

**Міністерство освіти і науки України
Національний технічний університет України «КПІ» імені Ігоря
Сікорського
Кафедра обчислювальної техніки ФІОТ**

**ЗВІТ
з лабораторної роботи №1
з навчальної дисципліни «Аналіз даних для завдань електронної
комерції»**

Тема:
**ДОСЛІДЖЕННЯ СТАТИСТИЧНИХ ХАРАКТЕРИСТИК
ЕКСПЕРИМЕНТАЛЬНИХ ДАНИХ**

Виконав:
Куш Родіон

Перевірив:
Писарчук
Олексій
Олександрович

Київ 2021

I. Мета:

виявити дослідити та узагальнити особливості застосування методів статистичного аналізу для задач визначення статистичних характеристик вхідного потоку експериментальних даних з використанням спеціалізованих пакетів мови програмування Python.

II. Завдання:

Лабораторія провідної IT-компанії реалізує масштабний проєкт розробки універсальної платформи з обробки Big Data масиву експериментальних даних поточного спостереження для виявлення закономірностей і прогнозування розвитку контрольованого процесу. Платформа передбачає розташування back-end компоненти на власному хмарному сервері з наданням повноважень користувачам заздалегідь адаптованого front-end функціоналу універсальної платформи.

Замовниками ресурсів платформи є: державні та комерційні компанії валютного трейдингу для прогнозування динаміки зміни курсу валют та ціни інших товарів; метеорологічні служби для прогнозування параметрів метеоумов; департаменти охорони здоров'я для прогнозування зміни показників епідеміологічних ситуацій.

Вам, як Data Science Engineer поставлене наступне завдання.

I. Розробити універсальний скрипт мовою Python що реалізує:

1. Модель генерації випадкової величини – похибки вимірювання за заданим у таблиці Д1 додатку 1 закону розподілу;
2. Модель зміни досліджуваного процесу за заданим у таблиці Д1 додатку 1 закону;
3. Адитивну модель експериментальних даних (вимірів досліджуваного процесу) відповідно до синтезованих в п.1,2 моделей випадкової (стохастична) і не випадкової складових.

Параметри закону розподілу та зміни досліджуваного процесу обрати самостійно.

4. Метод Монте-Карло для дослідження статистичних характеристик експериментальних даних, сформованих у п.3;

5. Визначення статистичних (числових) характеристик експериментальних даних (дисперсію, середньоквадратичне відхилення математичне сподівання, гістограми закону розподілу похибки та експериментальних даних).

II. Провести дослідження зміни статистичних характеристик сформованих моделей при зміні стохастичних характеристик помилки експериментальних даних.

III. Довести адекватність сформованих моделей та працездатність розробленого скрипта.

Розроблений код повинен бути раціональним та відповідати вимогам до чистого коду.

Завдання II рівня складності – максимально 8 балів. Варіант 1.

Закон зміни похибки – рівномірний,
нормальний;
Закон зміни досліджуваного процесу – постійна,
квадратичний.

