

# Исследование свойств случайных графов на выборках из распределений и проверка гипотез

## Цель

Построить случайные графы на основе выборок из различных распределений и исследовать поведение числовых характеристик графа в зависимости от параметров построения. Кроме того, реализовать проверку гипотезы  $H_0$  против  $H_1$  на основе статистик графа.

## Используемые распределения

- Экспоненциальное распределение:  $\text{Exp}(\lambda)$ , где  $\lambda = \frac{1}{\sqrt{e^2 - e}}$ ;
- Логнормальное распределение:  $\text{LogNormal}(0, \sigma = 1)$ .

## Функции и методы

### 1. Построение графов

- `build_knn_graph(data, k)` — построение графа ближайших соседей (kNN): каждая вершина соединяется с  $k$  ближайшими по расстоянию.
- `build_distance_graph(data, d)` — построение графа расстояния: вершины соединяются ребром, если  $|x_i - x_j| \leq d$ .

### 2. Характеристики графа

- $\delta(G)$  — минимальная степень вершины в графе;
- $\chi(G)$  — приближенное хроматическое число (оценка с помощью жадной раскраски).

## Эксперименты

### Грид-серч по параметрам графа

Были проведены переборы по параметрам:

- $k \in \{3, 4, 5, 6, 7\}$  для kNN-графов;

- $d \in \{1.0, 1.5, 3.0\}$  для dist-графов.

Для каждого значения параметра строились графы на выборках из  $\text{Exp}(\lambda_0)$  и  $\text{LogNormal}(0, 1)$ , после чего вычислялись  $\delta(G)$  и  $\chi(G)$ .

## Результаты

Графовые характеристики существенно различаются для разных распределений, особенно при росте  $k$  или  $d$ . Логнормальное распределение, как правило, даёт более плотные графы с большими  $\chi(G)$  и  $\delta(G)$ .

## Проверка гипотез

### Описание

Рассматривается задача проверки гипотезы:

$$H_0 : \xi \sim f(x, \theta) \quad \text{vs} \quad H_1 : \xi \sim h(x, \nu),$$

где  $f$  и  $h$  — плотности экспоненциального и логнормального распределений соответственно.

### Методика

1. Генерируется  $N = 1000$  выборок из  $H_0$  ( $\text{Exp}$ ), строятся графы и считается  $T(G)$ ;
2. Вычисляется критическое значение  $T^* = \text{quantile}(T_{H_0}, \alpha)$  при уровне значимости  $\alpha = 0.05$ ;
3. Считается доля выборок из  $H_1$  ( $\text{LogNormal}$ ), у которых  $T(G) < T^*$  — это оценка мощности критерия.

## Результаты проверки

При  $k = 5$  и  $n = 100$  для kNN-графа обе гипотезы не отвергаются.