

Анализ 1 части задания

Шевцов Лев и Дильдин Илья ПАДИИ

Содержание

1	Анализ функций по их параметрам	1
2	Анализ функций по k и d	1
3	Анализ функций по выборке n	1
4	Пункт 3	3

1 Анализ функций по их параметрам

1)stud распределение

Максимальная степень графа не влияет на числовую характеристику при изменении параметра

2)lap распределение

Размер максимального независимого множества к числовой характеристике стремиться к прямой зависимости, то есть чем больше параметр тем больше размер максимального независимого множества

2 Анализ функций по k и d

1)stud распределение

Максимальная степень графа имеет линейную зависимость при изменении параметра k при создании gk

2)lap распределение

Размер максимального независимого множества к d проявляет примерно обратную зависимость

3 Анализ функций по выборке n

1)stud распределение

Максимальная степень графа не влияет на числовую характеристику при изменении выборки

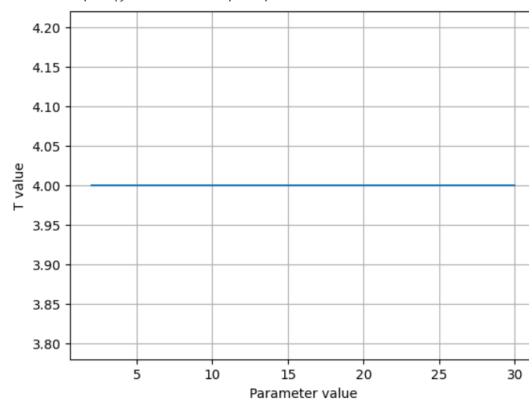


Рис. 1: Анализ по параметрам - stud распределение

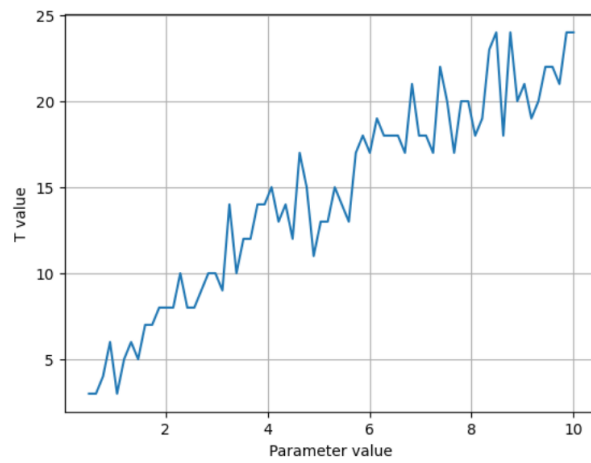


Рис. 2: Анализ по параметрам - лар распределение

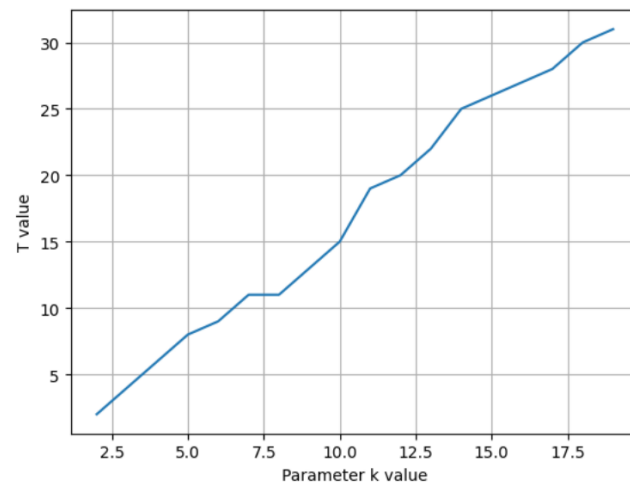


Рис. 3: Анализ по k - stud распределение

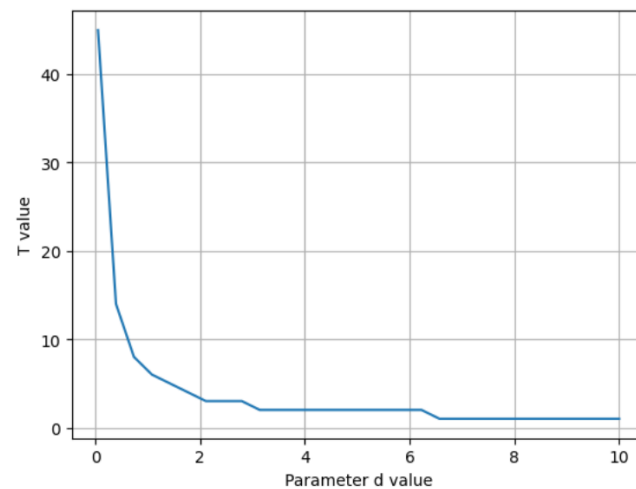


Рис. 4: Анализ по d - лар распределение

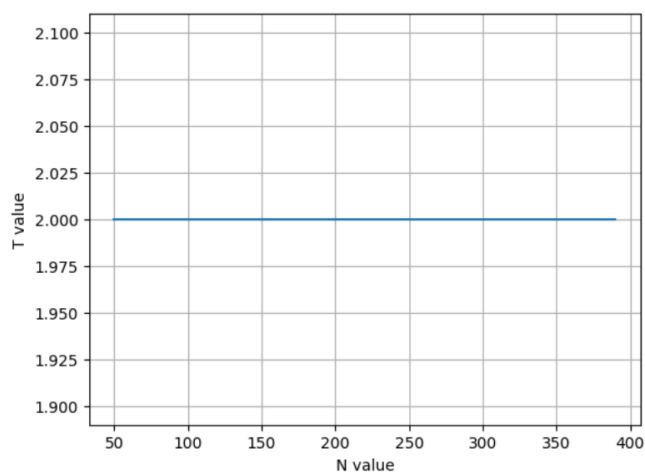


