МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ДЕРЖАВНИЙ УНІВЕРСИТЕТ «КИЇВСЬКИЙ АВІАЦІЙНИЙ ІНСТИТУТ»

Факультет комп’ютерних наук та технологiй

Кафедра прикладної математики

Лабораторна робота №2

З предмету

«Математична статистика»

Студент-виконавець:

Шапран Олег

ІІ курс, денна форма навчання

Перевірив:

Доцент Жултинська А.К.

Київ 2025

# ЗМІСТ

[ЗМІСТ 2](#_Toc197437932)

[1.Постановка задачі 3](#_Toc197437933)

[Мета: 3](#_Toc197437934)

[2.Теоретична частина 4](#_Toc197437935)

[1.Статистичні характеристики 4](#_Toc197437936)

[2. Довірчі інтервали 4](#_Toc197437937)

[3. Гістограма 4](#_Toc197437938)

[4. Емпірична функція розподілу 4](#_Toc197437939)

[5. Експоненціальний розподіл 5](#_Toc197437940)

[6. Аналіз типу дзвінків 5](#_Toc197437941)

[3.Опис програми 6](#_Toc197437942)

[Три основні модулі 6](#_Toc197437943)

[Основні об’єкти 6](#_Toc197437944)

[Схема взаємодії модулів 7](#_Toc197437945)

[Інтерфейс складається з чотирьох вкладок 7](#_Toc197437946)

[1. Основний аналіз 7](#_Toc197437947)

[2. Функції розподілу 8](#_Toc197437948)

[3. Експоненціальний розподіл 8](#_Toc197437949)

[4. Аналіз за типами 8](#_Toc197437950)

[Порядок роботи з програмою 8](#_Toc197437951)

[Формат вхідних даних 8](#_Toc197437952)

[Додаткові можливості 9](#_Toc197437953)

[4. Реалізація 10](#_Toc197437954)

[Вхідні дані 10](#_Toc197437955)

[Вихідні результати 10](#_Toc197437956)

[Коментарі та пояснення 14](#_Toc197437957)

[5.Діаграми Блок-схема 15](#_Toc197437958)

[6. Висновки 19](#_Toc197437959)

# 1.Постановка задачі

Написати програму, яка б дозволила користувачу провести аналіз статистичних даних, що передбачає реалізацію таких обчислювальних процедур:

1) відтворення розподілів (нормального, експоненціального, Вейбулла та рівномірного – обов’язково та ще одного – за варіантом), що включає:

* знаходження оцінок параметрів розподілів та оцінку їх точності;
* довірче оцінювання теоретичної функції розподілу;
* перевірку вірогідності відтворення на основі критеріїв згоди (Пірсона та Колмогорова).
* Провести тестування програмного забезпечення на реальних даних.

# 2.Теоретична частина

Нормальний розподіл

Нормальний (гаусівський) розподіл є одним із найпоширеніших у статистиці, придатним для даних із симетричною формою розподілу.

де:

* µ — математичне сподівання (середнє),
* σ — стандартне відхилення.

Експоненціальний розподіл

Експоненціальний розподіл моделює час між подіями у пуассонівському процесі, наприклад, час очікування дзвінка.

де:

* λ — параметр інтенсивності (λ > 0).

Розподіл Вейбулла

Розподіл Вейбулла використовується для моделювання часу до відмови або тривалості подій із змінною інтенсивністю.

де:

* k — параметр форми (k > 0),
* λ — параметр масштабу (λ > 0).

Рівномірний розподіл

Рівномірний розподіл застосовується, коли всі значення в інтервалі однаково ймовірні.

де:

* a — нижня межа,
* b — верхня межа.

Розподіл Релея

Розподіл Релея використовується для моделювання величин, таких як відстані або амплітуди, і є частковим випадком розподілу Вейбулла при k = 2.

де:

* σ — параметр масштабу

Для оцінки відповідності даних теоретичним розподілам застосовуються два тести.

Тест Колмогорова-Смірнова

Тест Колмогорова-Смірнова порівнює емпіричну функцію розподілу (EDF) з теоретичною (CDF).

де:

* Fn(x) — емпірична функція розподілу,
* F(x) — теоретична функція розподілу.

Тест Пірсона (Хі-квадрат)

Тест Пірсона оцінює відповідність гістограми даних теоретичним частотам.

де:

* Oi — спостережувана частота в i-му інтервалі,
* Ei — очікувана частота, Ei = n · (F(bi) − F(ai)),
* k — кількість інтервалів

# 3.Опис програми

## Три основні модулі

Програма складається з трьох основних модулів:

1. main.py:

* Точка входу програми.
* Ініціалізує головне вікно Tkinter, викликає функцію створення графічного інтерфейсу з gui.py та ініціалізує логіку обробки даних із logic.py.

2. gui.py:

* Модуль для створення графічного інтерфейсу.
* Містить функцію create\_gui, яка створює п’ять вкладок із елементами керування: кнопки, текстові поля, таблиці, поля введення (наприклад, для кількості класів гістограми), а також графічні полотна Matplotlib для відображення гістограм і графіків розподілів.
* На п’ятій вкладці («Гістограма та розподіли») додано прапорці для вибору розподілів (нормальний, експоненціальний, Вейбулла, рівномірний, Релея) і текстове поле для виведення результатів аналізу (оцінки параметрів, довірчі інтервали, тести відповідності).

