

УДК 512.44

DOI [https://doi.org/10.24144/2616-7700.2025.46\(1\).104-108](https://doi.org/10.24144/2616-7700.2025.46(1).104-108)**Г. І. Сливка-Тилищак¹, М. В. Маді², М. О. Тилищак³**

¹ ДВНЗ «Ужгородський національний університет»,
професор кафедри теорії ймовірностей і математичного аналізу,
доктор фізико-математичних наук
anna.slyvka@uzhnu.edu.ua
ORCID: <https://orcid.org/0000-0002-7129-0530>

² ДВНЗ «Ужгородський національний університет»,
аспірант кафедри теорії ймовірностей і математичного аналізу
mykhailo.madi@uzhnu.edu.ua
ORCID: <https://orcid.org/0009-0000-9356-3408>

³ Ужгородський науковий ліцей,
здобувач
marichkatlk@gmail.com
ORCID: <https://orcid.org/0009-0002-4379-1423>

ПОБУДОВА ВИБІРКИ З МНОЖИНИ МІРИ НУЛЬ ТА ЇЇ ХАРАКТЕРИСТИКИ

Робота присвячена дослідженню методів побудови вибірок з множин, що мають міру нуль. Розглянуто два способи побудови вибірок з множини Кантора та проаналізовано їх числові характеристики, зроблено висновки про принципову подібність чи відмінність отриманих вибірок.

Отримано, що вибіркове середнє у всіх вибірках обох серій для множини Кантора практично не відрізняються, а послідовність середньоквадратичних відхилень вибірок першої серії (ітераційної) стабілізується на другому кроці і прямує до середньоквадратичних відхилень усіх вибірок другої серії.

Ключові слова: множини міри нуль, множина Кантора, числові характеристики вибірки.

1. Вступ. Поняття міри є узагальненням таких числових характеристик множин як довжина, площа, об'єм. Ці характеристики застосовуються в різних просторах, для різних класів множин, але їх значення мають спільні риси: вони невід'ємні, і для об'єднання двох множин, що не перетинаються, дорівнюють сумі значень цих двох множин. Як це часто робиться в математиці, замість окремого вивчення довжини, площі, об'єму тощо, ми досліджуємо загальну числову функцію множин — міру, що визначена на певному абстрактному наборі множин і задовольняє дві вказані властивості. Можна побудувати багато цікавих множин, які не мають міри, тобто міра яких рівна нулю. Робота присвячена дослідженню методів побудови вибірок з множин, що мають міру нуль. Дана тема є актуальною тим, що вона допомагає уявити, що таке множина міри нуль. Методи побудови вибірок із складно сконструйованих математичних множин є слабо досліджені.

2. Основний результат. Множина Кантора — підмножина відрізка $[0, 1]$, яку вперше ввів німецький математик Георг Кантор. Множина утвориться з відрізка, якщо послідовно вилучити такі інтервали: спочатку на кроці $n = 1$ вилучаємо інтервал $(\frac{1}{3}, \frac{2}{3})$ (тобто центральну третину відрізка). Отже залишиться $A_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$. Далі на кроці $n = 2$ вилучаємо відрізки $(\frac{1}{9}, \frac{2}{9})$ та $(\frac{7}{9}, \frac{8}{9})$

(тобто центральні третини тих двох відрізків $[0, \frac{1}{3}]$ та $[\frac{2}{3}, 1]$, які залишилися. Отже залишиться $A_2 = [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1]$, і так далі. Множина K , яка залишилася, називається множиною Кантора [1]. Таким чином $K = \bigcap_{i=0}^{\infty} A_i$.

Для описання множин дійсних чисел скористаємось інтуїтивним поняттям міри. Цікаво, що міра відрізків, які залишаються на кожному кроці формування множини Кантора зменшується. Оскільки на кожному кроці відкидаються центральні третини тих відрізків, які залишилися, то після кожного кроку залишається $\frac{2}{3}$ від попередньої міри. Таким чином при $n = 1$ це $\frac{2}{3}$, при $n = 2$ це $\frac{4}{9}$, а для довільного n — $(\frac{2}{3})^n$. Якщо n спрямувати до нескінченності, то міра відрізків, які залишаються, прямує до 0. Таким чином міра множини Кантора рівна 0.