Рис. 5: Анализ по n - stud распределение

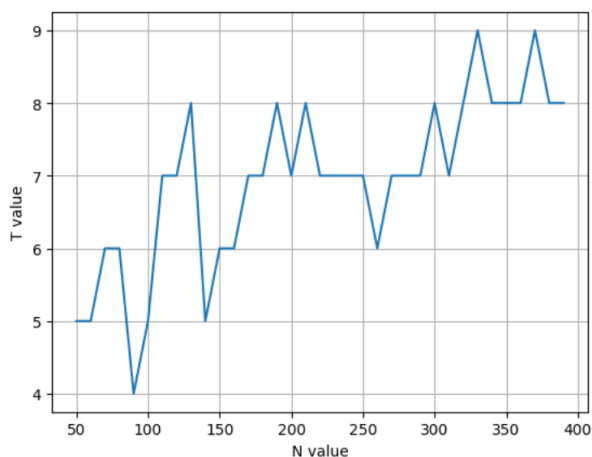


Рис. 6: Анализ по n - lap распределение

2)lap распределение

Размер максимального независимого множества к размеру выборки n стремиться к прямой зависимости, то есть чем больше d тем больше размер максимального независимого множества

4 Пункт 3

Только для Лапласа и Стьюденса:

После запуска функции мощность A вышла 0.13, а ошибка 1.0. Вероятность неправильно принять H_1 составляет не более 13 процентов.

Анализ 2 части задания

Шевцов Лев и Дильдин Илья ПАДИИ

Содержание

1	Исследование важность характеристик	1
2	Исследование метрики	1

1 Исследование важность характеристик

Важность признаков у stud и lap при постоянном выбранном нами n:

`max_degree: 0.6129`

`size_max_independent_set: 0.3871`

Важность признаков у exp и weib при постоянном выбранном нами n:

`number_of_connectivity_components: 0.4483`

`size_max_clique: 0.5517`

Посмотрим как выглядит график при различных n:

Для stud и lap - синяя линия - `max_degree`, а желтая - `size_max_independent_set`. Из этого можно сделать вывод, что для определения большую роль играет максимальная степень

Для exp и weib - синяя линия - `number_of_connectivity_components`, а желтая - `size_max_clique`. Из этого можно сделать вывод, что для определения большую роль играет число компонент связности.

