**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 «Вступ до технології Data Science»**

**Тема:**

**ДОСЛІДЖЕННЯ СТАТИСТИЧНИХ ХАРАКТЕРИСТИК   
ЕКСПЕРИМЕНТАЛЬНИХ ДАНИХ**

**Виконав:**

Студент X курсу кафедри ОТ ФІОТ,

Навчальної групи ІА-73

Петренко П.П.

**Перевірив:**

Професор кафедри ОТ ФІОТ

Писарчук О.О.

**Київ 2023**

**І. 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.

**ІІ. Завдання:**

Лабораторія провідної ІТ-компанії реалізує масштабний проект 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 трейдінгу 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](https://jobs.dou.ua/companies/gravitum/vacancies/147764/) поставлене наступне завдання.

**І. Розробити універсальний скрипт мовою Python що реалізує:**

Технічні умови реалізації завдання наведені у таблиці Д.1 додатку 1.

| **Варіант** (порядковий номер в списку групи) | **І рівень складності** | **ІІ рівень складності** |
| --- | --- | --- |
| 7, 22 | Закон зміни похибки – рівномірний;  Закон зміни досліджуваного процесу – квадратичний. | Закон зміни похибки – рівномірний, нормальний;  Закон зміни досліджуваного процесу – квадратичний, лінійний. |

1. Модель генерації випадкової величини – похибки вимірювання за заданим у таблиці Д1 додатку 1 закону розподілу

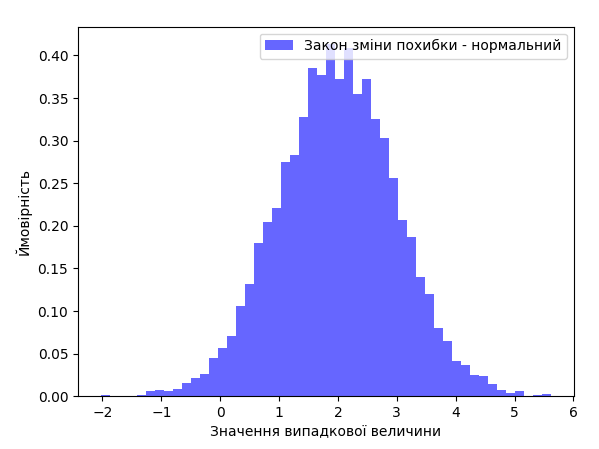


Рис. 1 Розподіл нормального закону похибки ( нормальний розподіл )

2. Модель зміни досліджуваного процесу за заданим у таблиці Д1 додатку 1 закону;

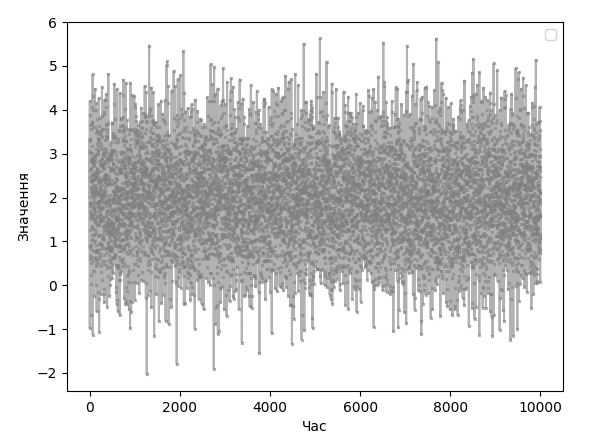


Рис.2 Графік значень моделі зміни досліджуваного процесу

3. Адитивну модель експериментальних даних (вимірів досліджуваного процесу) відповідно до синтезованих в п.1,2 моделей випадкової (стохастична) і невипадкової складових.

