DOI 10.31891/2307-5732-2023-329-6-258-265
УДК 004.94

ОДЕГОВ МИКОЛА
Державний університет інтелектуальних технологій і зв'язку
https://orcid.org/ 0000-0001-5526-2487
e-mail: onick_64@ukr.net
ГАДЖИЄВ МАТІН
Державний університет інтелектуальних технологій і зв'язку
http://orcid.org/0000-0001-7280-3863
e-mail: gadjievmm@ukr.net
КАЛІНІНА ТЕТЯНА
Державний університет інтелектуальних технологій і зв'язку
https://orcid.org/0000-0002-3184-3604
e-mail: kalininat384@gmail.com
ПЕТРОВИЧ ЯННА
Державний університет інтелектуальних технологій і зв'язку
https://orcid.org/0009-0008-8939-2333
yanna-petrovich@ukr.net
АНДРІЙЧЕНКО КИРИЛО
Державний університет інтелектуальних технологій і зв'язку
e-mail: almelovetm@gmail.com

ПОРІВНЯЛЬНИЙ АНАЛІЗ МЕТОДІВ ТЕСТУВАННЯ НЕЗАЛЕЖНОСТІ ПОСЛІДОВНОСТЕЙ ПСЕВДОВИПАДКОВИХ ЧИСЕЛ

У задачах статистичного та імітаційного моделювання можуть використовуватись послідовності псевдовипадкових чисел дуже великої довжини. При цьому виникає необхідність вирішення задач класу Big Data. Вирішення таких задач іноді потребує дещо нехтувати точністю рішень, щоб отримати практично прийнятні результати досліджень у прийнятний час.

На даний час розроблена значна кількість систем тестування послідовностей псевдовипадкових чисел (ПВЧ) на відповідність досить умовному поняттю «випадковості», тобто неможливості прогнозувати їх окремі значення. Хоча, ПВЧ при цьому генеруються за допомогою регулярних алгоритмів. Парадигмою даної роботи є вимога відповідності емпіричних функцій розподілення ПВЧ теоретичним функціям розподілення.

Для одновимірних (маргінальних) розподілень дана задача вирішується досить просто. Більш складною є задача встановлення характеристик статистичної залежності або незалежності ПВЧ. Для пари ПВЧ довжини N найбільш логічним є метод повної перевірки їх незалежності (МПВ). Суть даного методу полягає у встановленні відхилення добутків значень ймовірностей від теоретичних. Даний метод зводиться до алгоритмів порядку NxN.

У попередніх роботах розглядались алгоритми методу критерія сум (МКС), який зводиться до аналізу сум значень ПВЧ. При цьому у випадку статистичної незалежності розподілення суми матиме вид згортки маргінальних розподілень. Порядок даного алгоритму лише N.

У даній роботі виконано порівняльний аналіз МПВ та МКС за показниками надійності та швидкості. Штучна залежність ПВЧ при цьому моделювалась внесенням певного рівня кореляції. Порівняльний аналіз показав, що за показником надійності обидва методи приблизно однакові. За показником швидкості МКС на порядки (пропорційно N) є більш ефективним, ніж МПВ.

Остаточно зроблено висновок, що для задач моделювання класу Big Data слід віддавати перевагу МКС.

Ключові слова: випадкові числа, імітаційне моделювання, кореляція, статистична незалежність, генератори випадкових чисел, рівномірне розподілення, критерій Колмогорова.

NICK ODEGOV, MATIN HADZHYIEV, TETIANA KALININA, YANNA PETROVICH, KYRYLO ANDREICHENKO
State University of Intellectual Technologies and Communication

COMPARATIVE ANALYSIS OF METHODS FOR TESTING THE INDEPENDENCE OF PSEUDO-RANDOM NUMBER SEQUENCES

Very long sequences of pseudorandom numbers can be used in statistic and simulation modeling tasks. At the same time, there is a need to solve problems of the Big Data class. Solving such problems sometimes requires a slight disregard for accuracy in order to obtain practically acceptable research results in an acceptable time.

Currently, a significant number of systems for testing sequences of pseudo-random numbers (PRN) has been developed for compliance with the rather conventional concept of "randomness", that is, the impossibility of predicting their individual values. However, PRNs are generated using regular algorithms. The paradigm of this work is the requirement to match the empirical distribution functions of PRN with the theoretical distribution functions.

For one-dimensional (marginal) distributions, this problem is solved quite simply. The task of establishing the characteristics of statistical dependence or independence of PRN is more difficult. For a pair of PRNs of length N, the most logical method is the method of complete verification of their independence (CVM). The essence of this method is to establish the deviation of the products of the imperial values of the probabilities from the theoretical ones. This method is reduced to NxN algorithms.

