



# Статически оптимальное дерево отрезков

Артем Комендантян, 4 курс МФТИ, кафедра анализа данных

**Научный руководитель:** Михаил Тихомиров, к.ф.-м.н.



# Определения

Пусть дан массив из  $n$  элементов

**Дерево отрезков** - это двоичное дерево, в котором есть

- $n$  листьев, соответствующих отрезкам единичной длины
- Вершины с двумя сыновьями. Правый сын соответствует отрезку, следующему сразу за отрезком левого сына. Вершина соответствует объединению отрезков сыновей

Корень дерева соответствует всему массиву (отрезку  $[1; n]$ )



## Определения

**Запрос сверху** на отрезке  $[L, R]$  начинается в корне.

Если сейчас рассматривается вершина, отрезок которой не лежит полностью в отрезке  $[L, R]$ , то запрос рекурсивно вызывается от тех сыновей, отрезки которых пересекаются с  $[L, R]$ . Иначе рекурсивных вызовов от сыновей не происходит.

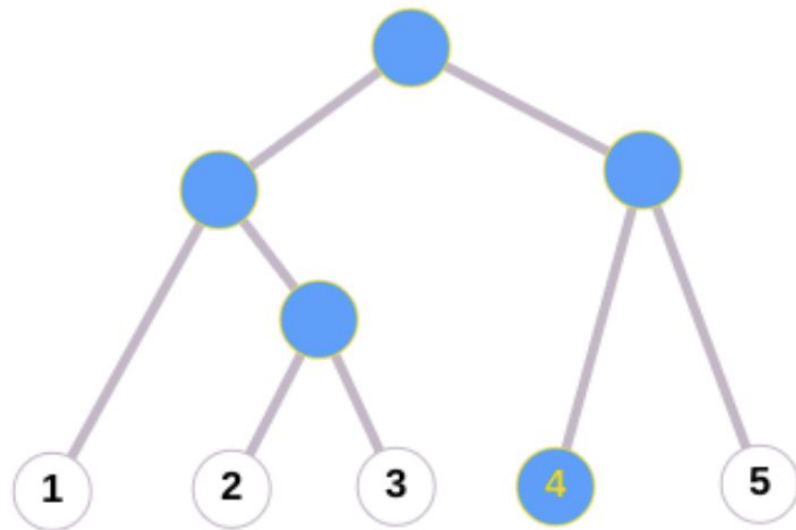
В обоих случаях вершина считается посещенной и в ней выполняются какие-то действия, специфичные для запроса.

# Пример

Дерево отрезков на пяти элементах

Запрос на отрезке  $[2; 4]$

Будет посещено пять выделенных вершин





## Постановка задачи

**Дано:** Распределение вероятностей на запросах-отрезках с границами из  $[1; n]$

**Необходимо:** построить дерево отрезков, для которого минимально среднее количество посещенных вершин при запросах сверху.

Интересует как точное решение за как можно более лучшую асимптотику, так и приближенное за сложность нахождения  $O(n + S)$  или  $O(n + S \log S)$ , где  $S$  - количество отрезков с ненулевой вероятностью



# Мотивация

- Дерево отрезков - мощная структура, позволяющая эффективно решать множество задач. Примеры: наименьший общий предок двух вершин в дереве, площадь объединения прямоугольников со сторонами, параллельными осям координат
- Потенциальное ускорение в реальных задачах
- Аналогичная задача для деревьев поиска хорошо изучена



## Статически оптимальное двоичное дерево поиска

**Дано:** множество из  $n$  упорядоченных элементов и  $2n + 1$  вероятностей, что запрос будет равен элементу из множества, либо находиться между двух соседних элементов

**Необходимо:** построить двоичное дерево поиска, для которого минимально среднее количество посещенных вершин при запросах



## Статически оптимальное двоичное дерево поиска

**Точное решение:** динамическое программирование за  $O(n^2)$  времени и памяти (Knuth, 1971)

**Приближенное решение:** жадное решение за  $O(n)$ , хуже оптимального не более чем в константу раз, где константа примерно равна 2.283... (Mehlhorn, 1975)

Идея решения - выбирать такой корень, чтобы максимально уравнивать сумму вероятностей в левом и правом поддереве, затем перейти к двум меньшим подзадачам





## Текущий прогресс и планы

**Точное решение:** динамическое программирование по подотрезкам  
 $dp[L][R]$  - стоимость оптимального дерева отрезков, построенного на точках  
с  $L$  по  $R$ . Считается за  $O(n^3)$  времени и  $O(n^2)$  памяти

Хочется научиться решать с лучшей асимптотикой



## Текущий прогресс и планы

**Приближенное решение:** количество вершин при запросе  $[L, R]$  можно приблизить суммой глубин запросов  $[1, R]$  и  $[L, n]$ . Новая задача сводится к построению оптимального дерева поиска. Решение хуже оптимального не более чем в  $4 \cdot$  константу от приближения дерева поиска (оценка сверху) и работает за  $O(n + S)$

Это решение асимптотически оптимально, но хочется уменьшить константу приближения



**Спасибо за внимание!**