Analysis of Selection Algorithm using Partition

ESO207

Indian Institute of Technology, Kanpur

The variables

- Consider a call to RAND-SELECT$(A, 1, n, k)$: find the kth smallest element in $A[1 \ldots n]$.
- Goal: Show a linear time bound on the expected (average) cost of Rand-Select.
- Source of randomness: The pivot may be any one of the n elements of $A[1 \ldots n]$ with equal chance.
- Let $q =$ index returned by call to PARTITION$(A, 1, n)$.
- For each $k = 1, 2, \ldots, n-1$, let

$$X_k = \begin{cases} 1 & \text{if the subarray } A[p \ldots q] \text{ has exactly } k \text{ elements} \\ 0 & \text{otherwise.} \end{cases}$$

- Hence,

$$\mathbb{E}[X_k] = 1 \cdot \Pr\{X_k = 1\} + 0 \cdot \Pr\{X_k = 0\} = 1/n$$

The variables

- Consider a call to RAND-SELECT$(A, 1, n, k)$: find the kth smallest element in $A[1 \ldots n]$.
- Goal: Show a linear time bound on the expected (average) cost of Rand-Select.
- Source of randomness: The pivot may be any one of the n elements of $A[1 \ldots n]$ with equal chance.
- Let $q =$ index returned by call to PARTITION$(A, 1, n)$.
- For each $k = 1, 2, \ldots, n-1$, let

$$X_k = \begin{cases} 1 & \text{if the subarray } A[p \ldots q] \text{ has exactly } k \text{ elements} \\ 0 & \text{otherwise.} \end{cases}$$

- Hence,

$$\mathbb{E}[X_k] = 1 \cdot \Pr\{X_k = 1\} + 0 \cdot \Pr\{X_k = 0\} = 1/n$$

The variables

- Consider a call to RAND-SELECT$(A, 1, n, k)$: find the kth smallest element in $A[1 \dots n]$.
- Goal: Show a linear time bound on the expected (average) cost of Rand-Select.
- Source of randomness: The pivot may be any one of the n elements of $A[1 \dots n]$ with equal chance.
- Let $q =$ index returned by call to PARTITION$(A, 1, n)$.
- For each $k = 1, 2, \dots, n-1$, let

$$X_k = \begin{cases} 1 & \text{if the subarray } A[p \dots q] \text{ has exactly } k \text{ elements} \\ 0 & \text{otherwise.} \end{cases}$$

- Hence,

$$\mathbb{E}[X_k] = 1 \cdot \Pr\{X_k = 1\} + 0 \cdot \Pr\{X_k = 0\} = 1/n$$

The variables

- Consider a call to RAND-SELECT$(A, 1, n, k)$: find the kth smallest element in $A[1 \dots n]$.
- Goal: Show a linear time bound on the expected (average) cost of Rand-Select.
- Source of randomness: The pivot may be any one of the n elements of $A[1 \dots n]$ with equal chance.
- Let $q =$ index returned by call to PARTITION$(A, 1, n)$.
- For each $k = 1, 2, \dots, n-1$, let

$$X_k = \begin{cases} 1 & \text{if the subarray } A[p \dots q] \text{ has exactly } k \text{ elements} \\ 0 & \text{otherwise.} \end{cases}$$

- Hence,

$$\mathbb{E}[X_k] = 1 \cdot \Pr\{X_k = 1\} + 0 \cdot \Pr\{X_k = 0\} = 1/n$$

The variables

- Consider a call to RAND-SELECT$(A, 1, n, k)$: find the kth smallest element in $A[1 \dots n]$.
- Goal: Show a linear time bound on the expected (average) cost of Rand-Select.
- Source of randomness: The pivot may be any one of the n elements of $A[1 \dots n]$ with equal chance.
- Let $q =$ index returned by call to PARTITION$(A, 1, n)$.
- For each $k = 1, 2, \dots, n-1$, let

$$X_k = \begin{cases} 1 & \text{if the subarray } A[p \dots q] \text{ has exactly } k \text{ elements} \\ 0 & \text{otherwise.} \end{cases}$$

- Hence,

$$\mathbb{E}[X_k] = 1 \cdot \Pr\{X_k = 1\} + 0 \cdot \Pr\{X_k = 0\} = 1/n$$

The variables

- Consider a call to RAND-SELECT$(A, 1, n, k)$: find the kth smallest element in $A[1 \dots n]$.
- Goal: Show a linear time bound on the expected (average) cost of Rand-Select.
- Source of randomness: The pivot may be any one of the n elements of $A[1 \dots n]$ with equal chance.
- Let $q =$ index returned by call to PARTITION$(A, 1, n)$.
- For each $k = 1, 2, \dots, n-1$, let

$$X_k = \begin{cases} 1 & \text{if the subarray } A[p \dots q] \text{ has exactly } k \text{ elements} \\ 0 & \text{otherwise.} \end{cases}$$