In previous works, the algorithms of the sum criterion method (SCM) were considered, which is reduced to the analysis of the sums of PRN values. At the same time, in the case of statistical independence, the sum distribution will have the form of a convolution of marginal distributions. The order of this algorithm is only N.

In this paper, a comparative analysis of CVM and SCM was performed based on reliability and speed indicators. At the same time, the artificial dependence of PRN was modeled by introducing a certain level of correlation. Comparative analysis showed that both methods

are approximately the same in terms of reliability. In terms of speed, the SCM is orders of magnitude (in proportion to N) more efficient than the CVM one.

Finally, it was concluded that ISS should be preferred for Big Data modeling tasks.

Keywords: random numbers, simulation modeling, correlation, statistical independence, random number generators, uniform distribution, Kolmogorov criterion.

Аналіз джерел та постановка проблеми

На даний час генератори послідовностей псевдовипадкових чисел (ГВЧ) розроблені для більшості операційних систем та різних мов програмування. Втім, актуальною залишається задача дослідження: а чи ці послідовності є випадковими? Відомі системи тестів «достатньої випадковості» послідовностей ГВЧ [1-3]. Проте, для моделювання реальних систем з випадковими характеристиками важливо інше розуміння «випадковості», а саме: відповідність статистичних характеристик послідовностей ГВЧ (ПГВЧ) заданим теоретичним значенням або функціям. Наприклад, маргінальні розподілення ПГВЧ повинні узгоджуватись із заданою функцією розподілення якоїсь випадкової величини.

Задача моделювання суттєво ускладнюється, якщо генерується не окрема випадкова величина, а випадковий процес. Такий процес, в залежності від задачі, може моделюватись як системою незалежних випадкових величин (типу «білого шуму»), так і більш складними системами величин з визначеною залежністю, наприклад із заданою кореляційною функцією. Відносно проста задача: сукупність незалежних випадкових величин перетворити у сукупність величин із заданими характеристиками залежності. Тому базовою задачею тестування ПГВЧ є аналіз статистичної незалежності цих послідовностей.

Для моделювання надвеликих масивів даних вирішуються задачі класу Dig Data. При цьому іноді приходиться нехтувати точністю методів на користь достатньої швидкості рішення задач [4]. Для тестування незалежності ПГВЧ розглянуто один із експрес-методів [5, 6] де запропоновано тестування за критерієм сум. Суть цього методу полягає в тому, що аналізується відповідність сум двох ПГВЧ теоретичному вигляду згортки їх функцій розподілення. Певним недоліком цих робіт є те, що результати тестування не порівнювались з тестами саме за визначенням незалежності випадкових величин.

Метою даної роботи є порівняння методу перевірки незалежності ПГВЧ за визначенням (МПВ) та методу за критерієм сум (МКС) по показникам надійності та швидкості.

Теоретичні основи методики порівняльного аналізу

Для незалежності випадкових величин (ВВ) X та Y необхідно і достатньо, щоб їх сумісне розподілення ймовірностей дорівнювало добутку маргінальних розподілень, що можна записати у такому загальному вигляді:

$$P(X,Y)=P(X)\cdot P(Y);\ F(x,y)=F_X(x)\cdot F_Y(y);\ f(x,y)=f_X(x)\cdot f_Y(y), \tag{1}$$

де P, F, f – відповідно розподілення ймовірностей, ФР ймовірностей та щільності розподілення, якщо останні визначені для даних ВВ. Для суми $Z=X+Y$ незалежних ВВ справедливі імплікації для ФР:

$$F(x,y)=F_X(x)\cdot F_Y(y)\cdot F_Z(z)=\int_{-\infty}^{\infty}F_Y(z-x)\,dF_X=F_X * F_Y; \tag{2}$$

Таким чином, розподілення сум виражаються як ***згортки*** маргінальних розподілень. Важливо, що згортки, на відміну від двовимірних розподілень, є функціями лише ***однієї*** змінної – суми значень ВВ.

Вирішення задач перевірки незалежності методом порівняння теоретичних ФР сум ВВ з емпіричними ФР (ЕФР) сум пар ПВЧ, таким чином, дозволяє суттєво зменшити порядок алгоритмів аналізу: вдвічі скорочується розмірність простору рішень. Втім, умови виду (2) є ***необхідними***, але ***не достатніми*** для вирішення задачі тестування ГВЧ на незалежність, тобто у загальному випадку зворотна імплікація не є справедливою. Тому методика порівняльного аналізу МПВ та МКС з необхідністю повинна передбачати:

– формалізацію критеріїв для порівняльної оцінки надійності цих методів;

– розробку алгоритмів генерації тестових ПВЧ;

– виконання та аналіз результатів багатьох обчислювальних експериментів.

