ВІЗУАЛІЗАЦІЯ ДЕЯКИХ ТЕОРЕМ ТЕОРІЇ ЙМОВІРНОСТЕЙ ТА МАТЕМАТИЧНОЇ СТАТИСТИКИ ЗАСОБАМИ МОВИ PYTHON

VISUALIZATION OF SOME THEOREMS OF THE THEORY OF PROBABILITY AND MATHEMATICAL STATISTICS USING THE TOOLS OF THE PYTHON LANGUAGE

УДК 378
DOI https://doi.org/10.32782/2663-6085/2024/74.40

Тегза А.М.,
канд. фіз.-мат. наук,
доцент кафедри теорії ймовірностей і математичного аналізу
Державного вищого навчального закладу «Ужгородський національний університет»

Імре Ю.Ю.,
ст. викладач кафедри фізико-математичних дисциплін
Українсько-угорського навчально-наукового інституту Державного вищого навчального закладу «Ужгородський національний університет»

Кенгерц С.С.,
ст. викладач кафедри фізико-математичних дисциплін
Українсько-угорського навчально-наукового інституту Державного вищого навчального закладу «Ужгородський національний університет»

Стаття присвячена проблемам та методам ефективної підготовки спеціалістів в області вищої математики. Зокрема у цій статті досліджено методи візуалізації деяких ключових теорем теорії ймовірностей і математичної статистики за допомогою мови програмування Python. Сучасна наука і освіта вимагають ефективних інструментів для наочного представлення складних математичних структур та концепцій, і Python, завдяки своїй гнучкості та потужним бібліотекам, таким як Matplotlib, Random і NumPy, надає широкі можливості для таких візуалізацій. Особливу увагу приділено поясненню основної теореми закону великих чисел, центральній граничній теоремі, точковим оцінкам та довірчим інтервалам невідомих параметрів нормального розподілу. Тобто основною метою дослідження є демонстрація того, як графічні методи можуть бути використані для покращеного розуміння і навчання студентами деяких важливих теорем і понять. В статті наведено кілька конкретних прикладів візуалізації, що ілюструють практичне застосування теоретичних понять та знань в цілому. Зокрема, графічно і статистично перевірено виконання теореми Хінчина, тобто доведено, що при збільшенні об'єму вибірки розподіл середнього арифметичного однаково розподілених випадкових величин ущільнюється навколо свого теоретичного математичного сподівання. Особлива увага зосереджена поясненню точності та інтерпретації графічних результатів. Наприклад, при дослідженні двох методів побудови точкових оцінок невідомого параметру розподілу одержали графіки і основні статистичні розрахунки цих оцінок, аналізуючи які, зробили висновок, що оцінка, одержана методом максимальної вірогідності, є точнішою у порівнянні з оцінкою, одержаною методом моментів.
Стаття загалом буде корисною для дослідників, викладачів, студентів та всіх, хто цікавиться математикою, статистикою і програмуванням, і може слугувати практичною рекомендацією для тих, хто хоче освоїти сучасні методи візуалізації математичних даних.
Ключові слова: *закон великих чисел, центральна гранична теорема, точкові оцінки парамет-рів розподілу, довірчі інтервали, візуалізація, Python.*

This article is dedicated to the challenges and methods of effective training for specialists in the field of higher mathematics. In particular, the article explores methods of visualizing key theorems in probability theory and mathematical statistics using the Python programming language. Modern science and education demand effective tools for the visual representation of complex mathematical structures and concepts, and Python, with its flexibility and powerful libraries such as Matplotlib, Random, and NumPy, provides ample opportunities for such visualizations. Special attention is given to explaining the Law of Large Numbers, the Central Limit Theorem, point estimates, and confidence intervals for unknown parameters of the normal distribution. The primary objective of the study is to demonstrate how graphical methods can enhance students' understanding and learning of important theorems and concepts. The article presents several specific examples of visualizations that illustrate the practical application of theoretical concepts and knowledge. In particular, the Khintchine's theorem is graphically and statistically verified, showing that as the sample size increases, the distribution of the arithmetic mean of identically distributed random variables becomes concentrated around its theoretical expected value. Special emphasis is placed on explaining the accuracy and interpretation of graphical results. For instance, when investigating two methods for constructing point estimates of an unknown distribution parameter, the graphs and main statistical calculations of these estimates were obtained. Analyzing these, the conclusion was drawn that the estimate obtained by the maximum likelihood method is more accurate compared to the estimate obtained by the method of moments. Overall, the article will be useful for researchers, educators, students, and anyone interested in mathematics, statistics, and programming. It can serve as a practical guide for those who wish to master modern methods of visualizing mathematical data.
Key words: *law of large numbers, central limit theorem, point estimates of distribution parameters, confidence intervals, visualization, Python.*