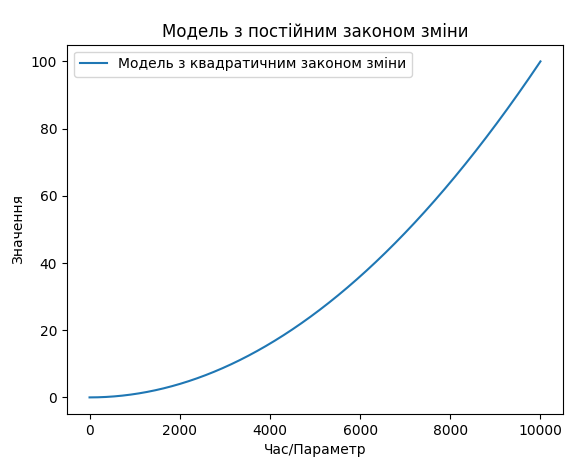


Рис.3 Модель квадратичного закону зміни

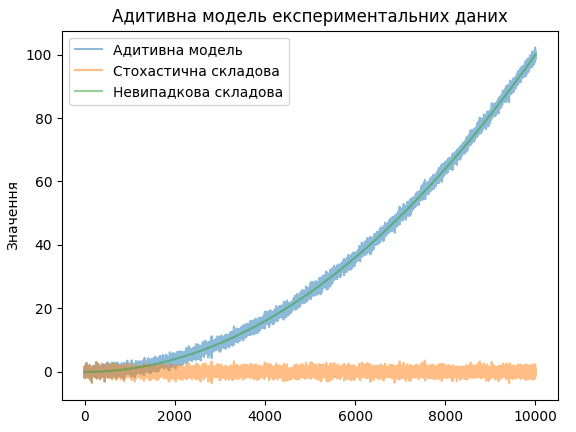


Рис.4 Адитивна модель

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

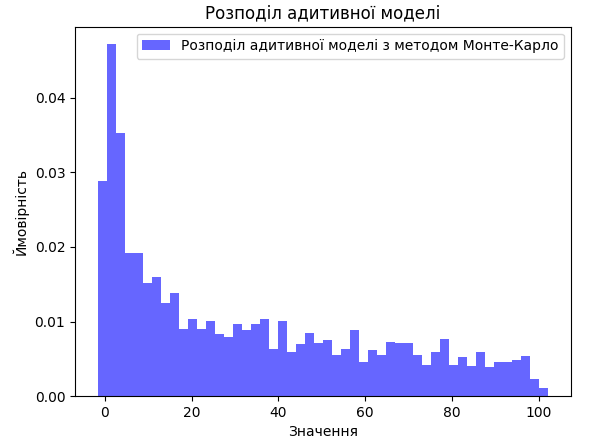
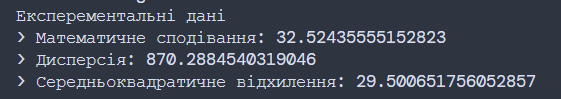


Рис.5 Розподіл адитивної моделі методом монте-карло



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

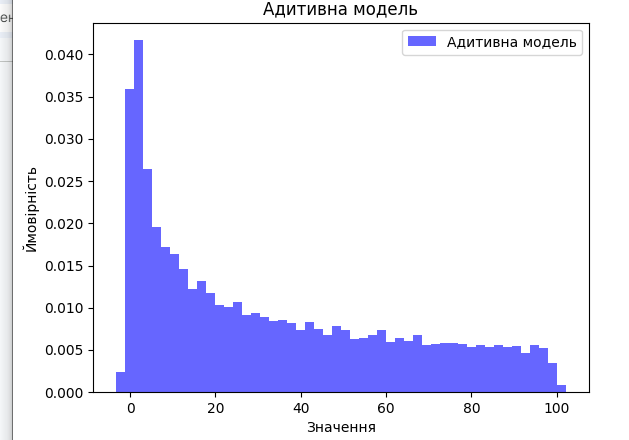


Рис. 6. Адитивна модель

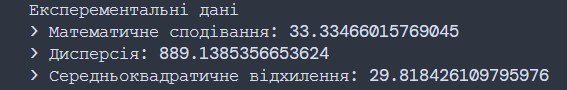


Рис. 7. Статистичні дані адитивної моделі

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

Метод Монте-Карло використовує випадкові величини для апроксимації математичного сподівання, і його значення може коливатися в залежності від конкретної реалізації випадкових подій.

Математичне сподівання:

Адитивна модель: 32.5244

Монте-карло: 33.3347

Дисперсія:

Адитивна модель: 870.2885

Монте-карло: 889.1385

Середньоквадратичне відхилення:

Адитивна модель: 29.5007

Зміни в значеннях могли виникнути через випадковий характер методу Монте-Карло та кількість ітерацій чи симуляцій, яку ви використовували. Також важливо врахувати, що результати методу Монте-Карло будуть точнішими при збільшенні кількості випробувань чи ітерацій.

