

# Выбор pivot при помощи алгоритма медианы медиан. Время работы быстрой сортировки и поиска $k$ -ой порядковой статистики при такой стратегии выбора pivot

ceagest

## 1 Медиана медиан

### Алгоритм (Блума-Флойда-Прата-Ривеста-Тарьяна)

- Разобьем массив на пятерки элементов.
- В каждой пятерке выберем медиану. Получили массив медиан.
- Для массива медиан ищем медиану (с помощью алгоритма поиска  $k$ -порядковой статистики), и используем ее как опорный элемент.
- Обратите внимание, что в алгоритме используется взаимная рекурсия — поиск медианы в массиве медиан, тоже будет использовать медиану медиан.

**Лемма 1.1.** Медиана медиан гарантированно делит массив в соотношении не хуже, чем  $3 : 7$ .

*Доказательство.* Сначала определим нижнюю границу для количества элементов, превышающих по величине опорный элемент  $x$ . В общем случае как минимум половина медиан, найденных на втором шаге, больше или равны медиане медиан  $x$ . Таким образом, как минимум  $\frac{n}{10}$  групп содержат по 3 элемента, превышающих величину  $x$ , за исключением группы, в которой меньше 5 элементов и ещё одной группы, содержащей сам элемент  $x$ . Таким образом получаем, что количество элементов больших  $x$  не менее  $\frac{3n}{10}$ .  $\square$

**Лемма 1.2.** Алгоритм нахождения  $k$ -ой порядковой статистики с использованием медианы медиан работает за линейное время.

*Доказательство.* У нас есть три составляющих работы алгоритма на каждом шаге:

1. Время на разделение массива на пятерки и их сортировка (+ Partition по медиане медиан):  $An$ , где  $A \in \mathbb{R}$  — некоторая константа.
2. Время на поиск медианы медиан  $T\left(\frac{n}{5}\right)$ .
3. Время на поиск  $k$ -ой порядковой не превзойдет времени его поиска в большей доле, то есть  $T\left(\frac{7n}{10}\right)$

Таким образом:

$$T(n) \leq T\left(\frac{n}{5}\right) + T\left(\frac{7n}{10}\right) + An$$

Пусть  $T(n) \leq Cn$  для некоторой константы  $C \in \mathbb{R}$  и  $\forall n \in \mathbb{N} : n \leq 140$ . Подставим это соотношение и получим, что:

$$T(n) \leq T\left(\frac{n}{5}\right) + T\left(\frac{7n}{10}\right) + An \leq \frac{Cn}{5} + \frac{7Cn}{10} + An = \frac{9Cn}{10} + An = Cn + \left(-\frac{Cn}{10} + An\right)$$

Тогда неравенство верно, если взять  $C \geq 10A$ . Поскольку такое  $C$  действительно можно выбрать, медиана медиан работает за линейное время.  $\square$

**Лемма 1.3.** Если использовать медиану медиан в качестве опорного элемента в быстрой сортировке, время работы сортировки будет  $\mathcal{O}(n \log n)$  в худшем случае.