

Міністерство освіти і науки України
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «КИЇВО-МОГИЛЯНСЬКА АКАДЕМІЯ»
Факультет інформатики



Протокол до лабораторної роботи №1
З дисципліни «Математичні методи машинного навчання»

Виконав
студент 4 курсу
факультету інформатики
Миколайчик Я. А.

1. Завдання лабораторної роботи

- 1) Сформувати тестову вибірку зображень з вихідного пакету;
- 2) Для зеленого каналу кольору тестових зображень обчислити наступні характеристики розподілу значень яскравості пікселів:
 - a) Математичне очікування та дисперсію;
 - b) Медіану та інтерквартильний розмах;
 - c) Коефіцієнти асиметрії та ексцесу (нормалізований);
 - d) Гістограму значень яскравості пікселів (нормалізовану);
- 3) Провести обробку отриманих гістограм:
 - a) Провести апроксимацію гістограм з використанням імовірнісних розподілів:
 - i) Нормального (гаусового) розподілу;
 - ii) Розподілу Лапласа;
 - iii) Розподілу Стюдента;
 - iv) Бета розподілу;
 - b) Для кожної гістограми визначити найкращий тип апроксимуючого розподілу за критерієм мінімізації середньо квадратичного відхилення;
 - c) Побудувати розподіл отриманих апроксимацій за видом апроксимуючого розподілу;
- 4) Побудувати багатовимірні гаусові моделі:
 - a) Сформувати вектори параметрів розподілу значень яскравості пікселів тестових зображень;
 - i) Математичне очікування;
 - ii) Математичне очікування та дисперсія;
 - iii) Математичне очікування, дисперсія та коефіцієнт асиметрії;
 - iv) Математичне очікування, дисперсія, коефіцієнти асиметрії та ексцесу;
 - b) Визначити параметри багатовимірних гаусових моделей для кожної групи векторів, відобразити їх графічно та у вигляді таблиць.
- 5) Підготувати звіт за отриманими результатами лабораторної роботи.

2. Порядок виконання роботи та отримані результати

Робота була виконана на мові Python в форматі Jupyter Notebook та за допомогою середовища JupyterLab.

1) Формування тестової вибірки зображень з вихідного пакету.

Для формування вибірки було використано модуль random з заданим відповідним діапазоном значень. У якості зображення, що використовуватиметься як приклад, на якому буде демонструватись робота програми, було вибране перше зображення з масиву випадково створеної вибірки зображень.

