку-

умних приладах зі схрещеними полями [1 – 3]. Експериментальні дослідження коливальних станів електровакуумних приладів зі схрещеними полями вимагають проведення великої кількості багаторазових вимірювань частоти. Тому під час комп'ютерної обробки результатів експериментальних досліджень постало питання відповідності статистичних характеристик, отриманих за допомогою математичних пакетів, характеристикам, розрахованим за теоретичними формулами.

За необхідності отримання числових результатів вимірювання аналіз помилок повинен бути невід'ємною частиною будь-якого серйозного розрахунку. Вхідна інформація часто не достатньо точна, оскільки дані, що використовують, найчастіше є експериментальними даними або базуються на наближеній оцінці. Крім того, обчислювальні процеси самі можуть додавати до результатів деякі помилки. Коли вирішують конкретну проблему, ми маємо справу з трьома основними типами помилок: помилки, що містяться у вхідній інформації; помилки, обумовлені обмеженням нескінченного математичного процесу кінцевою кількістю операцій (помилки обмеження); помилки, що виникають внаслідок необхідності подавати число у формі кінцевої послідовності цифр (помилки округлення).

Помилки в початковій інформації виникають в результаті неточності вимірювань або через неможливість подати необхідне значення кінцевим дробом. Помилки, що містяться у вхідній інформації, визначають точність результатів розрахунку, незалежно від методу, за допомогою якого ці розрахунки здійснюються. Два інших типи помилок – помилки обмеження та помилки округлення – визначають числовими методами, що використовують для вирішення проблеми. Навіть якщо ми припустили, що вхідна інформація не містить жодних помилок, а всі обчислювальні процеси є скінченними і не призводять до помилок обмеження, то у цьому випадку існує третій тип помилок – помилки округлення. Під час чисельного аналізу однією з найважливіших проблем є питання про те, як помилка, що виникає в певному місці під час обчислення, розподіляється в майбутньому, незалежно від того, стає її вплив більшим чи меншим під час виконання наступних операцій.

Проаналізуємо процес накопичення похибок при обчисленні найбільш часто обчислюваного параметра – дисперсії

$$D = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2. \quad (1)$$

Наведена формула складається з таких операцій: віднімання $x_i - \bar{x}$; множення (піднесення до ступеня) $(x_i - \bar{x})^2$; додавання $\sum_{i=1}^n (x_i - \bar{x})^2$; кінцеве ділення. Добре відомо, якщо потрібно додавати або віднімати послідовності довгих чисел (підсумовування) з метою зменшення похибки обчислення, спочатку необхідно знайти суму найменших чисел і потім переходити до більших. Це вимагає додаткових операцій під час написання коду програми. Ці операції призначені для того, щоб попередньо відсортувати значення перед отриманням суми. У таких випадках відносна похибка підрахунку суми не перевищуватиме $5 \cdot 10^{-t}$, де t – кількість значущих цифр. Якщо це можливо, слід уникати двох майже однакових чисел (обчислення віднімання). Формули, що містять таке віднімання, дуже часто можуть бути перетворені таким чином, щоб уникнути такої операції. Ця вимога, як і попередня, вимагає додаткових операцій при написанні програмного коду. У будь-якому випадку, щоб зменшити накопичення помилок під час обчислень, бажано зменшити кількість необхідних арифметичних операцій. Таким чином, під час обробки експериментальних результатів за допомогою комп'ютерних програм необхідно враховувати вищезазначені аспекти, якщо створюють оригінальну обчислювальну програму. Якщо використовують спеціальне програмне забезпечення, варто ознайомитися з алгоритмами обчислення необхідних характеристик, щоб зменшити накопичення похибок розрахунку [4].

Авторами розглянуто обробку даних за допомогою найпопулярніших пакетів: електронних таблиць Excel та систем комп'ютерної математики (СКМ) Maple, MatLab та MathCad.