Постановка проблеми у загальному вигляді. При сучасному вивченні у вищій школі математичних дисциплін, зокрема теорії ймовірностей та математичної статистики, використання візуальних методів є дуже важливим для кращого розуміння та засвоєння матеріалу студентами. Технічні засоби візуалізації, такі як комп'ютерні програми, графічні калькулятори та інтерактивні платформи, дозволяють не тільки унаочнити абстрактні математичні концепції, але й демонструвати виконання понять, суджень і теорем на конкретних прикладах, тобто в дії показують їх зміст.

Аналіз останніх досліджень і публікацій. Багато досліджень присвячено візуалізації статистичних даних та теоретичних результатів з використанням різних інструментів, зокрема, пакетів Statistica [3], R [2], мови програмування Python [1, 4] та платформ WolframAlpha, Desmos та GeoGebra. Проте, не всі аспекти візуалізації теорем теорії ймовірностей і математичної статистики були повністю розкриті, особливо в контексті їх застосування в освітніх цілях та для підтримки наукових досліджень.

Виділення невирішених раніше частин загальної проблеми. Незважаючи на значний прогрес у візуалізації математичних та статистичних понять, невирішеними залишаються ще деякі питання стосовно ефективного сприйняття студентами деяких фундаментальних теорем теорії ймовірностей (зокрема, теореми Хінчина, центральної граничної теореми Ляпунова для однаково розподілених величин) та понять математичної статистики (наприклад, точкових та інтервальних оцінок невідомих параметрів розподілу). Також потребують уваги методи візуалізації, які дозволяють інтерактивно досліджувати ці теореми.

Мета статті це:

1. Демонстрація візуалізації основних теорем теорії ймовірностей і математичної статистики за допомогою засобів Python.
2. Створення наочних графіків та діаграм, що сприяють кращому розумінню матеріалу.
3. Рекомендації щодо використання візуалізації в навчальному процесі.

Виклад основного матеріалу.

1. Закон великих чисел

Визначення: Нехай маємо послідовність випадкових величин $\xi_1, \xi_2, ...$ та випадкових величин $\eta_n = \frac{1}{n}(\xi_1 + ... + \xi_n)$. Кажуть, що для послідовності $\xi_1, \xi_2, ...$ має місце закон великих чисел, якщо існує числова послідовність $a_1, a_2, ...$, що послідовність η_n збігається до послідовності a_n за ймовірністю при $n \to \infty$ Іншими словами ймовірність того, що середнє вибіркове відхилиться від математичного сподівання на задану величину, зменшується зі збільшенням вибірки.

Є ряд теорем, які досліджують різні умови, при яких виконується ЗВЧ. Зокрема теорема Хінчина стверджує, що якщо випадкові величини незалежні та однаково розподілені і мають скінченне однакове математичне сподівання, то середнє арифметичне цих величин збігається до свого математичного сподівання за ймовірністю.

Засоби візуалізації:

– **Використання комп'ютерних програм: Python та R.** Програми, написані на Python або R, можуть генерувати великі набори даних і будувати відповідні графіки та діаграми.

– **Інтерактивні платформи WolframAlpha, Desmos та GeoGebra:** Ці онлайн-платформи дозволяють створювати інтерактивні графіки, які студенти можуть досліджувати самостійно.

Приклад візуалізації.

Для візуалізації цієї теореми можна використовувати графіки побудовані за допомогою Python, які демонструють, як зростання розміру вибірки зменшує різницю між вибірковим середнім та математичним сподіванням.

Наприклад, можна сформулювати студентам таке завдання: застосовуючи мову програмування Python, перевірити виконання теореми Хінчина при умові, що випадкові величини мають гама розподіл з параметром 7.

За однією вибіркою очевидно, що не можна перевірити виконання ЗВЧ. Тому для демонстрації даної теореми згенеруємо k вибірок об'єму n і тоді за допомогою графіків та числових характеристик можна переконатись, що при зростанні об'єму вибірки розподіл середніх значень вибірок ущільнюється.

