University of Bristol

COMS21103: Data Structures and Algorithms

Problem Set 8

Problem 1: Show that Randomised-Quicksort's expected runtime is $\Omega(n \log n)$.

Solution: The analysis for lower bound of Randomised-Quicksort is basically the same as the upper bound analysis, but we need to lower bound the Harmonic number H_i, instead of upper bounding H_i. $\square$

Problem 2: An alternative analysis of the running time of Randomised-Quicksort focuses on the expected running time of each individual recursive call to Quicksort, rather than on the number of comparisons performed.

1. Argue that, give an array of size n, the probability that any particular element is chosen as the pivot is $1/n$. Use this to define indicator random variables

   $$X_i = \mathbf{1}\{i\text{th smallest element is chosen as the pivot}\}.$$

   What is $\mathbb{E}[X_i]$?

2. Let $T(n)$ be a random variable denoting the running time of quicksort on an array of size n. Argue that

   $$\mathbb{E}[T(n)] = \mathbb{E}\left[\sum_{q=1}^{n} X_q(T(q-1) + T(n-q) + \Theta(n))\right].$$

3. Show that the equation above simplifies to

   $$\mathbb{E}[T(n)] = \frac{2}{n}\sum_{q=0}^{n-1} \mathbb{E}[T(q)] + \Theta(n). \tag{1}$$

4. Show that

   $$\sum_{k=1}^{n-1} k \lg k \leq \frac{1}{2}n^2 \lg n - \frac{1}{8}n^2, \tag{2}$$

   for large enough n.

5. Using the bound from (2), show that the recurrence in equation (1) has the solution $\mathbb{E}[T(n)] = \Theta(n \lg n)$.

Solution: (1) By definition, we have that

$$\mathbb{E}[X_i] = 1 \cdot \mathbb{P}[i\text{th smallest element is chosen as the pivot}] = 1/n.$$

(2) If we pick the qth smallest element as the pivot, then in the recursive step the size of the input becomes $q-1$ and $n-q$ respectively. Noticing the partitioning routine runs in time $\Theta(n)$, we have that $T(n) = T(q-1) + T(n-q) + \Theta(n)$. Since the pivot is chosen uniformly at random, applying the expectation on $T(n)$ gives the desired statement.

(3) From the solution of (1) and (2), we have that

$$\begin{aligned}
\mathbb{E}[T(n)] &= \mathbb{E}\left[\sum_{q=1}^{n} X_q(T(q-1)+T(n-q)+\Theta(n))\right] \\
&= \sum_{q=1}^{n} \frac{1}{n}\left(\mathbb{E}\left[T(q-1)\right]+\mathbb{E}\left[T(n-q)\right]+\Theta(n)\right) \\
&= \frac{2}{n}\sum_{q=0}^{n-1}\mathbb{E}[T(q)]+\Theta(n).
\end{aligned}$$

(4) We have that

$$\begin{aligned}
\sum_{k=1}^{n-1} k\lg k &= \sum_{k=1}^{\lceil n/2\rceil-1} k\lg k + \sum_{k=\lceil n/2\rceil}^{n-1} k\lg k \\
&\leq \sum_{k=1}^{\lceil n/2\rceil-1} k\lg\frac{n}{2} + \sum_{k=\lceil n/2\rceil}^{n-1} k\lg n \\
&= \sum_{k=1}^{\lceil n/2\rceil-1} (k\lg n - k) + \sum_{k=\lceil n/2\rceil}^{n-1} k\lg n \\
&= \sum_{k=1}^{n-1} k\lg n - \sum_{k=1}^{\lceil n/2\rceil-1} k \\
&\leq \frac{(n-1)n}{2}\cdot\lg n - \frac{n}{4}\left(\frac{n}{2}-1\right) \\
&= \frac{n^2}{2}\cdot\lg n - \frac{n}{2}\cdot\lg n - \frac{n^2}{8} + \frac{n}{4} \\
&\leq \frac{n^2}{2}\cdot\lg n - \frac{n^2}{8},
\end{aligned}$$

where the last inequality holds for any $n \geq 4$.

(5) We can use the substitution method, and assume that $\mathbb{E}[T(n)] \leq an\lg n - bn$ for some positive constant a and b. □

Problem 3: Describe an algorithm that, given n integers in the range 0 to k, preprocess its input and then answer any query about how many of the n integers fall into a range $[a..b]$ in $O(1)$ time. Your algorithm should use $\Theta(n+k)$ preprocessing time.

Solution: Remember that in the Counting sort there is an array C, in which $C[i]$ is the number of elements less than or equal to i. Moreover, such array can be obtained in $\Theta(n+k)$ time.

Assuming we have such array C, the number of n integers in a range $[a..b]$ is $C[b]-C[a-1]$. Hence, with $\Theta(n+k)$ preprocessing time, we can output every query in $O(1)$ time. □