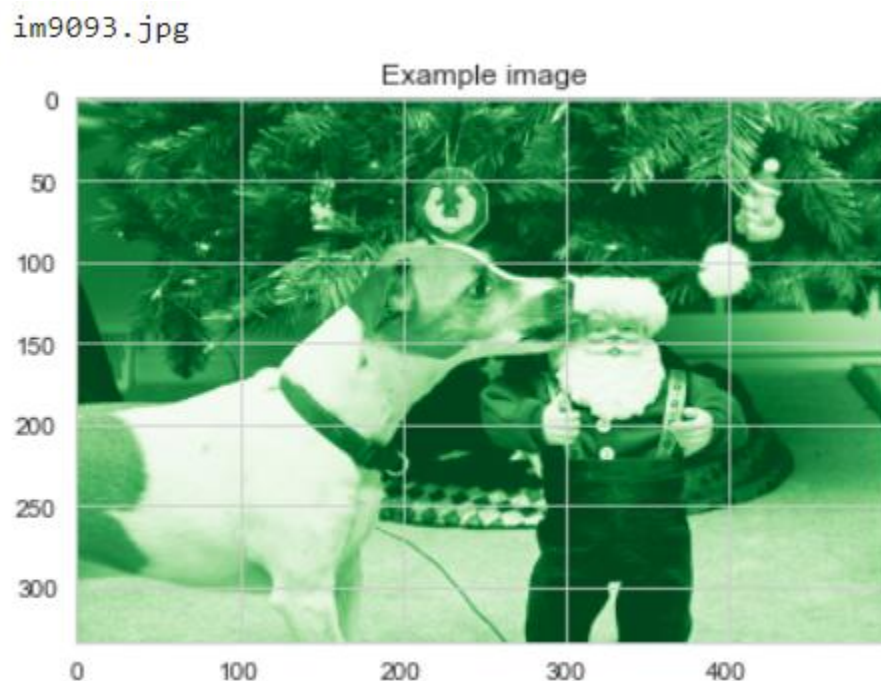


Рис. 1. Зображення для прикладу

2) Обчислення характеристик розподілу значень яскравості пікселів зеленого каналу кольору.

Для обчислення характеристик розподілу зеленого каналу конкретного зображення було формується масив значень яскравості усіх пікселів даного зображення. Для обчислення загальних характеристик розподілу зеленого каналу цілої вибірки було сформовано масив (сконкатенований з масивів окремих зображень) значень яскравості усіх пікселів усіх зображень цієї вибірки.

На основі цього отримали такі діаграми розподілів:

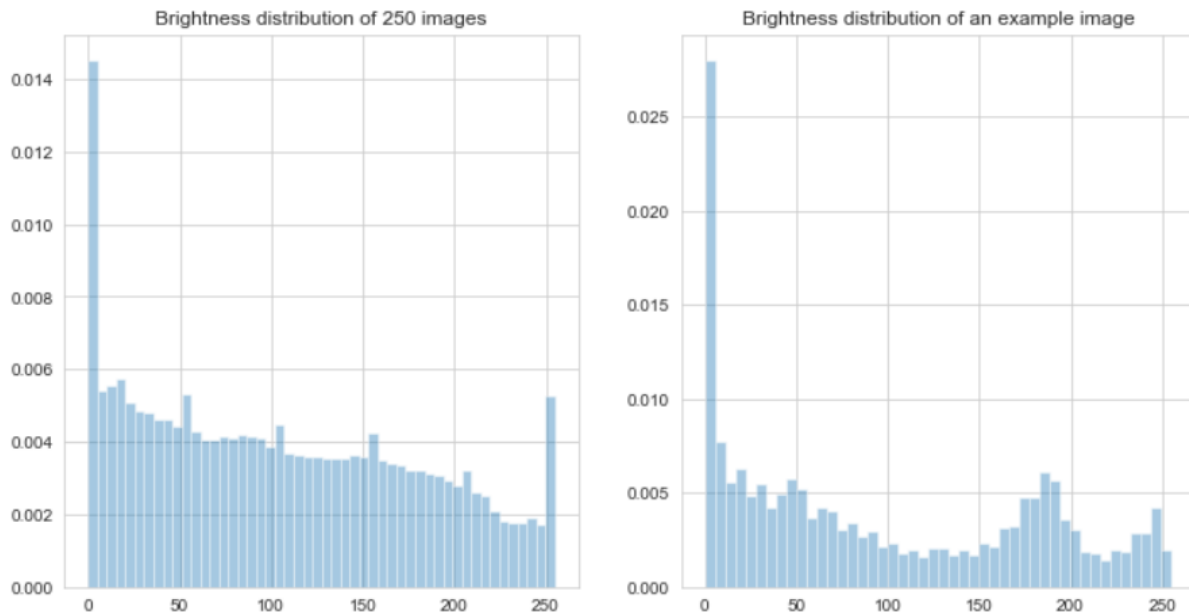


Рис. 2. Розподіли значень яскравості зеленого каналу пікселів.
Ліворуч - загальновибіркової, праворуч - зображення для прикладу

Числові характеристики загальновибіркового розподілу:

Математичне очікування	105.04
Дисперсія	5517.8
Медіана	97.0
Інтерквартильний розмах	126.0
Коефіцієнт ексцесу	-1.05
Коефіцієнт асиметрії	0.3

Пік розподілу є гладкішим за нормальний, оскільки маємо від'ємний коефіцієнт ексцесу. Також маємо великий показник дисперсії, що є закономірним, зважаючи на різноманіття змісту зображень вибірки.