Дві останні системи базуються на ядрі Maple. Більшість існуючих математичних пакетів дозволяють користувачам працювати з випадковими величинами, включаючи популярні пакети. У цих пакетах розділ статистики має власну розроблену систему команд для обслуговування прикладних завдань. Статистичні команди призначені для тих категорій користувачів, яким потрібне середовище, що дозволяє легко переходити від однієї математичної спеціалізації до іншої.

Порівняння результатів

Особливості обробки статистичних даних, отриманих під час дослідження коливальних станів електровакуумних приладів зі схрещеними полями, досліджено шляхом обробки вибірки з 80 значень частоти генерації магнетрону (табл. 1) за допомогою популярних математичних пакетів Excel, Maple, Matlab та MathCad та порівняння отриманих результатів з розрахунками за теоретичними формулами.

Таблиця 1

Sample (частота GHz)

13,39	13,46	13,26	13,59	13,54	13,42
13,42	13,53	13,33	13,36	13,37	13,45
13,38	13,55	13,43	13,44	13,31	13,32
13,53	13,29	13,50	13,34	13,37	13,44
13,51	13,24	13,44	13,33	13,33	13,58
13,30	13,34	13,53	13,25	13,54	13,50
13,40	13,54	13,48	13,28	13,32	13,36
13,28	13,55	13,48	13,49	13,26	13,40
13,53	13,43	13,34	13,33	13,26	13,36
13,57	13,50	13,52	13,58	13,30	13,62
13,43	13,37	13,39	13,66	13,50	13,40
13,42	13,40	13,23	13,38	13,31	13,47
13,57	13,28	13,45	13,34	13,64	13,56
13,40	13,31				

Результати статистичної обробки наведено в табл. 2.

Таблиця 2

Результати обробки

Математичний пакет	Математичне сподівання, ГГц	Дисперсія, ГГц ²	Стандартне відхилення, ГГц	Коефіцієнт асиметрії	Коефіцієнт ексцесу
Теоретичний	13,42	0,01136203	0,10659280	0,194202	2,0440198
Excel	13,42	0,0113620	0,1065928	0,201703	0,884131
Maple	13,42	0,011362	0,106593	0,196660	2,069893
Matlab	13,42	0,0114	0,1066	0,1979	2,0961
MathCad	13,42	0,011	0,107	0,202	0,884

Результати розрахунків за допомогою усіх пакетів дають однакові результати для математичного сподівання, дисперсії та стандартного відхилення. Щодо коефіцієнтів асиметрії та ексцесу, то більшість результатів не збігаються.

Результати побудови гістограми для зразкових значень наведено на рис. 1 – 5.

На рисунках видно, що гістограми, які побудовані вручну та за допомогою СКМ Maple, Matlab та MathCad, однакові. А гістограма, яка побудована за допомогою пакету Excel, має багато відмінностей.