Для порівняння результатів тестування МПВ та МКС необхідний сумісний критерій відповідності теоретичних розподілень та емпіричних розподілень. У даному дослідженні обираємо критерій Колмогорова-Смірнова (ККС), який заснований на відстані Чебишева:

$$D_M=\max_x\left|F(x)-F_M(x)\right|, \tag{3}$$

де $F(x)$ та $F_M(x)$ відповідно теоретична (ТФР) та емпірична (ЕФР) функції розподілення значень ВВ при вибірки об'єму N.

При $M\to\infty$ та за умови незалежності вимірів функція розподілення значень відстаней (3) прагне до функції Колмолгорова $K(p)$:

$$P\left(\sqrt{N}D_M\le p\right)=K(p)=\sum_{j=-\infty}^{\infty}(-1)^j\exp\left(-2j^2p^2\right). \tag{4}$$

де p – ймовірнісна міра (рівень довіри).

Суттєвим позитивним зауваженням відносно цієї роботи є те, що метрика Чебишева та заснована на неї статистика Колмогорова не залежать від масштабу аргументу x ТФР та ЕФР у формулах (3) та (4). Втім, треба зробити важливі, на наш погляд, методичні зауваження:

– як враховувати «точні» значення функції Колмогорова, якщо у формулі (4) треба сумувати щось таке від мінус безкінечності до плюс безкінечності?

– а що визначає, якщо подумати, «рівень довіри p»?

На перше питання є досить проста відповідь, якщо значення параметру p є досить близьким до одиниці. Тоді функція Колмогорова добре апроксимується більш простою залежністю:

$$K(p) = \sqrt{-\frac{1}{2}\ln\left(\frac{1-p}{2}\right)}. \quad (5)$$

Із залежності (5) можна встановити звороту залежність, графік якої дано на рис. 1.

$$p(K) = 1 - 2\exp\left(-2K^2\right). \quad (6)$$

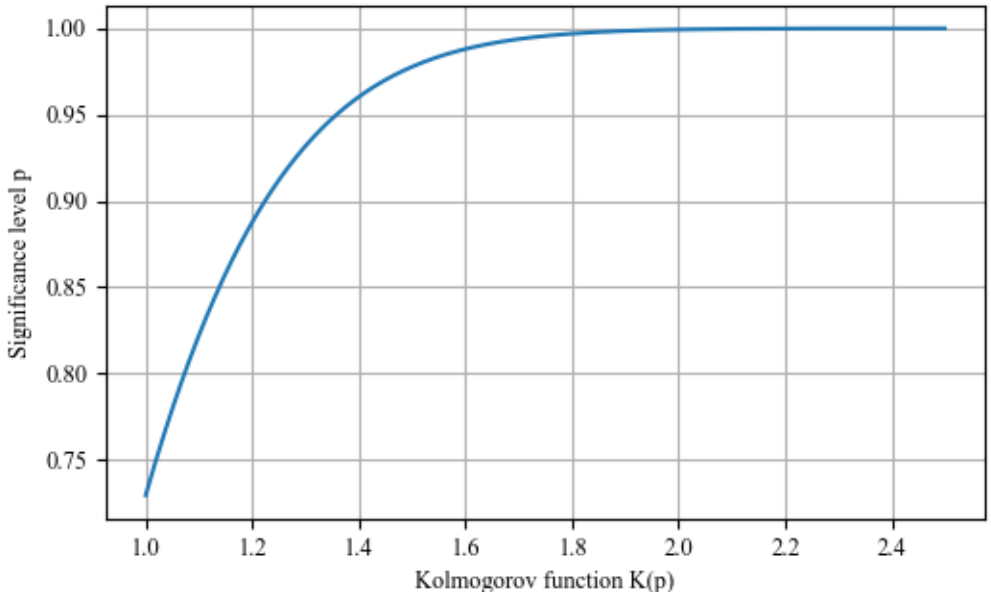

Рис. 1. Функція *p(K)*, зворотна до функції Колмогорова

Наведемо непрості міркування з приводу так званого «рівня довіри». Іноді дослідники за даними вибірок об'ємом N=10 або N =20 роблять висновки, що розподілення їх значень має, наприклад, характер нормального розподілення з певним рівнем значущості (довіри). На наш погляд, до висновків з подібних досліджень треба відноситись з певною обережністю. Як мінімум, цей рівень «довіри» визначає не хто-небудь, а сам дослідник. А найголовніше, що всі критерії значущості теоретично коректні, якщо вони застосовані при виконанні певних умов, а саме: статистичної незалежності окремих вимірів та кількості цих вимірів, яка прагне до нескінченності. У даній роботі досліджуються статистичні залежності або незалежності лише двох ВВ, які моделюються двома відповідними ПГВЧ. При цьому статистична залежність або незалежність окремих вимірів ВВ залишається великою методичною загадкою, гіпотезою, адекватність якої перевірити неможливо.