III. Результати виконання лабораторної роботи.

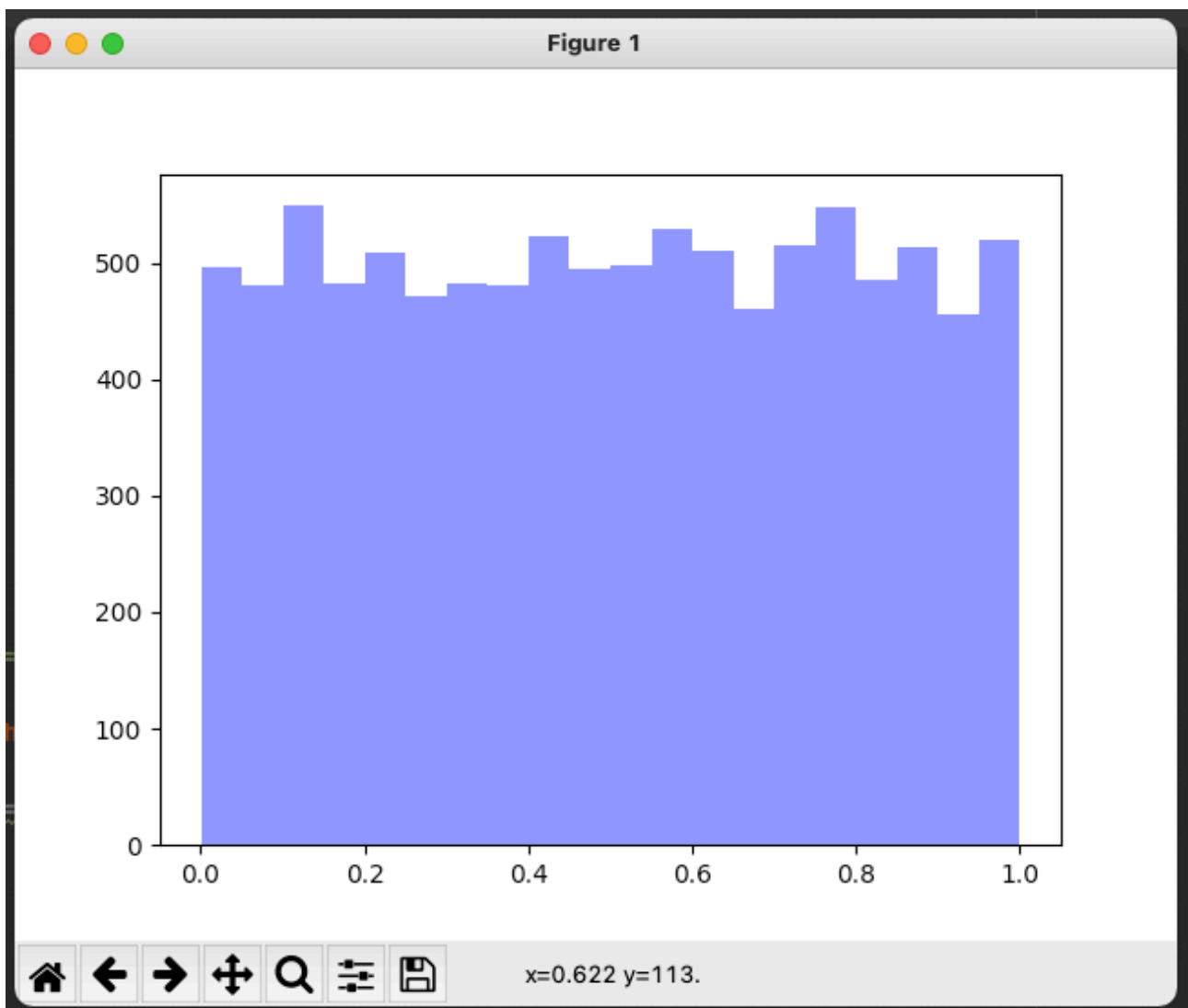
Для демонстрації результатів роботи програми була сгенерована вибірка з **10000** елементів.

Генерація похибок за рівномірним законом розподілу.

```
uniformS = np.random.rand(iter)

mUniformS = np.median(uniformS)
dUniformS = np.var(uniformS)
scvUniformS = mt.sqrt(dUniformS)

матриця реалізацій BB= [0.63070273 0.71163702 0.00743452 ... 0.28349118 0.22358192 0.26710315]
матиматичне сподівання BB= 0.4999006507146557
дисперсія BB = 0.08287381499153734
СКВ BB= 0.2878781252397225
=====
```



Генерація похибки за нормальним законом.

```
dm = 5
dsig = 5

normals = np.random.normal(dm, dsig, iter)

mNormals = np.median(normals)
dNormals = np.var(normals)
scvNormals = mt.sqrt(dNormals)
```

