Short Notes: Analysis of Randomized Selection Algorithm

Palash Dey
Indian Institute of Technology, Kharagpur
palash.dey@cse.iitkgp.ac.in

In the selection problem, the input is an integer array $\mathcal{A}[1,\ldots,n]$ and an integer $k \in [n]$, and the goal is to compute the k-th smallest number of $\mathcal{A}[1,\ldots,n]$. We will prove that the expected number of comparisons that the randomized selection algorithm performs on any input array of size n is $\mathcal{O}(n)$ irrespective of the value of k.

Let ALG is the random variable denoting the number of comparisons and $\mathcal{B}[1,\ldots,n]$ be the sorted array $\mathcal{A}[1,\ldots,n]$. For every i, j with $1 \leqslant i < j \leqslant n$, let $X_{i,j}$ be the indicator random variable for the event that $\mathcal{B}[i]$ and $\mathcal{B}[j]$ are compared by the algorithm. Then we have

$$\begin{aligned}
\text{ALG} &= \sum_{1\leqslant i<j\leqslant n} X_{i,j} \\
\Rightarrow \mathbb{E}[\text{ALG}] &= \sum_{1\leqslant i<j\leqslant n} \mathbb{E}[X_{i,j}] && \text{[Linearity of expectation]} \\
&= \sum_{1\leqslant i<j\leqslant n} \mathbb{P}[\mathcal{B}[i] \text{ and } \mathcal{B}[j] \text{ are compared}]
\end{aligned}$$

We now consider the following cases.

Case I: Suppose we have $k < i < j$. In this case, we have

$$\begin{aligned}
\mathbb{P}[\mathcal{B}[i] \text{ and } \mathcal{B}[j] \text{ are compared}] &= \mathbb{P}[\text{The first pivot chosen from } \mathcal{B}[k,\ldots,j] \text{ is either } \mathcal{B}[i] \text{ or } \mathcal{B}[j]] \\
&= \frac{2}{j-k+1}
\end{aligned}$$

Case II: Suppose we have $i \leqslant k < j$. In this case, we have

$$\begin{aligned}
\mathbb{P}[\mathcal{B}[i] \text{ and } \mathcal{B}[j] \text{ are compared}] &= \mathbb{P}[\text{The first pivot chosen from } \mathcal{B}[i,\ldots,j] \text{ is either } \mathcal{B}[i] \text{ or } \mathcal{B}[j]] \\
&= \frac{2}{j-i+1}
\end{aligned}$$

Case III: Suppose we have $i < j \leqslant k$. In this case, we have

$$\begin{aligned}
\mathbb{P}[\mathcal{B}[i] \text{ and } \mathcal{B}[j] \text{ are compared}] &= \mathbb{P}[\text{The first pivot chosen from } \mathcal{B}[i,\ldots,k] \text{ is either } \mathcal{B}[i] \text{ or } \mathcal{B}[j]] \\
&= \frac{2}{k-i+1}
\end{aligned}$$

Hence, we have

$$\mathbb{E}[\text{ALG}] = \sum_{1\leqslant k<i<j\leqslant n} \frac{2}{j-k+1} + \sum_{1\leqslant i\leqslant k<j\leqslant n} \frac{2}{j-i+1} + \sum_{1\leqslant i<j\leqslant k} \frac{2}{k-i+1}$$

Statutory warning: This is a draft version and may contain errors. If you find any error, please send an email to the author.

$$
\begin{aligned}
&= \sum_{j=2}^{n}(j-k+1)\frac{2}{j-k+1} + \sum_{\ell=1}^{n-1}\sum_{1\leqslant i\leqslant k<j\leqslant n:j-i=\ell}\frac{2}{\ell+1} + \sum_{i=1}^{k-1}(k-i+1)\frac{2}{k-i+1} \\
&\leqslant n + \sum_{\ell=1}^{n-1}\min\{k,\ell\}\frac{2}{\ell+1} + n \\
&\leqslant 3n
\end{aligned}
$$