Тому дослідження у даній роботі ми будемо засновувати скоріше на логічних і зрозумілих відстанях між ТВР та ЕФР (3). Значення критеріальних функцій (4-6) будемо використовувати, скоріше, для «інтуїтивного аналізу», щоб відповісти на запитання «а з якою ймовірністю p щось там трапилось»? Тому з урахуванням обмежень формул (5) та (6) визначимо умовну межу ймовірності p_0. Якщо $p(K) \geq p_0$, то будемо робити висновок, що ЕФР не узгоджується з ТФР з рівнем довіри p_0. У протилежному випадку, якщо $p(K) < p_0$, то робитимемо висновок, що можливо ЕФР більш-менш відповідає ТФР.

З урахуванням того, що функція (5) приблизно апроксимує функцію Колмогорова (4), встановлюємо критичний рівень ймовірності p_0 = 0,95. Знову підкреслимо, що дане значення обрано «інтуїтивно»: ані якої теорії або методу для цього вибору не існує.

Квантильний алгоритм визначення емпіричних функцій розподілення

Алгоритми визначення ЕФР для МПВ та МКС принципово відрізняються, оскільки в першому випадку задача вирішується у двовимірному просторі, у другому – в одновимірному. Втім, для рівномірно розподілених чисел на інтервалі [0,1] (РРЧ) першу задачу можна звести до другої.

Нехай розігруються дві ПВЧ виду РРЧ X_M та Y_M, де M – об'єм вибірок. Поділимо інтервал [0,1] на рівні частини в кількості $N < M$. Тоді можна сформувати квадратну матрицю можливих випадків потрапляння значень ПВЧ, де окреме значення x_i буде відповідати певному рядку цієї матриці r (Row), а значення другої ПВЧ y_i – стовбцю цієї матриці c (Column).

При цьому кожна пара значень $\{x_i, y_i\}$ буде потрапляти у певний осередок матриці з номером $\{r, c\}$. Парну нумерацію осередків можна перетворити у наскрізну (лінійну) нумерацію, як це показано у табл. 1, тобто розгорнути матрицю розмірності N x N у вектор-строку розмірності 1 x N^2 Унікальність РРЧ полягає в тому, що для незалежних ПВЧ функція розподілення по осередкам цієї вектор-строки також буде мати вигляд РРЧ.

Аналіз саме РРЧ має фундаментальне значення, оскільки з цих ПВЧ можна за допомогою лінійних або нелінійних перетворень (типу методом зворотних функцій [7]) відтворити послідовність з будь-яким законом розподілення.

Таблиця 1

Принцип перетворення нумерації осередків

Парна нумерація осередків

	c=0	c=1	c=2	c=3
r=0	0,0	0,1	0,2	0,3
r=1	1,0	1,1	1,2	1,3
r=2	2,0	2,1	2,2	2,3
r=3	3,0	3,1	3,2	3,3

Наскрізна нумерація осередків

	c=0	c=1	c=2	c=3
r=0	0	1	2	3
r=1	4	5	6	7
r=2	8	9	10	11
r=3	12	13	14	15

Математично перетворення парних номерів у наскрізну нумерацію здійснюється за допомогою формули:

$$U = r \cdot N + c. \tag{7}$$

Для обробки даних класу Big Data за допомогою формули (7) можна використовувати швидкий алгоритм обчислення номерів осередків.

Номери строк та стовбців матриць виду, який показано для простого прикладу у табл. 1 обчислюються за формулами:

$$r = \text{trunc}\left(N \cdot x_i\right), \;\; c = \text{trunc}\left(N \cdot y_i\right), \tag{8}$$

де trunc – функція відкидання дробовий частини числа. Надалі номер осередка вектор-строки визначається за формулою (7).

Номери осередків U утворюють нову ПВЧ, яка за умови незалежності ПВЧ X_M та Y_M буде приблизно рівномірно розподілена на інтервалі $\left[0, N^2 - 1\right]$. Для остаточного аналізу виконується просте перетворення:

$$z_i = \frac{U_i}{N^2}, \;\; i = \overline{1, M} \tag{9}$$

Для МКС квантильний алгоритм також зводиться до визначення осередків, але у цьому випадку (аналізується одновимірна послідовність сум ПВЧ) кількість цих осередків лише N.

Планування порівняльного обчислювального експерименту для МПВ та МКС

Якщо ВВ X та Y статистично незалежні, то ПВЧ, що їх моделюють будуть мати ТФР при $N \to \infty$ вигляд для МПВ:

$$F(z) = \begin{cases} 0, & z < 0 \\ z, & 0 \le z \le 1 \\ 1, & z > 1 \end{cases}, \tag{10}$$

а для МКС вигляд ТФР для трикутного розподілення:

$$F(z) = \begin{cases} 0, & z < 0 \\ 0{,}5 \cdot z^2, & 0 \le z < 1 \\ -1 + 2 \cdot z - 0{,}5 \cdot z^2, & 1 \le z \le 2 \\ 0, & z > 2 \end{cases}. \tag{11}$$

Графіки функцій (10) та (11) для ТФР «незалежних» ПВЧ показані на рис. 2 водночас з ЕФР для досить великих вибірок.