2.1. Генерація вибірки з множини Кантора методом хаоса. Для унаочнення множини Кантора K , скористаємось рівномірно розподіленою випадковою величиною над якою проводимо деякі обчислення та відповідними вибірками. Нехай x — рівномірно розподілена на множині $[0; 1]$ випадкова величина, тобто ймовірність попадання у підмножину цієї множини рівна її мірі, бо довжина відрізка $[0; 1] \in 1$. Перемістимо x втричі ближче до 0 або втричі ближче до 1. Тобто для дискретної рівномірно розподіленої випадкової величини δ , що приймає значення 0 або 1, кожне з однаковою ймовірністю $\frac{1}{2}$, розглянемо

$$y = f(x) = \begin{cases} \frac{x}{3}, & \text{якщо } \delta = 0 \\ \frac{x+2}{3}, & \text{якщо } \delta = 1. \end{cases}$$

Зрозуміло, що y є рівномірно розподіленою випадковою величиною на множині $A_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$, якщо при кожному зверненні до $f(x)$ незалежно генерується δ , що приймає значення 0 або 1, кожне з однаковою ймовірністю $\frac{1}{2}$. Якщо б x була рівномірно розподіленою випадковою величиною на множині A_1 , то y була б рівномірно розподіленою випадковою величиною на множині $A_2 = [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1]$. І так далі.

Побудуємо в деякому наближенні вибірку з множини Кантора. Використаємо метод хаосу. Скористаємось вибіркою $x_{01}, x_{02}, \dots, x_{0k}$ випадкової величини x_0 рівномірно розподіленої на множині $[0; 1]$. Вона буде ілюструвати множину $A_0 = [0, 1]$ при зображенні. Далі будемо випадкові величини $x_1 = f(x_0), x_2 = f(x_1), \dots$ та їх вибірки $x_{i+1,j} = f(x_{i,j})$ ($j = 1, \dots, k$). Зрозуміло, що $x_{1j} = f(x_{0j})$ для кожного j обчислюється за однією з формул $x_{1j} = \frac{x_{0j}}{3}$, або $x_{1j} = \frac{x_{0j}+2}{3}$. Оскільки $x_{0j} \in [0; 1]$, то в першому випадку $x_{1j} \in [0, \frac{1}{3}]$, в другому випадку $x_{1j} \in [\frac{2}{3}, 1]$, тоді $x_{1j} \in [0, \frac{1}{3}] \cup [\frac{2}{3}, 1] = A_1$. Продовжуючи далі, з умови $x_{1j} \in [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$ та того, що $x_{2j} = f(x_{1j})$ для кожного j обчислюється за однією з формул $x_{2j} = \frac{x_{1j}}{3}$, або $x_{2j} = \frac{x_{1j}+2}{3}$. Оскільки в першому випадку $x_{2j} \in [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}]$, в другому випадку $x_{2j} \in [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1]$, тоді $x_{2j} \in [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1] = A_2$. І так далі. Кожна з вибірок $x_{i,j}$, ($j = 1, \dots, k$) ілюструватиме множину A_i , до якої, очевидно, належать всі її елементи.

Оскільки, кожна множина A_i симетрична відносно 0.5, то середнє значення

$$\bar{x}_i = \frac{x_{i1} + x_{i2} + \dots + x_{ik}}{k}$$

має наближатися до 0,5 при зростанні k . Середньоквадратичне відхилення бу-

демо обчислювати за формулою:

$$D_i = \sqrt{(x_i - \bar{x}_i)^2} = \sqrt{\frac{(x_{i1} - \bar{x}_i)^2 + (x_{i2} - \bar{x}_i)^2 + \dots + (x_{ik} - \bar{x}_i)^2}{k}}.$$

Запуск коду, що починає з вибірки в $k = 1000$ елементів рівномірно розподіленої випадкової величини на проміжку $[0; 1]$, генерує вибірки поданих характеристик та зображує кожну вибірку точками на одному горизонтальному відрізку, а різні вибірки на різних горизонтальних відрізках.