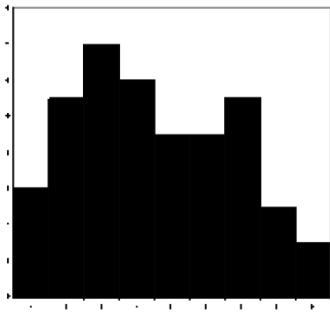


Рис. 1. Теоретична гістограма

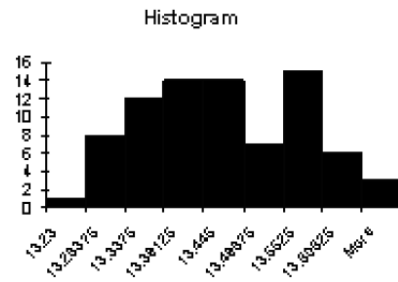


Рис. 2. Гістограма Excel

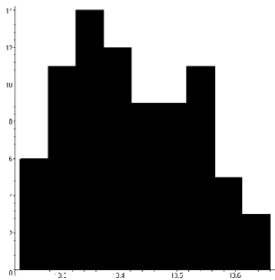


Рис. 3. Гістограма Maple

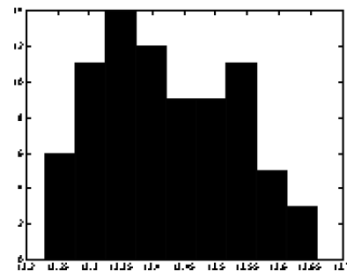


Рис. 4. Гістограма Matlab

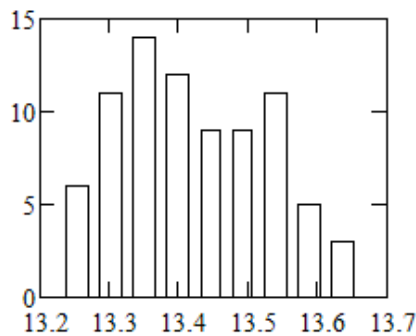


Рис. 5. Гістограма MathCad

Обговорення

Аналіз результатів розрахунків показав, що відмінність отриманих значень коефіцієнтів асиметрії та ексцесу обумовлений різними визначеннями цих показників вищезазначеними пакетами.

А. Коефіцієнт асиметрії

В теорії коефіцієнт асиметрії, який характеризує асиметричність функції розподілу, визначають як

$$\frac{m_3}{\sigma^3}; \quad (2)$$

де m_3 – центральний момент третього порядку, що визначають як $m_3 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3$, де n – розмір вибірки; x_i – виміряне значення; \bar{x} – математичне сподівання (середнє арифметичне) $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$; σ – стандартне відхилення.

В пакетах Excel і MathCad коефіцієнт асиметрії розраховуються таким чином [5, 8]:

$$\frac{n}{(n-1)(n-2)} \frac{m_3}{\sigma^3}. \quad (3)$$

В СКМ Maple коефіцієнт асиметрії розраховують таким чином [6]:

$$\frac{n}{(n-1)} \frac{m_3}{\sigma^3}. \quad (4)$$

В СКМ Matlab коефіцієнт асиметрії розраховують за теоретичною формулою [7].

Б. Коефіцієнт ексцесу

В теорії коефіцієнт ексцесу, який характеризує площинність функції розподілу, розраховують як

$$\frac{m_4}{\sigma^4} - 3; \quad (5)$$

де m_4 – центральний момент четвертого порядку, який визначають за формулою $m_4 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4$; 3 – враховує той факт, що коефіцієнт ексцесу нормального закону розподілу дорівнює 3.

В пакетах Excel і MathCad коефіцієнт ексцесу розраховують за формулою [5, 8]

$$\frac{n^2(n+1)}{(n-1)(n-2)(n-3)} \frac{m_4}{\sigma^4} - \frac{3(n-1)^2}{(n-2)(n-3)}. \quad (6)$$

В СКМ Maple коефіцієнт ексцесу розраховуються як [6]

$$\frac{n}{(n-1)} \frac{m_4}{\sigma^4}. \quad (7)$$

В СКМ Matlab коефіцієнт ексцесу розраховують за теоретичною формулою без приведення до нормального закону розподілу [7].

В. Гістограма.

Для з'ясування причин невідповідності гістограми було проаналізовано межі інтервалів варіаційного ряду вибірки, яку досліджували. Результати визначення меж інтервалів, розраховані за допомогою Excel, Maple, Matlab і MathCad, наведено в табл. 3.

Таблиця 3

Межі інтервалів

Excel									
Bin	13,23	13,28375	13,3375	13,39125	13,445	13,49875	13,5525	13,60625	More
Frequency	1	8	12	14	14	7	15	6	3
Maple and Matlab									
Bin	13,23000 ...13,277 78	13,27778 ...13,325 56	13,32556 ...13,373 33	13,37333 ...13,421 11	13,42111 1...13,46 889	13,46889 ...13,516 67	13,51667 ...13,564 44	13,56444 ...13,612 222	13,61222 ...13,66
Frequency	6	11	14	12	9	9	11	5	3
MathCad									
Average	13,254	13,302	13,349	13,397	13,445	13,493	13,541	13,588	13,636
Frequency	6	11	14	12	9	9	11	5	3

Порівняння меж інтервалів з табл. 2 показало, що в Excel межі інтервалів обчислюються з помилками. Це призводить до неправильного визначення кількості елементів у цих інтервалах.