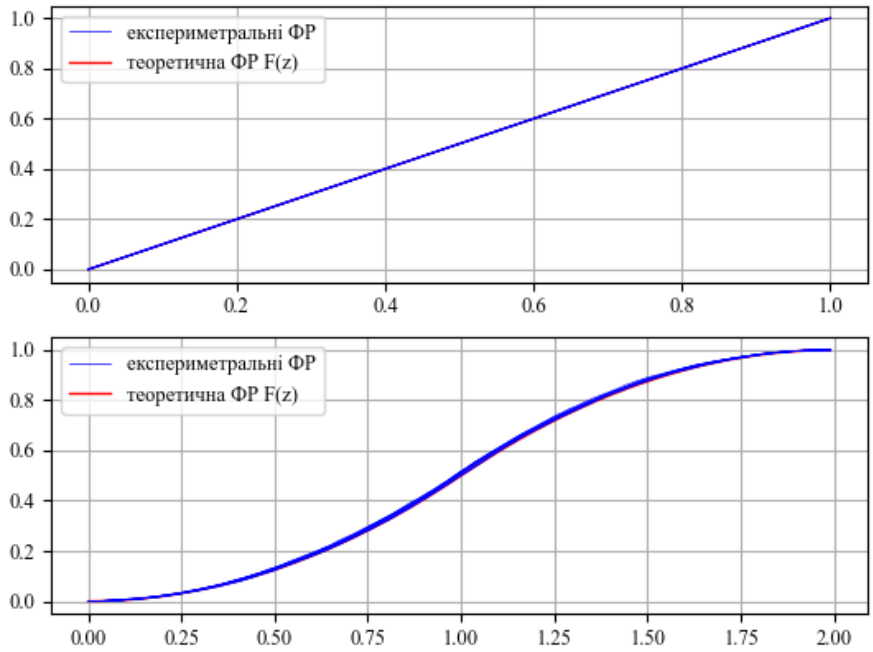

Рис. 2. Розраховані графіки ТФР та ЕФР для МПВ (зверху) та МКС (знизу) за умови незалежності ВВ

Основною задачею порівняльного аналізу МПВ та МКС є визначення характеристик чутливості до малих або великих значень параметрів статистичної залежності.

Штучна статистична залежність за методикою [6] моделюється кореляційними залежностями. Для цього якась доля α ($0 \le \alpha \le 1$) елементів ПВЧ Y_M замінюється елементами ПВЧ X_M. Не важко довести, що при $M \to \infty$ параметр α дорівнюватиме коефіцієнту кореляції $R(X,Y)$ двох ПВЧ X_M та Y_M.

При значенні $R = 0$ ЕФР ПВЧ за обома методами повинні мати розподілення, близькі до ТФР (10) для МПВ та (11) для МКС. Якщо коефіцієнт кореляції матиме граничне (максимальне) значення $R = 1$, то ТФР суттєво змінюються. Так, для МПВ ймовірні випадки будуть розміщуватись на головній діагоналі матриць виду, що показано у табл. 1. Приклад розподілення ймовірностей для незалежних та гранично корельованих ВВ для грубої сітки значень $N = 4$ дано у табл. 2. У випадку R=1 значущі ймовірності зосереджуються на головній діагоналі матриці.

Таблиця 2

Ймовірності потрапляння значень ВВ в осередки матриці у випадках незалежності та залежності

Ймовірності для незалежних ВВ

	c=0	c=1	c=2	c=3
r=0	1/16	1/16	1/16	1/16
r=1	1/16	1/16	1/16	1/16
r=2	1/16	1/16	1/16	1/16
r=3	1/16	1/16	1/16	1/16

Ймовірності для ВВ з коефіцієнтом кореляції R=1

	c=0	c=1	c=2	c=3
r=0	1/4	0	0	0
r=1	0	1/4	0	0
r=2	0	0	1/4	0
r=3	0	0	0	1/4

Для МКС ЕФР при R=1 та за умов значного об'єму вибірок мають приблизно співпадати з РРЧ на інтервалі [0,2]. Варіанти отриманих графіків для МПВ та МКС саме за умови R=1 показані на рис. 3.

Взагалі окремий обчислювальний експеримент (ОЕ) визначається набором параметрів:

$$N, Q, M, R,$$

де N – кількість осередків маргінальних (одновимірних) розподілень;

Q – середня кількість значень ПВЧ, які потрапляють у певний осередок;

M – об'єм вибірок, який для МПВ дорівнює $M = Q \cdot N^2$, а для МКС $M = Q \cdot N$;

R – коефіцієнт кореляції.