3.logic.py:

* Містить логіку обробки даних, статистичного аналізу та побудови графіків.
* Основні функції:
* load\_data: Завантаження даних із файлів Excel або текстових файлів, обробка пропущених значень.
* update\_histogram: Побудова гістограми часу очікування з автоматичним вибором кількості класів.
* update\_characteristics: Обчислення та відображення статистичних характеристик (середнє, дисперсія, асиметрія тощо) із довірчими інтервалами
* plot\_distribution\_functions: Порівняння емпіричного та нормального розподілів із тестом Колмогорова-Смірнова.
* plot\_exponential\_distribution: Побудова імовірнісної сітки для експоненціального розподілу.
* analyze\_call\_types: Аналіз часу очікування за типами дзвінків із рекомендаціями.
* update\_distribution\_plot: Побудова гістограми з накладеними теоретичними розподілами (нормальний, експоненціальний, Вейбулла, рівномірний, Релея), оцінка параметрів, обчислення довірчих інтервалів і виконання тестів Колмогорова-Смірнова та Пірсона для перевірки відповідності.
* Функції редагування даних: standardize\_data, log\_transform, shift\_data, remove\_outliers, reset\_data для стандартизації, логарифмування, зсуву, видалення викидів і скидання даних.

## Схема взаємодії модулів

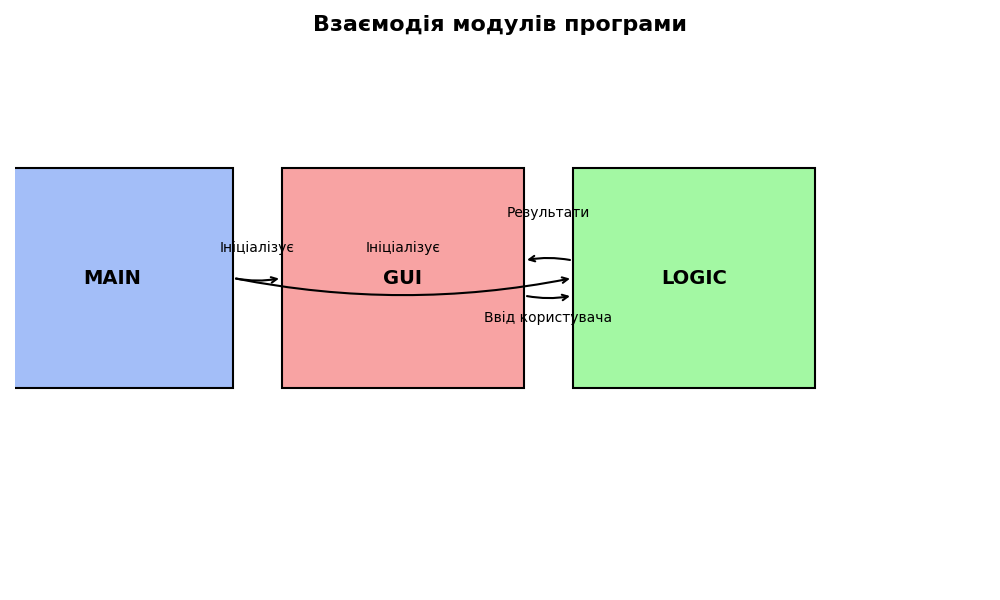


Рис.1

1. main.py викликає create\_gui із gui.py для створення інтерфейсу.
2. create\_gui повертає словник gui\_objects, який передається в initialize\_logic із logic.py.
3. initialize\_logic прив’язує функції обробки (load\_data, analyze\_call\_types тощо) до кнопок і подій.
4. При взаємодії користувача (наприклад, натискання «Завантажити дані») викликаються функції з logic.py, які оновлюють gui\_objects (таблиці, графіки, текстові поля).

Інтерфейс складається з п’яти вкладок:

### 1. Основний аналіз

* Кнопка «Завантажити дані».
* Поле для введення кількості класів гістограми.
* Таблиця характеристик (середнє, дисперсія тощо).
* Текстове поле для редагування даних.
* Кнопки редагування (стандартизація, логарифмування, зсув, видалення аномалій, скидання).
* Поле для границь даних і рівня довіри.
* Гістограма з кривою щільності.

### 2. Функції розподілу

* Графік емпіричного та нормального розподілів із довірчим інтервалом.
* Результати тесту Колмогорова-Смірнова.

### 3. Експоненціальний розподіл

* Імовірнісна сітка експоненціального розподілу.
* Результати тесту Колмогорова-Смірнова.

### 4. Аналіз за типами

* Таблиця характеристик для кожного типу дзвінка (кількість, середнє, медіана, стандартне відхилення).
* Гістограми часу очікування за типами.
* Рекомендації для оптимізації.

5.Гістограма та розподіли

* Гістограма
* Чек-бокси до 5 розподілі: Нормальний, Експоненціальний, Вейбула, Рівномірний і Релея.

## Порядок роботи з програмою

1. Запустити програму (main.py).
2. На вкладці «Основний аналіз» натиснути «Завантажити дані» та вибрати Excel-файл.
3. Ввести кількість класів для гістограми (або залишити 0 для автоматичного вибору).
4. Переглянути гістограму, характеристики та рекомендації.
5. За потреби відредагувати дані (стандартизація, видалення аномалій тощо).
6. Перейти до вкладки «Функції розподілу» для порівняння з нормальним розподілом.
7. На вкладці «Експоненціальний розподіл» перевірити відповідність експоненціальному розподілу.
8. На вкладці «Аналіз за типами» дослідити залежність часу очікування від типу дзвінків.
9. Зберегти відредаговані дані за допомогою кнопки «Зберегти дані».

