Статистический анализ и визуализация сиракузских последовательностей

Содержание

1 Введение

Гипотеза Коллатца ($3n + 1$ гипотеза, сиракузская проблема) — одна из нерешенных задач теории чисел, с довольно простой формулировкой. Названа в честь немецкого математика Лотара Коллатца, который сформулировал схожую задачу в 1932 году. Несмотря на всю простоту формулировки, задача обладает потрясающими иллюстрациями, построение и исследование которых будут приведены в данной статье.

2 Постановка задачи

Для объяснения сути гипотезы рассмотрим следующую последовательность чисел, называемую сиракузской последовательностью. Берём любое натуральное число n. Если оно четное, то делим его на 2, а если нечетное, то умножаем на 3 и прибавляем 1 (получаем $3n + 1$). Над полученным числом выполняем те же самые действия, и так далее.

Гипотеза Коллатца заключается в том, что какое бы начальное число n мы ни взяли, рано или поздно мы получим единицу.

3 Визуализация и описание поведения сиракузских последовательностей

Для начала определим что такое сиракузская последовательность

Определение 1. *Сиракузской последовательностью будем называть последовательность чисел, полученную действиями умножения числа на 3 с прибавлением единицы, если число нечетное, и делением на 2 если число четное.*

Приведем пример сиракузской последовательности от числа 3:

- 3 нечетное $\Rightarrow 3 \cdot 3 + 1 = 10$
- 10 четное $\Rightarrow \frac{10}{2} = 5$
- 5 нечетное $\Rightarrow 5 \cdot 3 + 1 = 16$
- 16 четное $\Rightarrow \frac{16}{2} = 8$
- 8 четное $\Rightarrow \frac{8}{2} = 4$
- 4 четное $\Rightarrow \frac{4}{2} = 2$
- 2 четное $\Rightarrow \frac{2}{2} = 1$
- 1 нечетное $\Rightarrow 1 \cdot 3 + 1 = 4$

Можем видеть что последовательность пришла в цикл 4, 2, 1, следовательно, для числа 3 гипотеза верна. Полученная последовательность: $\{3, 10, 5, 16, 8, 4, 2, 1\}$

Попробуем изобразить поведение последовательность используя библиотеку Matplotlib для построения графиков на языке Python. Для начала напишем функцию, которая будет принимать некоторое натуральное число и генерировать сиракузскую последовательность, начиная с данного числа.

```
def kollats_hyp(n:int):
    number    = []
    iteration = []
    counter = 0
    number.append(n)
    iteration.append(counter)
    while (n != 1):
        if n % 2 == 0:
            n /= 2
        else:
            n = 3 * n + 1
        counter += 1
        number.append(int(n))
        iteration.append(counter)
    return (iteration, number)
```

В данном случае функция возвращает пару значений, для дальнейшего удобства построения графика.

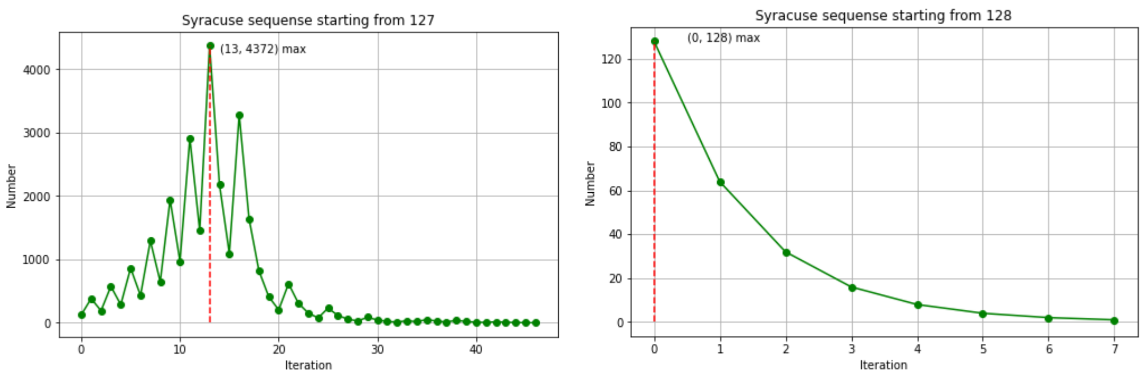

Рис. 1: Поведение сиракузских последовательности от чисел 127 и 128

Последовательности также можно представить как граф, однако он не позволяет рассмотреть конкретное их поведение, позволяет лишь зрительно сравнивать длины путей и наглядно показывает формулировку и экспериментальное подтверждение гипотезы на конкретных числах.