Для кожного з цих експериментів визначається відстань Чебишева для МПВ та МКС. По кількості таких експериментів E визначається середня відстань Чебишева $\overline{D_M(E)}$, яке надалі використовується для визначення функції Колмогорова $K(M) = \sqrt{M}\,\overline{D_M(E)}$ та «рівня довіри p» за формулою (6). При цьому враховується час T, с виконання тестів обома методами: МПВ та МКС.

Для розрахунків по рис. 3 використано план обчислювальних експериментів: N=20, Q=100, R=1, E=10.

Зауважимо, що кількість експериментів при E >10 практично не впливає на кінцеві висновки. Тому усюди для проведення тестування обрано значення E >10.

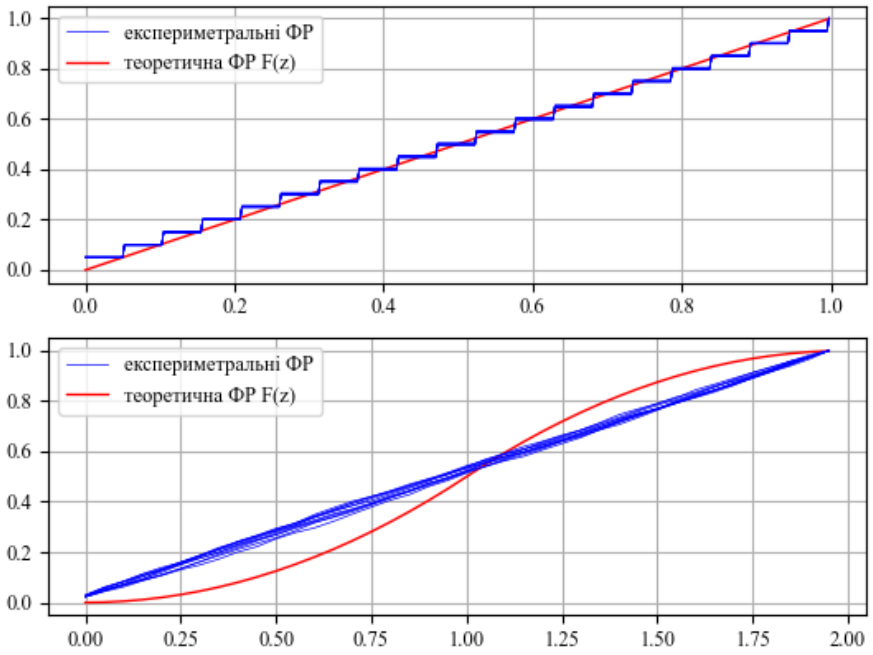

Рис. 3. Розраховані графіки ТФР та ЕФР для МПВ (зверху) та МКС (знизу) за умови сильної кореляції при *R*=1

Порівняння МПВ та МКС здійснювалось на прикладі ГВЧ бібліотек мови Python. При цьому використані бібліотеки:

– Numpy для генерації ПВЧ та математичних розрахунків;
– Matplotlib для графічного відображення результатів аналізу;
– Time для визначення порівняльної тривалості процесів тестування для МПВ та МКС.

Результати порівняльного аналізу МПВ та МКС

Порівняльний аналіз МПВ та МКС здійснено для випадків:

– малих вибірок та великих значень коефіцієнта кореляції (табл. 3);
– малих вибірок та малих значень коефіцієнта кореляції (табл. 4);
– великих вибірок та малих значень коефіцієнта кореляції (табл. 5);
– надвеликих вибірок та дуже малих значень коефіцієнта кореляції (табл. 6).

Таблиця 3

Результати обчислювальних експериментів для малих вибірок та великих значень коефіцієнта кореляції

План ОЕ			Дані аналізу МПВ			Дані аналізу МКС		
N	*Q*	*R*	*M*	*p*(*K*)	*T*, с	*M*	*p*(*K*)	*T*, с
10	10	0,50	1000	0.99797	0.090	100	0.99108	0.050
10	10	0,75	1000	0.99999	0.089	100	0.99423	0.060
10	10	1,00	1000	1.00000	0.080	100	0.99620	0.052

Як видно з табл. 3, навіть при відносно малих об’ємах вибірок обидва методи МПВ та МКС з великим рівнем довіри відкидають гіпотезу про незалежність пари ПВЧ, якщо наявна суттєва статистична залежність (коефіцієнт кореляції $R \geq 0{,}5$).