Правий хвіст розподілу більший за лівий, що підтверджується додатнім коефіцієнтом асиметрії. Звідси також маємо те, що мат. очікування та медіана знаходяться ближче до мінімального значення яскравості, ніж до максимального. Тобто у більшості зображень вибірки переважають або темні кольори, або такі, що не включають у себе зелений.

Також бачимо піки біля значень 0 та 255, що означає, що чорний та білий кольори є більш поширеними за інші, що інтуїтивно є логічним.

3) Обробка та апроксимація отриманих гістограм.

Були проведені апроксимації отриманих гістограм зображень з використанням імовірнісних розподілів: нормального(гаусового), Лапласа, Стьюдента та бета.

Для зображення з прикладу отримані апроксимації мали такий вигляд:

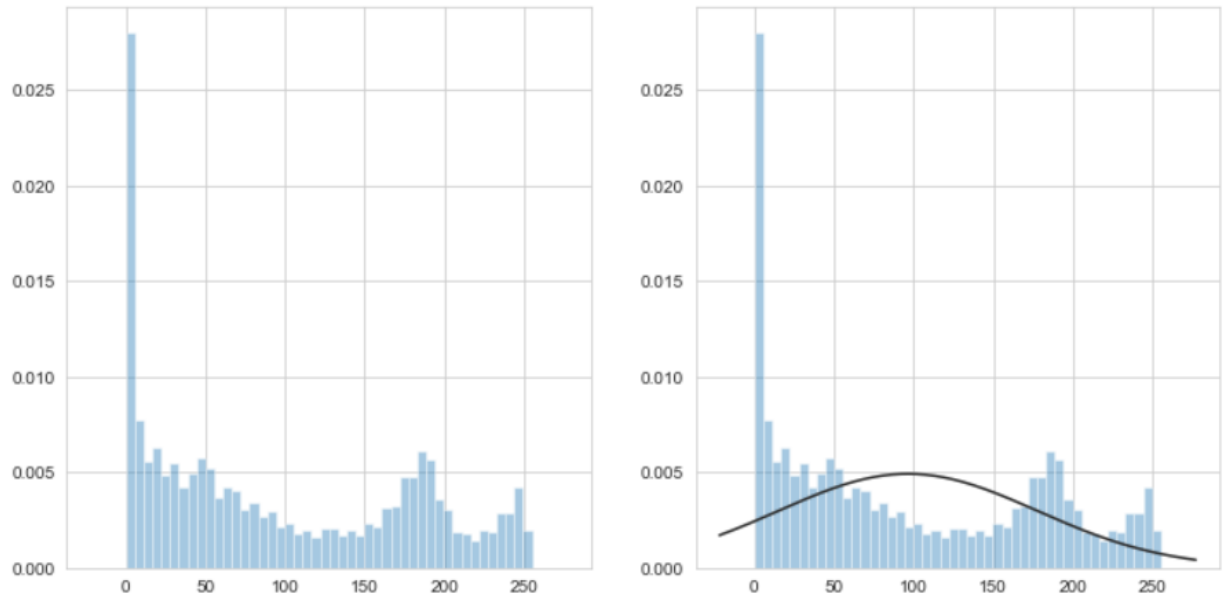


Рис. 3. Апроксимація нормальним розподілом

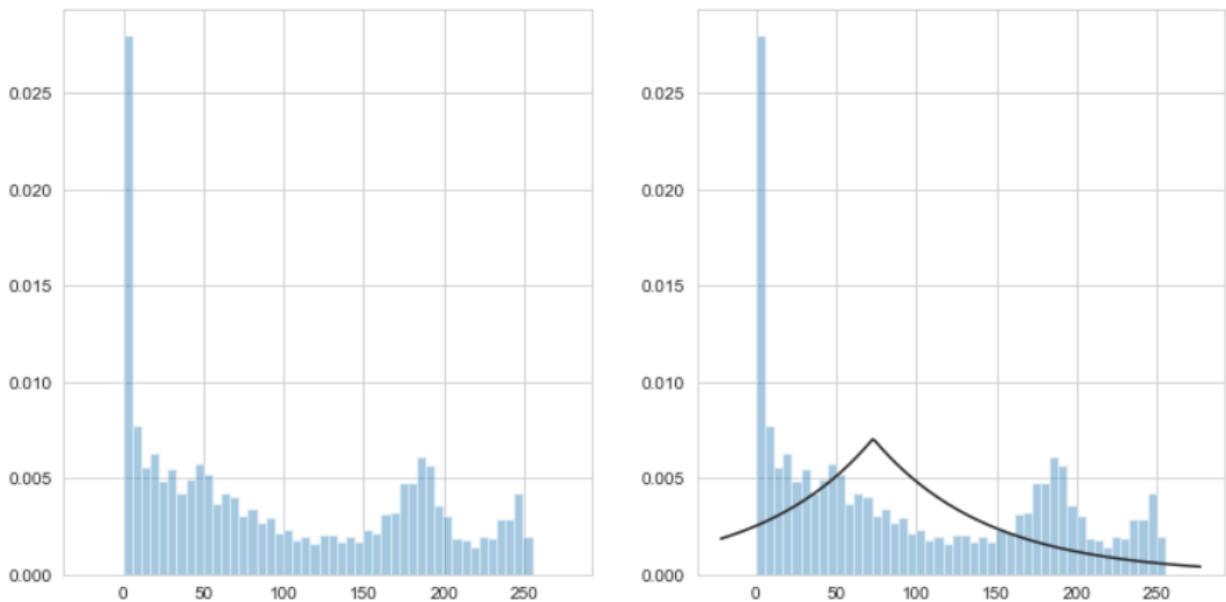


Рис. 4. Апроксимація розподілом Лапласа

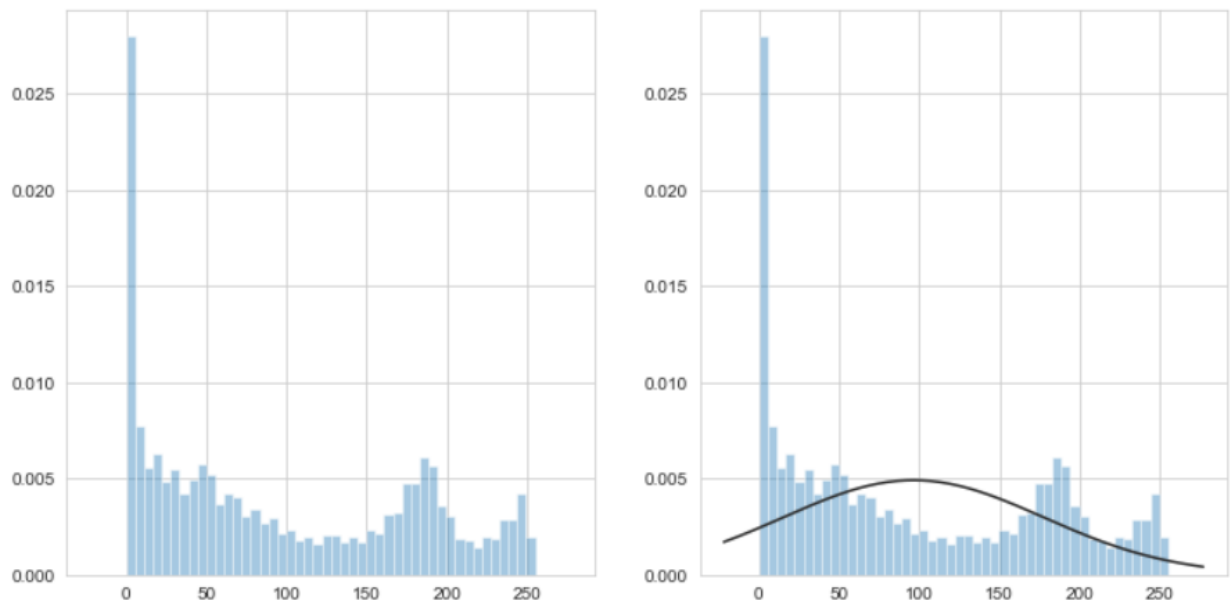


Рис. 5. Апроксимація розподілом Стюдента

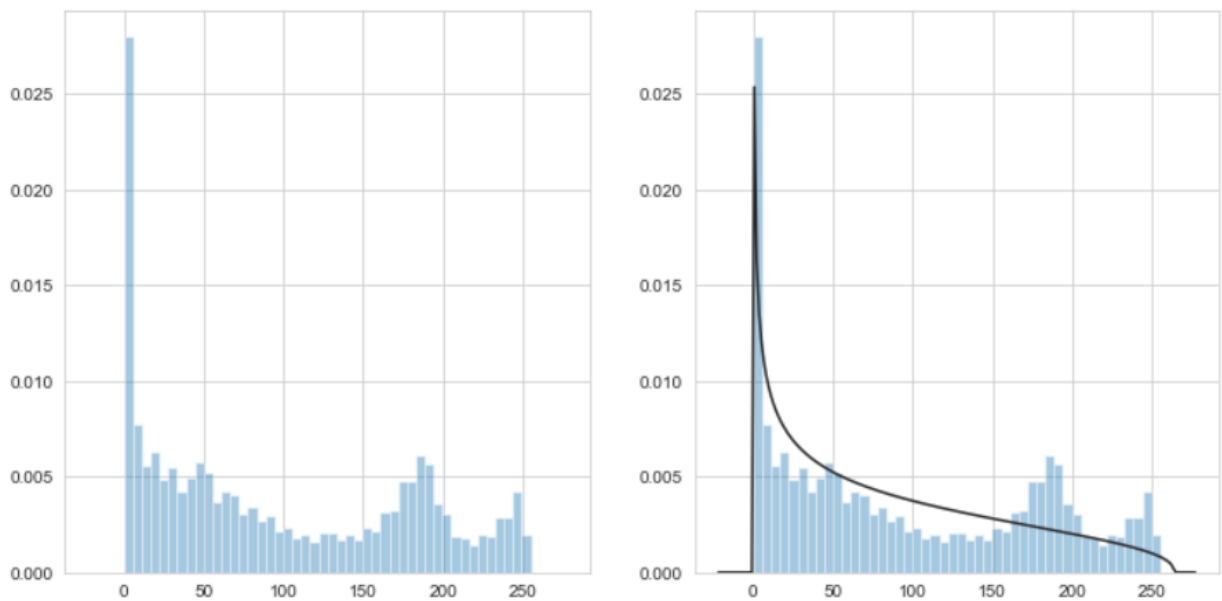


Рис. 6. Апроксимація бета-розподілом

Далі для кожного зображення вибірки були знайдені дані апроксимації і з них обрані розподіли, за якими апроксимація була найточніша за критерієм мінімізації середньоквадратичного відхилення. Нижче наведено приклад для трьох зображень з виведенням його характеристик та отриманого найкращого середньоквадратичного відхилення:

122
im13678.jpg
Math expectancy: 111.76171733333334
Variance: 3395.4458667040994
Median: 100.0
IQR: 74.0
Kurtosis: -0.08509335130554252
Skewness: 0.6700722648703121
Best approximation distribution type: beta
Squared error: 0.0006830018057152328

