УДК 004.5 (519.2)

[1]В.І. Масол, д-р фіз.-мат. наук, проф.,
[2]Н.О. Маслова, к.т.н., доц.,
[3]Є. Котереу, магістрант
Донецький національний технічний університет, м. Луцьк, Україна
[1]volodymyr.masol@donntu.edu.ua
[2]nataliia.maslova@donntu.edu.ua
[3]ekotereu@gmail.com

Ймовірнісний аналіз одночасного виникнення кількох подій в контексті схеми Бернуллі

Однією з фундаментальних моделей у ймовірнісному аналізі бітових послідовностей є схема Бернуллі. У роботі пропоновано застосування дискретного розподілу для аналізу ймовірності одночасної появи декількох подій у контексті схеми Бернуллі, описана математична модель та програмна реалізація методу досягнення максимальної ймовірності, отримано якісні оцінки базового та пропонованого методів, наведено приклад застосування дискретного розподілу. Аналіз одночасного виникнення кількох подій в контексті схеми Бернуллі є важливою сучасною задачею, застосовною в криптографії, стеганографії, інформаційній безпеці та інших галузях.

Ключові слова: бітові послідовності, дискретний розподіл, оцінки якості, ймовірнісний аналіз, криптографія

DOI: 10.31474/1996-1588-2024-1-38-12-18

Вступ

Схема Бернуллі є однією з фундаментальних моделей у ймовірнісному аналізі. Вона дозволяє моделювати послідовність незалежних випробувань. Аналіз бітових послідовностей, зокрема у контексті схеми Бернуллі, є важливим у криптографії, математиці, фінансовому аналізі, контролі якості виробництва, телекомунікаціях, при тестуванні апаратного забезпечення та інших галузях.

Так, у криптографії вирішальним моментом для захисту конфіденційності даних є застосування надійних бітових послідовностей для генерації ключів. Відхилення у бітових послідовностях можуть свідчити про спроби несанкціонованого доступу або атаки на систему, а тестування на випадковість сприяють виявленню аномалій та потенційних загроз.

При контролі апаратного забезпечення інформаційних та телекомунікаційних систем певні характеристики випадковості та послідовності бітів використовуються для перевірки на відповідність стандартам.

А у телекомунікаціях застосування бітових послідовностей – це інструмент вдосконалення методів передачі даних, виявлення помилок та оптимізації пропускної здатності каналів зв'язку. Тож необхідність й актуальність аналізу бітових послідовностей полягає у забезпеченні безпеки, стабільності та ефективності різноманітних систем у сучасному цифровому світі, який невпинно рухається у бік максимальної цифровізації.

Мета роботи – опис авторського методу аналізу бітових послідовностей з використанням дискретного розподілу ймовірності появи декількох подій у контексті схеми Бернуллі, його математична модель та програмна реалізація.

В контексті схеми Бернуллі у роботі наведено опис методу досягнення максимальної ймовірності і приклад застосування дискретного розподілу, створено програмний продукт аналізу бітових послідовностей.

Огляд попередніх досліджень

Питання ймовірнісного аналізу бітових послідовностей не є новим в математиці та статистиці, бо саме випадкову природу мають більшість об'єктів і явищ, що знаходяться та відбуваються навкруги нас. Поняття й базові алгоритми обробки бітових послідовностей застосовуються у основоположних роботах Д. Кнута, К. Дейла, Томаса Джефферсона та багатьох інших науковців, що розвинули напрямки теорії алгоритмів, криптології та криптографії, обробки даних. У сучасності актуальність й необхідність проведення поглибленого аналізу виникла у поєднанні з проблемами вирішення складних обчислювальних задач, розширенням кола задач, пов'язаних з розвитком ІТ-технологій, у тому числі проблем захисту даних.

Сучасні напрямки аналізу бітових послідовностей з точки зору ймовірнісного аспекту пов'язані з роботами науковців Національного інституту стандартів і технологій (National Institute of Standards and Technology,

NIST, США) [1]. Роботи у цьому напрямку продовжують й вітчизняні, українські науковці [2-4]. Так, у роботі [2] особлива увага приділяється статистичному дослідженню бітових послідовностей, наведено тести для виявлення послідовностей з використанням багатовимірної статистики. У роботі [3] встановлено явний вид сумісного розподілу дво- та три- елементних ланцюгів різних фіксованих варіантів бітових послідовностей 0 та 1 довжини n. Надано математичну модель та таблиці, які застосовані в якості основи й тестових прикладів даної роботи. Робота [4] – це всебічний статистичний аналіз бітових послідовностей. В ній отримано явний вигляд розподілу спеціально обраних дво- та тривимірних статистик, що характеризують розміщення 0 та 1 бітової послідовності.