2 Исследование метрики

Для stud и lap - при минимальном n - лучшим алгоритмом будет К-ближайших соседей, при остальных n, чем больше n тем лучше результат, а при максимальном n точность будет равна 1 для всех алгоритмов.

Для stud и lap - при минимальном n - лучшим алгоритмом будет Логистическая регрессия и К-ближайших соседей, при остальных n, чем больше n тем лучше результат, а при максимальном n Дерево и Логистическая регрессия будут лучшими алгоритмами.

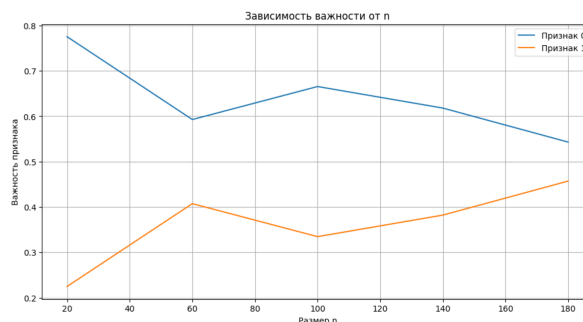


Рис. 1: Важность признаков у stud и lap

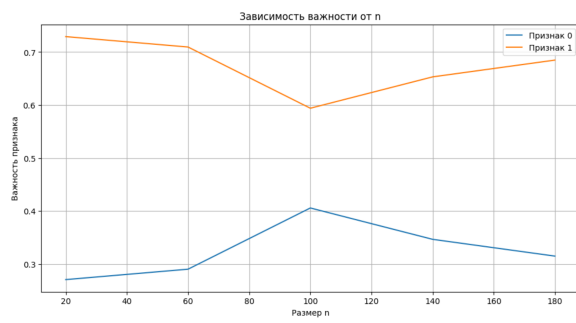


Рис. 2: Важность признаков у exp и weib

Подготовил Дильдин И. Н.

Пункт 1:

Для проведения эксперимента фиксировалась выборка размером 100, k равное 5 и d равное 0.2. α варьировалась от 0.5 до 10 с делением на 60 значений. Усреднение шло по 10 различным значениям, так как такое уже позволило понять форму большинства распределений.

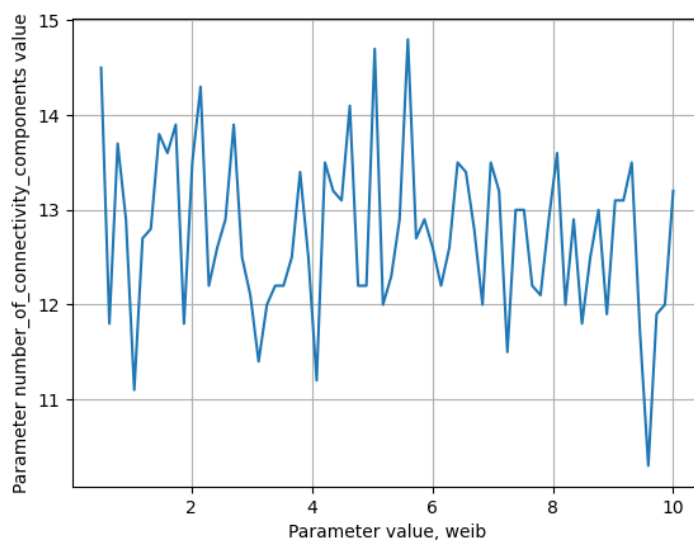


Рисунок 1 – Зависимость числа компонент связности от α при распределении weibull

В случае анализа числа компонент связности в knn при обоих распределениях (рис. 1 и рис. 2) их распределение судя по всему независимо от α и имеет при наших условиях среднее около 12.5 в случае weibull и 11 в случае с exp.

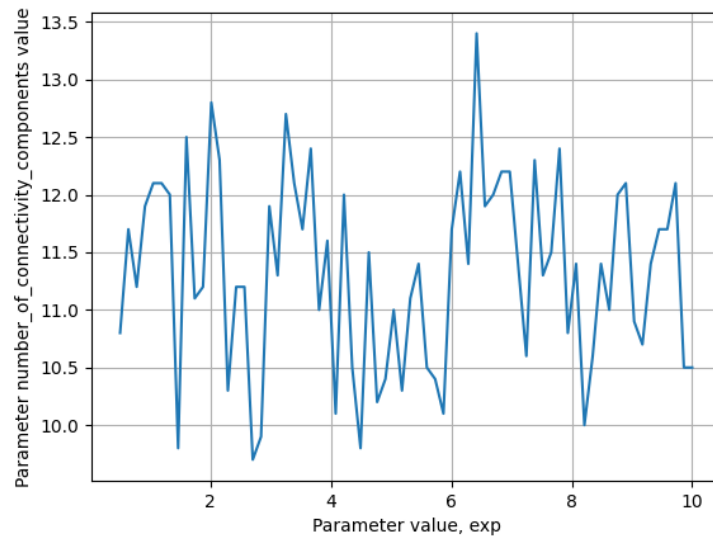


Рисунок 2 – Зависимость числа компонент связности от α при распределении *exp*

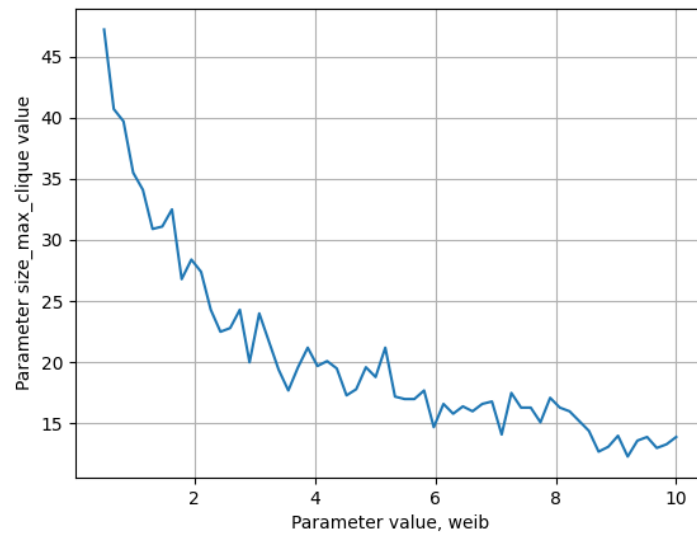


Рисунок 3 – Зависимость максимальной клики от α при распределении *weibull*

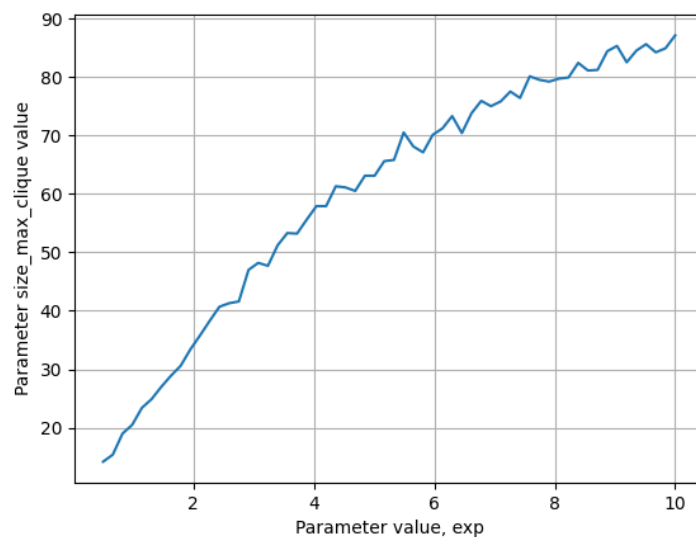


Рисунок 4 – Зависимость максимальной клики от α при распределении exp

В случае же с максимальной кликой видно (рис. 3 и рис. 4), что распределение напоминает степенную функцию, но с совершенно разными степенями. По моим расчетам при наших условиях степень составляет около в случае weibull и $2/5$ в случае с exp.

Пункт 2:

Для проведения эксперимента фиксировалась данное в задании α и значения k проходили от 2 до 20 с шагом 1, значения d проходили от 0.05 до 10 с делением на 60 участков и значения n проходили от 50 до 100 с шагом 2. Усреднение шло по 10 различным значениям аналогично первому пункту.

Тут можно отметить, что от k зависимость степенная с отрицательным коэффициентом в обоих случаях, от d зависимость степенная с коэффициентом меньше 0, а от n зависимость линейная положительная во всех случаях, при том с меньшей дисперсией при подсчете кликового числа.

Пункт 3:

После запуска функции мощность полученного A на выборке размером 300 и с количеством итераций 1000 составило $\text{power} = 0.9919999999999998$, $\text{error} = 0.9999999999999998$ для knn и $\text{power} = 0.5780000000000001$, $\text{error} = 1.0$ для dist. Это говорит о том, что в случае с числом компонент связности принимаемые значения похожи друг на друга и обеих плотностей, а вот кликовое число достаточно разнится, но все равно вероятность ошибиться можно оценить примерно как 50 на 50.

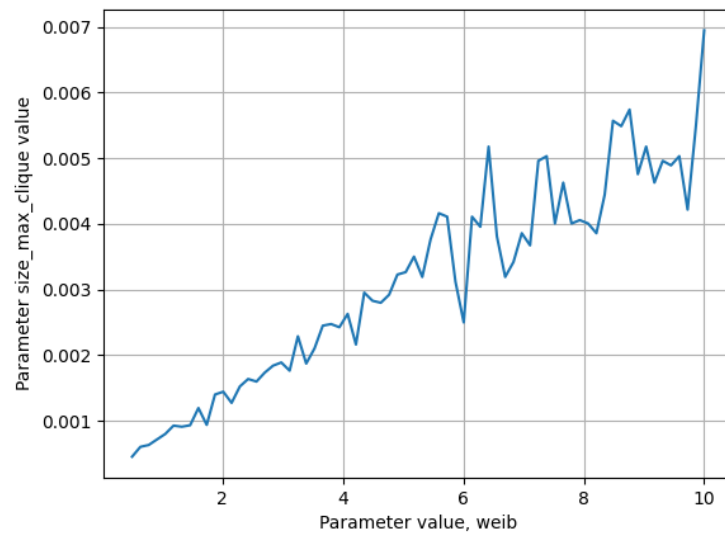


Рисунок 5 – Зависимость максимальной клики от α при распределении weibull после выравнивание возведением в степень

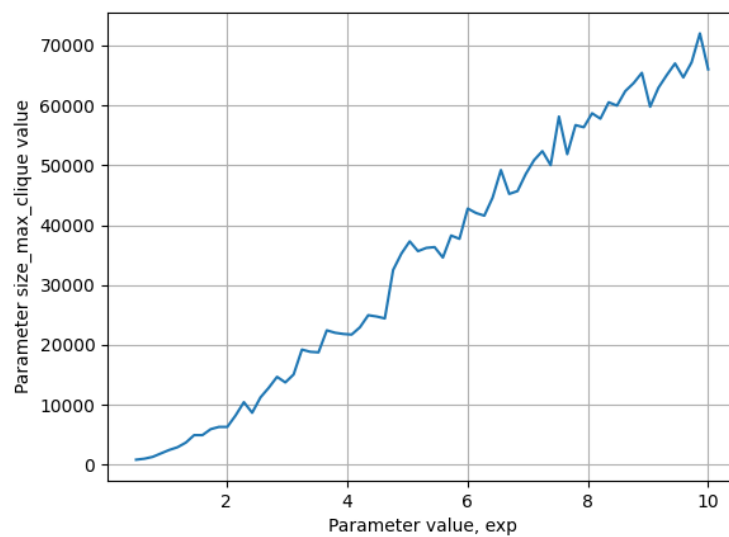


Рисунок 6 – Зависимость максимальной клики от α при распределении exp после выравнивание возведением в степень