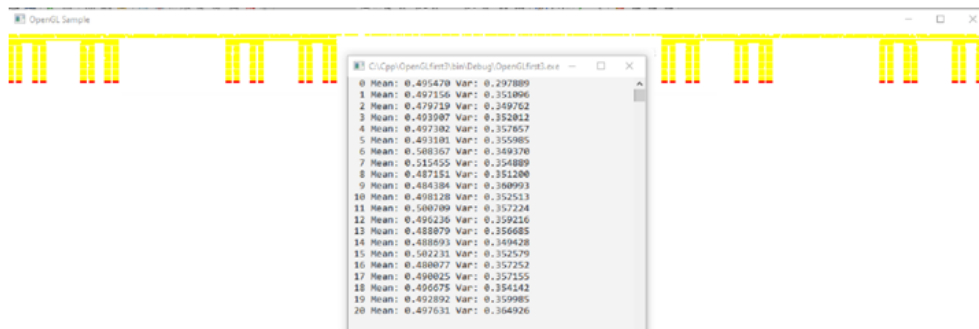


Рис. 1

Таблиця 1

i	Вибіркове середнє	Середньо квадратичне відхилення	i	Вибіркове середнє	Середньо квадратичне відхилення
0	0.504464	0.287646	10	0.508515	0.360488
1	0.491488	0.344500	11	0.509505	0.355088
2	0.511163	0.347142	12	0.513835	0.350003
3	0.509721	0.350557	13	0.504612	0.354932
4	0.503241	0.350744	14	0.522204	0.350737
5	0.501080	0.360044	15	0.520735	0.351285
6	0.495694	0.354741	16	0.529579	0.348256
7	0.487898	0.349755	17	0.483859	0.348135
8	0.496632	0.345271	18	0.494620	0.348737
9	0.487544	0.351190	19	0.483540	0.352806

Аналіз таблиці показує, що вибіркове середнє, дійсно знаходиться завжди біля 0.5, відрізняється від нього не більше ніж на 0.03, а середньо квадратичне відхилення, крім першого значення, знаходиться завжди біля 0.35 і відрізняється від нього не більше ніж на 0.011.

2.2. Генерація вибірки з множини Кантора методом перетворення цифр. Зменшимо множину Кантору, послідовно вилучаючи такі інтервали: спочатку на кроці $n = 1$ вилучаємо інтервал $[1/3, 2/3]$ разом з лівою межею. Подаватимемо числа в цифровому записі в 3-вій системі числення, вигляду

$0, a_1 a_2 a_3 a_4, \dots (a_i \in \{0, 1, 2\})$. Очевидно, $0, a_1 a_2 a_3 \dots a_i 22 \dots = 0, a_1 a_2 a_3 \dots (a_i + 1)00 \dots$, якщо $a_i < 2$. Тому, далі відмовляємося від представлення, які не закінчуються послідовністю 2-йок. Тобто $0, a_1 a_2 a_3 a_4 \dots = \frac{a_1}{3} + \frac{a_2}{3^2} + \frac{a_3}{3^3} + \frac{a_4}{3^4} + \dots$. Тоді на першому кроці вилучають числа, в описаному цифровому записі якого $a_1 = 1$. Отже залишаться $A_1 = [0, \frac{1}{3}) \cup [\frac{2}{3}, 1)$. Тобто на першому кроці залишаються числа, в описаному цифровому записі якого $a_1 = 0$ або $a_1 = 2$. Далі, на кроці $n = 2$ вилучаємо відрізки $[\frac{1}{9}, \frac{2}{9})$ та $[\frac{7}{9}, \frac{8}{9})$. Тоді на другому кроці вилучаються ще числа, в описаному цифровому записі якого $a_2 = 1$. Отже, залишаться $A_2 = [0, \frac{1}{9}) \cup [\frac{2}{9}, \frac{1}{3}) \cup [\frac{2}{3}, \frac{7}{9}) \cup [\frac{8}{9}, 1)$. Отже на другому кроці залишаються числа, в описаному цифровому записі якого $a_1, a_2 \in \{0, 2\}$. Таким чином, побудована підмножина Ω множини Кантора складається з чисел в цифровому записі в 3-вій системі числення, вигляду $0, a_1 a_2 a_3 a_4 \dots, (a_i \in \{0, 2\})$. Оскільки, відкидання чи приєднання межевої точки довільного інтервалу, ніяк не впливає на характеристики вибірки рівномірно розподіленої величини з цього інтервалу, то ми можемо знехтувати неможливістю розгляду чисел, що є лівою межею відкидуваних інтервалів. Очевидно ліва межа відкидуваних проміжків є числа вигляду $0, a_1 a_2 a_3 \dots a_i 1100 \dots$. Тобто для кожного фіксованого i множина чисел такого вигляду скінчена. Таким чином для довільного i множина чисел такого вигляду зліченна. Натомість, множина Ω незліченна, бо вона еквівалентна множині чисел вигляду $0, a_1 a_2 a_3 a_4 \dots, (a_i \in \{0, 2\})$, які не закінчуються послідовністю 2-йок в 3-й системі числення. Або множині чисел вигляду $0, a_1 a_2 a_3 a_4 \dots, (a_i \in \{0, 1\})$, які не закінчуються послідовністю 1-чок в 2-й системі числення. Тобто Ω еквівалентна інтервалу $[0, 1)$ і є незліченною.