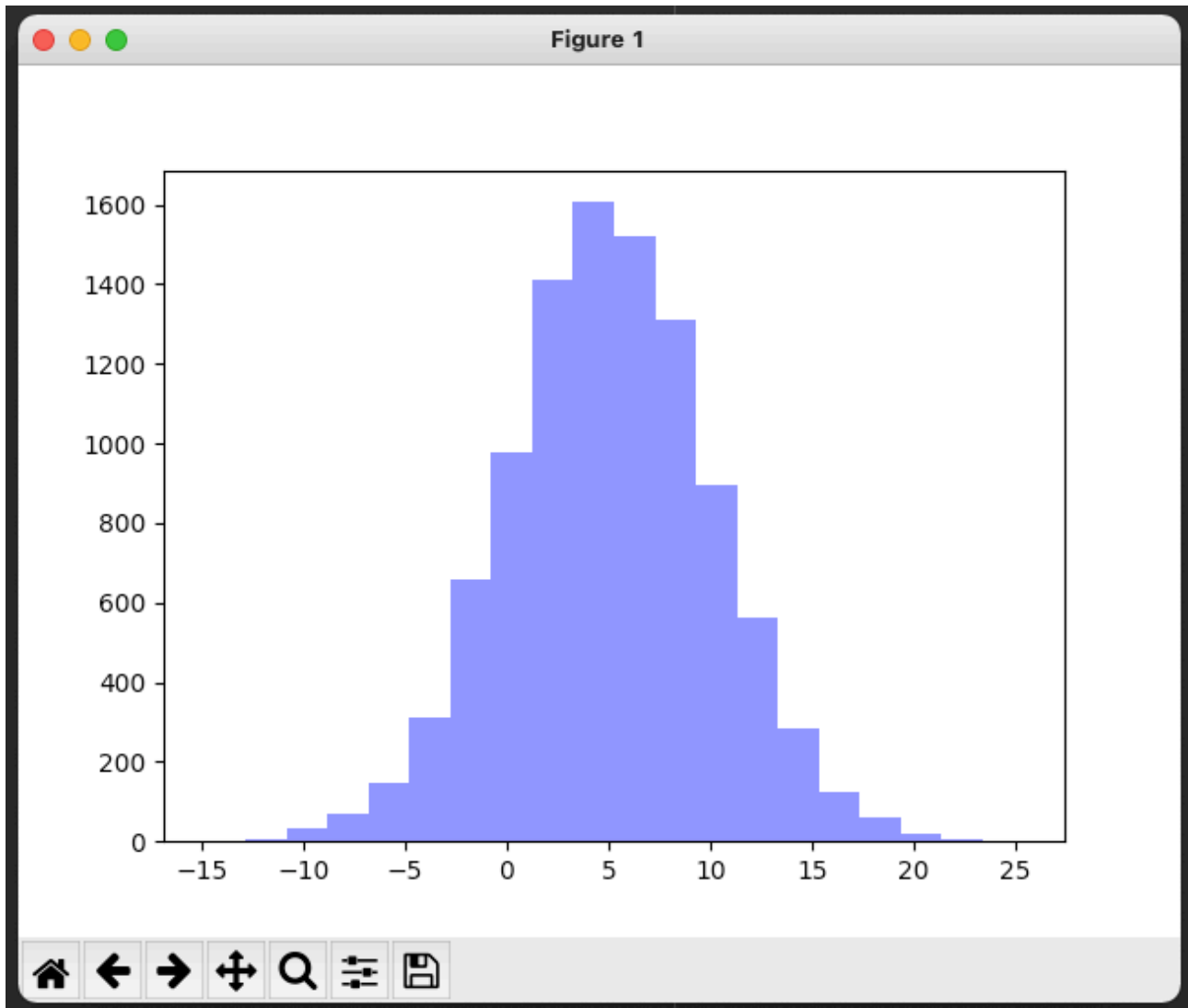
```
матриця реалізацій BB= [ 0.17312321  0.59238374  0.66714893 ...  7.30915073 11.82630609
 4.05104888]
```

```
матиматичне сподівання BB= 4.977098158647337
```

```
дисперсія BB = 25.340780846330368
```

```
СКВ BB= 5.033962737876629
```

```
=====
```



Генерація досліджуваного процесу за постійним законом зміни.

```
constantS = np.zeros(n)

for i in range(n):
    constantS[i] = 0.0000005

mConstantS = np.median(constantS)
dConstantS = np.var(constantS)
scvConstantS = mt.sqrt(dConstantS)
```

```
-----
матриця реалізацій BB= [5.e-07 5.e-07 5.e-07 ... 5.e-07 5.e-07 5.e-07]
матиматичне сподівання BB= 5e-07
дисперсія BB = 1.1210387714598537e-44
СКВ BB= 1.0587911840678754e-22
```