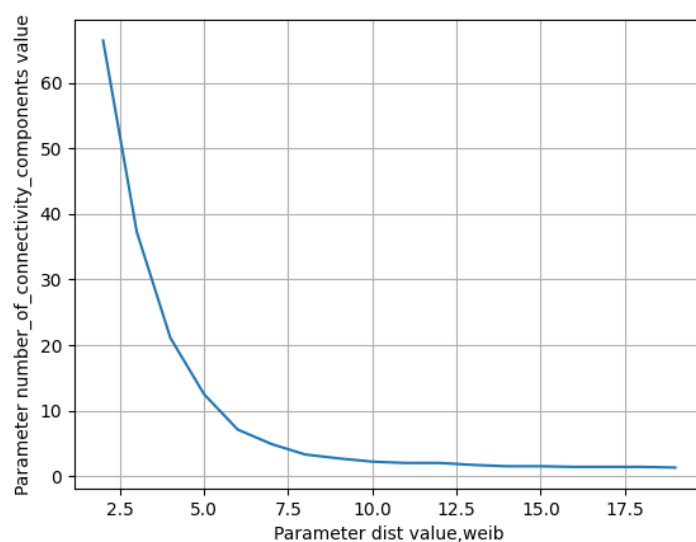


Рисунок 7 – Зависимость числа компонент связности от k при распределении weibull

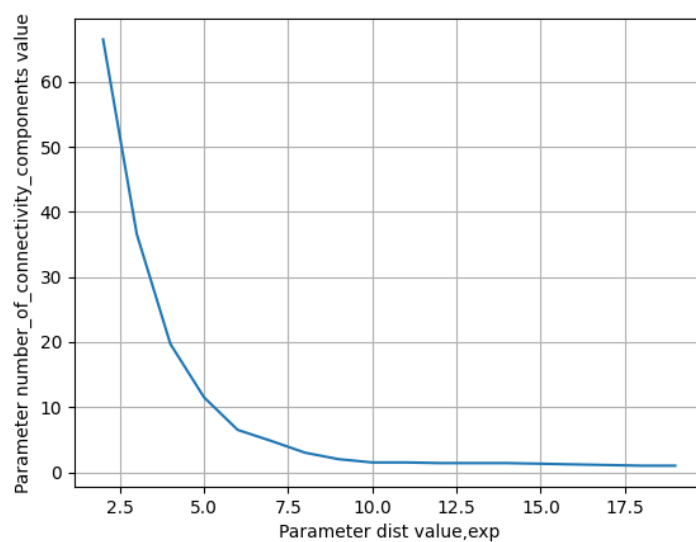


Рисунок 8 – Зависимость числа компонент связности от k при распределении exp

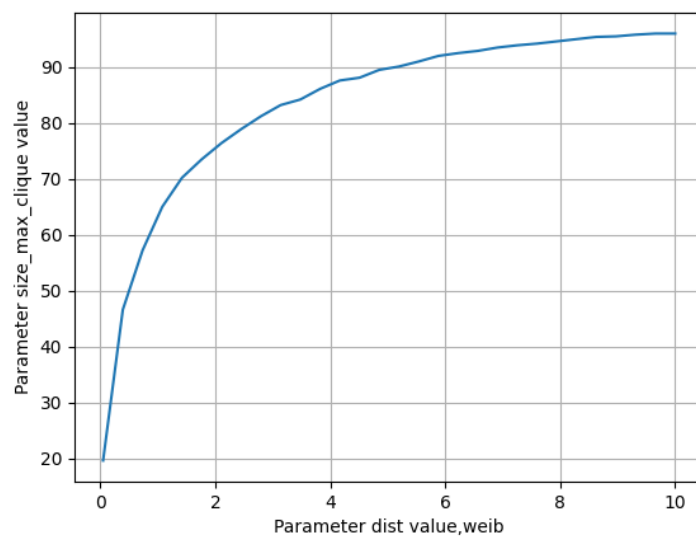


Рисунок 9 – Зависимость размера максимальной клики от d при распределении *weibull*