## Формат вхідних даних

Excel-файл

* Стовпець «Час очікування (хв)»: числові значення (наприклад, 2.5, 7.8).
* Стовпець «Тип дзвінка»: текстові значення (наприклад, «Технічна підтримка», «Консультація», «Інше»).
* Пропуски в часі очікування замінюються середнім значенням.

Текстовий файл

* Числові значення, розділені пробілами або комами.
* Типи дзвінків не підтримуються (заповнюються як «Невідомо»).

## Додаткові можливості

* Інтерактивне видалення аномалій із вибором значень.
* Налаштування точності виведення (знаки після коми).
* Автоматичне масштабування графіків із 10% запасом.
* Обробка некоректних назв стовпців у Excel (пошук за ключовими словами «час», «тип»).
* Рекомендації для кожного типу дзвінка окремо.

# 4. Реалізація

## Вхідні дані

Для тестування використано різні типи данних

## Нормальний розподіл norm.txt

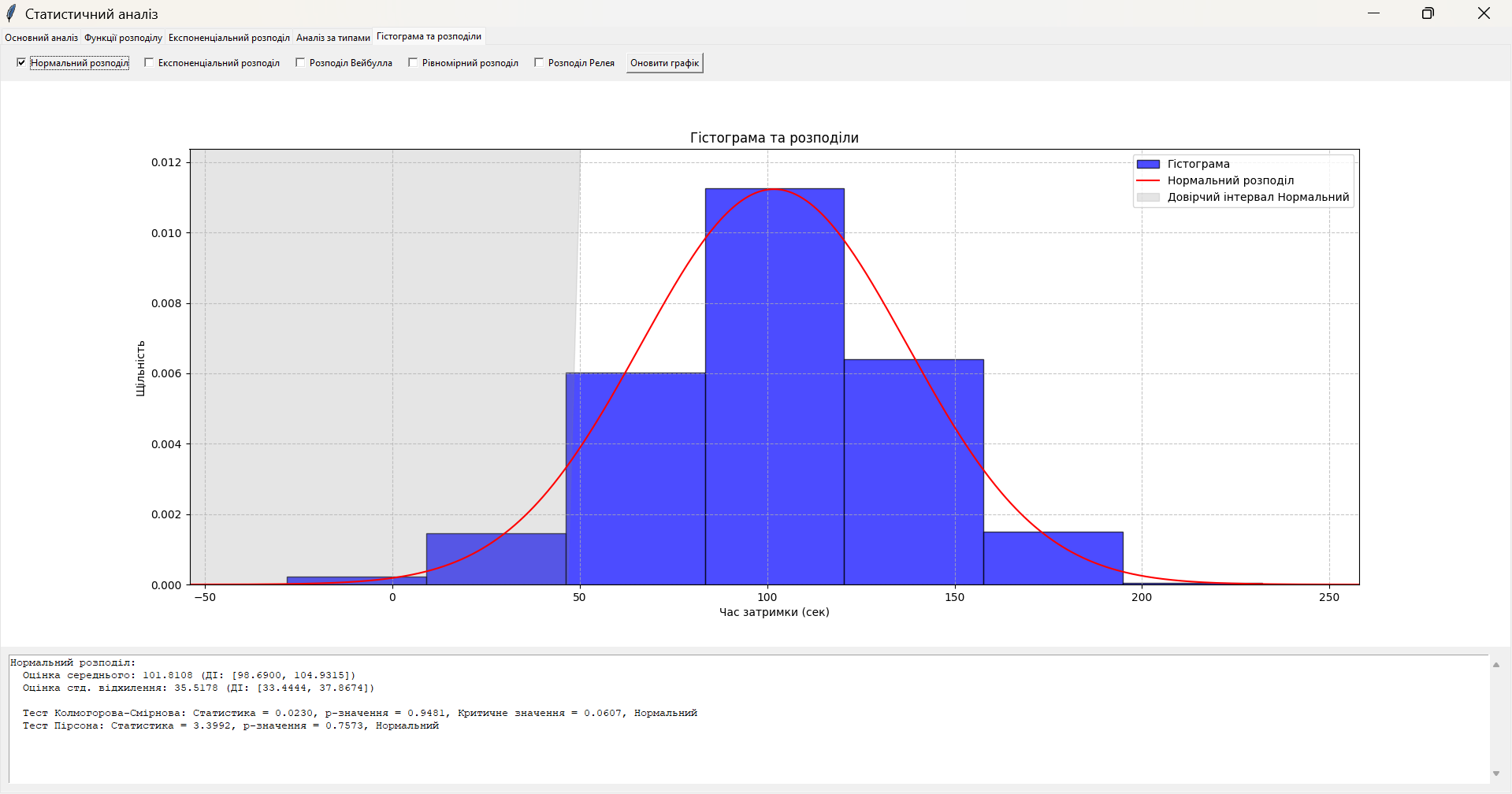


Рис.2

Експоненціальний розподіл exp.txt

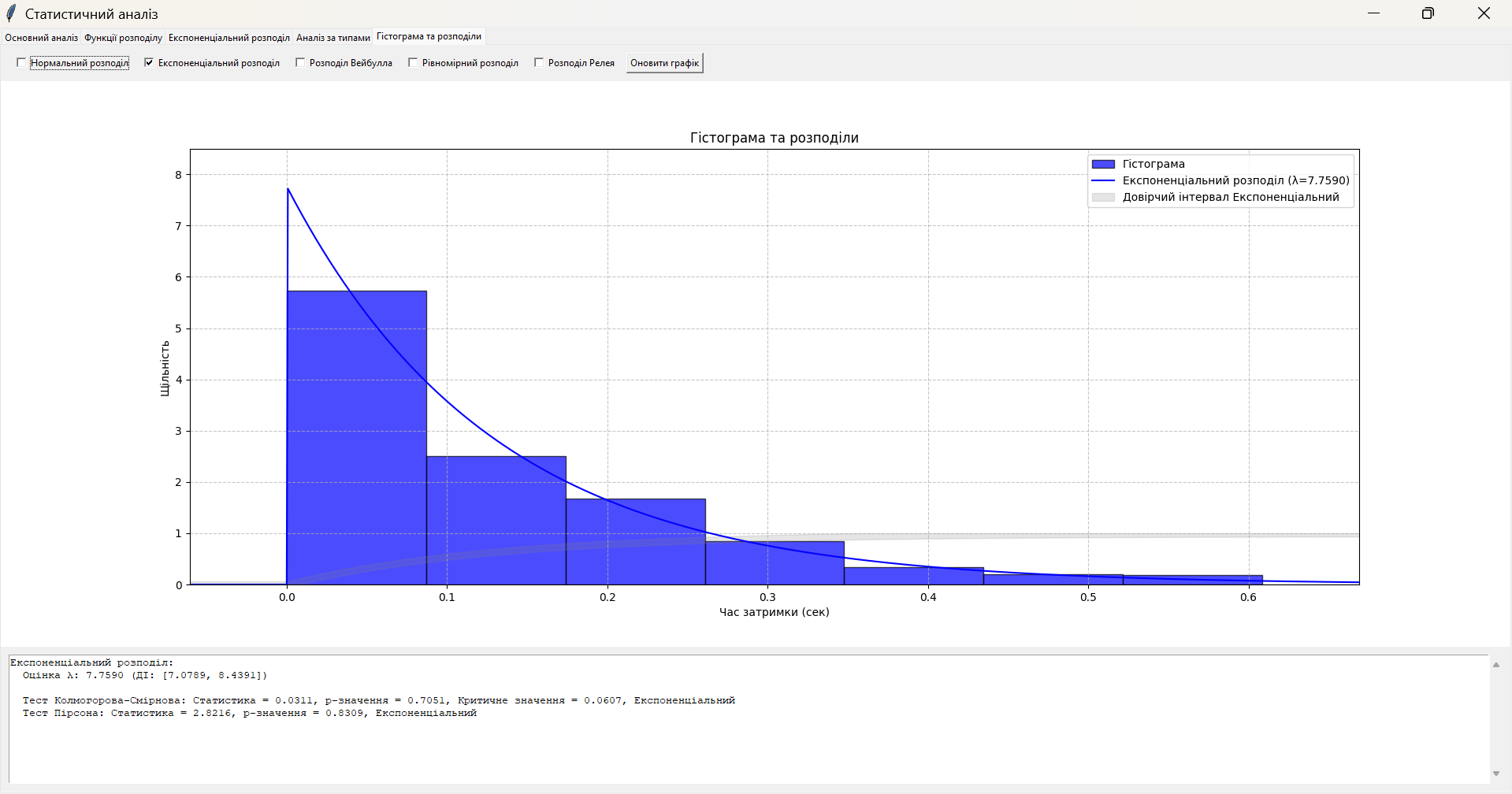


Рис.3

Розподіл Вейбулла veib.txt

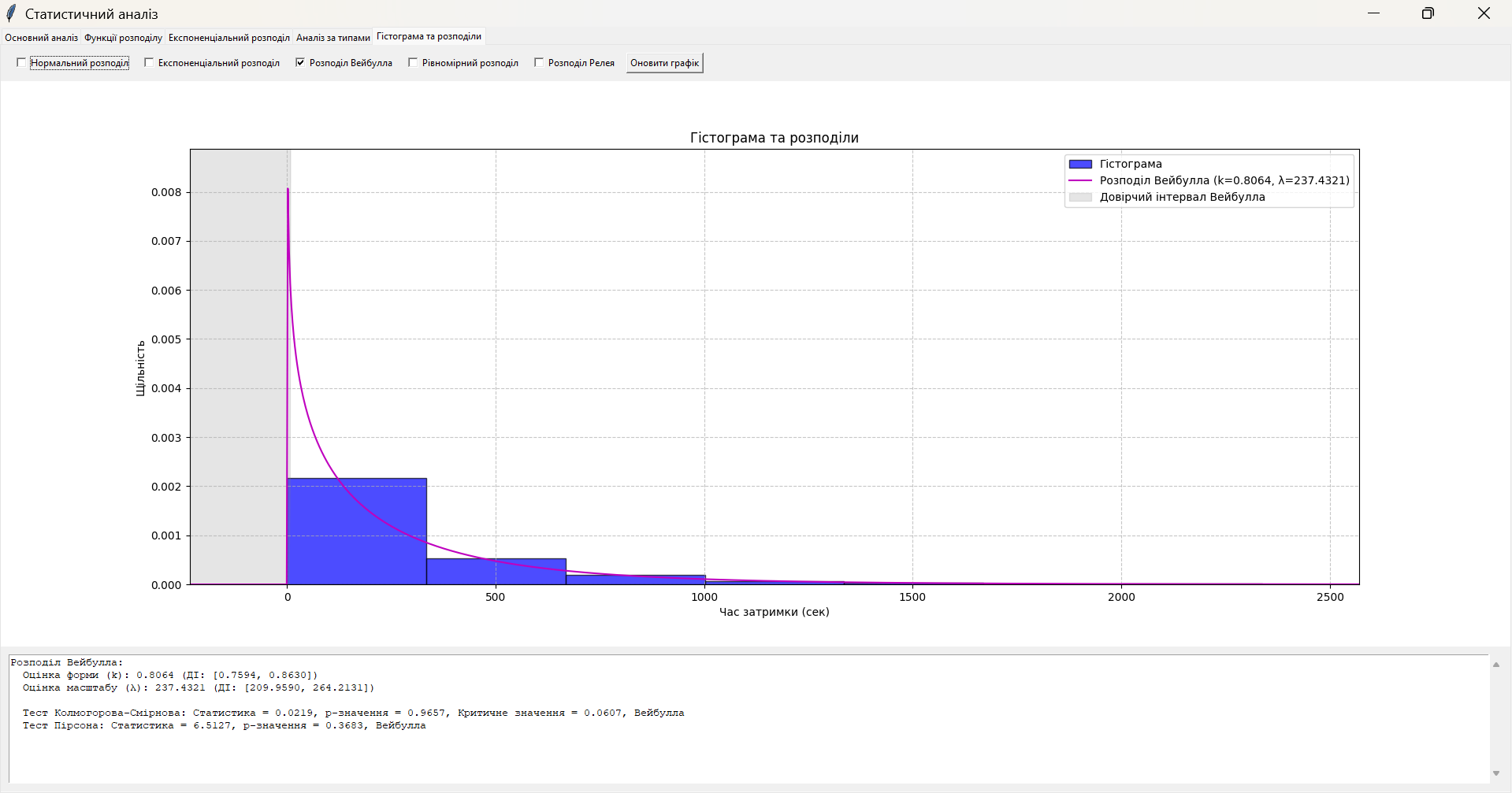


Рис.4

Рівномірний розподіл ravn.txt