- Hence,

$$\mathbb{E}[X_k] = 1 \cdot \Pr\{X_k = 1\} + 0 \cdot \Pr\{X_k = 0\} = 1/n$$

Notation

- Let $T(n)$ be a random variable that denotes the running time of RAND-SELECT$(A, 1, n, i)$, that is the time taken to perform any selection from an array of size n.
- We are interested in the expected value $\mathbb{E}[T(n)]$.
- When we call RAND-SELECT$(A, 1, n, i)$, we do not know if the recursion will terminate immediately (i.e., $k = q$, or will recurse on the left part $(A, 1, q-1, i)$ or on the right part $(A, q+1, n, i-q)$.
- We assume that $T(n)$ is monotonic in n. Upper bound $T(n)$ by assuming that we recurse over the larger of the two partitions.
- That is, we assume that the ith element is always on the side of the larger partition.

Notation

- Let $T(n)$ be a random variable that denotes the running time of RAND-SELECT$(A, 1, n, i)$, that is the time taken to perform any selection from an array of size n.
- We are interested in the expected value $\mathbb{E}[T(n)]$.
- When we call RAND-SELECT$(A, 1, n, i)$, we do not know if the recursion will terminate immediately (i.e., $k = q$, or will recurse on the left part $(A, 1, q - 1, i)$ or on the right part $(A, q + 1, n, i - q)$.
- We assume that $T(n)$ is monotonic in n. Upper bound $T(n)$ by assuming that we recurse over the larger of the two partitions.
- That is, we assume that the ith element is always on the side of the larger partition.

Notation

- Let $T(n)$ be a random variable that denotes the running time of RAND-SELECT$(A, 1, n, i)$, that is the time taken to perform any selection from an array of size n.
- We are interested in the expected value $\mathbb{E}[T(n)]$.
- When we call RAND-SELECT$(A, 1, n, i)$, we do not know if the recursion will terminate immediately (i.e., $k = q$, or will recurse on the left part $(A, 1, q - 1, i)$ or on the right part $(A, q + 1, n, i - q)$.
- We assume that $T(n)$ is monotonic in n. Upper bound $T(n)$ by assuming that we recurse over the larger of the two partitions.
- That is, we assume that the ith element is always on the side of the larger partition.

Notation

- Let $T(n)$ be a random variable that denotes the running time of $\text{RAND-SELECT}(A, 1, n, i)$, that is the time taken to perform any selection from an array of size n.
- We are interested in the expected value $\mathbb{E}[T(n)]$.
- When we call $\text{RAND-SELECT}(A, 1, n, i)$, we do not know if the recursion will terminate immediately (i.e., $k = q$, or will recurse on the left part $(A, 1, q-1, i)$ or on the right part $(A, q+1, n, i-q)$.
- We assume that $T(n)$ is monotonic in n. Upper bound $T(n)$ by assuming that we recurse over the larger of the two partitions.
- That is, we assume that the ith element is always on the side of the larger partition.

Basic Recurrence

- Recall $X_k = 1$ iff the left partition is of size $k - 1$.
- When $X_k = 1$, the left partition is of size k and right partition is of size $n - k$.
- Recall: For each $k = 1, 2, \ldots, n - 1$,

$$X_k = \begin{cases} 1 & \text{if the subarray } A[p \ldots q] \text{ has exactly } k \text{ elements} \\ 0 & \text{otherwise.} \end{cases}$$

- This gives the recurrence:

$$\begin{aligned} T(n) &\leq \sum_{k=1}^{n} X_k \cdot (T(\max(k-1, n-k)) + O(n)) \\ &= \sum_{k=1}^{n} X_k \cdot \max(T(k-1), T(n-k)) + O(n) \end{aligned}$$

since, $\sum_{k=1}^{n} X_k = 1$.

Basic Recurrence

- Recall $X_k = 1$ iff the left partition is of size $k-1$.
- When $X_k = 1$, the left partition is of size k and right partition is of size $n-k$.
- Recall: For each $k = 1, 2, \ldots, n-1$,

$$X_k = \begin{cases} 1 & \text{if the subarray } A[p \ldots q] \text{ has exactly } k \text{ elements} \\ 0 & \text{otherwise.} \end{cases}$$

- This gives the recurrence:

$$\begin{aligned} T(n) &\leq \sum_{k=1}^{n} X_k \cdot (T(\max(k-1, n-k)) + O(n)) \\ &= \sum_{k=1}^{n} X_k \cdot \max(T(k-1), T(n-k)) + O(n) \end{aligned}$$

since, $\sum_{k=1}^{n} X_k = 1$.