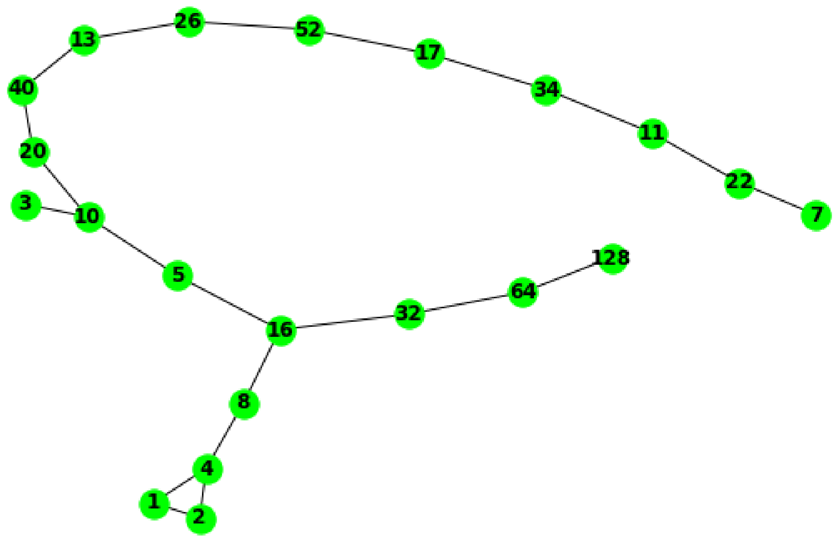

Рис. 2: Граф сиракузских последовательностей от чисел 3, 7 и 256

Графическое представление позволяет рассмотреть неоднозначность поведения последовательности. В данном случае для чисел 127 и 128, отличающихся на единицу, длина сиракузской последовательности отличается более чем в 4 раза. Закономерно возникает рассмотреть зависимость длины сиракузской последовательности от числа, от которого она начинается. Для этого запустим N раз функцию, которая генерирует последовательность, с аргументом от 1 до N. Получим довольно впечатляющую картину.

Можем увидеть и объяснение, почему же некоторые пути такие короткие. Так как действие деления на 2 "приближает" нас к 1, в отличие от действия умножения на 3 с прибавлением единицы, то кратчайшие пути до конечного цикла будут начинаться именно со степени двойки. Так как поделив степень двойки пополам мы получим степень двойки, нетрудно показать, что нижней границей длин является $m = \log_2(n)$ – длина последовательности от 2^m до единицы.

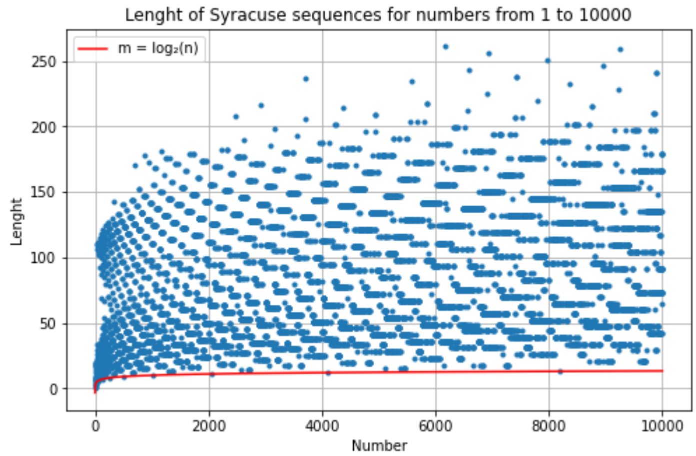

Рис. 3: Длины сиракузских последовательностей для чисел от 1 до 10000

4 Распределение длин сиракузских последовательностей

Рассмотрим поведение длин последовательностей статистически, а именно рассмотрим распределение длин последовательностей для чисел от 1 до N.

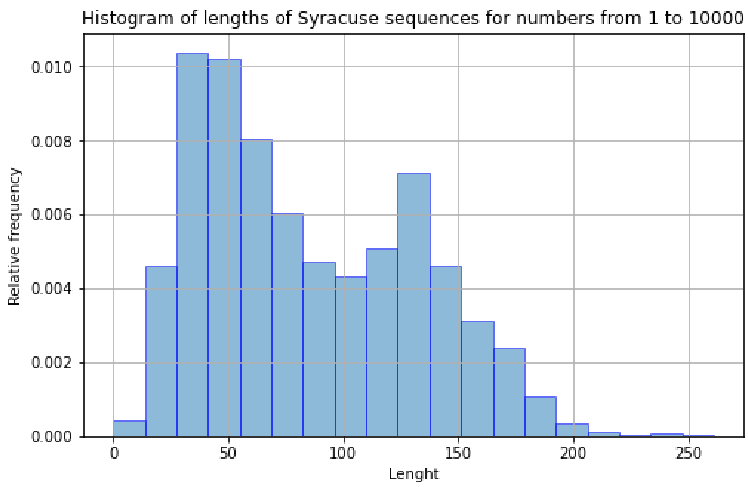

Рис. 4: Гистограмма сиракузских последовательностей для чисел от 1 до 10000

На данный момент распределение имеет непонятную форму, поэтому рассмотрим поведение средней длины последовательности на разных объемах выборки чисел, от которых построены последовательности.

