

Отчёт по проверке гипотез с использованием случайных графов

Часть I: $\text{Stable}(\alpha = 1)$ vs $\text{Normal}(0, 1)$

Хамаганов Ильдар

Введение

Целью данной части исследования было оценить, насколько топологические характеристики случайных графов позволяют различать выборки из двух распределений:

- H_0 : $\text{Stable}(\alpha = 1)$;
- H_1 : $\text{Normal}(0, 1)$.

Использовались два типа графов:

KNN-граф: характеристика $T^{\text{knn}} = \max \deg(G)$ (максимальная степень).

Дистанционный граф: характеристика $T^{\text{dist}} = \chi(G)$ (хроматическое число).

1 Настройка окружения и код

Импорт и автозагрузка

```
%load_ext autoreload
%autoreload 2
import sys, os
project_root = os.path.abspath(os.path.join(os.getcwd(), '..'))
sys.path.append(project_root)

import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from tqdm import tqdm

from src.data_utils import sample_stable, sample_normal
from src.build_graph import build_knn_graph, build_distance_graph
from src.graph_analyzer import GraphAnalyzer
from src.monte_carlo import monte_carlo_simulation
from src.visualization import plot_distributions, plot_critical_region
```

Параметры экспериментов

- Размер выборки: $n = 200$.
- Число MC-итераций: $N_{MC} = 500$.
- Параметры KNN-графа: $k \in \{3, 5, 7, 10, 12, 15, 20\}$.
- Параметры дистанционного графа: $d \in \{0.5, 1.0, 1.5, 2.0\}$.

2 Эксперимент1: зависимость от «»

Описание. Для каждой из двух распределений (Stable, Normal) вычисляли:

$$\bar{T}^{\text{knn}}(k) = \text{mean}[\max \deg(G)], \quad \bar{T}^{\text{dist}}(d) = \text{mean}[\chi(G)].$$

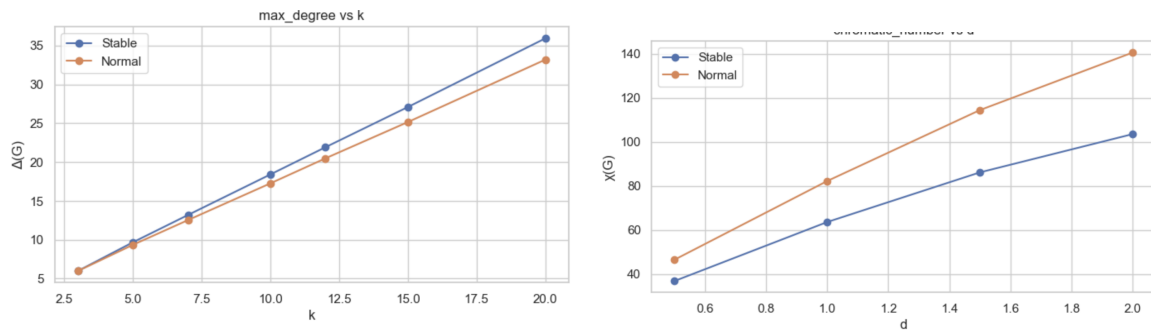


Рис. 1: Слева: $\bar{T}^{\text{knn}} = \Delta(G)$ vs k . Справа: $\bar{T}^{\text{dist}} = \chi(G)$ vs d .

Результаты.

Выводы.

- **KNN-граф** (T^{knn}): кривая почти горизонтальна, разрыв между Stable и Normal менее 1%, распределения перекрываются.
- **Дистанционный граф** (T^{dist}): $\chi(G)$ растёт с d , и для Normal значения значительно выше (до ~ 140 vs ~ 103 при $d = 2$). Статистика хорошо разделяет гипотезы.

3 Эксперимент 2: зависимость от k, d, n

Описание. Исследовали:

1. Зависимость $\bar{T}^{\text{knn}}(k)$ и $\bar{T}^{\text{dist}}(d)$ при $n = 200$.
2. Зависимость при фиксированных $k = 10, d = 1.0$ от $n \in \{100, 200, 300, 500\}$.

Сводные итоги.

Таблица 1: Отношение $\overline{T}^{H_1}/\overline{T}^{H_0}$

Параметр	KNN (k)	Dist (d)	Dist (n)
Минимум	$0.92\times$	$1.47\times$	$2.30\times$
Максимум	$1.06\times$	$2.80\times$	$3.10\times$

Выводы.

- $\Delta(G)$ увеличивается с k, n , но H_0/H_1 остаются близки (отношения ≈ 0.9 – 1.06).
- $\chi(G)$ показывает высокую чувствительность: отношение до $\sim 3\times$ при росте n .

4 Эксперимент 3: критические области и мощность

Условия. $n = 500$, $k = 10$, $d = 1.0$, уровень значимости $\alpha = 0.05$.

Таблица 2: Критические значения и характеристики теста

Граф	CV	FPR	TPR	AUC
KNN (Δ)	18.7	5.0%	4.8%	0.545
Distance (χ)	114	5.0%	100.0%	1.000

Результаты.**Выводы.**

- Тест по $\Delta(G)$ практически не различает гипотезы (мощность $\approx \alpha$).
- Тест по $\chi(G)$ идеально разделяет выборки (AUC=1, мощность=100%).

Заключение

- Для Stable(1) vs Normal(0, 1) лучшая статистика — $\chi(G)$ дистанционного графа.
- $\Delta(G)$ KNN-графа неинформативна и не подходит в качестве единственной характеристики.
- Рекомендуем: $d = 1.0$, $n \geq 500$ для надёжного критерия.
- Возможные улучшения: исследовать дополнительные метрики (центральность, доминирование).