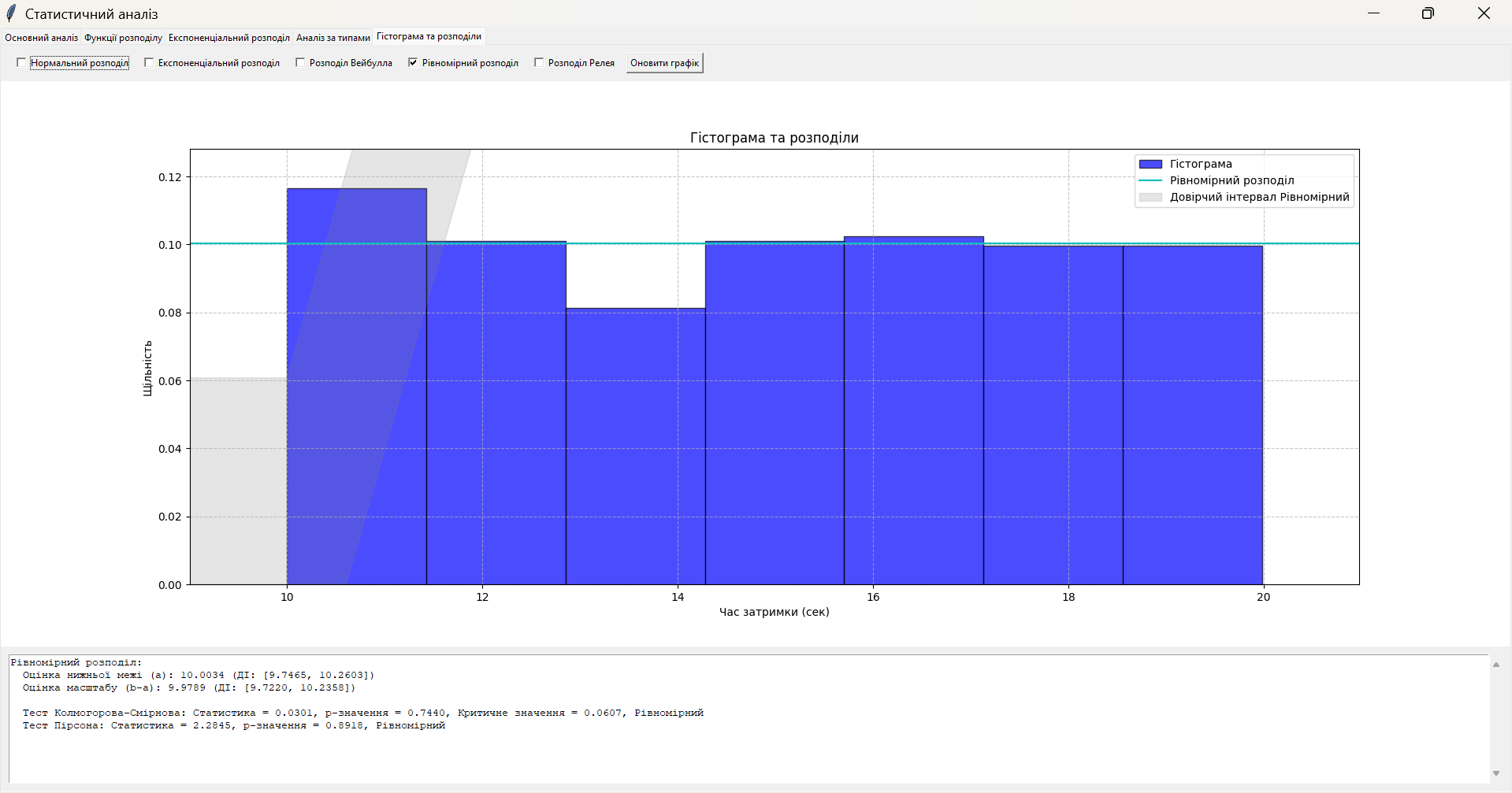


Рис.5

Розподіл Релея Rel.txt

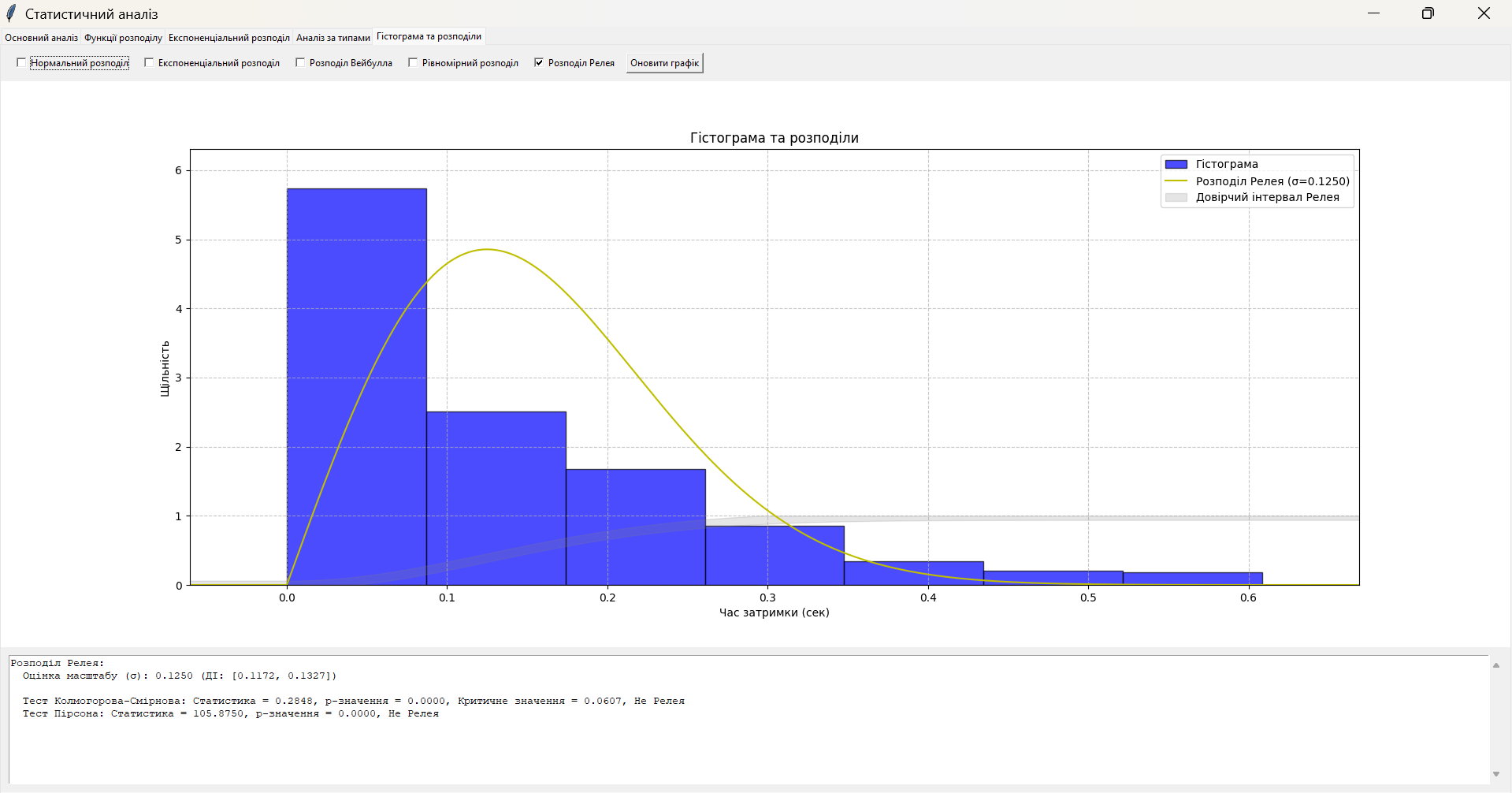


Рис.6

# 5.Діаграми Блок-схема

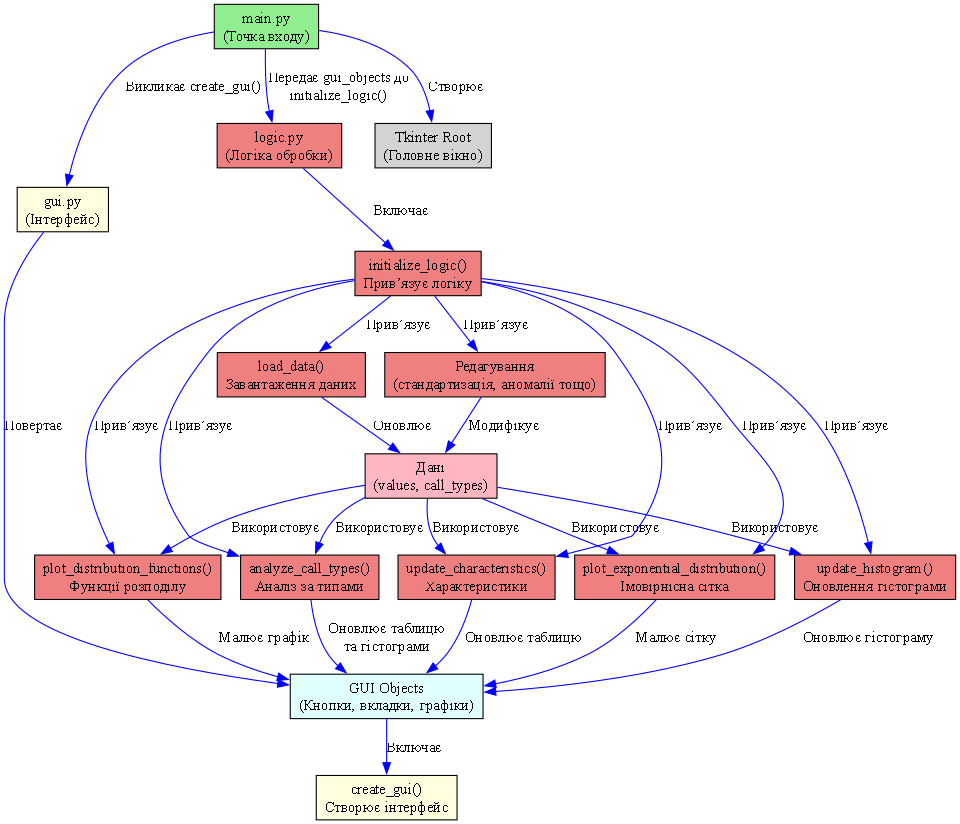
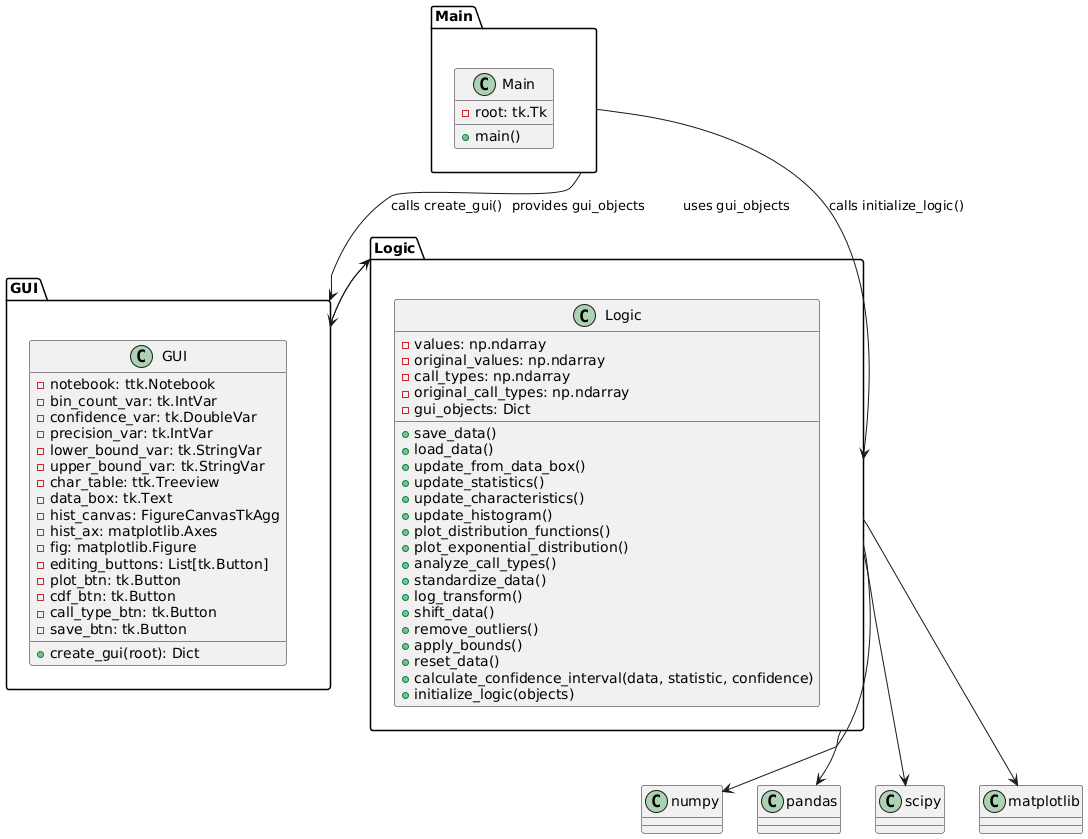
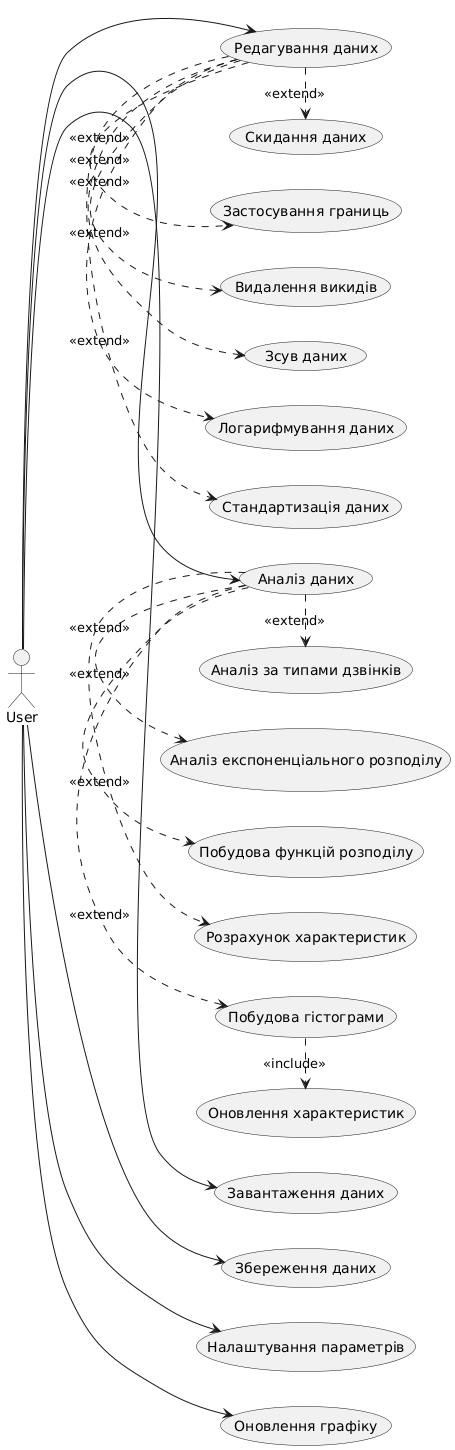


