«Разделяй и властвуй»: порядковые статистики

Александр Куликов

Онлайн-курс «Алгоритмы: теория и практика. Методы»
http://stepic.org/217

Постановка задачи

k-я порядковая статистика

Вход: массив $A[1 \dots n]$.

Выход: k-й элемент упорядоченного по неубыванию массива (то есть $A'[k]$).

Линейное в среднем время

Функция $\textsc{RandomSelect}(A, \ell, r, k)$

если $\ell \geq r$: вернуть $A[\ell]$

выбрать случайный элемент x из $A[\ell \dots r]$
разбить $A[\ell \dots r]$ на
  $A[\ell \dots m_1]$, $A[m_1 + 1 \dots m_2]$, $A[m_2 + 1 \dots r]$
  ($< x$, $= x$, $> x$ соответственно)
если $\ell \leq k \leq m_1$:
  вернуть $\textsc{RandomSelect}(A, \ell, m_1, k)$
иначе если $m_1 + 1 \leq k \leq m_2$:
  вернуть x
иначе:
  вернуть $\textsc{RandomSelect}(A, m_2 + 1, r, k)$

Вспомогательная лемма

Лемма

Математическое ожидание количества подбрасываний монетки до первой решки (включительно) равно 2.

Доказательство

- $E = \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot (1 + E)$ и, следовательно, $E = 2$
- или по определению:

$$E = \sum_{i=1}^{\infty} i \cdot \frac{1}{2^i} = 2$$

$\square$

Лемма

Среднее время работы алгоритма RandomSelect есть $O(n)$.

Доказательство

- Пусть $T(n)$ — среднее время работы.

Лемма

Среднее время работы алгоритма RandomSelect есть $O(n)$.

Доказательство

- Пусть $T(n)$ — среднее время работы.
- Тогда $T(n) \leq T(3n/4) + O(n)$:

$$
\begin{aligned}
&\text{время работы для массива размера } n \\
&\quad \leq (\text{время работы для массива размера } 3n/4) \\
&+(\text{время на уменьшение размера массива до} \leq 3n/4)
\end{aligned}
$$

Лемма

Среднее время работы алгоритма RANDOMSELECT есть $O(n)$.

Доказательство

- Пусть $T(n)$ — среднее время работы.
- Тогда $T(n) \leq T(3n/4) + O(n)$:

время работы для массива размера n

$\leq$ (время работы для массива размера $3n/4$)

$+$(время на уменьшение размера массива до $\leq 3n/4$)

- Значит, $T(n) = O(n)$ (убывающая геометрическая прогрессия). □

Факт

Существует алгоритм, находящий произвольную порядковую статистику за линейное в худшем случае время.