

## **РОЗРОБКА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ ПРИКЛАДНОГО СТАТИСТИЧНОГО АНАЛІЗУ ВИПАДКОВИХ ВЕЛИЧИН**

*Олексій Клименко*

ORCID: <https://orcid.org/0009-0009-5164-1688>

Дніпровський національний університет імені Олеся Гончара, Дніпро, Україна

### **Вступ**

Математична статистика — наука, яка розробляє математичні методи систематизації та використання статистичних даних для наукових і практичних висновків. Математична статистика базується на теорії ймовірностей, яка дозволяє оцінити достовірність і точність отриманих висновків, зроблених на основі статистичного матеріалу. Наприклад, для оцінки необхідного розміру вибірки для отримання результатів необхідної точності у вибірковому дослідженні. Перевірка статистичних гіпотез є змістом одного з широких класів задач математичної статистики. Статистична гіпотеза — гіпотеза про тип розподілу та властивості випадкової величини, яка може бути підтверджена або спростована шляхом застосування статистичних методів до вибіркових даних [1-6].

На сучасному етапі розвитку прикладної статистики важливе місце займає питання перевірки нормальності вимірювань. За даними *Statistical Science Journals*, понад 60% статистичних досліджень, проведених у 2020 році, використовували різні методи перевірки нормальності, що підкреслює актуальність цього питання. Розробка спеціалізованого програмного забезпечення, автоматизація процесів перевірки нормальності відкриває нові можливості для статистиків, дозволяючи проводити аналіз швидше та ефективніше. Зокрема, за даними *Statistical Analysis System Institute*, використання автоматизованих систем дозволяє скоротити час аналізу на 40-50% [1, 3].

Прикладна математична статистика передбачає, що нормальний закон розподілу ймовірностей є фундаментальним для експериментальних вимірювань, які широко використовуються в практичних задачах. Ці вимірювання дають важливу інформацію про стан і положення об'єктів. Аналіз зразків вимірювань, отриманих з часом, дозволяє глибше зрозуміти досліджуване явище [2, 6].

Отже, можна зробити висновок, що розробка програмного забезпечення та проведення досліджень використовуючи сучасні інформаційні технології є актуальним і має велике практичне та наукове значення для сучасної статистики.

## МЕТА І ЗАВДАННЯ

При проектуванні та реалізації генераторів псевдовипадкових послідовностей необхідно дотримуватись наступних загально-прийнятих вимог: простота апаратної або програмної реалізації; максимальна швидкодія; максимальна наближеність послідовності отриманої на виході генератора до теоретичного закону розподілу (не обов'язково рівномірного); можливість керування вихідними параметрами; можливість роботи генераторів псевдовипадкових послідовностей в широкому діапазоні частот; можливість швидкого переналадження його роботи в залежності від вибору вихідних параметрів. Метою статті є розробка програмного забезпечення для прикладного статистичного аналізу випадкових величин.

### АЛГОРИТМ ГЕНЕРАЦІЇ ВИБІРКИ ВИПАДКОВОЇ ВЕЛИЧИНИ

У процесі проведення прикладних досліджень часто виникає потреба аналізувати групу процесів або об'єктів, що володіють певними якісними або кількісними характеристиками [2, 4, 6]. При виборі генератора псевдовипадкових величин слід враховувати наступні критерії: якість випадковості; швидкодія; простота реалізації; безпека; можливість налаштування; відповідність застосування. Тобто генератор має проходити стандартні тести на випадковість (наприклад, тестування NIST, Diehard), а період генератора повинен бути достатньо довгим, щоб уникнути повторення послідовностей. На рис. 1 наведено алгоритм методу статистичних випробувань, який застосовано у розробленому програмному забезпеченні.

Генератор повинен бути достатньо швидким для конкретних застосувань, особливо якщо потрібно генерувати великі обсяги випадкових чисел, але при цьому має бути легко реалізованим у потрібному програмному середовищі. Для криптографічних застосувань генератор має бути стійким до атак, тобто задовольняти вимоги криптографічної безпеки. Генератор повинен дозволяти налаштування параметрів для відповідності специфічним вимогам завдань. Вибір генератора повинен відповідати конкретним потребам проекту, будь-то моделювання, криптографія, наукові обчислення чи інші завдання. Загалом, важливо тестувати вибраний генератор у контексті конкретних завдань і вимог, щоб переконатися в його придатності та надійності.

Перевірка програмного генератора псевдовипадкових величин здійснюється за допомогою кількох основних методів і тестів, які оцінюють його статистичні властивості та відповідність критеріям випадковості [5, 6]. По перше – це візуальна перевірка: гістограма

розподілу та статистичні характеристики. Тобто, перевірка відображення частоти значень, які генерує генератор псевдовипадкових величин, для візуальної перевірки рівномірності розподілу, а також графічне представлення кількох послідовностей для виявлення можливих закономірностей.