Таблиця 4

Результати обчислювальних експериментів для малих вибірок та малих значень коефіцієнта кореляції

План ОЕ			Дані аналізу МПВ			Дані аналізу МКС		
N	*Q*	*R*	*M*	*p*(*K*)	*T*, с	*M*	*p*(*K*)	*T*, с
10	10	0,1	1000	0.67727	0.100	100	0.84443	0.042
10	10	0,2	1000	0.83916	0.143	100	0.80173	0.084
10	10	0,3	1000	0.95456	0.082	100	0.91213	0.046

Аналіз даних у табл. 4 показує, що при відносно малих рівнях кореляції та при малих вибірках МПВ може більш впевнено відкидати гіпотезу незалежності ПВЧ: навіть при значенні $R = 0{,}3$ рівень довіри до цього висновку більше встановленого раніше критичного значення $p_0 = 0{,}95$.

Таким чином, на малих вибірках МКС дещо поступається МПВ за показником надійності відхилення гіпотези про незалежність. За показником швидкості у цьому випадку МПВ лише незначно

поступається МКС: час тестування для МКС лише приблизно вдвічі більший.

Таблиця 5

Результати обчислювальних експериментів для великих вибірок та малих значень коефіцієнта кореляції

План ОЕ			Дані аналізу МПВ			Дані аналізу МКС		
N	*Q*	*R*	*M*	*p*(*K*)	*T*, с	*M*	*p*(*K*)	*T*, с
100	100	0,05	1000000	0.87837	17.230	10000	0.90516	0.179
100	100	0,1	1000000	0.97823	18.561	10000	0.99970	0.174
100	100	0,2	1000000	0.99994	17.753	10000	1.00000	0.189

Аналіз даних у табл. 5 показує, що обидва методи впевнено відкидають гіпотезу незалежності при достатньо великих об'ємах вибірок. Навіть за умов малої кореляції $R = 0,1$ на підставі даних обробки ПВЧ в обох випадках можна зробити висновок про наявність статистичної залежності. Як не дивно, але за показником надійності МКС вже переважає МПВ. Втім, МКС для даного плану ОЕ має вже суттєву перевагу: дані аналізуються майже у 100 разів скоріше, ніж методом МПВ.

У табл. 6 наведено приклади для зовсім малих рівнів кореляції та для дуже значних об'ємів вибірок. Аналіз цієї таблиці показує значну перевагу МКС над МПВ як за показником надійності, так і за показником швидкості. Так, МКС дозволяє робити висновок щодо наявності статистичної залежності ПВЧ навіть, якщо рівень кореляції ледь помітний ($R \geq 0,03$). Різниця у часі тестування у даному випадку і зовсім вражаюча: при тому, що аналіз за допомогою МПВ займає приблизно 3 хвилини, МКС дає результати лише за половину секунди.

Таблиця 6

Результати обчислювальних експериментів для надвеликих вибірок та дуже малих значень коефіцієнта кореляції

План ОЕ			Дані аналізу МПВ			Дані аналізу МКС		
N	*Q*	*R*	*M*	*p*(*K*)	*T*, с	*M*	*p*(*K*)	*T*, с
200	100	0,01	4000000	0.50024	71.124	20000	0.76166	0.330
200	100	0,03	4000000	0.64483	73.148	20000	0.91289	0.322
200	100	0,05	4000000	0.79821	71.343	20000	0.99581	0.320
300	100	0.01	9000000	0.61222	157.452	30000	0.87613	0.542
300	100	0.03	9000000	0.73985	175.044	30000	0.96817	0.567
300	100	0.05	9000000	0.77941	171.969	30000	0.99416	0.512

Висновки та рекомендації

1. Обидва методи, що порівнювались вказують не те, що послідовності випадкових чисел мови Python можуть впевнено використовуватись для вирішення задач статистичного моделювання.

2. За показником надійності обидва методи, що порівнювались дають приблизно однакові результати. Лише у деяких випадках за цим показником МПВ є більш надійним, ніж МКС.

3. За показником швидкості МКС є суттєво більш ефективним, оскільки для нього порядок алгоритму має значення N, тоді як для МПВ цей порядок складає N^2.

4. Для вирішення задач статистичного та імітаційного моделювання для задач класу Big Data беззаперечно можна рекомендувати використання МКС.

Література

1. Andrew Rukhin, JuanSoto, James Nechvatal, Miles Smid, ElaineBarker, Stefan Leigh, MarkLevenson, Mark Vangel, DavidBanks, Alan Heckert, Jame sDra and San Vo, " A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications: NIST Special Publication 800-22 Revision 1a", National Institute of Standards and Technology Gaithersburg, MD 20899-8930, Revised: April 2010. – 131 pp.

2. Roman Kochana, Lyudmila Kovalchuk, Oleksandr Korchenko and Nataliia Kuchynska, "Statistical Tests Independence Verification Methods", Procedia Computer Science, Volume 192, 2021, Pages 2678-2688, ISSN 1877-0509, DOI: https://doi.org/10.1016/j.procs.2021.09.038.

