# Исследование характеристик случайных графов для различения распределений

Куценко Дмитрий, Шатурный Алексей

2025

## 1 Постановка задачи

В данной работе рассматривается задача классификации двух пар параметрических распределений (Laplace$(0, \beta)$ с Normal$(0, \sigma^2)$ и Pareto$(\alpha)$ с Exp$(\lambda)$) с использованием конструкций случайных графов. Основная цель - исследовать, как числовые характеристики графов, построенных на выборках из этих распределений, зависят от параметров распределений и могут быть использованы для построения статистического критерия.

### 1.1 Математическая формулировка

Пусть задана выборка $\hat{\Xi} = (\xi_1, \ldots, \xi_n)$ независимых реализаций случайной величины $\xi$. Требуется проверить две гипотезы:

- $H_0 : \xi \sim \mathcal{N}(0, \sigma^2)$ - нормальное распределение (для Алексея Pareto$(\alpha)$)
- $H_1 : \xi \sim$ Laplace$(0, \beta)$ - распределение Лапласа (для Алексея Exp$(\lambda)$)

Для решения задачи используются две конструкции случайных графов:

1. **KNN-граф** $\mathcal{GK}(\hat{\Xi}, k)$:
   - Вершины: индексы наблюдений $V = \{1, \ldots, n\}$
   - Рёбра: $(i, j) \in E$ если $\xi_j \in \text{KNN}(\xi_i, k)$ или $\xi_i \in \text{KNN}(\xi_j, k)$
2. **Дистанционный граф** $\mathcal{GD}(\hat{\Xi}, d)$:
   - Вершины: индексы наблюдений $V = \{1, \ldots, n\}$
   - Рёбра: $(i, j) \in E$ если $|\xi_i - \xi_j| \leq d$

## 2 Исследование характеристик графов

В первой части работы исследовалось поведение числовых характеристик графов в зависимости от параметров распределений.

## 2.1 Используемые характеристики

Для анализа были выбраны следующие характеристики графов:

1. Для пары из распределений Normal и Laplace
   - Число треугольников - для KNN-графа
   - Хроматическое число - для дистанционного графа
2. Для пары из распределений Pareto и Exp
   - Число компонент связности - для KNN-графа
   - Размер минимального кликового покрытия - для дистанционного графа

## 2.2 Методология исследования

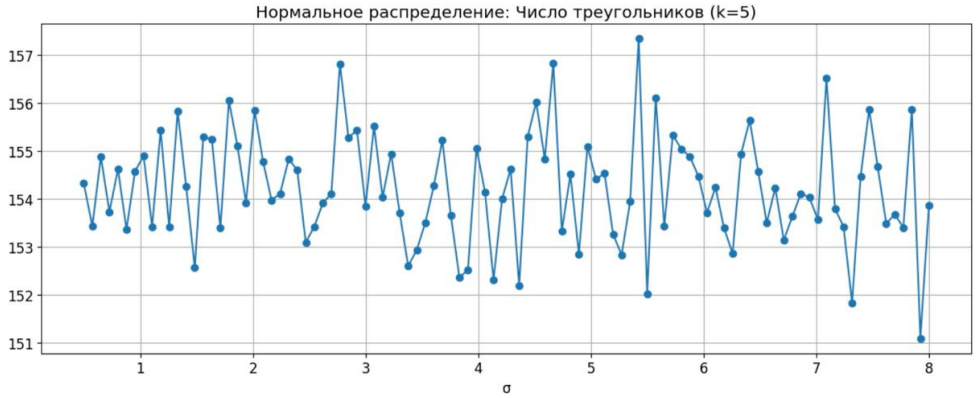
Для каждого типа графа и характеристики проводилось:

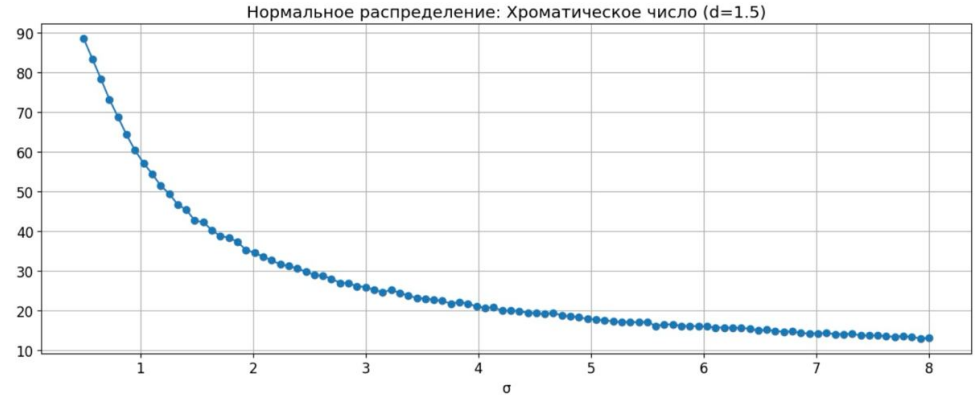
1. Фиксация размера выборки $n$ и параметра построения графа ($k$ или $d$)
2. Вариация параметра распределения:
   - Для нормального: $\sigma \in [0.5, 8.0]$
   - Для Лапласа: $\beta \in [0.5, 8.0]$
3. Для каждого набора параметров выполнялось 100 симуляций Монте-Карло
4. Усреднение значений характеристики по симуляциям

## 2.3 Результаты

### 2.3.1 Зависимость характеристик от параметра $\sigma$ нормального распределения

Ниже представлены зависимости характеристик от параметров распределений при фиксированных $n = 100$, $k = 5$, $d = 1.5$.

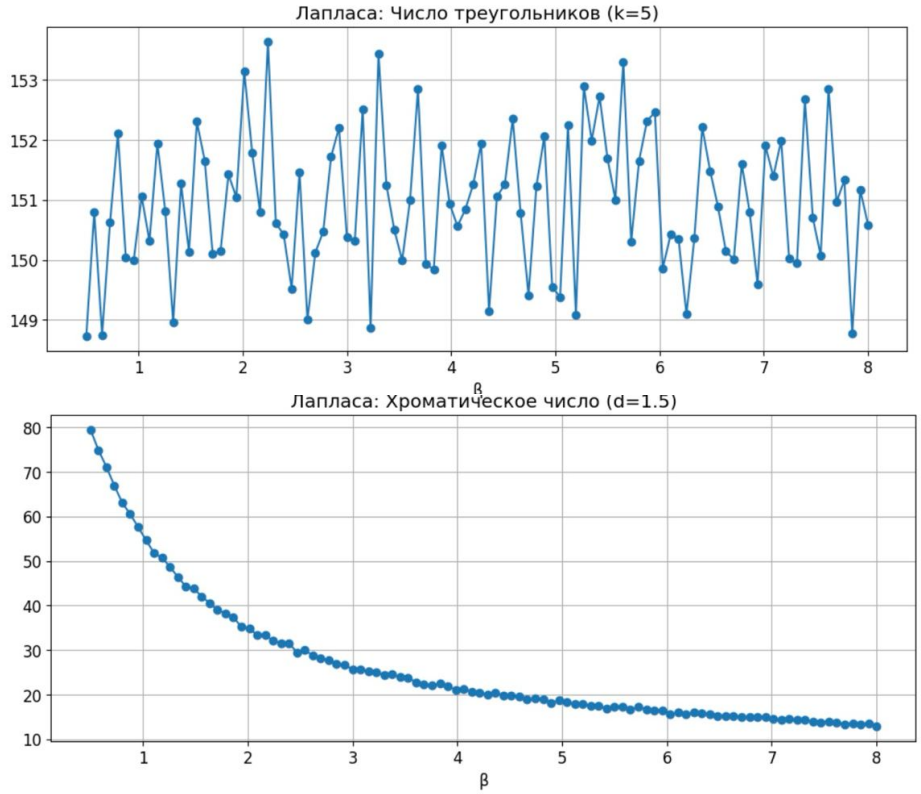



Основные наблюдения:

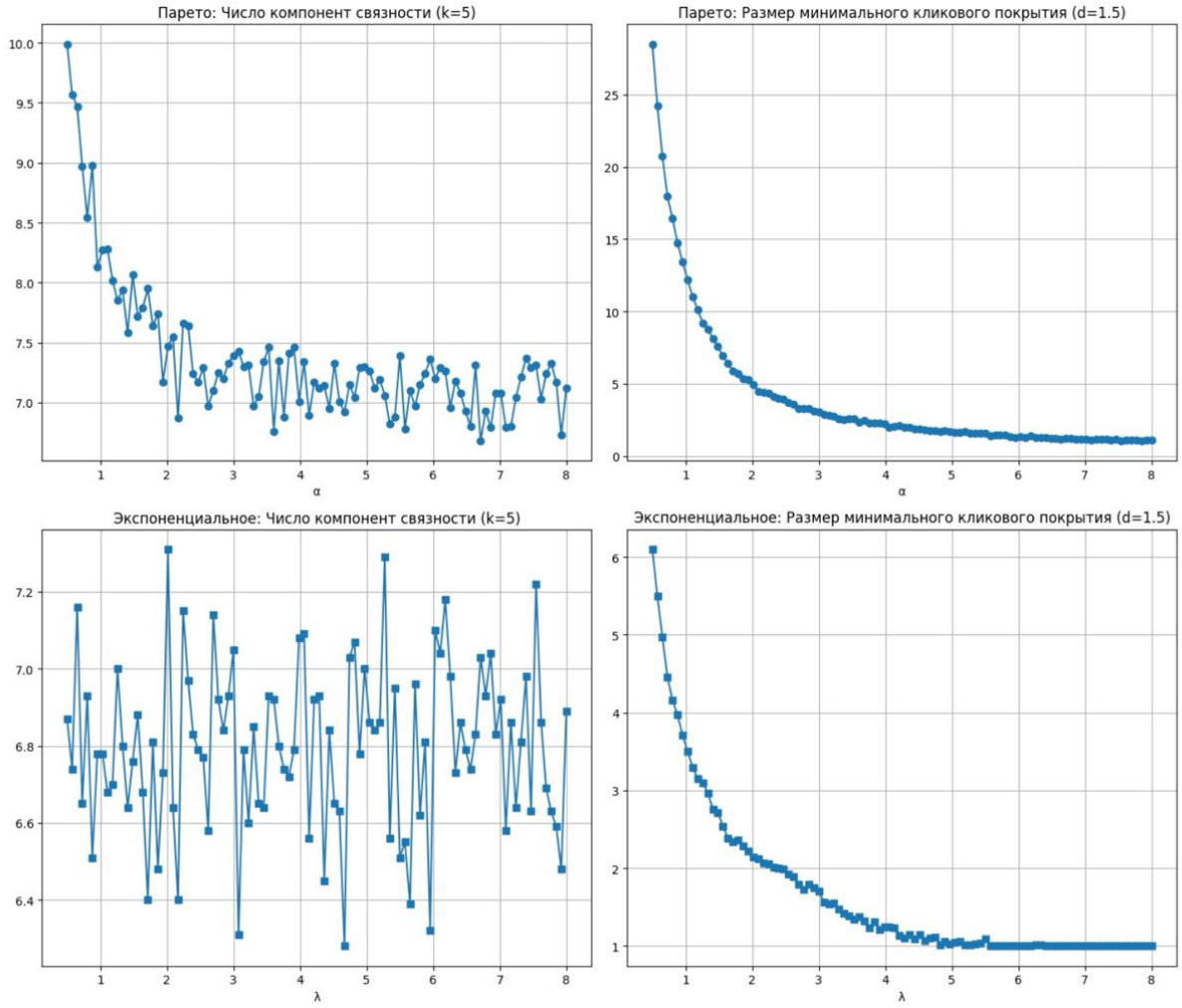
- **Число треугольников (KNN-граф)**:
  - Тяжело установить явную зависимость числа треугольников в получаемомо KNN-графе в зависимости от параметра $\sigma$ нашего распределения. В среднем количество треугольников колеблется около 154
- **Хроматическое число (дистанционный граф)**:
  - Резко убывает с ростом $\sigma$
  - Объяснение: вероятно увеличение разброса приводит к разрежению графа

### 2.3.2 Зависимость характеристик от параметра $\beta$ распределения Лапласа




Можем наблюдать ситуацию, похожую на нормальное распределение – видна явная зависимость хроматического числа от параметра $\beta$, в то время как число треугольников в KNN-графе колеблется вокруг значения 151

### 2.3.3 Зависимость характеристик от параметров $\lambda$ экспоненциального распределения и $\alpha$ распределения Парето




Основные наблюдения:

1. Аналогично паре из нормального распределения и распределения Лапласа числовая характеристика дистанционного графа выглядит более информативной и менее шумной

2. С увеличением $\lambda$ и $\alpha$ для обоих распределения размер минимального кликового покрытия дистанционного графа резко падает

### 2.3.4 Исследование поведения числовых характеристик в зависимости от параметров процедуры построения графа и размера выборки при фиксированных параметрах распределений

Будем симулировать выборки при фиксированных параметрах распределений

1. Laplace $\left(0, \sqrt{\frac{1}{2}}\right)$

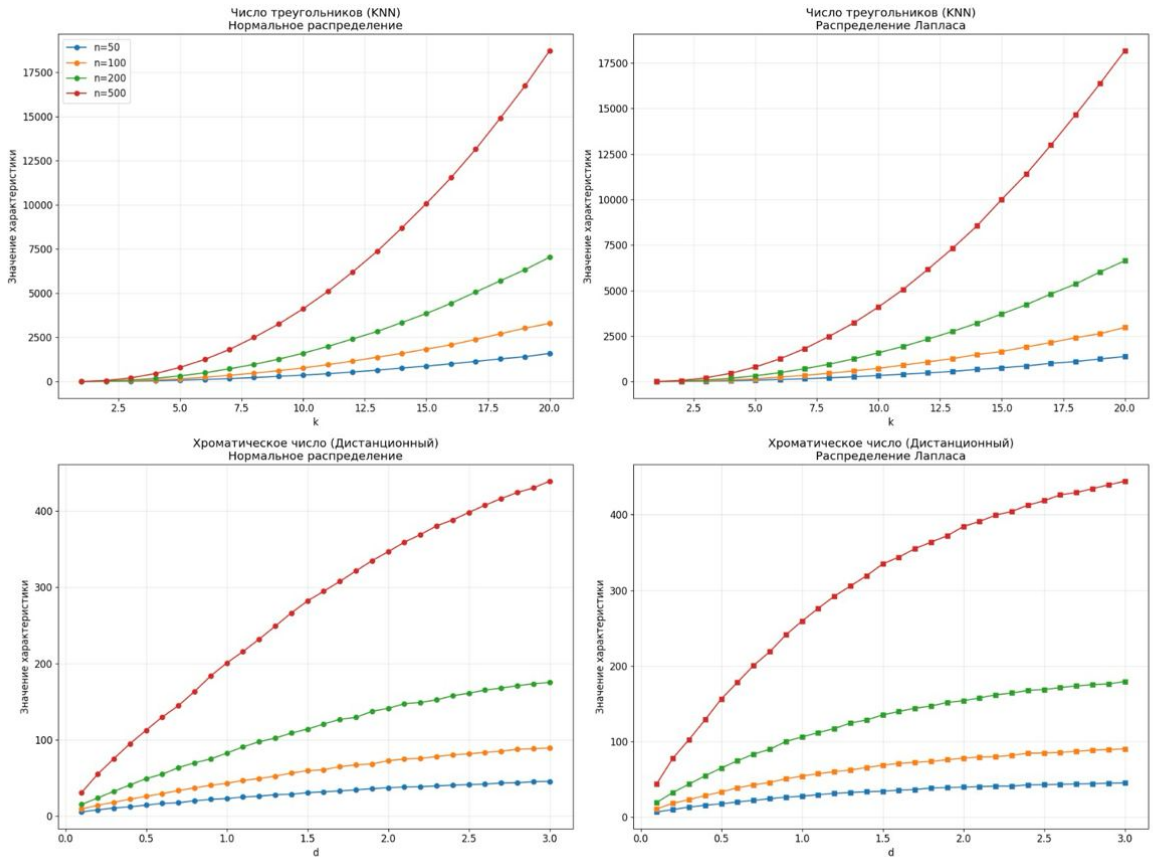
2. Normal $(0, 1)$

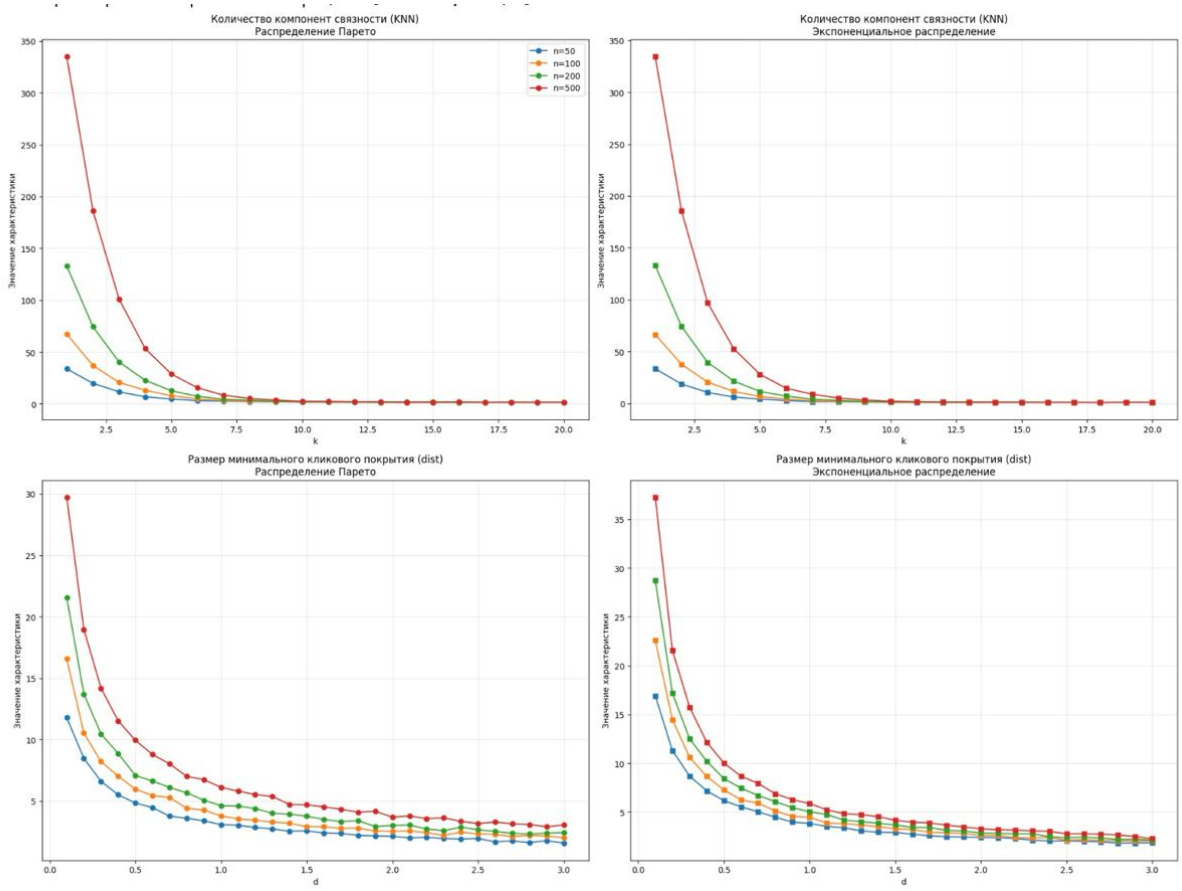
3. Pareto(3)

4. Exp $\left(\frac{2}{\sqrt{3}}\right)$

Рассмотрим следующие параметры процедуры построения графов

1. $k = 1, 2, 3, ..., 20$
2. $d \in [1, 3]$
3. $n = 50, 100, 200, 500$






Основные наблюдения:

1. При исследовании пары из нормального распределения и распределения Лапласа было установлено, что при увеличении параметров $k$ и $d$ вне зависимости от величины $n$, обе числовые характеристики получаемых случайных графов растут.

2. В паре из распределения Парето и экспоненциального распределения наоборот при увеличении параметров $k$ и $d$ исследуемые числовые характеристики падали

3. В обоих случаях логично с увеличением размера выборок (то есть параметра $n$) значение числовых характеристик росло, но характер роста (или наоборот падения) оставался прежним

### 2.3.5 Построение критических областей

# 3 Заключение первой части

Проведенное исследование показало:

1. Числовые характеристики случайных графов чувствительны к параметрам распределений

2. Пронаблюдали род зависимости каждой из выбранных числовых характеристик KNN-графа и дистанционного графа в зависимости от различных параметров распределений

3. Более информативную и явную зависимость удалось выявить при исследовании дистанционных графов

- Для пары Laplace и Normal хорошо показало себя хроматическое число
- Для пары Pareto и Exp хорошо показал себя размер минимального кликового покрытия

# 4 Применение нескольких характеристик для проверки гипотезы

По итогам исследования поведения числовых характеристик при изменении параметров распределений и параметров процедуры построения графа было принято решение работать именно с дистанционным графом, так как числовые характеристики дистанционного графа проявляли себя как более информативные.