**ІІІ. Довести адекватність сформованих моделей та працездатність розробленого скріпта.**

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

**import pandas as pd**

**import numpy as np**

**import datetime as dt**

**import matplotlib.pyplot as plt**

**import math as mt**

**import scipy as sci**

**def normal\_d(size=100000, mean=2, std\_dev=2):**

**""" Normal Distribution array generator """**

**mean = 0**

**std\_dev = 4**

**return np.random.normal(mean, std\_dev, size)**

**def histogram(data, bins=50, label="", xlabel="", ylabel="", title=""):**

**plt.hist(data, bins=bins, density=True, alpha=0.6, color='b', label=label)**

**plt.xlabel(xlabel)**

**plt.ylabel(ylabel)**

**plt.title(title)**

**plt.legend()**

**# Show the plot**

**plt.show()**

**def plot(data, label="", xlabel="", ylabel="", title=""):**

**plt.plot(data, color="gray", alpha=0.6, marker="x", markersize=2)**

**plt.xlabel(xlabel)**

**plt.ylabel(ylabel)**

**plt.title(title)**

**plt.legend()**

**plt.show()**

**def quadratic\_law(size=1000, show=False):**

**data = np.linspace(0, 10, size)\*\*2**

**if show:**

**plt.plot(data, label='Модель з квадратичним законом зміни')**

**plt.xlabel('Час/Параметр')**

**plt.ylabel('Значення')**

**plt.title('Модель з постійним законом зміни')**

**plt.legend()**

**plt.show()**

**return data**

**def additive\_model(data, show=False):**

**size = len(data)**

**stochastic\_component = normal\_d( size, 0, 1)**

**deterministic\_component = data**

**experimental\_data = stochastic\_component + deterministic\_component**

**if show:**

**plt.plot(experimental\_data, label='Адитивна модель', alpha=0.5)**

**plt.plot(stochastic\_component, label='Стохастична складова', alpha=0.5)**

**plt.plot(deterministic\_component, label='Невипадкова складова', alpha=0.5)**

**plt.ylabel('Значення')**

**plt.title('Адитивна модель експериментальних даних')**

**plt.legend()**

**plt.show()**

**return experimental\_data**

**def monte\_carlo\_method(data, num\_samples=10000):**

**""" Метод монте карло ( випадкова вибірка з даних ) """**

**return np.random.choice(data, size=num\_samples, replace=True)**

**def variance(data):**

**""" Дисперсія """**

**return np.var(data)**

**def mean(data):**

**""" Математичне сподівання """**

**return np.mean(data)**

**def deviation(data):**

**""" Відхилення """**

**return (data - np.mean(data))**

**def mean\_squared\_deviations(data):**

**""" Середньоквадратичне відхилення """**

**return np.sqrt(np.mean(deviation(data) \*\* 2))**

**def overall\_analysis(data, label="", xlabel="", ylabel="", title=""):**

**print(f" Експерементальні дані \n"**

**+ f" > Математичне сподівання: {mean(data)}\n"**

**+ f" > Дисперсія: {variance(data)}\n"**

**+ f" > Середньоквадратичне відхилення: {mean\_squared\_deviations(data)}\n"**

**)**

**histogram(data, label=label, xlabel=xlabel, ylabel=ylabel, title=title)**

**def normal\_d(size=1000, mean=0, std\_dev=1):**

**""" Генератор нормального розподілу """**

**return np.random.normal(loc=mean, scale=std\_dev, size=size)**

**if \_\_name\_\_ == '\_\_main\_\_':**

**size, lambda\_ = 10000, 2**

**normal\_distribution\_model = normal\_d(size, lambda\_) #normal\_d(size, mean=0, std\_dev=0.1) #**

**""" 1. Модель генерації випадкової величини – похибки вимірювання за заданим у таблиці Д1 додатку 1 закону розподілу """**

**histogram(normal\_distribution\_model, 50, 'Закон зміни похибки - нормальний', 'Значення випадкової величини', 'Ймовірність')**

**""" 2. Модель зміни досліджуваного процесу за заданим у таблиці Д1 додатку 1 закону """**

**plot(normal\_distribution\_model, label='Модель зміни досліджуваного процесу нормального розподілу', xlabel='Час', ylabel='Значення')**

**""" 3. Адитивну модель експериментальних даних (вимірів досліджуваного процесу) відповідно до синтезованих в п.1,2 моделей випадкової (стохастична) і невипадкової складових. """**

**#additive\_model\_ = additive\_model(linear\_law(size, show=True), show=True)**

**additive\_model\_ = additive\_model(quadratic\_law(size, show=True), show=True)**

**""" 4. Метод Монте-Карло для дослідження статистичних характеристик експериментальних даних"""**

**overall\_analysis(monte\_carlo\_method(additive\_model\_, len(additive\_model\_)//4), 'Розподіл адитивної моделі з методом Монте-Карло', 'Значення', 'Ймовірність', 'Розподіл адитивної моделі')**

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

**overall\_analysis(additive\_model\_, 'Адитивна модель', 'Значення', 'Ймовірність', 'Адитивна модель')**

**IV. Висновки.**

виявив дослідив та узагальнив особливості застосування методів статистичного аналізу для задач визначення статистичних характеристик вхідного потоку експериментальних даних з використанням спеціалізованих пакетів мови програмування Python.

Виконав: студент Петренко П.П.