Згенеруємо, до прикладу, k=20 вибірок розподілу Г(7) об'ємами n=15, 100, 500, 1000. Зауважимо, що теоретичне математичне сподівання цих випадкових величин рівне параметру розподілу, тобто числу 7. Далі для всіх 20 колонок обчислимо їх середні значення. Для всіх n покажемо ці результати на графіках ламаних (рис. 1) та точок скупчень середніх значень (рис. 2).

З графіків видно, що при збільшенні об'єму вибірки розподіл середнього арифметичного однаково розподілених випадкових величин ущільнюється навколо свого теоретичного математичного сподівання. Статистичні характеристики цих чотирьох масивів (рис. 3) підтверджують ці результати.

Таким чином через графіки і розрахункові характеристики студенти мають можливість краще усвідомити прочитаний ними теоретичний матеріал.

2. Центральна гранична теорема

Визначення: Під ЦГТ розуміють твердження, у яких іде мова про збіжність розподілів сум до нормального закону розподілу. Зокрема, ЦГТ для однаково розподілених випадкових величин стверджує, що нормована сума великої кількості незалежних випадкових величин з однаковим розподілом, незалежно від їх початкового розподілу, має розподіл, близький до стандартного нормального.

Методологія візуалізації на приклад:

Застосовуючи мову програмування Python, можна перевірити виконання ЦГТ для нормованої суми, до прикладу, k=15 незалежних випадкових величин розподілу Лапласа $L(2,1)$ об'ємом вибірки 700. Зауважимо, що теоретичні математичне сподівання і дисперсія рівні $M\xi = a = 2$, $D\xi = 2b^2 = 2$.

Для демонстрації виконання ЦГТ подамо два рисунки: гістограму згенерованої однієї випадкової величини розподілу Лапласа (рис. 4) і гістограму нормованої суми 15-ти випадкових величин цього ж розподілу (рис. 5). З гістограм наглядно видно, що розподіл нормованої суми вже близький до стандартного нормального закону $N(0,1)$.

3. Точкові оцінки параметрів розподілу

Визначення: Метод моментів і метод максимальної вірогідності є двома широко

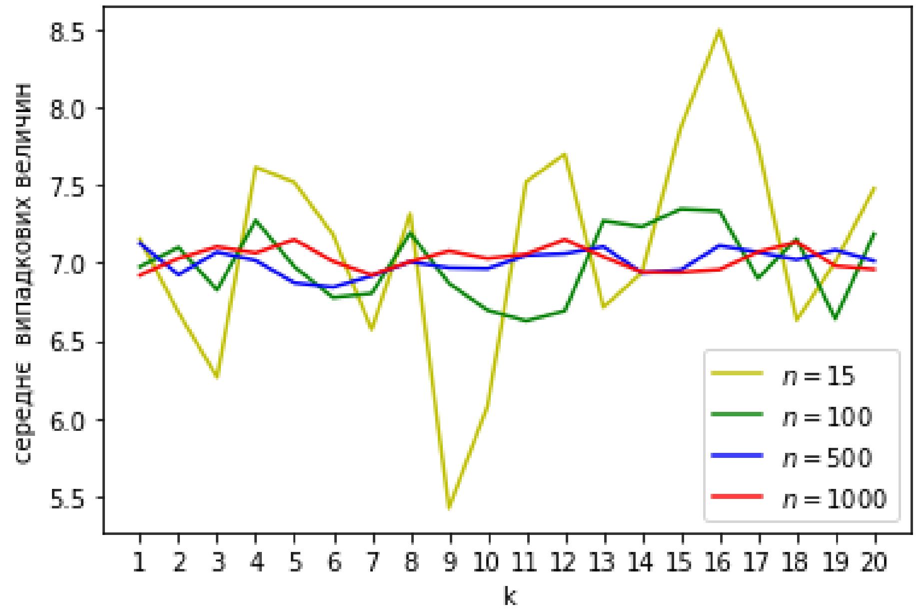

Рис. 1.

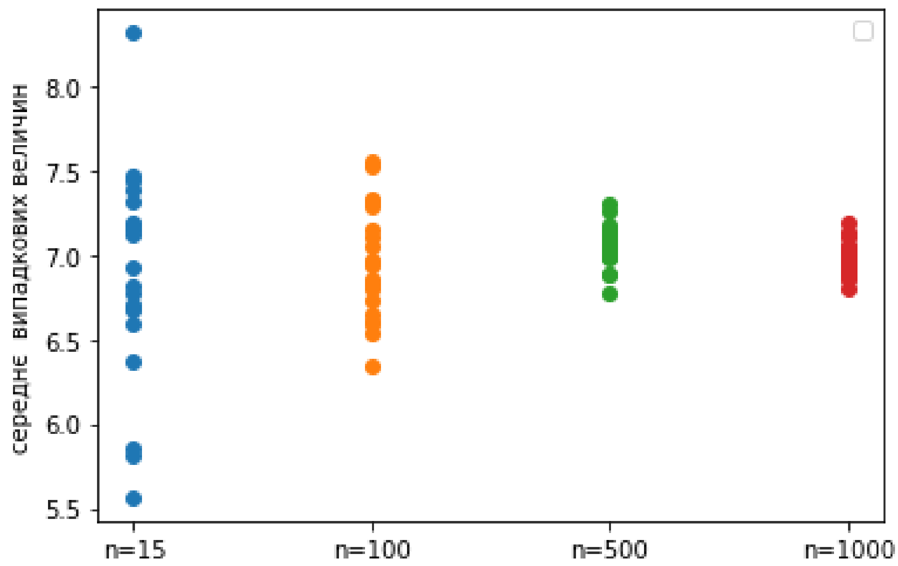

Рис. 2.

```
---------------------------------
Статистика результатів
---------------------------------
             min       max         S
n=15    5.564830  8.314222  0.642636
n=100   6.336708  7.546515  0.334241
n=500   6.778941  7.304554  0.123376
n=1000  6.805145  7.187392  0.099198
```

Рис. 3.

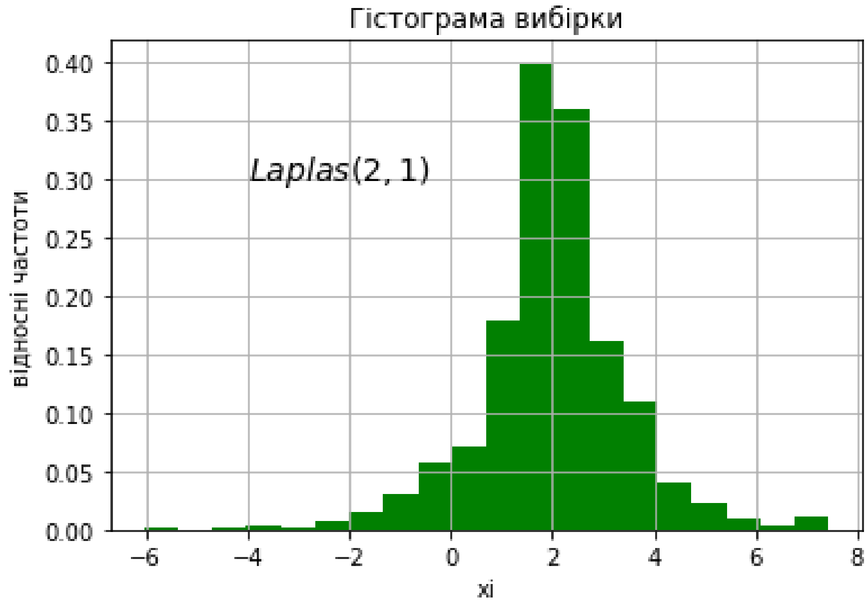

Рис. 4.