Рис.7

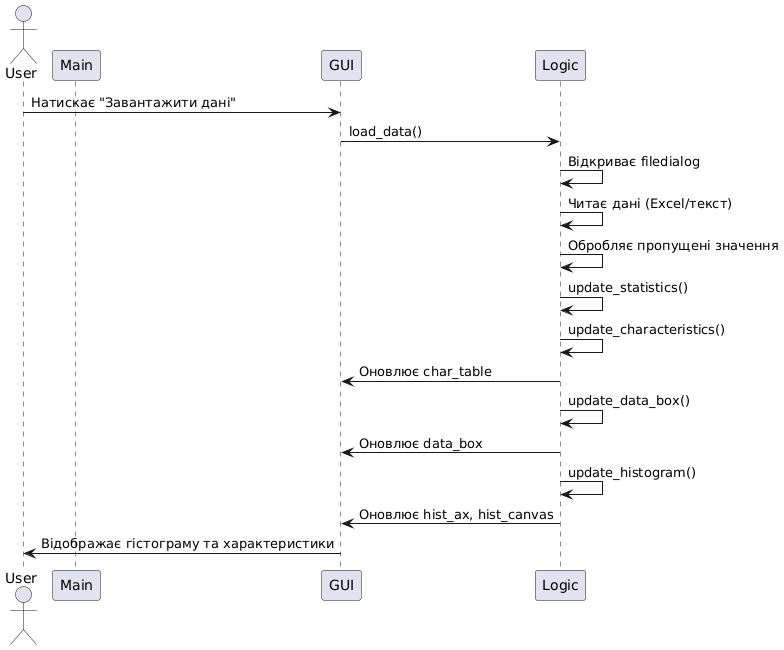
UML Діаграма класів

  
Рис.8

UML Діаграма варіантів використання

  
Рис.9

UML Діаграма послідовності

  
Рис.10

# 6. Висновки

Запроваджено п’ять розподілів:

* Нормальний: для симетричних даних, оцінюємо середнє (середнє вибірки) та стандартне відхилення (корінь із вибіркової дисперсії).
* Експоненціальний: для часу між подіями, параметр інтенсивності — обернена величина середнього.
* Вейбулла: для даних зі змінною інтенсивністю, параметри форми та масштабу оцінюються методом максимальної правдоподібності.
* Рівномірний: для однаково ймовірних значень, межі — мінімальне та максимальне значення вибірки.
* Релея: для невід’ємних величин, масштаб обчислюється як корінь із половини середнього квадрата даних.

Перевірки відповідності:

* Тест Колмогорова-Смірнова:
* Порівнює емпіричну та теоретичну функції розподілу, обчислюючи максимальну відстань між ними.
* Критичне значення залежить від розміру вибірки та рівня довіри.
* Додано довірчі смуги для теоретичних функцій розподілу.
* Результати: статистика тесту, p-значення, висновок (чи відповідає розподіл).
* Тест Пірсона (Хі-квадрат):
* Порівнює спостережувані частоти гістограми з очікуваними.
* Очікувані частоти коригуються, щоб бути не менше 5.
* Відповідність підтверджується, якщо p-значення > 0.05.