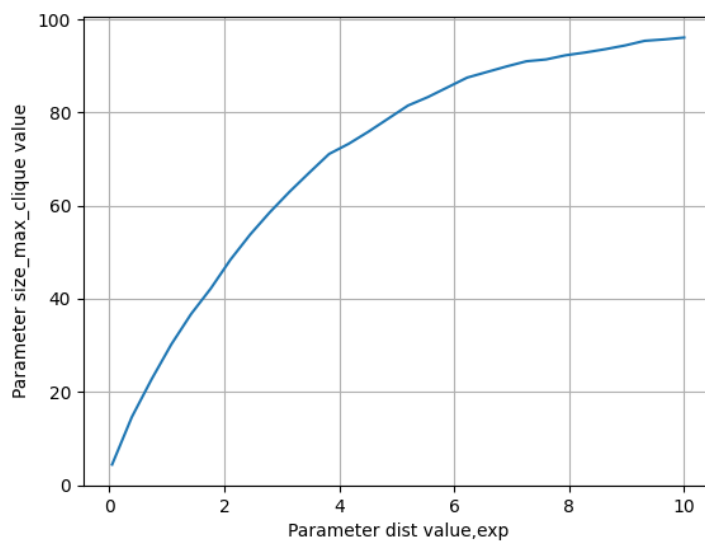


Рисунок 10 – Зависимость размера максимальной клики от d при распределении *exp*

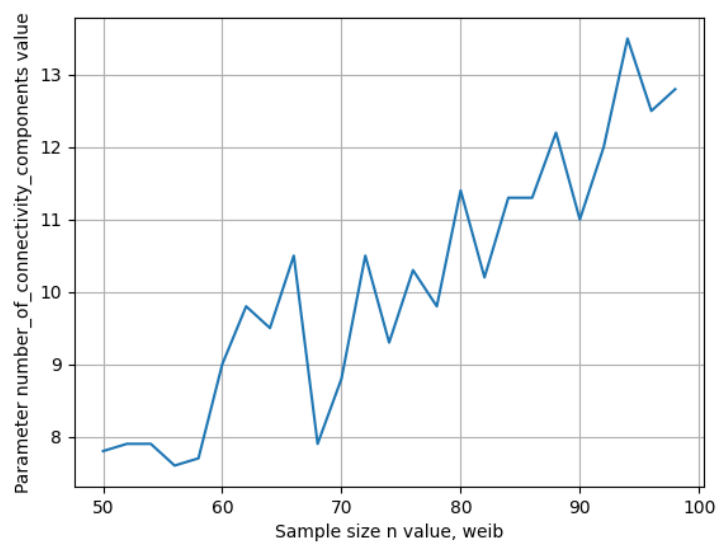


Рисунок 11 – Зависимость числа компонент связности n при распределении *weibull*

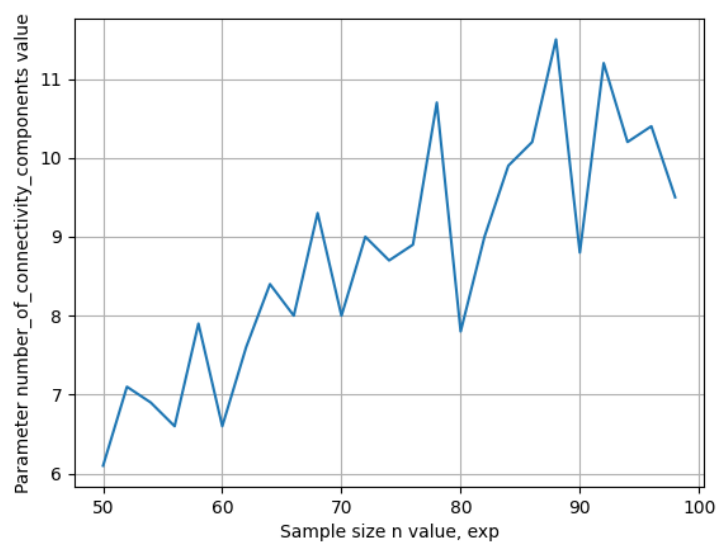