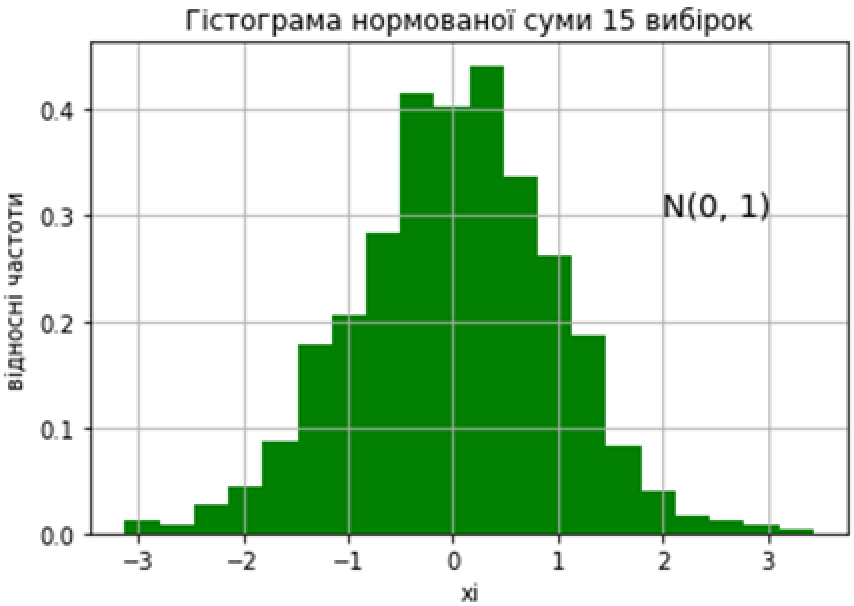

Рис. 5.

використовуваними методами для точкової оцінки невідомих параметрів розподілу в математичній статистиці. Метод моментів полягає у прирівнюванні теоретичних і емпіричних моментів відповідних порядків. Розв’язавши рівняння чи систему рівнянь, одержують значення шуканих параметрів розподілу. Метод максимальної вірогідності полягає у відшуканні таких значень параметрів, які роблять спостережувані дані найбільш ймовірними, тобто ідея методу полягає в максимізації ймовірності того, що наявні дані мають розподіл з цими параметрами.

Методологія візуалізації на прикладі:

Методом моментів та методом максимальної вірогідності отримано наступні оцінки невідомого параметра *a*, до прикладу, рівномірного закону розподілу $R(a=4;14)$: $\hat{a}=2\overline{x}_в-14$ та $\hat{a}=\frac{n-1}{n}min\{x_1,x_2,...,x_n\}$.

Використовуючи Python, продемонструємо поведінку одержаних оцінок, до прикладу, для 15-ти вибірок при збільшенні об’єму вибірки: n=10, n=100, n=500. Для цього згенеруємо k=15 вибірок рівномірного закону розподілу об’ємом n=10. Для кожної колонки побудуємо дві оцінки, згадані вище. Одержимо графік ламаних (рис. 6).

Аналогічно можна одержати графіки для більших об’ємів вибірки, зокрема при *n*=500 матимемо (рис. 7).

Аналізуючи графіки, бачимо динаміку розсіювання двох оцінок $\hat{a}_1$ і $\hat{a}_2$ при збільшенні об’єму

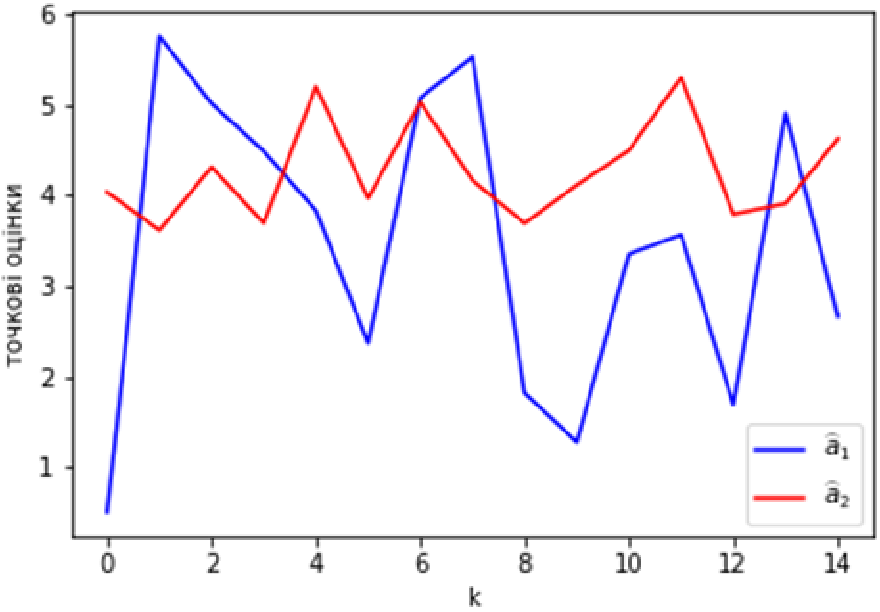

Рис. 6.