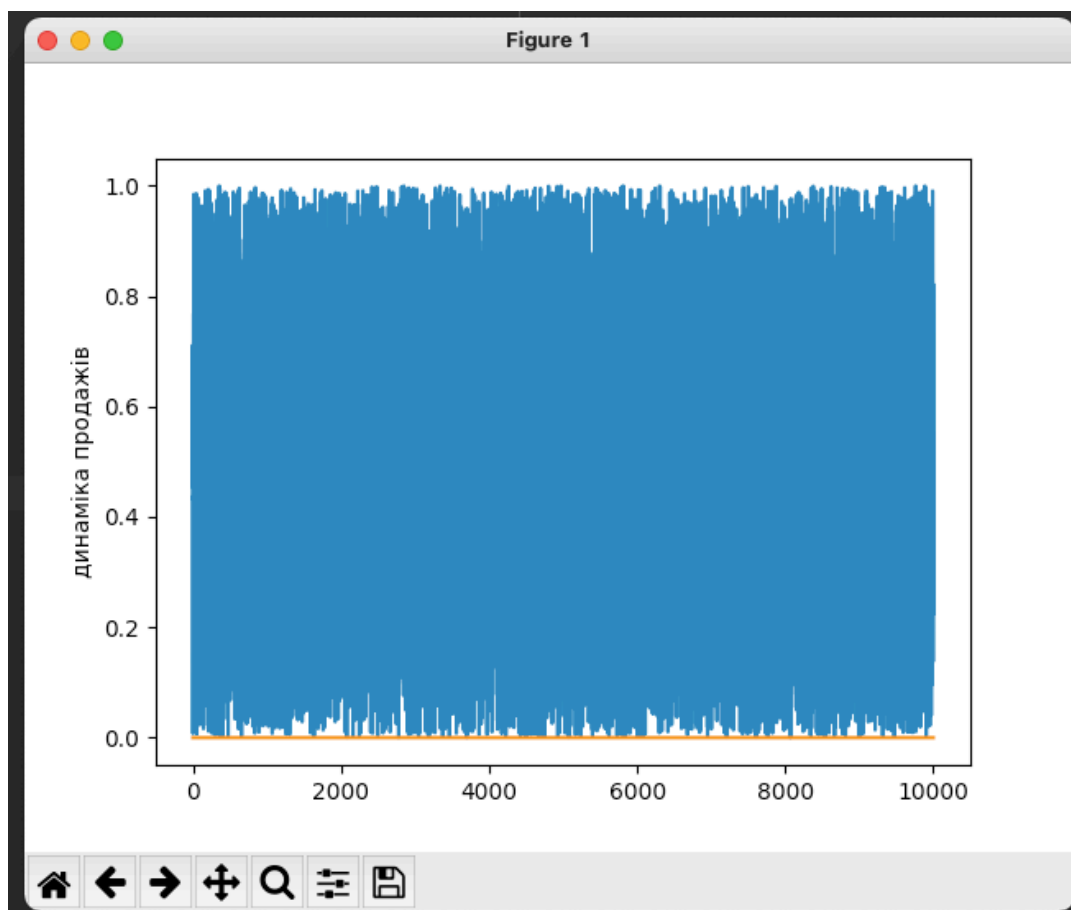
Генерація досліджуваного процесу за постійним законом зміни з урахуванням похибки, згенерованої за рівномірним законом.

```
constantS1 = np.zeros(n)

for i in range(n):
    constantS1[i] = constantS[i] + uniformS[i]

mConstantS1 = np.median(constantS1)
dConstantS1 = np.var(constantS1)
scvConstantS1 = mt.sqrt(dConstantS1)
```

```
матриця реалізацій BB= [0.44856444 0.5422404 0.28919231 ... 0.32419824 0.33807453 0.02173855]
матиматичне сподівання BB= 0.49553758741783166
дисперсія BB = 0.0839719497002532
СКВ BB= 0.2897791395187949
```



Генерація досліджуваного процесу за квадратичним законом зміни.

```
quadraticS = np.zeros(n)

for i in range(n):
    quadraticS[i] = (0.0000005 * i * i)

mQuadraticS = np.median(quadraticS)
dQuadraticS = np.var(quadraticS)
scvQuadraticS = mt.sqrt(dQuadraticS)

матриця реалізацій BB= [0.00000000e+00 5.00000000e-07 2.00000000e-06 ... 4.99700045e+01
 4.99800020e+01 4.99900005e+01]
матиматичне сподівання BB= 12.49750025
дисперсія BB = 222.18055486152772
СКВ BB= 14.90572221871613
```