Аналіз випадкових послідовностей виконується на базі різноманітних статистик [1], зокрема, це можуть бути s-ланцюги та $(s+1)$-ланцюги [5] або s-ланцюги, що перетинаються [6]. Під s-ланцюгом, де $1 \le s \le n$, розуміється довільна підпослідовність $\gamma_i, \gamma_{i+1}, \dots, \gamma_{i+s-1}, i = 1, 2, \dots, n-s+1$, послідовності випадкових величин:

$$\gamma_1, \gamma_2, \dots, \gamma_n, \quad (1),$$

де $\gamma_k, k = 1, 2, \dots, n,$ приймає значення з деякої дискретної множини потужності N.

У роботі [7] наведено дослідження щодо оцінки точності пуасонівської апроксимації розподілу загального числа серій, що утворені k-кратними $(k \ge 2)$ повтореннями s-ланцюгів у послідовності (1), а у статті [3] досягнуто розвиток підходу до встановлення явних сумісних розподілів двовимірних та тривимірних ланцюгів заданого виду.

Постановка задачі

Одним з інструментів опису послідовності випробувань, в яких можливі два результати: успіх або невдача, 1 або 0, є стохастична модель, яка має назву схема Бернуллі. Послідовність незалежних випробувань називають рядом Бернуллі, який може бути скінченним або нескінченним.

Кількість успіхів у схемі Бернуллі – це випадкова величина, яка має біноміальний розподіл. Імовірність успіху позначають як p, імовірність невдачі – q. Оскільки результати випробувань є взаємопов'язаними, справедлива рівність $p + q = 1$. Імовірність того, що кількість успіхів у серії з n випробувань за схемою Бернуллі дорівнює k, розраховується за формулою Бернуллі:

$$P_n(k) = C_n^k p^k q^{n-k} = \frac{n!}{k!(n-k)!} p^k q^{n-k},$$

де C_n^k – коефіцієнт біноміального розподілу, що позначає кількість способів вибрати k успіхів з n випробувань.

Математична модель, яка описує ймовірності усіх можливих значень у конкретній дискретній випадковій послідовності або випадковому експерименті носить назву дискретного розподілу.

Використання дискретного розподілу ймовірностей в аналізі стохастичних процесів дозволяє моделювати та вивчати випадкові події у різних системах, аналізувати і передбачати результати випадкових явищ.

Одне з основних завдань роботи можна сформулювати так – отримати та програмно реалізувати знаходження із заданою точністю eps параметр p, для якого максимізується ймовірність (P) появи трьох подій у схемі Бернуллі й оцінити кількість точок у проміжку $(p - eps;\ p + eps)$, для кожної з яких досягається P.

Позначимо число усіх s-ланцюгів у послідовності (1) як $\eta(t_1, t_2, \dots, t_s)$, де $t_i \in \{0, 1\}, i = 1, 2, \dots, s$.

Теорема. Якщо послідовність (1) складається з $n > 0$ незалежних випадкових величин, які мають розподіл:

$$P\{\gamma_k = 1\} = p, P\{\gamma_k = 0\} = q, p + q = 1, \quad k = 1, 2, \dots, n; \quad (2),$$

де k_1, k_2, k_3, t, t_1 – цілі числа, які задовольняють умовам $k_1 \ge 0, k_2 \ge 0, k_3 \ge 0, t, t_1 \in \{0, 1\}$, то рівності (3) та (4) є коректними:

$$P\{\eta(tt) = k_1, \quad \eta(t_1 t t_1^*) = k_2, \quad \eta(t_1 t^* t_1^*) = k_3\} =$$
$$= \sum_{m_1=0}^{n} p^{m_1} q^{m_0} \{C_{m_t - k_1}^{k_2} C_{m_t - k_1}^{k_3} C_{m_{t^*} + k_1}^{k_3} Z(k_1; k_2) +$$
$$+ C_{m_t - k_1 - 1}^{k_2} C_{m_t - k_1 - 1}^{k_3} C_{m_{t^*} - m_t + k_1 + 1}^{k_3} C_{k_1}^{k_2}\}, \quad (3)$$

де $m_0 = n - m_1, t^* = 1 - t,$

$$Z(a; b)^{def} = \begin{cases} C_{a-1}^{b-1}, & \text{якщо } a \ge b \ge 1, \\ 1, & \text{якщо } a = b = 0, \\ 0 \text{ в інших випадках;} \end{cases}$$

$$P\{\eta(tt) = k_1, \quad \eta(t_1 t t_1^*) + \eta(t_1 t^* t_1^*) = k_2\} =$$
$$= \sum_{m_1=0}^{n} p^{m_1} q^{m_0} \sum_{i \in \{k_1, k_1+1\}} \sum C_{m_t - i}^{\delta_t} C_{m_t - i}^{\delta_{t^*}} *$$
$$* C_{m_{t^*} - m_t + i}^{m_t - \delta_{t^*} - i} Z(i; m_t - k_1 - \delta_t). \quad (4)$$