3. Герасимчук О.І. Генератори псевдовииадкових чисел, їх застосування, класифікація, основні методи побудови і оцінка якості / О.І.Герасимчук, В.М.Максимович // Науково - технічний журнал "Захист інформації", № 3, 2003. С. 29-36. DOI: https://doi.org/10.18372/2410-7840.5.4270.

4. Одегов М.А. Обґрунтування швидких алгоритмів класифікації на множинах BIG DATA за критеріями надійності і продуктивності / М.А. Одегов, М.М. Гаджиєв, Л.М. Буката, Л.В. Глазунова, М.В. Кочеткова // Інфокомунікаційні та комп'ютерні технології. - №1, 2023. - С. 148 - 160. DOI: https://doi.org/10.36994/2788-5518-2023-01-05-16.

5. М. Одегов, Ю. Бабіч, Д. Багачук, М. Кочеткова, Я. Петрович . Методика критеріїв сум у задачах тестування незалежності послідовностей випадкових чисел // Інфокомунікаційні технології та електронна

інженерія: Львів, №2, 2023. С. 20-22. DOI: https://doi.org/10.23939/ictee2023.02.020.

6. Одегов М.А. Методика двокомпонентного експрес-тестування незалежності послідовностей псевдовипадкових чисел / М.А. Одегов, Ю.О. Бабіч, Д.Г. Багачук, М.В. Кочеткова, Я.О. Петрович // Міжнародний наук.-техн. журнал "Вимірювальна та обчислювальна техніка в технологічних процесах". - Хмельницькій. - 2023. - № 4. С. 64 - 73. DOI: https://doi.org/10.31891/2219-9365-2023-76-8.

7. Буртняк І.В. Імітаційне моделювання / І.В. Буртняк. Прикарпатський національний університет ім. Василія Стефаника, 2019. – 97 с.

References

1. Andrew Rukhin, JuanSoto, James Nechvatal, Miles Smid, ElaineBarker, Stefan Leigh, MarkLevenson, Mark Vangel, DavidBanks, Alan Heckert, Jame sDra and San Vo, " A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications: NIST Special Publication 800-22 Revision 1a", National Institute of Standards and Technology Gaithersburg, MD 20899-8930, Revised: April 2010. – 131 rr.

2. Roman Kochana, Lyudmila Kovalchuk, Oleksandr Korchenko and Nataliia Kuchynska, "Statistical Tests Independence Verification Methods", Procedia Computer Science, Volume 192, 2021, Pages 2678-2688, ISSN 1877-0509, DOI: https://doi.org/10.1016/j.procs.2021.09.038.

3. Gerasimchuk O.I. Generatori psevdoviiadkovih chisel, yih zastosuvannya, klasifikaciya, osnovni metodi pobudovi i ocinka yakosti / O.I.Gerasimchuk, V.M.Maksimovich // Naukovo - tehnichnij zhurnal "Zahist informaciyi", № 3, 2003. S. 29-36. DOI: https://doi.org/10.18372/2410-7840.5.4270.

4. Odegov M.A. Obgruntuvannya shvidkih algoritmiv klasifikaciyi na mnozhinah BIG DATA za kriteriyami nadijnosti i produktivnosti / M.A. Odegov, M.M. Gadzhiyev, L.M. Bukata, L.V. Glazunova, M.V. Kochetkova // Infokomunikacijni ta komp'yuterni tehnologiyi. - №1, 2023. - S. 148 - 160. DOI: https://doi.org/ 10.36994/2788-5518-2023-01-05-16.

5. M. Odegov, Yu. Babich, D. Bagachuk, M. Kochetkova, Ya. Petrovich . Metodika kriteriyiv sum u zadachah testuvannya nezalezhnosti poslidovnostej vipadkovih chisel // Infokomunikacijni tehnologiyi ta elektronna inzheneriya: Lviv, №2, 2023. S. 20-22. DOI: https://doi.org/10.23939/ictee2023.02.020.

6. Odegov M.A. Metodika dvokomponentnogo ekspres-testuvannya nezalezhnosti poslidovnostej psevdovipadkovih chisel / M.A. Odegov, Yu.O. Babich, D.G. Bagachuk, M.V. Kochetkova, Ya.O. Petrovich // Mizhnarodnij nauk.-tehn. zhurnal "Vimiryuvalna ta obchislyuvalna tehnika v tehnologichnih procesah". - Hmelnickij. - 2023. - № 4. S. 64 - 73. DOI: https://doi.org/10.31891/2219-9365-2023-76-8.

7. Burtnyak I.V. Imitacijne modelyuvannya / I.V. Burtnyak. Prikarpatskij nacionalnij universitet im. Vasiliya Stefanika, 2019. – 97 s.