З описаного вище, можна запропонувати наступний алгоритм побудови вибірки з множини Кантора:

- 1) Генеруємо рівномірно розподілену випадкову величину з проміжку $[0, 1)$.
- 2) Кожне число вибірки записуємо в двійковій системі числення, щоб запис не закінчувався послідовністю 1-чок.
- 3) Міняємо запис замінюючи цифри 1 на 2.
- 4) Одержаний запис трактуємо в трійковій системі числення, що подасть число з підмножини Ω множини Кантора.

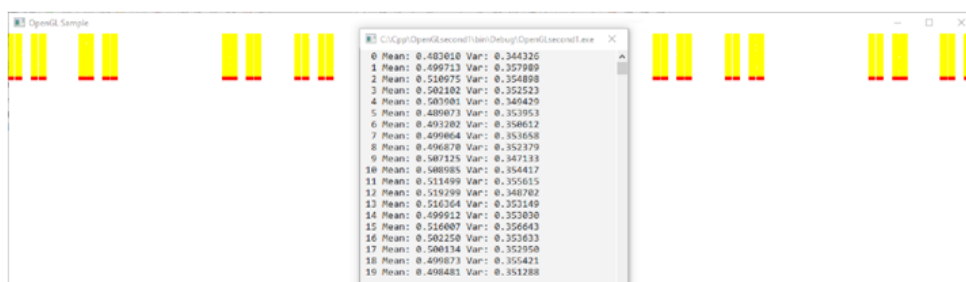


Рис. 2

Аналіз таблиці показує, що вибіркове середнє, дійсно наближено до 0.5 відрізняється від нього не більше ніж на 0.02, а середньо квадратичне відхилення знаходиться завжди біля 0.35 і відрізняється від нього не більше ніж на 0.001. Ці характеристики є кращими ніж аналогіні у вибірок згенерованих методом

Таблиця 2

i	Вибіркове середнє	Середньо квадратичне відхилення	i	Вибіркове середнє	Середньо квадратичне відхилення
0	0.483010	0.344326	10	0.511499	0.355615
1	0.499713	0.357989	11	0.519299	0.348702
2	0.510975	0.354898	12	0.516364	0.353149
3	0.502102	0.352523	13	0.499912	0.353030
4	0.503901	0.349429	14	0.516007	0.356643
5	0.489073	0.353953	15	0.502250	0.353633
6	0.493202	0.350612	16	0.500134	0.352950
7	0.499064	0.353658	17	0.499873	0.355421
8	0.496870	0.352379	18	0.498481	0.351288
9	0.507125	0.347133	19	0.511499	0.355615

хаоса, хоча середні значення отримуваних вибірових середніх та середньоквадратичних відхилень практично не відрізняються.

Список використаної літератури

1. G. Cantor. On the Power of Perfect Sets of Points (De la puissance des ensembles parfait de points). *Acta Mathematica*. 1884. Vol. 4. P. 381–392.
2. Огірко О. І., Галайко Н. В. Теорія ймовірностей та математична статистика: навчальний посібник. Львів : ЛьвДУВС, 2017. 292 с.
3. Бевз В. Г. Історія математики. Харків : Основа, 2006. 144 с.

Slyvka-Tylyshchak G. I., Madi M. V., Tylyshchak M. O. Construction of a sample from a set of measure zero and its characteristics.

This work is devoted to the study of methods for constructing samples from sets of measure zero. Two methods of constructing samples from the Cantor set are considered and their numerical characteristics are analyzed. Conclusions are drawn regarding the fundamental similarity or difference between the obtained samples.

It was found that the sample means in all samples from both series for the Cantor set are practically identical, and the sequence of sample standard deviations in the first (iterative) series stabilizes at the second step and converges to the standard deviations of all samples in the second series.

Keywords: sets of measure zero, the Cantor set, statistical characteristics of the sample.

References

1. Cantor, G. (1884). On the Power of Perfect Sets of Points (De la puissance des ensembles parfait de points). *Acta Mathematica*, 4, 381–392.
2. Ohirko, O. I., & Halaiko, N. V. (2017). *Probability Theory and Mathematical Statistics: A Textbook*. Lviv: Lviv State University of Internal Affairs [in Ukrainian].
3. Bevez, V. H. (2006). *History of Mathematics*. Kharkiv: Osnova [in Ukrainian].

Одержано 01.04.2025