Символ Σ позначає доданок за усіма цілими невід'ємними числами δ_t та δ_{t^*} такими, що

$$\delta_t + \delta_{t^*} = 2(m_t - i) - k_2.$$

Маємо:

$$P\{\eta(tt) = k_1, \quad \eta(t_1 t t_1^*) = k_2\} =$$

$$= \sum_{m_1=0}^{n} p^{m_1} q^{m_0} \{C_{m_t-k_1}^{k_2} C_{m_{t^*}}^{m_t-k_1} Z(k_1; k_2) + + C_{m_t-k_1-1}^{k_2} C_{m_{t^*}}^{m_t-k_1-1} C_{k_1}^{k_2}\}; \quad (5)$$

$$P\{\eta(tt) = k_1, \quad \eta(t_1 t^* t_1^*) = k_2\} = = \sum_{m_1=0}^{n} p^{m_1} q^{m_0} Z(m_t; m_t - k_1)\{C_{m_t-k_1}^{k_2} * * C_{m_{t^*}-m_t+k_1}^{k_2} + C_{m_t-k_1-1}^{k_2} C_{m_{t^*}-m_t+k_1+1}^{k_2}\}. \quad (6)$$

Надалі будемо вважати, що числова двійкова послідовність для проведення експерименту має вигляд: 100001001101101101 й на її основі отримаємо числові розрахунки. Необхідно порахувати кількість 2-ланцюгів виду 00 та кількість 3-ланцюгів виду 100 і виду 110. У наведеній послідовності кількість 2-ланцюгів становить $\eta(00) = 4$, а 3-ланцюгів – $\eta(100) = 2$ та $\eta(110) = 3$ відповідно.

Цифрові інструменти аналізу бітових послідовностей

Для аналізу схеми Бернуллі та обчислення ймовірностей успіху, невдачі та інших параметрів існує кілька готових програмних рішень аналізу бітових послідовностей (БП). Так, у таблиці 1 наведено перелік багатофункціональних систем, в які включено розділи для аналізу бітових послідовностей.

Таблиця 1 – Системи з функціями аналізу БП

	Назва
1	GNU Octave
2	JMP
3	MATLAB
4	Minitab
5	SAS (Statistical Analysis System)
6	SPSS (Statistical Package for the Social Sciences)
7	Wolfram Mathematica

Окрім багатофункціональних систем, існують пакети прикладних програм, розроблених для рішення окремих задач, які виникають в процесі аналізу бітових послідовностей. Прикладами таких програмних пакетів є набір програм Diehard Suite (статистичне тестування генераторів випадкових чисел (ГВЧ) та послідовностей); NIST Statistical Test Suite (статистичні тести бітових послідовностей, оцінка їх випадковості; TestU01 (тестування ГВЧ на рівень їх випадковості та стійкості).

Крім того, вбудовані функції для обчислення ймовірностей і використання біноміального розподілу має табличний калькулятор Microsoft Excel, а також ряд онлайн-калькуляторів, розробка й застосування яких достатньо поширені в наш час. Звертаючись до мов програмування, назвемо мову програмування R зі спеціальними пакетами (binom) – для роботи з біноміальним розподілом та аналізу схеми Бернуллі. А також застосовану у даному дослідженні для статистичного аналізу та обчислень мову програмування Python й бібліотеки для статистичного аналізу, зокрема і для роботи зі схемою Бернуллі – NumPy, SciPy та Matplotlib (візуалізація результатів).

Проєктування та особливості реалізації програмної системи

Основним завданням, яке має вирішити розроблюваний програмний продукт, є можливість розрахувати максимальну ймовірність одночасного виникнення трьох подій за схемою Бернуллі за сформульованими співвідношеннями. Програмний продукт передбачає застосування одразу 10 методів для розрахунків, серед них – 9 існуючих (метод градієнтного спуску; симплекс-метод Nealder-Mead; метод Powell; метод спряжених градієнтів (CG); метод Бройдена-Флетчера-Голдфарба-Шанно (BFGS); метод L-BFGS-B; метод Ньютона (TNC); метод найменших квадратів (SLSQP); метод умовної оптимізації в довірчій області (trust-constr). Десятим є авторська розробка. Структуру проєкту відображено за допомогою діаграми пакетів на рисунку 1.