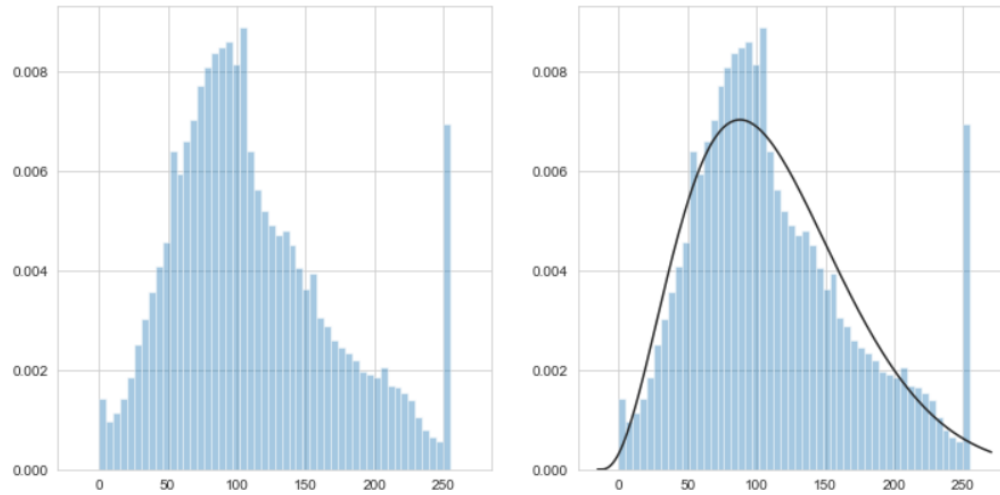


Рис. 7. Обробка даних зображення. Приклад 1

138
im20880.jpg
Math expectancy: 118.36754736842106
Variance: 4819.378835247756
Median: 118.0
IQR: 122.0
Kurtosis: -1.1403138279741718
Skewness: -0.08060536917416532
Best approximation distribution type: student
Squared error: 0.0012605499990109468

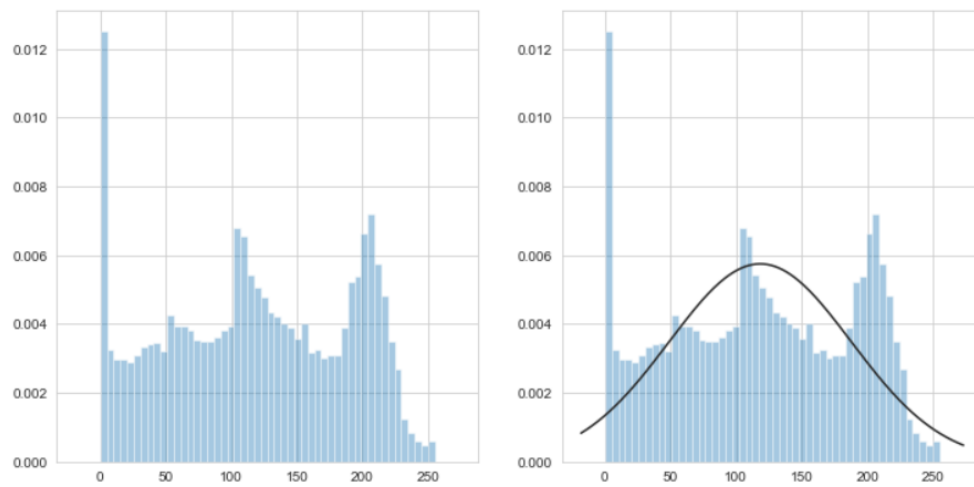


Рис. 8. Обробка даних зображення. Приклад 2

The figure consists of two side-by-side histograms. Both histograms have an x-axis ranging from 0 to 250 and a y-axis ranging from 0.0000 to 0.0200. The left histogram shows a bimodal distribution with a primary peak at approximately 25 nodes (height ~0.0195) and a secondary peak at approximately 130 nodes (height ~0.0075). The right histogram shows a similar bimodal distribution but includes a smooth black curve overlaid, which represents a theoretical model fit. The curve starts at 0, rises to a peak of about 0.007 at x=25, and then gradually declines, following the general shape of the distribution.

Після опрацювання всіх 250 зображень вибірки маємо наступний розподіл отриманих найкращих апроксимацій за видом апроксимуючого розподілу:

Рис. 10. Розподіл найкращих апроксимацій



Бачимо, що найкраще з апроксимацією зображень цієї вибірки впорались бета-розподіл та розподіл Лапласа, трохи гірше – Стюдента, і найгірше – нормальний розподіл.