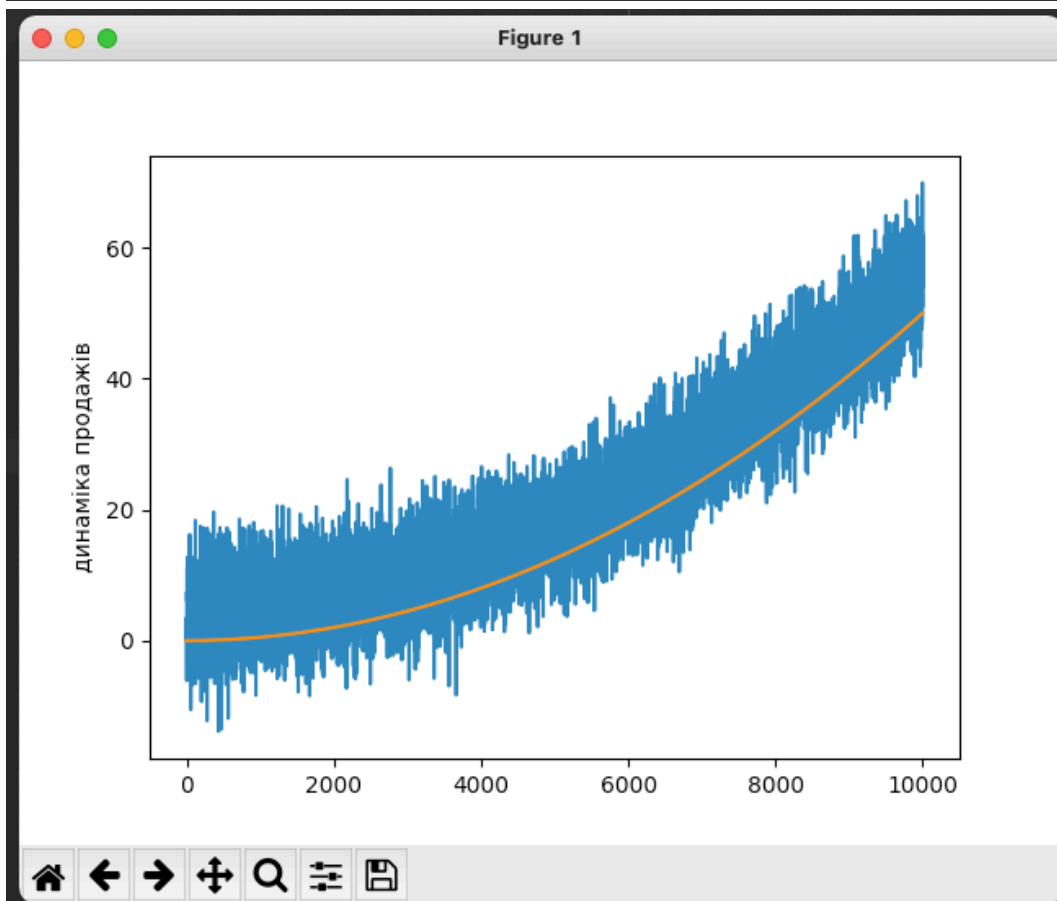
Генерація досліджуваного процесу за квадратичним законом зміни з урахуванням похибки, згенерованої за нормальним законом.