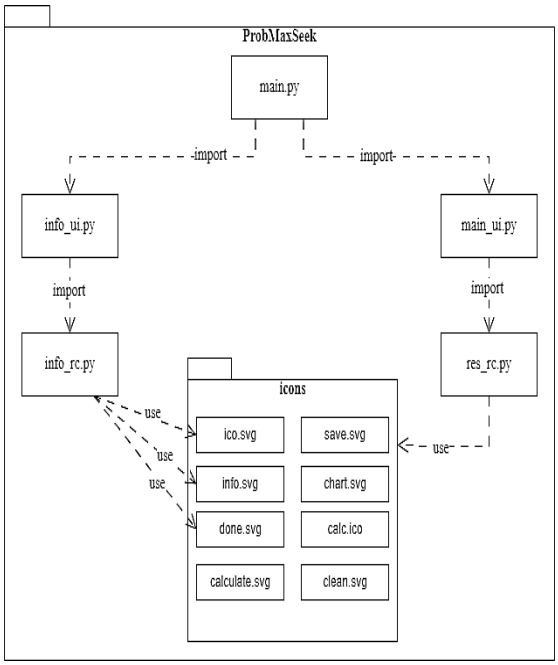

Рисунок 1 – Діаграма пакетів проєкту

Розробка архітектури системи є ключовим етапом у створенні будь-якого програмного продукту, тому ключові функції програми було модельовано за допомогою UML-діаграм. Доступні для користувача дії описані за допомогою різних типів діаграм. Так, важливим моментом роботи програми є збереження результатів, відповідна діаграми наведена на рисунку 2.

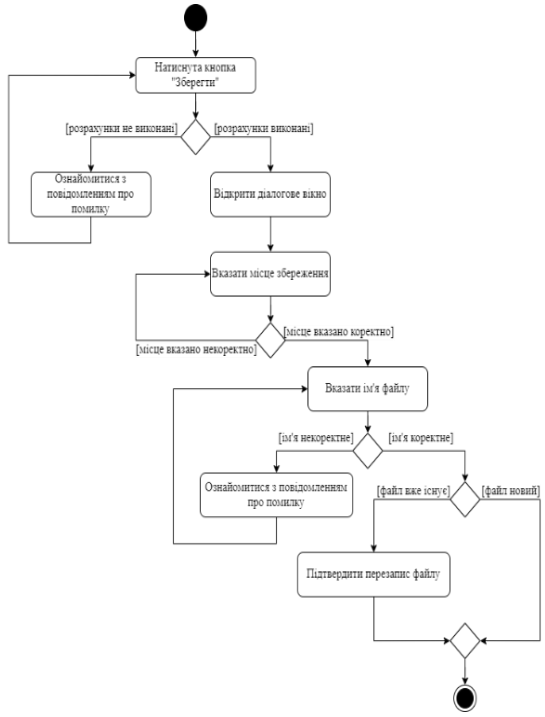

Рисунок 2 – Діаграма діяльності, процес збереження розрахунків

Сутність розробки полягає в тому, що:

– вказується точність розрахунків;

– обирається крок, кратний точності;

– виконується прохід від 0 до 1 з обраним кроком і обирається максимальне значення ймовірності.

Результатом роботи десятого блоку є максимальне значення ймовірності за схемою Бернуллі та кількість виконаних ітерацій.

Блок-схема десятого блоку представлена на рисунку 3.

При розробці програми було використано декілька багатовимірних масивів:

– масив вхідних даних;

– масив послідовностей Фур'є;

– масив з віджетами «Label» для відображення результатів розрахунків;

– масив з назвами методів та, відповідними результатами розрахунків, що призначений для зберігання отриманих результатів;

– масив з часом роботи кожного методу для побудови порівняльного графіку;

– масив для запису даних у файл.

При виконанні програмної розробки обрано мову Python (версія 3.11.5), а середовищем розробки стала IDE PyCharm з достатньо простим та інтуїтивно зрозумілим інтерфейсом, вбудованим терміналом та можливістю завантажувати необхідні бібліотеки безпосередньо з програми.

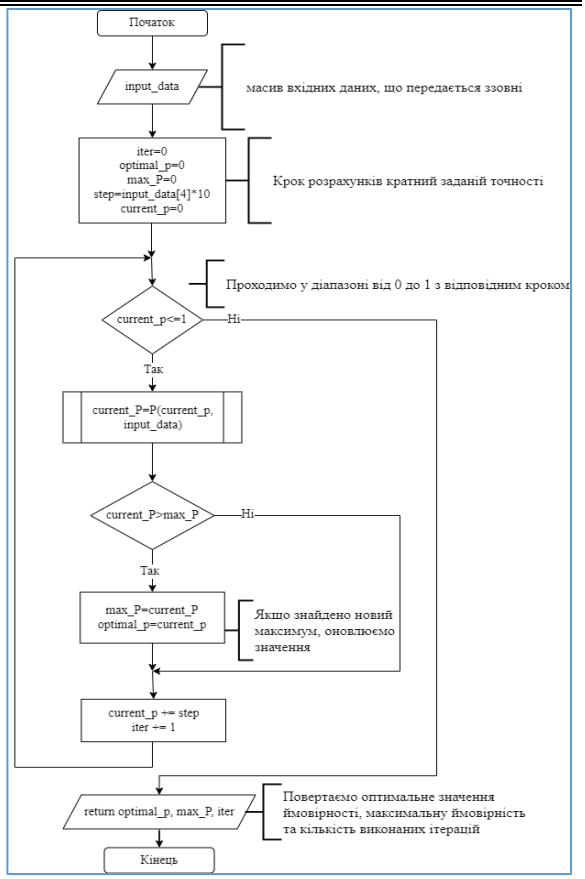

Рисунок 3 – Схема авторського метода

На рисунку 4 відображено головне вікно програми в процесі роботи з тестовими даними.

Рисунок 4 – Головне вікно програми

Для генерації двійкової послідовності використано метод «choice» з бібліотеки random, функція генерації (рис. 5).

```
def generate_random_binary_sequence(length):
    binary_sequence = ""
    for _ in range(length):
        bit = random.choice("01")
        binary_sequence += bit
    return binary_sequence.
```

Рисунок 5 – Функція генерації

Для підрахунку числа комбінацій використано функцію «comb» з бібліотеки math, але з врахуванням випадку, коли $n < k$. Для замірів часу виконання методів використано функцію «time» з однойменної бібліотеки. При збереженні у файл передбачено механізм перевірки, що дозволяє уникати переписування минулих розрахунків.

Результати експериментів

Для розрахунку ймовірності одночасної появи трьох подій було розроблено консольний прототип майбутнього програмного продукту. Програму було протестовано на послідовності, довжиною $n = 16$. Отримані результати ймовірностей наведено на рисунку 6.

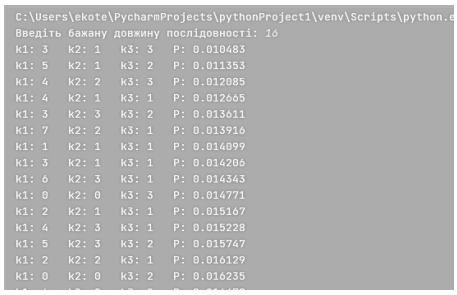

Рисунок 6 – Значення ймовірностей

Розраховані значення збігаються з отриманими у роботі [3]. Для виконання максимізації ймовірності появи трьох подій, розглянуто приклад двійкової послідовності довжиною $n = 5$: 01101.

У наведеній послідовності кількість 2-ланцюгів $\eta(00) = k_1 = 0$, кількість 3 ланцюгів $\eta(100) = k_2 = 0$ та $\eta(110) = k_3 = 1$ відповідно. За умови, що $p = q = 0{,}5$, загальна ймовірність складає $P(\eta(00) = 0, \eta(100) = 0, \eta(110) = 1) = 0{,}21875$.

Для досягнення максимального значення використано метод градієнтного спуску (функцію minimize з бібліотеки SciPy, з модулю optimize). Результат розрахунків за наведеним прикладом продемонстровано на рисунку 7.

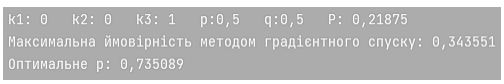

Рисунок 7 – Результати розрахунків максимальної ймовірності

Отримано, що оптимальне значення параметра $p = 0{,}735089$, при якому досягається локальне максимальне значення $P = 0{,}343551$.

У якості параметрів власний метод потребує окрім кількості входжень s-ланцюгів ще й точність розрахунків, тому маємо такі вхідні дані: $n = 5, k_1 = 0, k_2 = 0, k_3 = 1, eps = 0{,}0001$.

Щоб перевірити коректність запропонованого методу, було побудовано послідовність Фур'є, розраховано конкретні значення послідовностей на краях інтервалу (0;1).

Для обраного прикладу послідовності Фур'є мають такий вигляд:

$$fseq(0) = [0, 0, 24, -120, 120]$$

та

$$fseq(1) = [0, -3, -16, -36, 0, 120].$$

Кількість змін знаків у послідовності $fseq(0)$ – 2, а у $fseq(1)$ – 1. Різниця у кількості змін знаків 1. Це свідчить про те, що за теоремою Фур'є маємо не більше одного розв'язку, а за теоремою Ролля – не менше одного розв'язку. Отже, існує точно один розв'язок.

З метою підтвердження правильності результатів, розраховано значення першої похідної у знайденій точці. Тож якщо значення похідної у точці дорівнює нулю, це то це означає знаходження локального максимуму, що й продемонстровано результатами роботи програми, наведеними на рисунку 8. Оскільки для усіх методів значення похідної нуль, можна зробити висновок, що кожен метод знайшов локальний максимум.

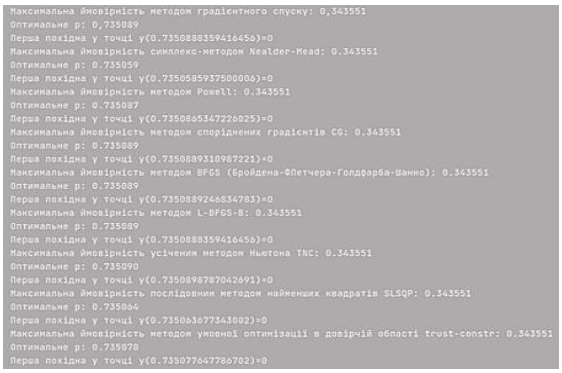

Рисунок 8 – Перевірка коректності роботи методів

Кінцевий розроблений програмний продукт було застосовано для вхідних даних з [3], що продемонстровано результатами роботи консольного варіанту програми на рисунку 3. Остаточні результати розрахунків записано у файл Excel й його частина відображена на рисунках 9-10.

				Послідовність Фур'є:				
Дов	Кіль	Кіль	Кіль	Згенерована послі	fseq(0):	fseq(1):	Ймо	Загальна ймовірність з	Точність
16	3	1	3	1010110101011110	[0 0 0 0 0 0	[0 0 0 0 0 0	0,5	0.0104827880859375	0,0001
16	5	1	2	0001001110000110	[0 0 0 0 0 0	[0 0 0 0 0 0	0,5	0.0113525390625	0,0001
16	4	2	3	1011001100101010	[0 0 0 0 0 0	[0 0 0 0 0 0	0,5	0.0120849609375	0,0001
16	4	1	1	1101010100011000	[0 0 0 0 0 0	[0 0 0 0 0 -	0,5	0.012664794921875	0,0001
16	3	3	2	1010011111001000	[0 0 0 0 0 0	[0 0 0 0 0 0	0,5	0.01361083984375	0,0001
16	7	2	1	1000010110100110	[0 0 0 0 0 6	[0 0 0 0 0 0	0,5	0.013916015625	0,0001
16	1	1	1	1101011111011000	[0 0 0 0 0 0	[0 0 26 75	0,5	0.01409912109375	0,0001
16	3	1	1	1101110111101011	[0 0 0 0 0 0	[0 0 0 0 26	0,5	0.0142059326171875	0,0001
16	6	3	1	0010000011001110	[0 0 0 0 0 3	[0 0 0 0 0 0	0,5	0.0143432617187 5	0,0001
16	0	0	3	0100000110010000	[0 0 0 0 0 0	[0 0 0 - 72 0	0,5	0.0147705078125	0,0001
16	2	1	1	0010000011011001	[0 0 0 0 0 0	[0 0 0 - 72 -	0,5	0.015167236328125	0,0001
16	4	3	1	0111101110001010	[0 0 0 0 0 0	[0 0 0 0 0 0	0,5	0.015228271484375	0,0001

Рисунок 9 – Отримані результати для послідовності Фур'є

Авторський метод

	Оптимальне значення p	Максимальна ймовірість за схемою Бернуллі	Час розрахунків
1	0.5570000000000004	0.011481846662245294	0.0229990482330322
1	0.44600000000000034	0.012277273122926912	0.02299976348876953
1	0.5030000000000003	0.012088426910518833	0.023000478744506836
1	0.4870000000000004	0.012707352136405656	0.0230002403259277
1	0.5260000000000004	0.013856143051560627	0.02299976348876953
1	0.33000000000000024	0.03341391909035702	0.02200055122375488
1	0.8110000000000006	0.06802613703478277	0.022999286651611328
1	0.5700000000000004	0.015566248672847952	0.0230000019073486
1	0.34600000000000025	0.02869485038919756	0.021999835968017578
1	0.7700000000000006	0.10064055314953306	0.02299976348876953
1	0.6800000000000005	0.025016366529449947	0.0239996910095214
1	0.43800000000000033	0.01691235915217643	0.02299976348876953

б)

Рисунок 10 – Отримані результати для авторського методу

Оцінки ефективності

Серед множини критеріїв для порівняння й оцінювання якості результатів застосовано часові критерії, дослідження точності отриманих результатів та складності обчислень.