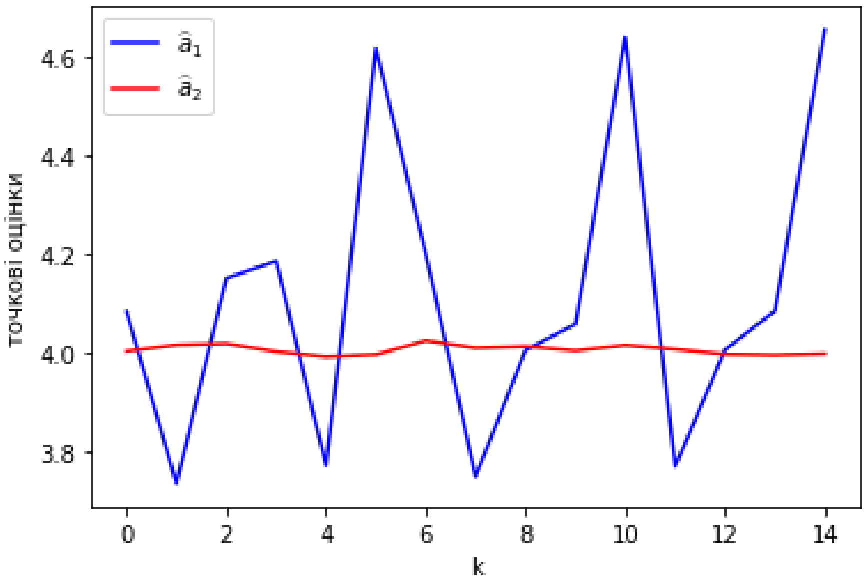

Рис. 7.

вибірки, за якою, можна зробити висновок, що оцінка $\hat{a}_2$ дає кращі результати.

За розмахом, середнім значенням та виправленим середньоквадратичним відхиленням робимо той самий висновок, тобто, що оцінка $\hat{a}_2$, одержана методом максимальної вірогідності, дає кращі результати у порівнянні з методом моментів.

4. Побудова довірчих інтервалів для параметрів розподілу

Визначення: Довірчі інтервали – це інтервали, побудовані на основі вибіркових даних, в яких з деякою ймовірністю (довірчим рівнем) міститься справжнє значення параметра розподілу. Вони використовуються для оцінки невідомих параметрів розподілу і дають інформацію про те, наскільки точною є ця оцінка.

Процес побудови довірчого інтервалу залежить від типу параметра (наприклад, середнє або дисперсія) і відомих властивостей розподілу. Наприклад, довірчий інтервал для оцінки математичного сподівання нормального розподілу при відомому σ має наступний формульний вигляд: $(\bar{x}-\frac{t_\alpha \sigma}{\sqrt{n}};\bar{x}+\frac{t_\alpha \sigma}{\sqrt{n}})$, де $\bar{x}$ – вибіркове середнє, t_α – розв'язок рівняння: $2\Phi_0(t_\alpha)=\alpha$, тобто t_α – є квантиллю, порядку $\frac{1+\alpha}{2}$ стандартного нормального розподілу.

Методологія візуалізації на прикладі:

Для наочності в середовищі Python побудуємо довірчий інтервал для оцінки математичного сподівання *a* нормального розподілу при відомому $\sigma = 1.5$ з надійністю α=0.95 і графічно представимо результати. Для цього генеруємо, наприклад, k=30 вибірок об'ємом n=35 нормального розподілу N(9, 1.5). Далі для всіх цих вибірок знайдемо їх вибіркові середні. Для надійності α=0.95 квантиль порядку $\frac{1+0.95}{2}=0.975$ стандартного нормального розподілу рівна 1.9599.

Далі для всіх 30 вибірок, використовуючи формулу довірчого інтервалу, одержуємо 30 лівих і 30 правих меж довірчого інтервалу. З графіку (рис. 8) видно, що всі 30 точок попали у межі довірчого інтервалу.

Рекомендації для викладачів щодо використання візуалізації в навчальному процесі.

1. Пояснення візуалізацій

• **Крок за кроком:** потрібно пояснювати кожен крок побудови графіка, щоб студенти розуміли, що відбувається на кожному етапі.

• **Зв'язок з теорією:** показувати, як візуалізація відображає теоретичні положення, на які спирається.