4) Багатовимірні гаусові моделі

Сформувавши вектори параметрів розподілу значень яскравості пікселів для тестових зображень, ми сформували наступні гаусові моделі:

а) Математичне очікування

Вектор середнього	104.88
Матриця коваріації	1689.84

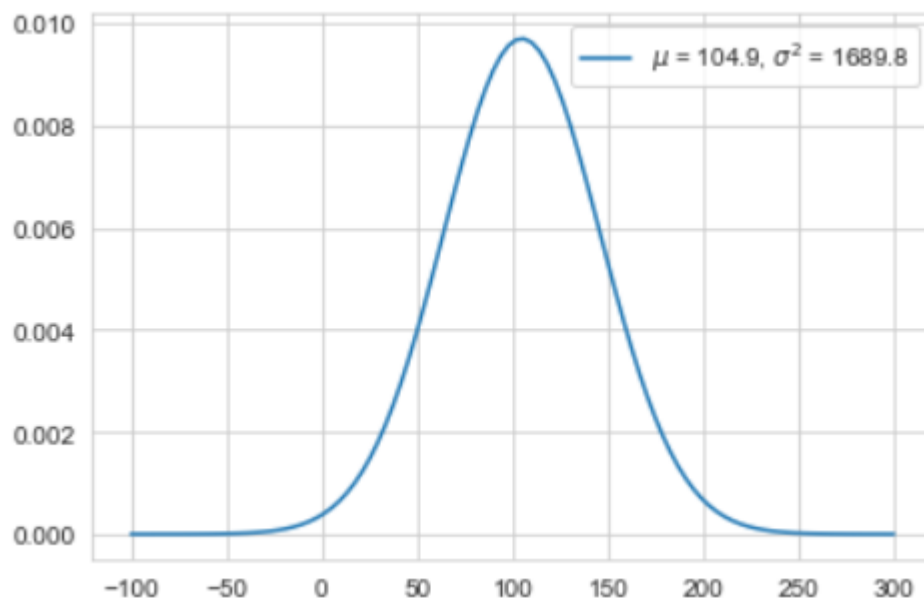


Рис. 12. Графік гаусової моделі за вектором [мат. очікування]

б) Математичне очікування та дисперсія

Вектор середнього	(104.88, 3850.91)
Матриця коваріації	[[1689.84, 10215.79] [10215.79, 3990494.03]]