Критерій 1. Часові характеристики. Програму було запущено з такими вхідними даними: $n = 16, k_1 = 3, k_2 = 2, k_3 = 2, eps = 0{,}001$. Рисунок 11 демонструє час розрахунків за різними методами.

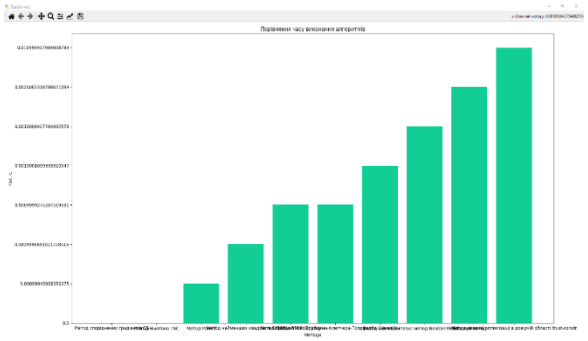

Рисунок 11 – Час розрахунків

Найшвидшими виявилися метод Ньютона та метод споріднених градієнтів. Час роботи авторського методу складає 0.001с., це повільніше, ніж більшість використаних методів, але різниця не є суттєвою. Але авторський метод дозволяє вказувати бажану точність розрахунків, й як результат – зменшити час розрахунків, або навпаки, отримати більш точні результати за довший час. Й це, є, безумовно, вагомою перевагою.

Критерій 2. Точність. Контрольні параметри експерименту – час та точність. Авторський метод запущено багатократно з однаковими вхідними даними, але з різними значеннями точності.

Виявлено, що при точності 0,0001 час зростає приблизно у 4 рази, а при наступних збільшеннях точності – у 10 разів. Експеримент показав, що значення максимальної ймовірності при точності 0,0001 майже повністю збігається зі значенням, отриманим при максимальній точності, тож, точність 0,0001 є достатньою для проведення розрахунків з оглядом на часові вимоги.

Критерій 3. Складність обчислень. Контрольний параметр – оцінка кількості елементарних арифметичних операцій. Зазначимо, що базова функція розрахунку ймовірності одночасної появи трьох подій за схемою Бернуллі має складність $O(n^3)$.

При розрахунках авторським методом використовувались операції піднесення до степеню (множення), комбінаторна формула обчислення кількості комбінацій з n по k (функції «combinations» та «z» з заміною операції розрахунку факторіалу множенням), функція «find_optimal_p», що безпосередньо є реалізацією методу. Піднесення до степеню відбувається двічі за одну ітерацію циклу. Виклики функцій «combinations» та «z» відбуваються n разів. З урахуванням описаного, оцінка операції піднесення до степеню – $O(n^2)$. Оцінка складності функцій «combinations» та «z» – $O(n)$. З урахуванням n – кратності виклику функцій, остаточна оцінка складності функції «find_optimal_p», що реалізовує запропонований метод $-O(n^3)$.

Висновки

Аналіз одночасного виникнення кількох подій в контексті схеми Бернуллі є важливою сучасною задачею, застосовною в криптографії, стеганографії, інформаційній безпеці та інших галузях, де застосовні випадкові події. Існує достатня кількість пакетів, що дозволяють виконувати розрахунки з бітовими послідовностями, але проведення експериментальних досліджень вимагають розробки окремих інструментів, зручних для дослідника, які є повнофункціональними з точки зору обраної методи дослідження, й таких, характеристики яких не нижче за існуючі аналоги.

Особливістю методу, запропонованого в роботі, є застосування дискретного розподілу ймовірностей появи подій.

Автори вважають застосування дискретного розподілу ймовірності появи трьох подій у схемі Бернуллі науковою новизною роботи й прогнозують, що дискретний розподіл імовірності появи трьох подій для аналізу бітових послідовностей у схемі Бернуллі відкриває нові можливості для точного й ефективного аналізу бінарних даних. Отримання максимального значення ймовірності досягається з урахуванням необхідної точності розрахунків та кроку, кратного заданій точності. Тож й результатом роботи програмної розробки є максимальне значення ймовірності за схемою Бернуллі та кількість виконаних ітерацій.

Додатковим функціоналом програмної реалізації є вбудована графічна інтерпретація

результатів, можливість порівнювати застосовані для розрахунків методи візуально. Також функціонал передбачає можливість збереження розрахунків у файл Excel з можливістю застосування отриманих результатів у подальших дослідженнях у інших програмних комплексах.

Список літератури

1. Rukhin À., Soto J., Nechvatal J., Smid M., Barker E., Leigh S., Levenson M., Vangel M., Banks D., Heckert A., Dray J., Vo S. A statistical test suite for random and pseudorandom number generators for cryptographic applications. National Institute of Standards and Technology. Special Publication, 2010. 131 p.
2. Поперешняк С.В Засіб для тестування бітової послідовності на випадковість. Вчені записки ТНУ імені В.І. Вернадського. Серія: технічні науки. 2020. Т. 31 (70) № 4. С. 112-120.
3. Masol, V.I., Popereshnyak, S.V. Checking the Randomness of Bits Disposition in Local Segments of the (0, 1)-Sequence. Cybernetics and Systems Analysis volume 56. 2020. P. 513–520.
4. Масол В.І., Поперешняк С.В. Явний вид розподілу обраних двовимірних та тривимірних статистик (0, 1)-послідовності. 2021, №5. С.72-81.
5. Billingsley P., Statistical Inference for Markov Processes, University of Chicago Pressm, 1961. 75 p.
6. Shannon C. E. Mathematical theory of communication. Bell Syst Tech J, 1948. – 55 p.
7. Mikhaylov V.G. Estimates of accuracy of the Poisson approximation for the distribution of number of runs of long string repetitions in a Markov chain. Discrete Math. Appl. 2016. Vol. 26, N 2. P. 105–113.

References

1. Rukhin, À., Soto, J., Nechvatal, J., Smid, M., Barker, E., Leigh, S., Levenson, M., Vangel, M., Banks, D., Heckert, A., Dray, J., Vo, S. (2010), "A statistical test suite for random and pseudorandom number generators for cryptographic applications", *National Institute of Standards and Technology*, 131 p.
2. Popereshnyak, S.V. (2020), " A tool for testing bit sequences for randomness", [Zasib dlia testuvannia bitovoi poslidovnosti na vypadkovist. Vcheni zapysky TNU imeni V.I. Vernadskoho. Seriia: tekhnichni nauky.] *Scientific Notes of Vernadsky Kyiv Polytechnic National University. Series: technical sciences*. T. 31 (70) № 4, pp. 112-120.
3. Masol, V.I., Popereshnyak, S.V. (2020), "Checking the Randomness of Bits Disposition in Local Segments of the (0, 1)-Sequence", *Cybernetics and Systems Analysis*. Vol. 56, pp. 513–520.
4. Masol, V.I., Popereshnyak, S.V. (2021), " Explicit type of distribution of the selected two-dimensional and three-dimensional statistics of the (0, 1)-sequence" [Yavnyy vyd rozpodilu obranykh dvovymirnykh ta tryvymirnykh statystyk (0, 1)-poslidovnosti]. Vol. 5, pp.72-81.
5. Billingsley, P. (1961), "Statistical Inference for Markov Processes", *University of Chicago Pressm*, pp. 19– 75.
6. Shannon, C.E. (1948), "Mathematical theory of communication", *Bell Syst Tech J*, 55 p.
7. Mikhaylov, V.G. (2016), "Estimates of accuracy of the Poisson approximation for the distribution of number of runs of long string repetitions in a Markov chain", *Discrete Math*. Vol. 26, N 2, pp. 105–113.

Надійшла до редакції 23.01.2024

V. MASOL, N. MASLOVA, Y. KOTEREU
Donetsk National Technical University, Lutsk, Ukraine
volodymyr.masol@donntu.edu.ua, nataliia.maslova@donntu.edu.ua, ekotereu@gmail.com

PROBABILISTIC ANALYSIS OF THE SIMULTANEOUS OCCURRENCE OF SEVERAL EVENTS IN THE CONTEXT OF THE BERNOULLI SCHEME

One of the fundamental models in the probabilistic analysis of bit sequences is the Bernoulli scheme. The paper reviews modern developments in the indicated direction, proposes the use of a discrete distribution for analyzing the probability of the simultaneous occurrence of several events in the context of the Bernoulli scheme, describes the mathematical model and software implementation of the method of achieving the maximum probability, and obtained qualitative evaluations of the basic and proposed methods. An example of the application of discrete distribution is given. The software product implemented and compared ten methods - nine existing optimization methods and the author's method. The correctness of the software implementation of the applied methods has been substantiated and proven, the working time, the complexity of the calculations and the dependence of the time characteristics on the given accuracy of the calculations were evaluated. The analysis of the simultaneous occurrence of several events in the context of the Bernoulli scheme is an important modern task applicable in cryptography, steganography, information security and other fields.

Key words: bit sequences, discrete distribution, quality assessments, probabilistic analysis, cryptography