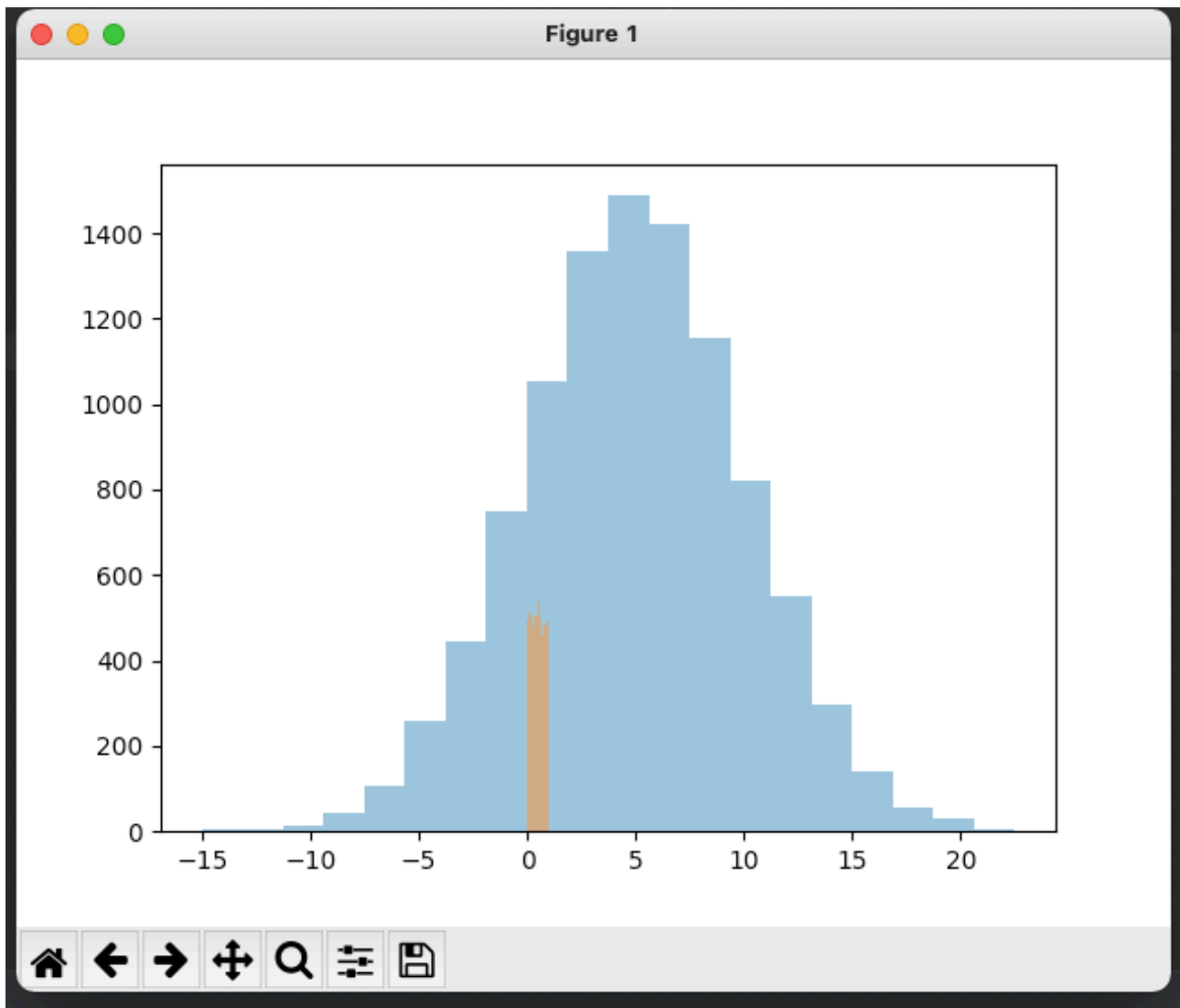
```
quadraticS1 = np.zeros(n)

for i in range(n):
    quadraticS1[i] = quadraticS[i] + normalS[i]

mQuadraticS1 = np.median(quadraticS1)
dQuadraticS1 = np.var(quadraticS1)
scvQuadraticS1 = mt.sqrt(dQuadraticS1)

матриця реалізацій BB= [ 0.52967845  9.55414362  5.70780348 ... 61.92260005 44.4037273
 54.0764013 ]
матиматичне сподівання BB= 18.06208793032779
дисперсія BB = 245.70824455223604
СКВ BB= 15.675083558062331
```





IV. Висновки.

В ході виконання лабораторної роботи були використані методи статистичного аналізу для оцінки та визначення характеристик вхідного потоку експериментальних даних. Для цього були створені дві випадкові величини, розподілені за постійним та квадратичним законами, на які були накладені випадкові величини похибок, сгенеровані за рівномірним та нормальним законом відповідно.

Виконав: