

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 3 курсу  
факультету інформатики  
Добровольський І. С.

## 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 Groovy, з використанням бібліотеки *Apache Commons Math*. Бібліотеку було використано для генератора випадкових чисел та побудови апроксимуючих розподілів.

- 1) Для формування вибірки було використано бібліотеку *Apache Commons Math*.

2) Для кожного зображення з вибірки було створено масив зі значеннями яскравості зеленого каналу кожного пікселя.

Потім для кожного масиву було обраховано характеристики розподілу та нормалізовану гістограму. Всі вихідні дані для кожного зображення було записано у окремий вихідний файл.

В середньому було отримано наступні значення для характеристик розподілів:

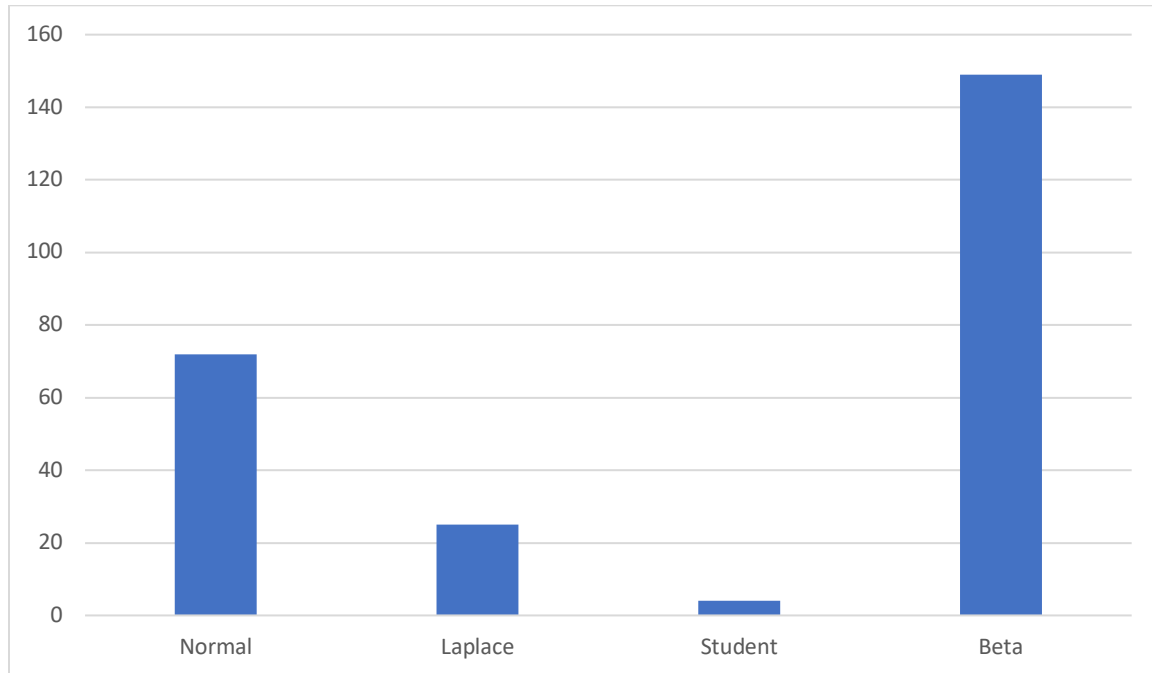
Математичне очікування - 157.42204939176364

Дисперсія - 3447.3198232942555

Коефіцієнт асиметрії – -0.544311016148974

Коефіцієнт ексцесу - 0.882267849578084

3) Після апроксимації отриманих гістограм для кожної гістограми було визначено найкращий тип апроксимуючий розподіл та побудовано розподіл отриманих апроксимацій за видом апроксимуючого розподілу:

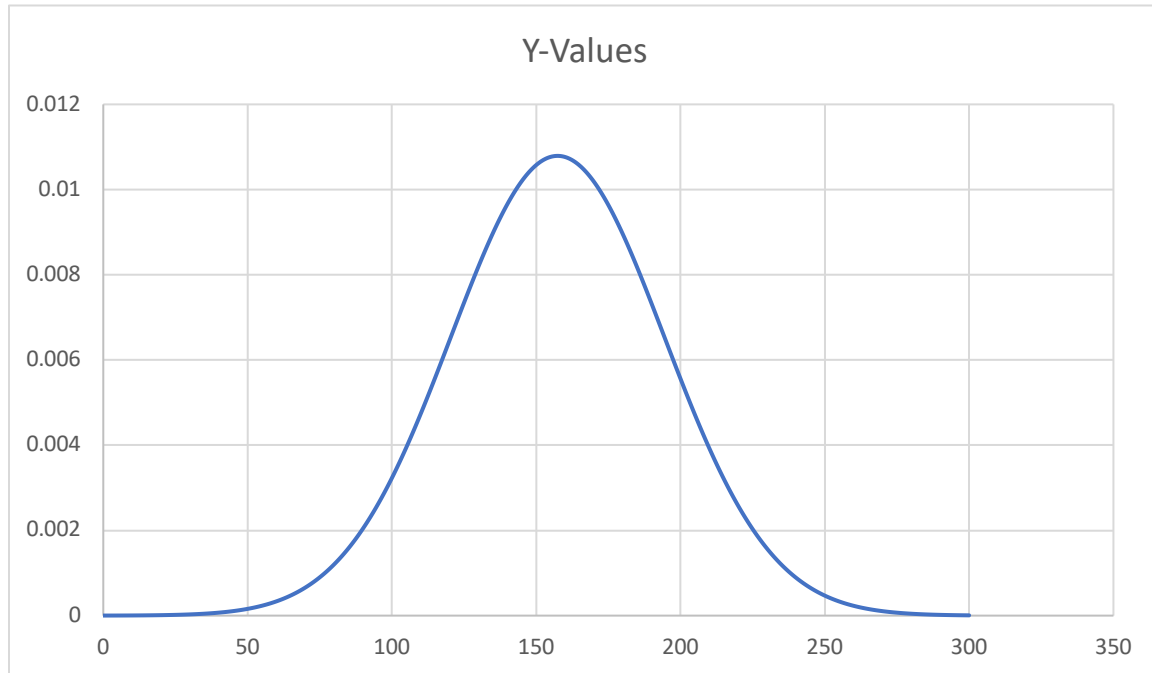


Для даної вибірки зображень найкраще підійшов бета-розподіл.

#### 4) Багатовимірні гаусові моделі:

- Математичне очікування

Вектор середнього	157.4221
Матриця коваріації	1366.7403



- Математичне очікування та дисперсія

Вектор середнього	157.4221	3447.31982
Матриця коваріації	1366.7403	-33561.6732
	-33561.6732	4191897.3060

- Математичне очікування, дисперсія та коефіцієнт асиметрії

Вектор середнього	157.4221	3447.3198	0.5443
Матриця коваріації	1366.7403	-33561.6732	-28.7551
	-33561.6732	4191897.3060	579.6401
	-28.7551	579.6401	1.1048

- Математичне очікування, дисперсія, коефіцієнти асиметрії та ексцесу

Вектор середнього	157.4221	3447.3198	0.5443	0.8823
Матриця коваріації	1366.7403	-33561.6732	-28.7551	89.2180
	-33561.6732	4191897.3060	579.6401	-4065.5181
	-28.7551	579.6401	1.1048	-4.5731
	89.2180	-4065.5181	-4.5731	33.2634

### **3. Висновки**

В ході лабораторної роботи було проаналізовано розподіл яскравості пікселів зеленого каналу сформованої вибірки зображень. Проаналізувавши кожне зображення з вибірки та обчисливши числові характеристики, ми намагалися апроксимувати гістограму кожного зображення до одного з заданих розподілів. З результатів обчислень можна зробити висновок, що однозначно визначити розподіл (вид розподілу), що описує яскравості пікселів зеленого каналу зображення, для всіх зображень не можна. Однак, мінімізація помилки показала, що Бета-розподіл частіше є найкращим, а розподіл Стюдента підходить лише для деяких (для малої кількості) зображень.