Рисунок 12 – Зависимость числа компонент связности n при распределении *exp*

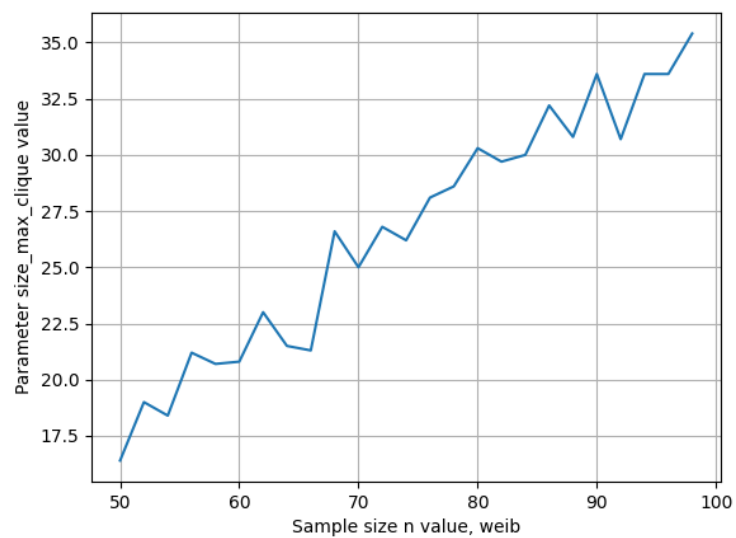


Рисунок 13 – Зависимость размера максимальной клики от n при распределении *weibull*

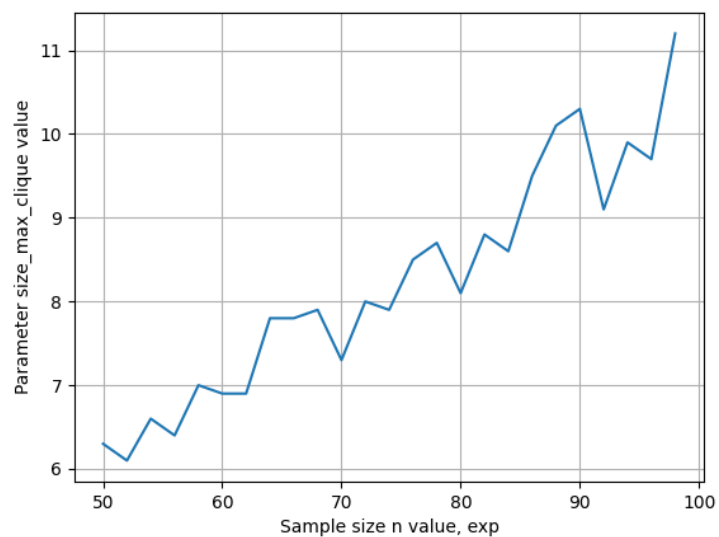


Рисунок 14 – Зависимость размера максимальной клики от n при распределении *exp*