2. Інтерактивні завдання

• **Практичні заняття:** організовувати заняття, на яких студенти самі будуть будувати графіки та діаграми. Зручно використовувати інструменти типу Jupyter Notebook, щоб студенти могли змінювати параметри і спостерігати за результатами.

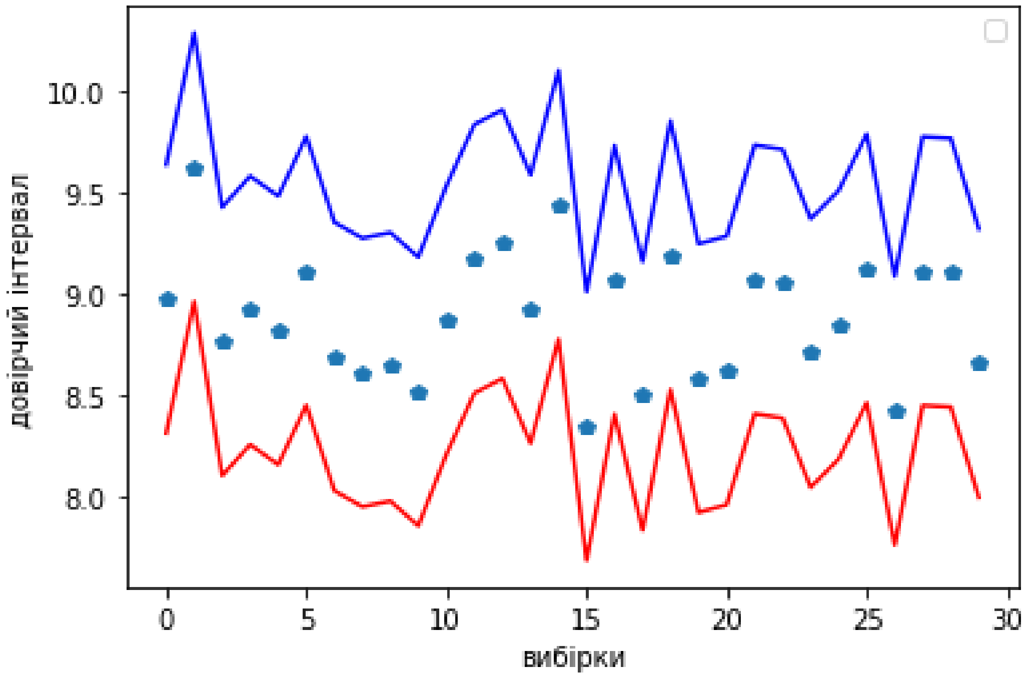

Рис. 8.

• **Проектні роботи:** пропонувати студентам створювати власні проекти, використовуючи візуалізацію для демонстрації теорем та понять.

3. Зворотний зв'язок

• **Опитування:** проводити опитування серед студентів щодо користі візуалізацій.

• **Аналіз результатів:** аналізувати результати опитувань та коригувати методи візуалізації на основі зворотного зв'язку.

Висновок. Візуалізація теорем теорії ймовірностей та математичної статистики за допомогою технічних засобів значно покращує засвоєння матеріалу студентами. Використання сучасних комп'ютерних програм та інтерактивних платформ дозволяє створювати наочні приклади, які допомагають студентам краще зрозуміти абстрактні математичні конструкції. Впровадження рекомендацій щодо візуалізації в навчальний процес сприятиме підвищенню якості навчання та глибшому розумінню студентами теорії ймовірностей та математичної статистики. Подальші перспективи у цьому напрямку включають розробку інтерактивних візуалізацій та використання більш складних моделей для вивчення статистичних явищ.

БІБЛІОГРАФІЧНИЙ СПИСОК:

1. Бен С. Python. Збірник вправ. Введення в мову Python із завданнями та рішеннями. – Харків, 2021. – 238 с.
2. Майборода Р.Є, Сугакова О.В. Аналіз даних за допомогою R.- Навчальний посібник. «Київський університет», 2015. – 65 с.
3. Мамчич Т.І., Оленко А.Я., Осипчук М.М., Шпортюк В.Г. (2006). Статистичний аналіз даних з пакетом Statistica. Навчально-методичний посібник. – Дрогобич: "Відродження", 2006. – 208 с.
4. Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering, 9(3), 90–95.