Для того щоб правильно побудувати гістограму за допомогою пакету Excel, необхідно заздалегідь розрахувати межі інтервалів.

Висновки

Таким чином, коли обчислюють статистичні характеристики за допомогою комп'ютерних пакетів, необхідно:

- здійснити попереднє порівняння результатів теоретичних розрахунків та розрахунків за допомогою визначеного комп'ютерного математичного пакету;
- за наявності відмінностей з'ясувати, за якими формулами розраховані необхідні параметри, та вжити відповідних заходів для усунення можливих розбіжностей.

Список літератури:

1. Nikitenko O.M., Volovenko M.V. Chaotic behavior of oscillations in crossed-field electron vacuum devices // 2008 IEEE International Vacuum Electronics Conference, IVEC with 9th IEEE International Vacuum Electron Sources Conference, IVESC. 2008. P. 257-258.
2. Moshchenko I., Nikitenko O., Chen Xin. Research of low frequency components of a magnetron oscillator spectrum // Ukrainian Metrological Journal. 2019. No 4. P. 29 – 32.
3. Chen Xin, Ruzhentsev I. V., Nikitenko O.M. Pokaznyky yakosti elektrovakuumnykh system zi skhreshchenymu poliamy [Quality indicators of crossed-field electron vacuum systems] // Information Processing Systems. 2011. No 6 (96). P. 72 – 77. (in Ukrainian).
4. Korchakova A.S., Nikitenko O.M. Osoblyvosti statystychnoi obrobky danykh za dopomohoiu kompiutera [Features of statistical data processing using a computer] // Metrology and Instruments. 2014. No 1. P. 138 – 142. (in Ukrainian).
5. Microsoft Official Academic Course MICROSOFT EXCEL 2016
<https://www.dit.ie/media/ittraining/msoffice/MOAC Excel 2016 Core.pdf>
6. Maple User Manual, Maplesoft, a division of Waterloo Maple Inc., 2014
<https://www.maplesoft.com/documentation center/maple18/UserManual.pdf>
7. MATLAB The Language of Technical Computing, <https://web.stanford.edu/class/ee254/software/using ml.pdf>
8. MathCAD Tutorial, Colorado State University Student, www.engr.colostate.edu/ECE562/mathcad.pdf

Надійшла до редколегії 28.08.2021

Відомості про авторів:

Мощенко Інна Олексіївна – канд. техн. наук, Харківський національний університет радіоелектроніки, старший викладач кафедри метрології та технічної експертизи; Україна; e-mail: inna.moshchenko@nure.ua; ORCID: <https://orcid.org/0000-0002-2738-0037>

Нікітенко Олександр Миколайович – канд. техн. наук, доцент, Харківський національний університет радіоелектроніки, доцент кафедри метрології та технічної експертизи; Україна; e-mail: [nixonipe@gmail.com](mailto:nikonxipe@gmail.com); ORCID: <https://orcid.org/0000-0002-1082-5247>

Козлов Юрій Валентинович – канд. техн. наук, доцент, Харківський національний університет радіоелектроніки, доцент кафедри метрології та технічної експертизи; Україна; e-mail: yurii.kozlov@nure.ua; ORCID: <https://orcid.org/0000-0002-6165-4978>

Жарко Юрій Григорович – канд. техн. наук, Державне підприємство "Харківський регіональний науково-виробничий центр стандартизації, метрології та сертифікації", провідний інженер із стандартизації відділу оцінки відповідності продукції машинобудування; Україна; e-mail: 090sert@gmail.com; ORCID: <https://orcid.org/0000-0002-1328-567X>