Рисунок 1 – Алгоритм методу статистичних випробувань

Використання статистичних тестів:

- тест частоти (Frequency Test), за допомогою якого визначається перевірка, чи рівномірно розподілені значення генератора псевдовипадкових величин;
- тест на послідовність (Runs Test), який визначає перевірку довжин послідовностей однакових значень для виявлення нерівномірності;
- критерій Колмогорова-Смирнова, за яким відбувається перевірка, чи розподіл значень відповідає теоретично рівномірному розподілу;
- тест кумулятивної суми (Cumulative Sums Test), перевірка відхилень сум послідовностей від очікуваних значень;
- Chi-square тест ( $\chi^2$ -тест), перевірка відповідності частот значень до очікуваних частот у рівномірному розподілі. Також

використовуються більш складні статистичні тести: критерій Шермана; критерій Морана; критерій Фроцині; критерій Грінвуда-Кесенберрі-Міллера, критерій Чанга та Спірінга.

## ОПИС РОЗРОБЛЕНОГО ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

Розроблено програмне забезпечення для генерації випадкових величин з рівномірним законом розподілу різними генераторами, а також можливістю аналізу вибірок випадкових величин за допомогою критеріїв оцінки інформативності вибірок рівномірних випадкових величин описаних вище. На рис. 2 наведено інтерфейс програмного забезпечення.

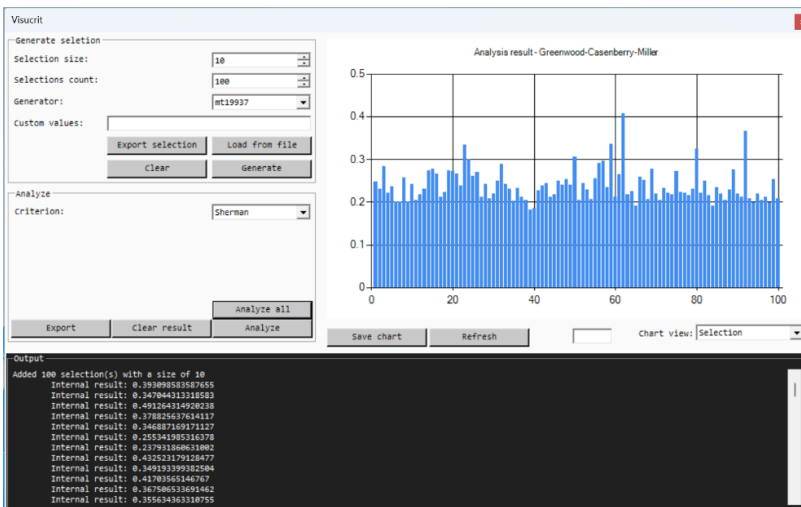


Рисунок 2 – Інтерфейс програмного забезпечення

Інтерфейс розробленого програмного забезпечення розділений на три вікна: 1) для вибору довжини вибірки, кількості експериментів та вибору самого генератора; 2) для графічного відображення вибірки чи результату перевірки критерію; 3) для виведення результатів дослідження. У першому вікні користувач вибирає розмір та кількість вибірок для генерації, а також сам генератор. У другому вікні користувач може спостерігати або зберігати графічне подання вибірки, яку він згенерував або завантажив у програму, а також результат оцінки критерію. У третьому вікні користувач може спостерігати результат оцінки кожної вибірки, яку він згенерував або завантажив за допомогою програмного забезпечення.

## ВИСНОВКИ

На сьогодні існує безліч стандартних генераторів псевдовипадкових величин, але слід пам'ятати, що більшість з запропонованих генераторів не видають якісну випадковість і не підходять для роботи з будь-якими прикладними завданнями (неруйнівний контроль, кібербезпека та захист інформації, тощо), яка вимагає точної статистики. Розроблене програмне забезпечення включає сучасні програмні генератори за допомогою яких можна проводити дослідження та порівнювати з власними генераторами. У запропонованому програмному забезпеченні застосовано сучасні алгоритми генерації випадкових величин, що надасть можливість застосовувати його для прикладних задач. У подальшому передбачається проведення додаткового дослідження щодо перевірки інформативності з застосуванням статистичних критеріїв **Шермана; Морана; Фроцині; Грінвуда-Кесенберрі-Міллера, Чанга та Спірінга**. Особливістю представленого програмного забезпечення є можливість завантаження вибірок випадкових величин з текстових файлів, що у більшості випадків є зручним при дослідженнях та подальшого аналізу даних у прикладних задачах.

## ПОСИЛАННЯ

1. Parthasarathy, H. (2022). *Advanced Probability and Statistics: Remarks and Problems*. CRC Press.
2. Malaichuk, V., Klymenko, S., & Astakhov, D. (2023). Computer processing of measurements in problems of observation of the condition of technical objects. *Journal of Rocket-Space Technology*, 30(4), 99-106. <https://doi.org/10.15421/452213>
3. Montgomery, D. C., & Runger, G. C. (2020). *Applied statistics and probability for engineers*. John Wiley & sons.
4. Kobzar, A.I. (2006) *Applied Mathematical Statistics. For Engineers and Scientists*. Fizmatlit, 816 p.
5. Rohrbeck, C. (2023). Christian Rohrbeck's contribution to the discussion of 'The First Discussion Meeting on Statistical aspects of climate change'. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 72(4), 852-852.
6. Malaichuk, V., Klymenko, S., & Astakhov, D. (2022, May). Study of informativity of the inversion criterion in testing the hypothesis about accidentality in problems control and cyber security. In International scientific and technical conference Information technologies in metallurgy and machine building (pp. 40-41).