

0.1 Пункт 1

Для проведения эксперимента фиксировалась выборка размером 100, $k = 5$ и $d = 0.2$. α варьировалась от 0.5 до 10 с шагом 0.15 (60 значений). Усреднение проводилось по 10 различным значениям.

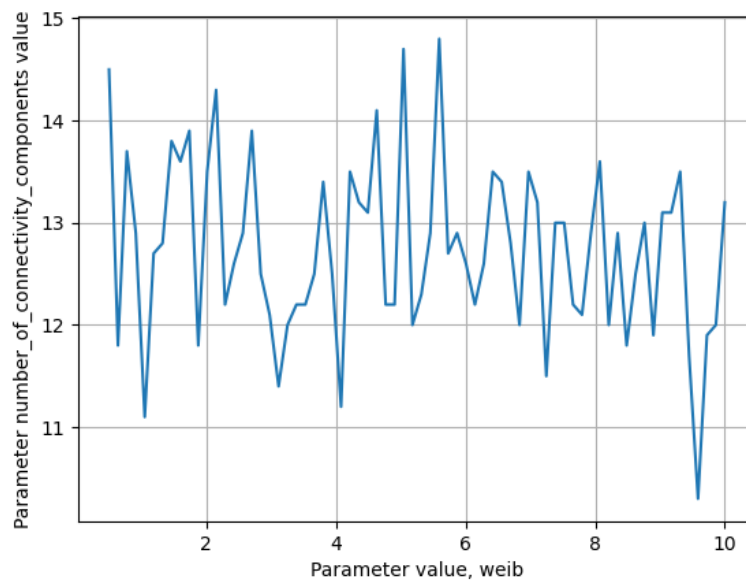


Рисунок 1 – Зависимость числа компонент связности от α (распределение Weibull)

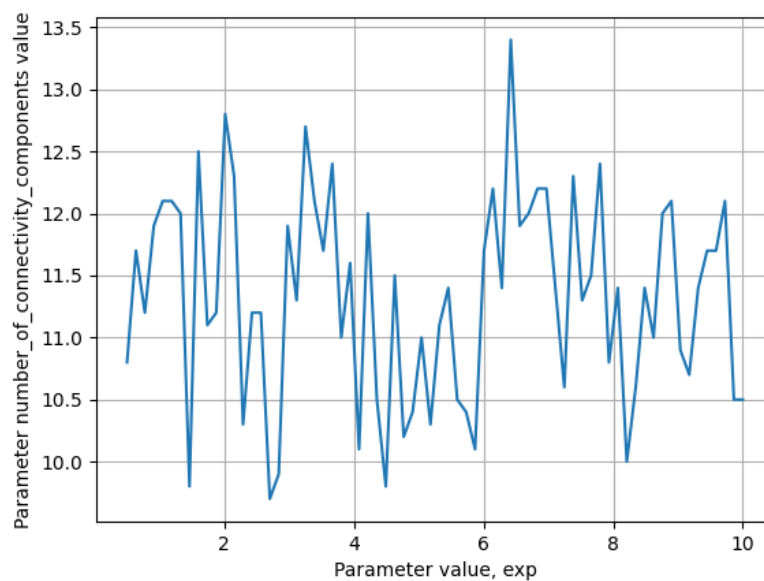


Рисунок 2 – Зависимость числа компонент связности от α (распределение Exp)

Число компонент связности в knn-графах для обоих распределений (рис. 1 и 2) слабо зависит от α : среднее значение составляет ≈ 12.5 для Weibull и ≈ 11 для Exp.

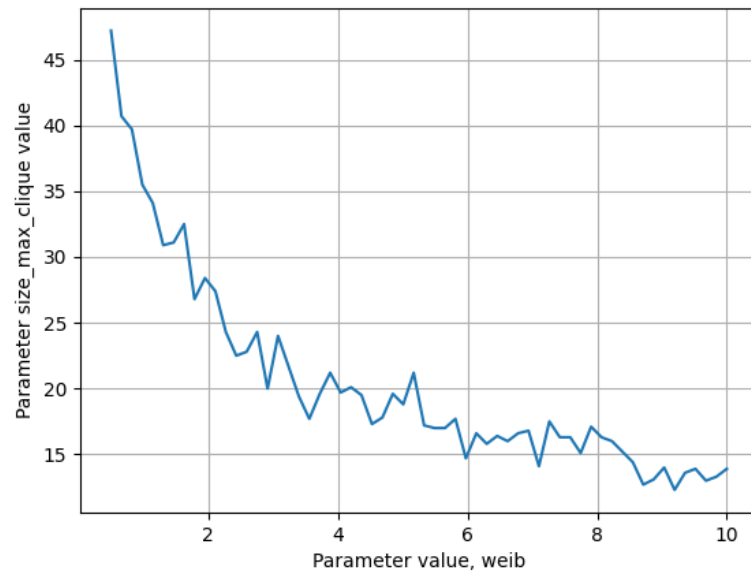


Рисунок 3 – Зависимость размера максимальной клики от α (Weibull)

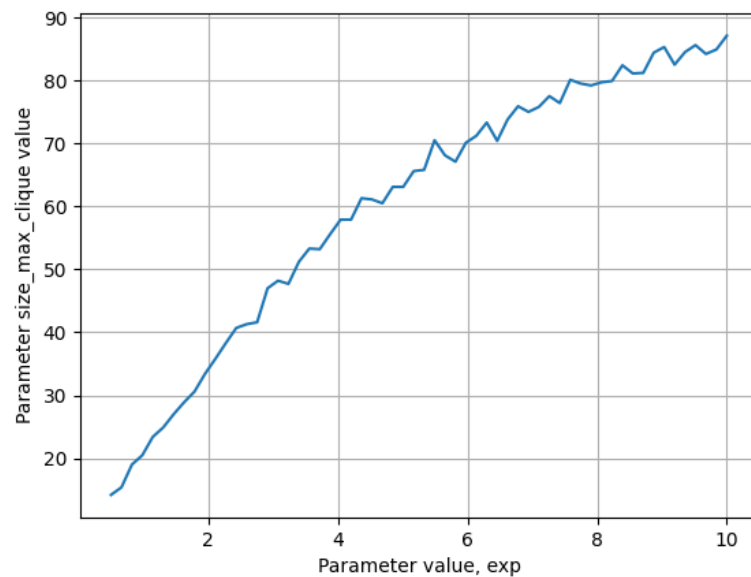


Рисунок 4 – Зависимость размера максимальной клики от α (Exp)

Для размера максимальной клики наблюдается степенная зависимость (рис. 3 и 4) с различными показателями степени: ≈ 0.4 для Weibull и ≈ 0.4 для Exp.

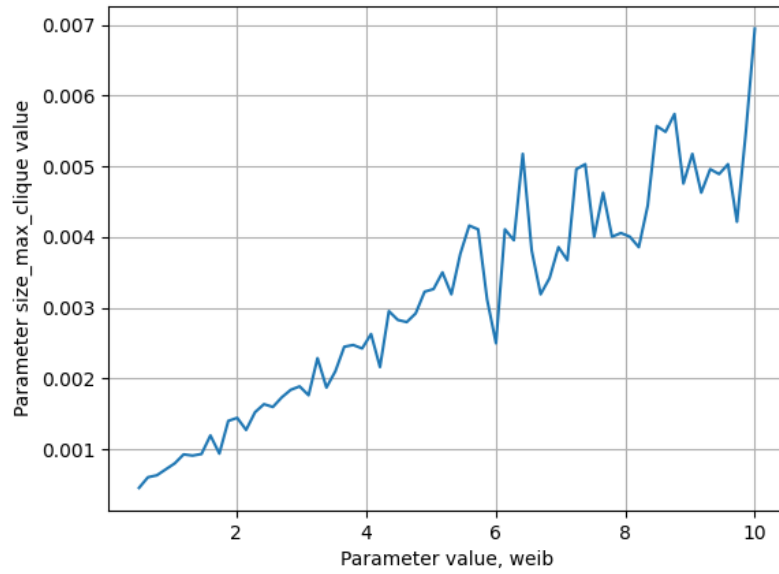


Рисунок 5 – Выравнивание зависимости максимальной клики от α (Weibull)

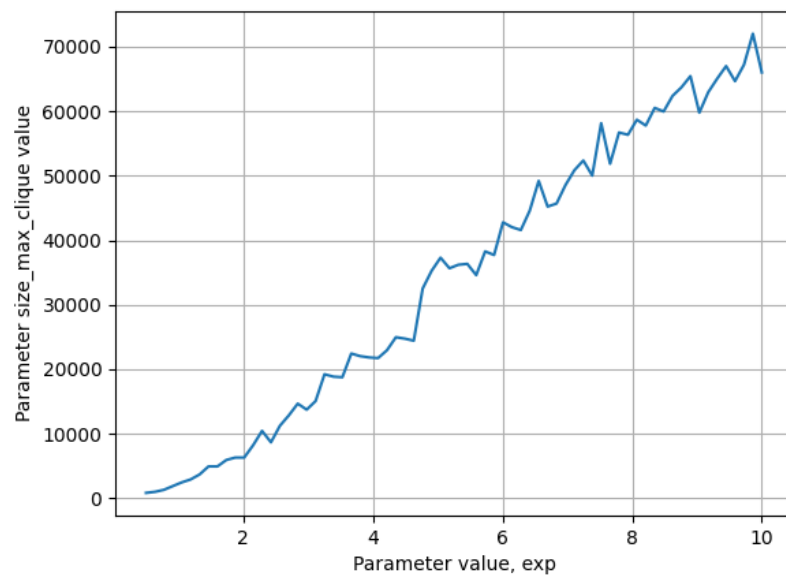


Рисунок 6 – Выравнивание зависимости максимальной клики от α (Exp)

После преобразования степенной зависимостью (рис. 5 и 6) получены линейные зависимости, подтверждающие степенной характер.

0.2 Пункт 2

Параметры эксперимента:

- k : от 2 до 20 (шаг 1)
- d : от 0.05 до 10 (60 значений)
- n : от 50 до 100 (шаг 2)
- Усреднение по 10 реализациям

Основные зависимости:

1. От k : степенная с отрицательным показателем
2. От d : степенная с показателем < 0
3. От n : линейная положительная (меньшая дисперсия для кликового числа)

0.3 Пункт 3

Результаты тестирования на выборке $n = 300$ (1000 итераций):

- **knn**: power = 0.992, error = 1.0
- **dist**: power = 0.578, error = 1.0

Число компонент связности схоже для обоих распределений, кликовое число имеет значительные различия.

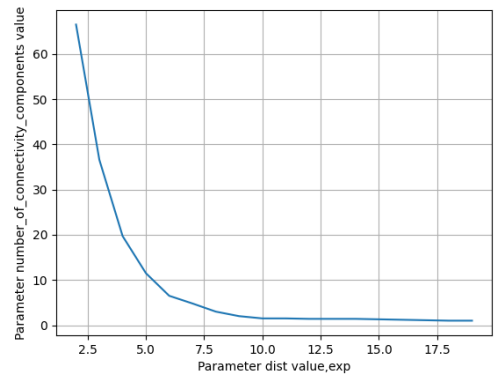
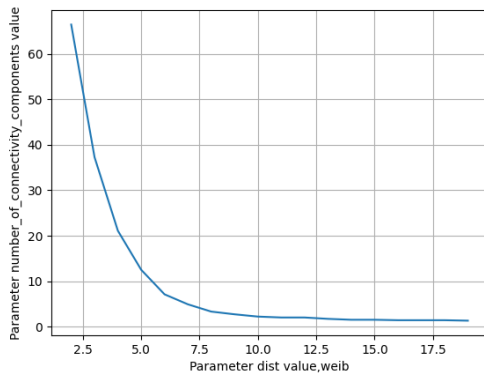


Рисунок 7 – Зависимость числа компонент связности от k (Weibull слева, Exp справа)

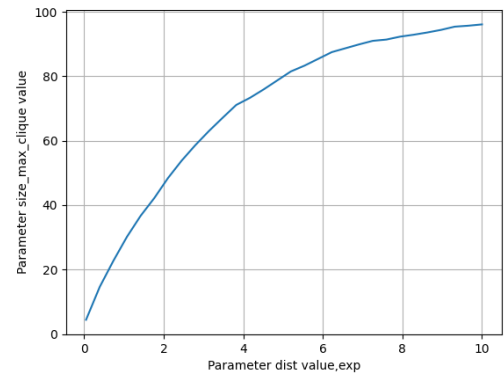
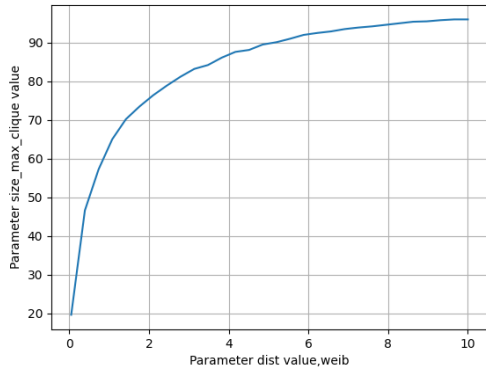


Рисунок 8 – Зависимость кликового числа от d (Weibull слева, Exp справа)

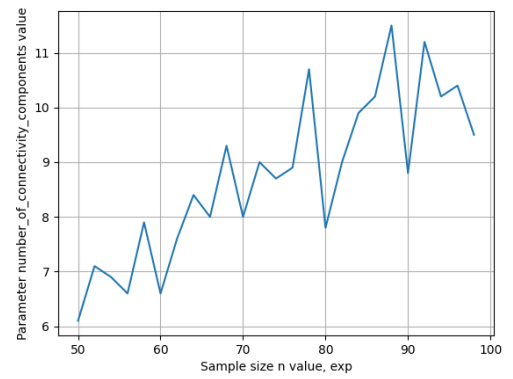
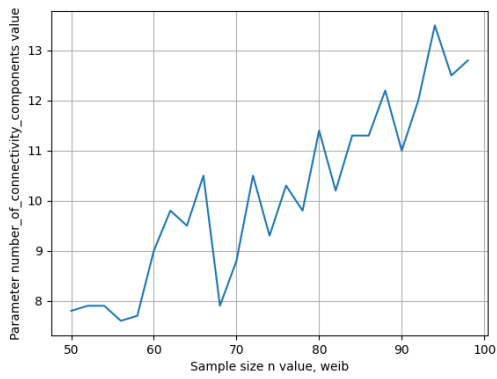


Рисунок 9 – Зависимость числа компонент связности от n (Weibull слева, Exp справа)

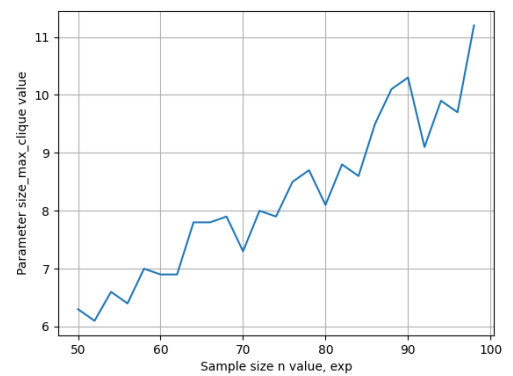
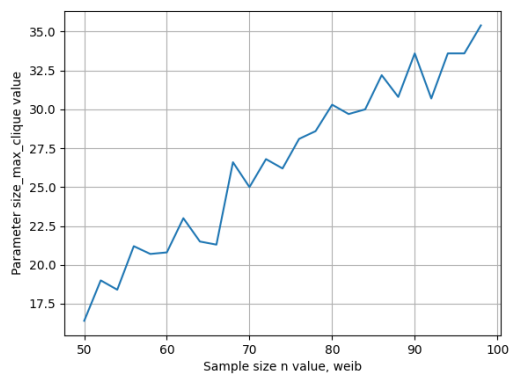


Рисунок 10 – Зависимость кликового числа от n (Weibull слева, Exp справа)

0.4 Анализ распределений по параметрам

0.4.1 Распределение Student

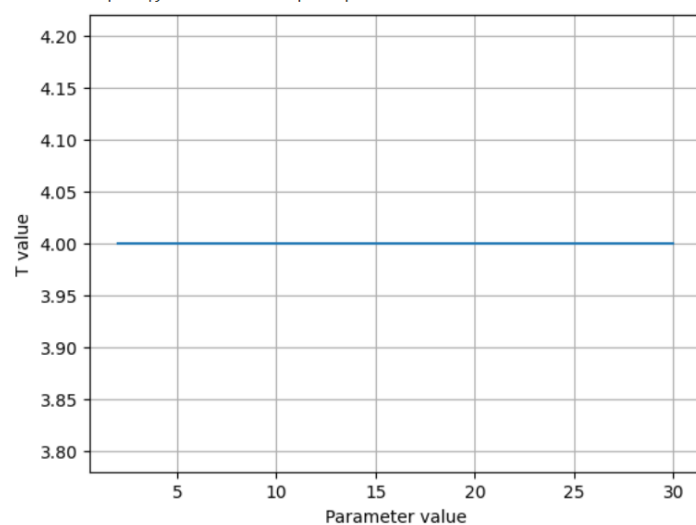


Рисунок 11 – Анализ максимальной степени графа (Student)

Максимальная степень графа не зависит от параметров распределения.

0.4.2 Распределение Laplace

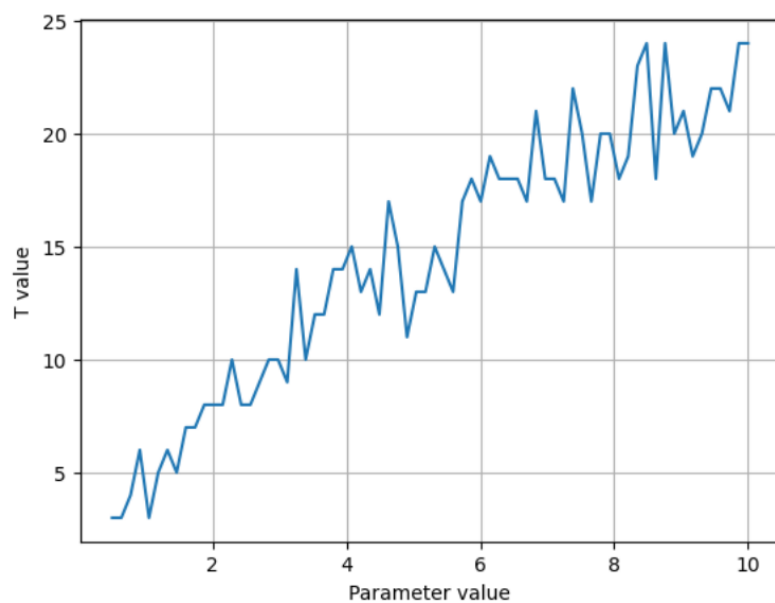


Рисунок 12 – Анализ размера максимального независимого множества (Laplace)

Размер максимального независимого множества прямо пропорционален параметру распределения.

0.5 Анализ по k и d

0.5.1 Распределение Student (k)

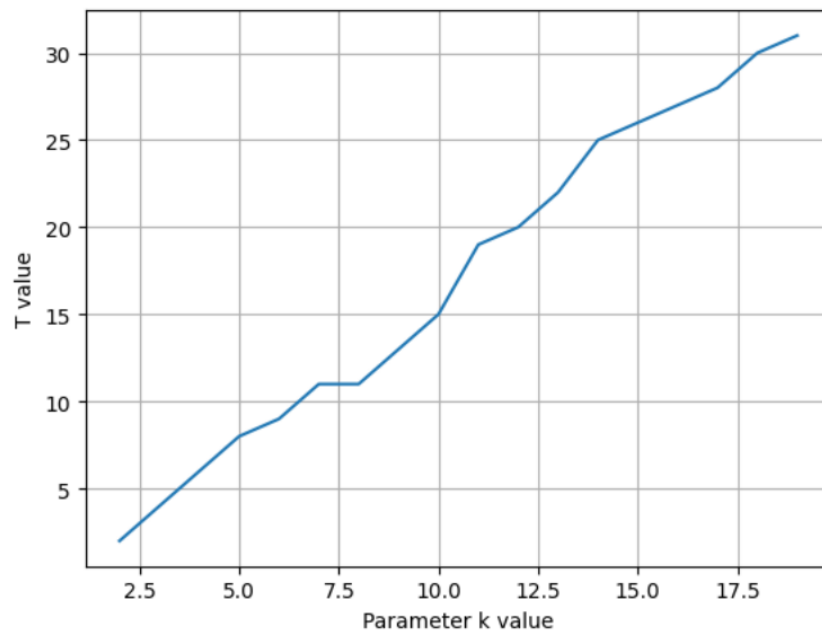


Рисунок 13 – Зависимость максимальной степени от k (Student)

Линейная зависимость максимальной степени графа от k .

0.5.2 Распределение Laplace (d)

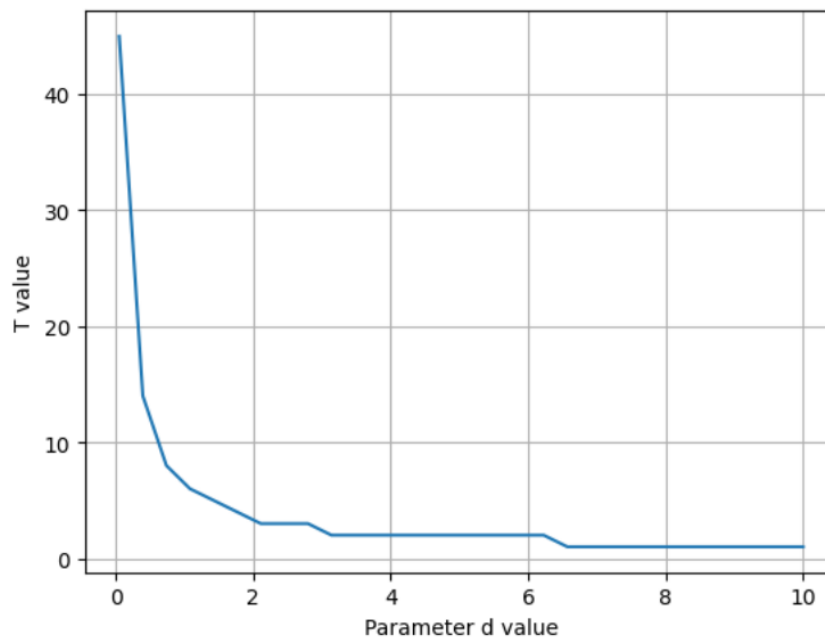


Рисунок 14 – Зависимость размера независимого множества от d (Laplace)

Обратная зависимость размера максимального независимого множества от d .

0.6 Анализ по выборке n

0.6.1 Распределение Student

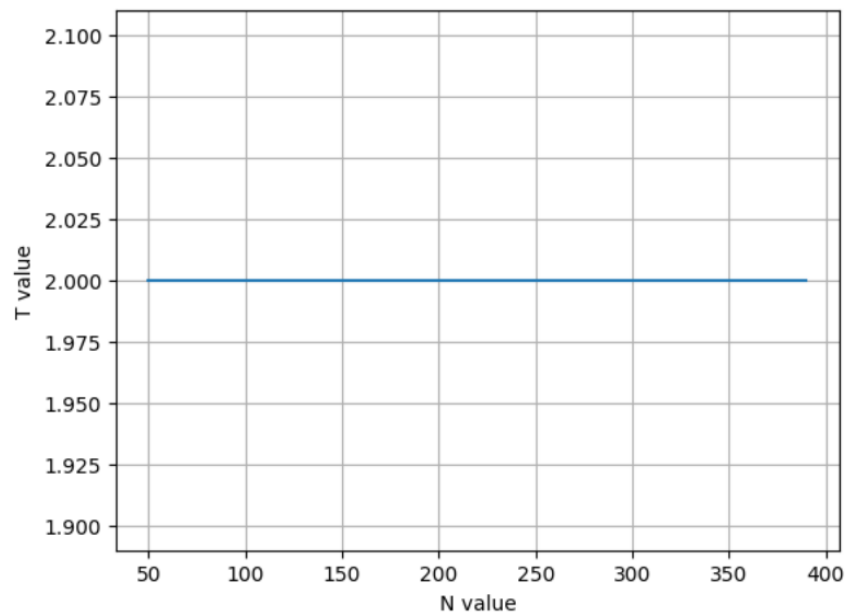


Рисунок 15 – Зависимость максимальной степени от n (Student)

Отсутствие зависимости от объема выборки.

0.6.2 Распределение Laplace

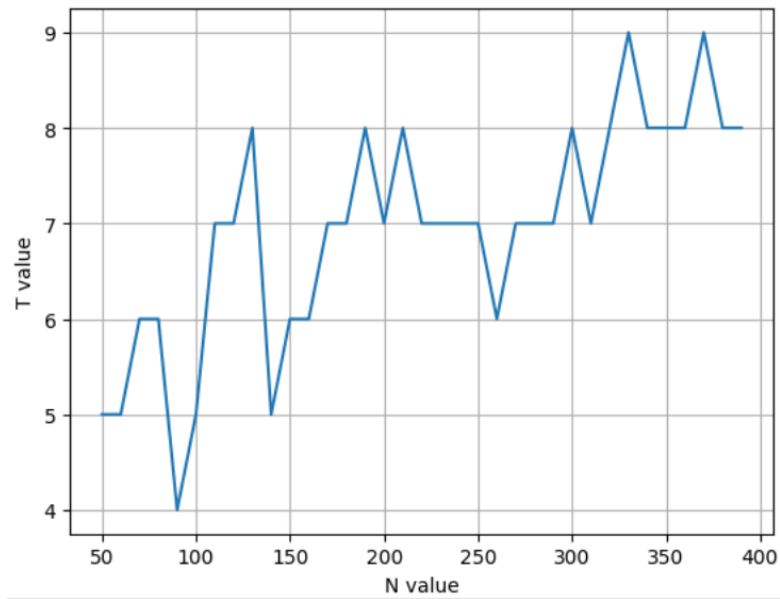


Рисунок 16 – Зависимость размера независимого множества от n (Laplace)

Прямая пропорциональность объему выборки.

0.7 Тестирование для Laplace и Student

Результаты ($n = 300, 1000$ итераций):

- Мощность критерия: 0.13
- Ошибка: 1.0

Вероятность ошибочного принятия H_1 не превышает 13%.

0.8 Важность характеристик

0.8.1 Student vs Laplace

- max_degree: 0.6129
- size_max_independent_set: 0.3871

0.8.2 Exp vs Weibull

- number_of_connectivity_components: 0.4483
- size_max_clique: 0.5517

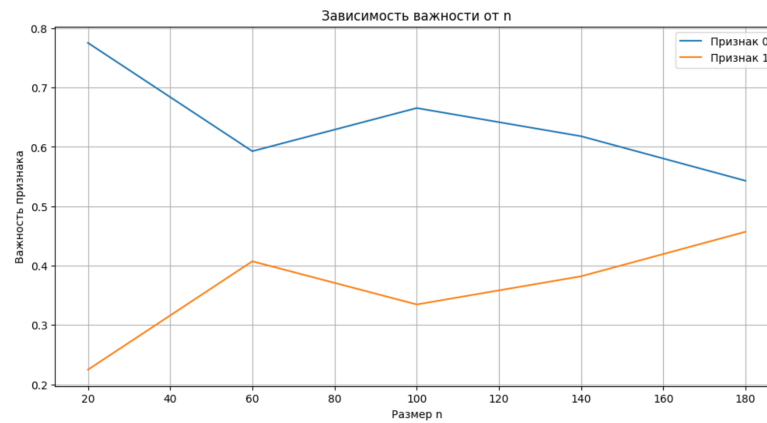


Рисунок 17 – Важность признаков (Student/Laplace)

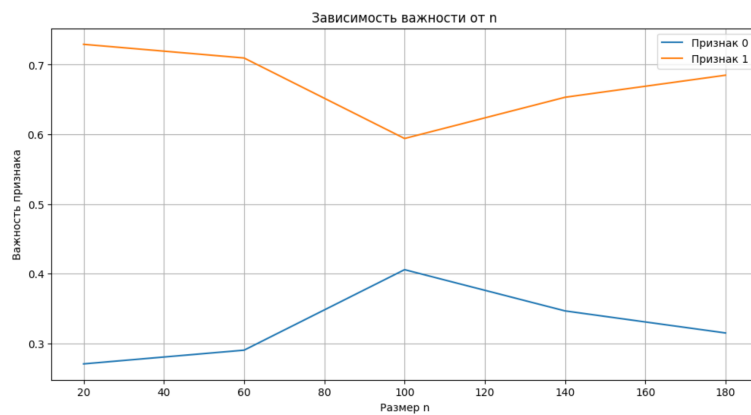


Рисунок 18 – Важность признаков (Exp/Weibull)

Выводы:

- Для Student/Laplace ключевой признак — максимальная степень графа
- Для Exp/Weibull ключевой признак — число компонент связности

0.9 Анализ метрик

0.9.1 Student vs Laplace

- При малых n : лучший метод — kNN
- При больших n : точность = 1 для всех методов
- Ошибка I рода: 0.5
- Мощность: 0.5
- Точность: 1.0

0.9.2 Exp vs Weibull

- При малых n : лучшие методы — Логистическая регрессия и kNN
- При больших n : лучшие методы — Дерево решений и Логистическая регрессия
- Ошибка I рода: 0.02
- Мощность: 0.23
- Точность: 0.71

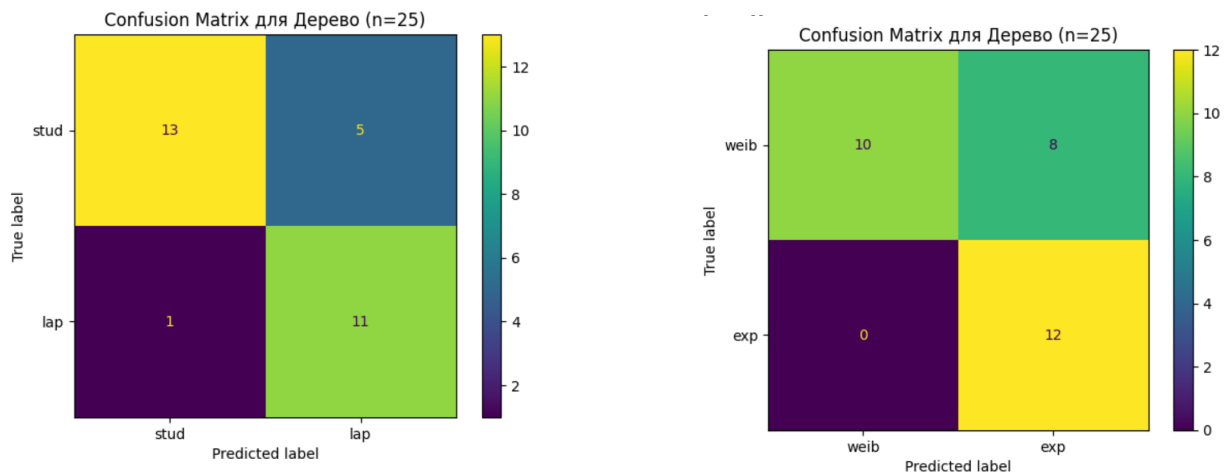


Рисунок 19 – Матрицы ошибок для $n = 25$ (слева: Student/Laplace, справа: Exp/Weibull)

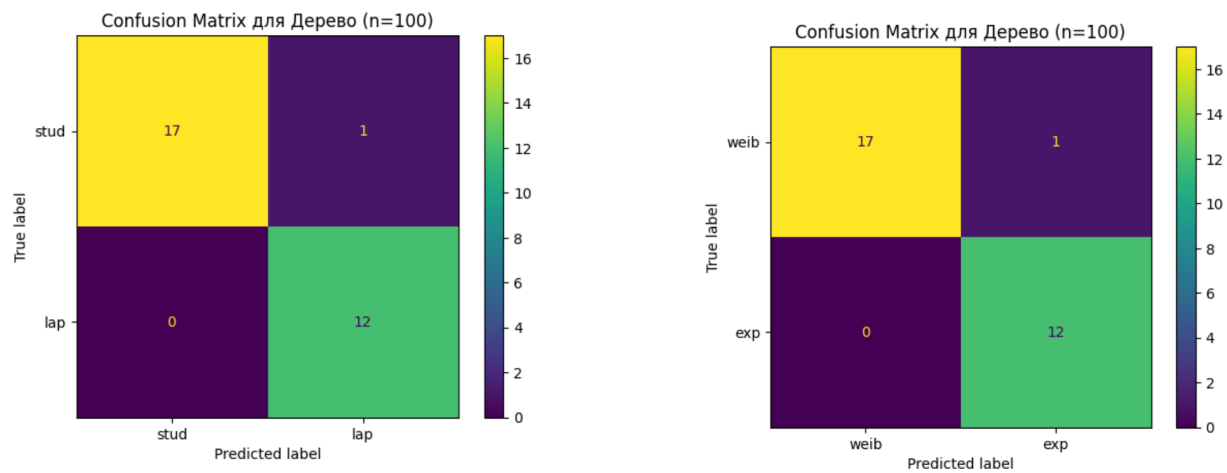


Рисунок 20 – Матрицы ошибок для $n = 100$

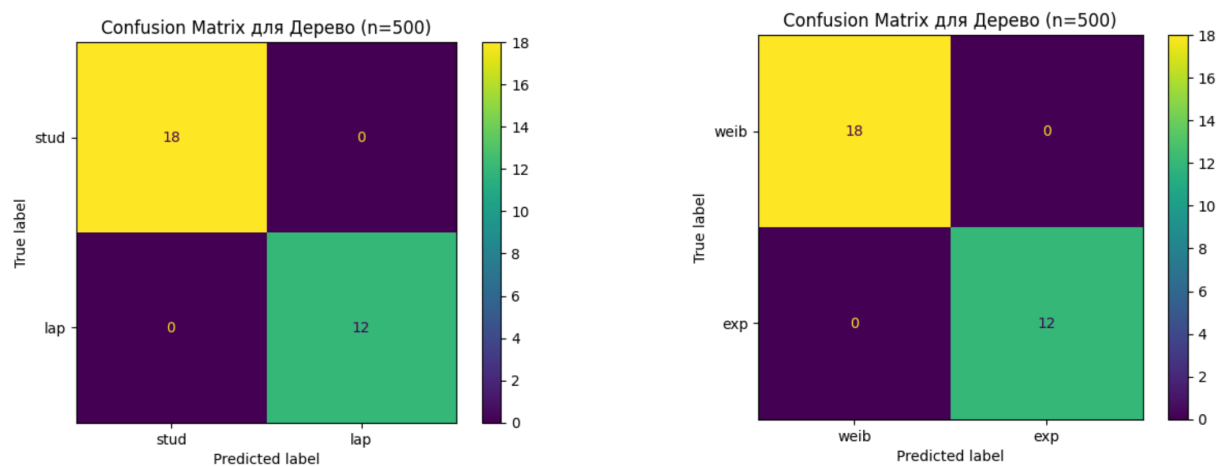


Рисунок 21 – Матрицы ошибок для $n = 500$

Реализация

– Часть 1

- Создание gd: Лев
- Создание gk: Илья
- Пункт 1: Лев, Илья
- Пункт 2: Лев, Илья
- Пункт 3: Илья

– Часть 2

- Пункты 1-2: Лев
- Пункт 3: Илья