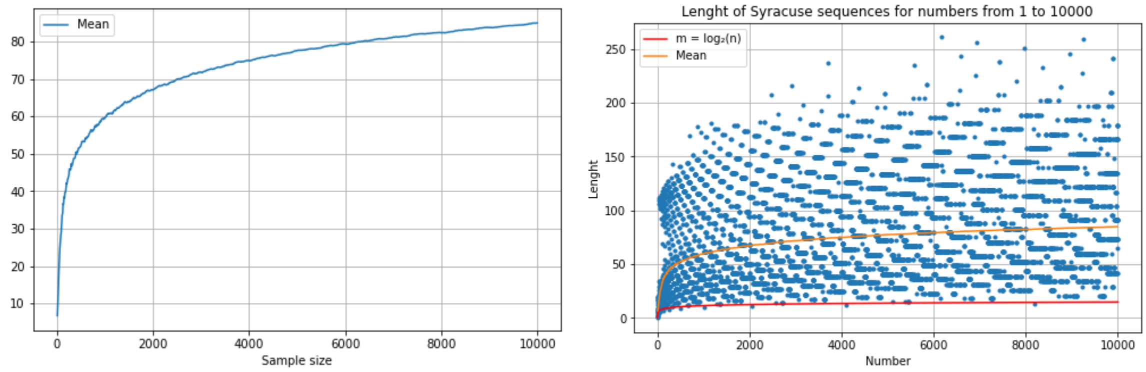

Рис. 5: Поведение среднего значения длины сиракузских последовательностей для чисел от 1 до 10000

Определим вид функции поведения средней длины. Для простоты вычисления уменьшим количество точек, после чего применим регрессию.

Number	1000	4000	9000	16000	25000	36000
Mean	59.4905	74.8092	83.7247	89.3601	93.7652	97.3137

Запишем уравнение регрессии

$$\hat{y} = a \ln x - b$$

$$a = \frac{n \sum\limits_{i=1}^{n} y_i \ln x_i - \sum\limits_{i=1}^{n} \ln x_i \cdot \sum\limits_{i=1}^{n} y_i}{n \sum\limits_{i=1}^{n} (\ln x_i)^2 - \left(\sum\limits_{i=1}^{n} (\ln x_i) \right)^2}$$

$$b = \frac{1}{n} \sum_{i=1}^{n} y_i - \frac{a}{n} \sum_{i=1}^{n} \ln x_i$$

где y - средняя длина последовательности, $\hat{y}$ - оценка средней длины последовательности, а x - число, до которого проверены последовательности.

Для приведенного выше набора данных, аппроксимация функции имеет вид

$$y = 10.3639 \cdot \ln x - 11.3936$$

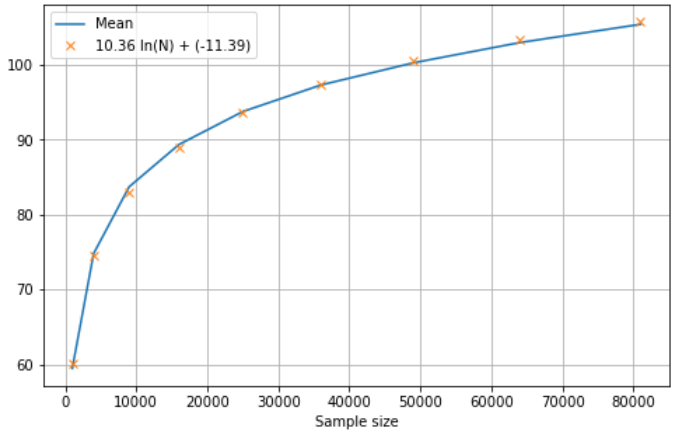

Рис. 6: Аппроксимация функции средней длины

Проверив функцию на большем объеме данных, коэффициенты приняли следующие значения

$$y = 10.3888 \cdot \ln x - 11.7752$$

функция не сильно отличается от предыдущей, следовательно, она посчитана довольно точно.

5 Заключение

Визуализация тех или иных процессов позволяет по новому взглянуть на различные задачи. На примере гипотезы Коллатца показано, что статистический анализ может применяться не только к процессам, имеющим случайную природу, но и для детерминированных. Исходя из статистики можно сделать множество предположений, которые в последствии стоит попытаться доказать. Например, при сохранении на множестве натуральных чисел закона поведения средней длины сиракузской последовательности, имеющего вид $y = \alpha \cdot \ln x - \beta$, $\alpha, \beta \in \mathbb{Z}$, $x, y \in \mathbb{N}$, можем предположить, что все сиракузские последовательности конечны. Доказательство данного факта подтверждает половину гипотезы.