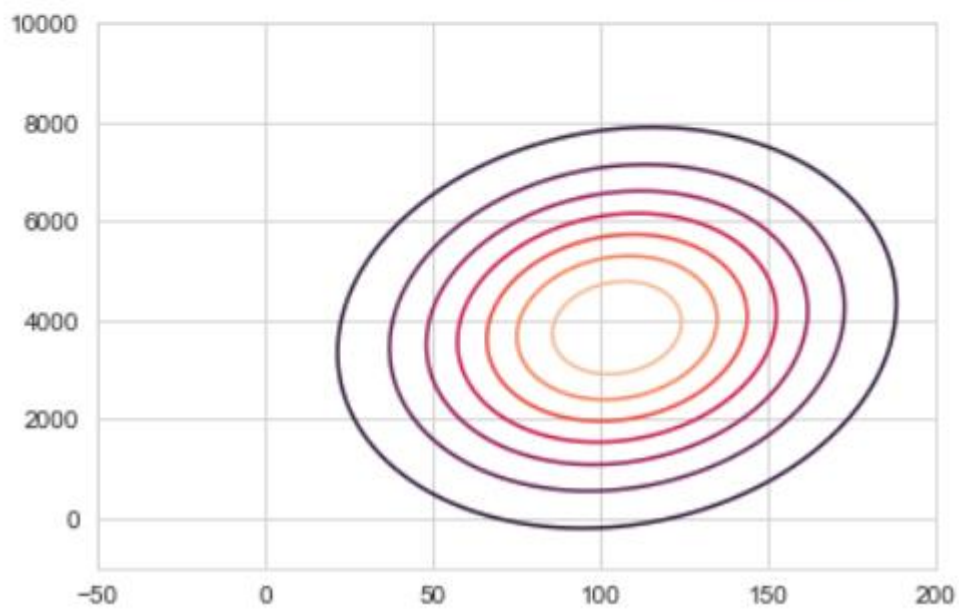
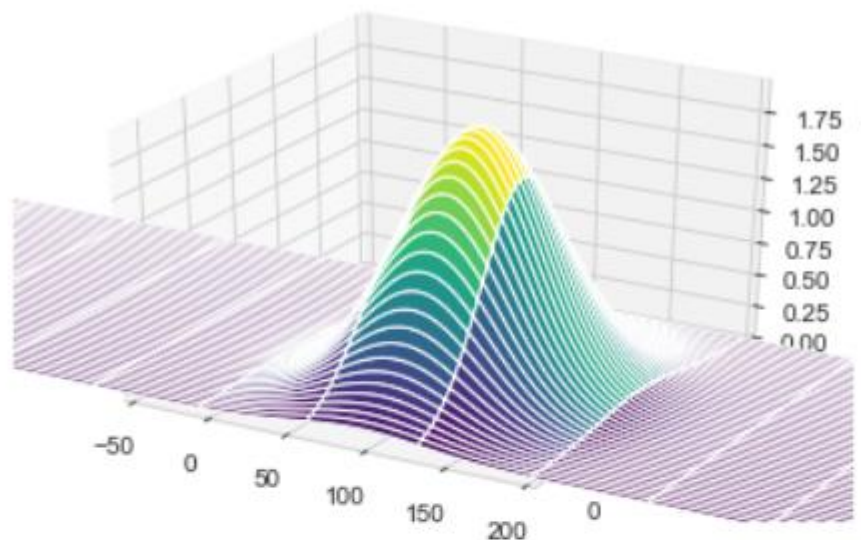


Рис. 13. Графік гаусової моделі за вектором [мат. очікування, дисперсія]

с) Математичне очікування, дисперсія та коефіцієнт асиметрії

Вектор середнього	(104.88, 3850.91, 0.88)
Матриця коваріації	[[1689.84, 10215.79, -67.6] [10215.79, 3990494.03, -3961.88] [-67.6, -3961.88, 34.68]]

d) Математичне очікування, дисперсія, коефіцієнти асиметрії та ексцесу

Вектор середнього	(104.88, 3850.91, 0.88, 0.35)
Матриця коваріації	[[1689.84, 10215.79, -67.6, -40.09] [10215.79, 3990494.03, -3961.88, -189.1] [-67.6, -3961.88, 34.68, 4.2] [-40.09, -189.1, 4.2, 1.43]]

3. Висновки

В ході лабораторної роботи ми проаналізували розподіли яскравості зеленого каналу пікселів випадково згенерованої вибірки зображень з пакету.

Сформувавши вибірку та обчисливши числові характеристики розподілів яскравості відповідних картинок, ми побачили, що розподіл їхніх значень істотно відрізняється поміж картинками та в загальному розподіл яскравості має велику дисперсію, всі значення є поширеними і немає якогось домінуючого діапазону значень (окрім піків, що свідчать про перевагу чорного та білого кольорів), хоча і є додатна асиметрія, яка свідчить про незначну перевагу темних відтінків або ненасичених зеленим кольорів у зображеннях.

Апроксимація зображень вибірки заданими розподілами і мінімізація їх похибок за критерієм середньоквадратичного відхилення показала, які розподіли краще апроксимують отримані гістограми, а саме бета-розподіл та розподіл Лапласа. Найгірше ж зображення отриманої вибірки апроксимував нормальний розподіл.

Також по векторах параметрів розподілів яскравості зеленого каналу ми побудували багатовимірні гаусові моделі та визначили їхні вектори середнього та матриці коваріації.