Basic Recurrence

- Recall $X_k = 1$ iff the left partition is of size $k - 1$.
- When $X_k = 1$, the left partition is of size k and right partition is of size $n - k$.
- Recall: For each $k = 1, 2, \ldots, n - 1$,

$$X_k = \begin{cases} 1 & \text{if the subarray } A[p \ldots q] \text{ has exactly } k \text{ elements} \\ 0 & \text{otherwise.} \end{cases}$$

- This gives the recurrence:

$$\begin{aligned} T(n) &\leq \sum_{k=1}^{n} X_k \cdot (T(\max(k-1, n-k)) + O(n)) \\ &= \sum_{k=1}^{n} X_k \cdot \max(T(k-1), T(n-k)) + O(n) \end{aligned}$$

since, $\sum_{k=1}^{n} X_k = 1$.

Taking Expectation

$$
\begin{aligned}
&\mathbb{E}\left[T(n)\right] \\
&\leq \mathbb{E}\left[\sum_{k=1}^{n} X_k \cdot \left(T(\max(k-1, n-k))\right) + O(n)\right] \\
&= \sum_{k=1}^{n} \mathbb{E}\left[X_k \cdot T(\max(k-1, n-k))\right] + O(n), \text{ by linearity of expectation} \\
&= \sum_{k=1}^{n} \mathbb{E}\left[X_k\right] \mathbb{E}\left[T(\max(k-1, n-k))\right] + O(n), \text{ by independence} \\
&= \sum_{k=1}^{n} \frac{1}{n} \mathbb{E}\left[T(\max(k-1, n-k))\right] + O(n)
\end{aligned}
$$

We have used X_k to be independent of $T(\max(k-1, n-k))$

Analysis contd.

- Clearly,

$$\max(k-1, n-k) = \begin{cases} k-1 & \text{if } k > \lceil n/2 \rceil \\ n-k & \text{if } k \leq \lceil n/2 \rceil \end{cases}$$

- Recall expression:
  $\mathbb{E}\left[T(n)\right] = \sum_{k=1}^{n} \frac{1}{n} \mathbb{E}\left[T(\max(k-1, n-k))\right] + O(n)$.
- If n is even, then each of the terms $T(\lceil n/2 \rceil) \ldots T(n-1)$ appears exactly twice in the expression.
- If n is odd, then the same holds, and in addition, $T(\lfloor n/2 \rfloor)$ appears once.
- We get

$$\mathbb{E}\left[T(n)\right] \leq \frac{2}{n} \sum_{k=\lfloor n/2 \rfloor}^{n-1} \mathbb{E}\left[T(k)\right] + O(n)$$

Recurrence Equation

- The general recurrence equation we get is

$$\mathbb{E}[T(n)] \leq \frac{2}{n} \sum_{k=\lfloor n/2 \rfloor}^{n-1} \mathbb{E}[T(k)] + an$$

- Solve it by substitution: Let $\mathbb{E}[T(n)] \leq 6an$. Then,

$$\begin{aligned}
&\frac{2}{n} \sum_{k=\lfloor n/2 \rfloor}^{n-1} \mathbb{E}[T(k)] + an \\
&= \frac{2}{n} \left(\sum_{k=1}^{n-1} 6ak - \sum_{k=1}^{\lfloor n/2 \rfloor - 1} 6ak \right) + an \\
&\leq \frac{2}{n} \left(\frac{6an(n-1)}{2} - \frac{6a(n/2-1)(n/2-2)}{2} \right) + an \\
&= \frac{2a}{n} \left(3n(n-1) - 3(n/2-1)(n/2-2)\right) + an
\end{aligned}$$

Analysis contd.

$$
\begin{aligned}
&\frac{2}{n}\sum_{k=\lfloor n/2 \rfloor}^{n-1} \mathbb{E}\left[T(k)\right] + an \\
&\leq \frac{2a}{n}\left(3n(n-1) - 3(n/2-1)(n/2-2)\right) + an \\
&\leq \frac{2a}{n}\left(2.25n^2 + 1.5n - 6\right) + an \\
&= 4.5an + 3a - \frac{12a}{n} + an \\
&\leq 5.5an + 3a \\
&\leq 6an, \quad \text{for } n \geq 6.
\end{aligned}
$$

Hence it is shown that $\mathbb{E}\left[T(n)\right] \leq 6an$, or that $\mathbb{E}\left[T(n)\right] = O(n)$.