CS5330: Week 2 Solutions

February 9, 2020

Here are solution sketches to the Week 2 problems. If anything is unclear, please talk to me or your TA.

1. Recall the definition of a treap in Lecture 1, and our observation that once the priorities are assigned to the keys, the structure of the treap is the same no matter what the insertion order is.

   Suppose n keys $x_1, \ldots, x_n$ are inserted into a treap with random priorities (assumed to be distinct). By the above, without loss of generality $x_1 < x_2 < \cdots < x_n$. Let D_i be the depth of x_i in the tree. In other words, D_i is the number of ancestors x_i has in the tree. Let D_{ij} be the indicator variable that x_j is an ancestor of i in the tree.

   Suppose $j > i$ for parts (a) through (c).

   (a) What is D_{ij} if x_i has the highest priority among $\{x_i, \ldots, x_j\}$?

   Solution sketch: $D_{ij} = 0$ since x_j being an ancestor of x_i will violate the heap property on the priorities.

   (b) What is D_{ij} if for some $k \in \{i+1, \ldots, j-1\}$ has the highest priority among $\{x_i, \ldots, x_j\}$?

   Solution sketch: In this case also, $D_{ij} = 0$. During the construction of the treap, x_k is the first pivot chosen among $\{x_i, \ldots, x_j\}$ since it has the highest priority. x_i and x_j fall on different sides of x_k, so x_j cannot be an ancestor of x_i.

   (c) What is D_{ij} if x_j has the highest priority among $\{x_i, \ldots, x_j\}$?

   Solution sketch: $D_{ij} = 1$. During the construction of the treap, x_j is the first pivot chosen among $\{x_i, \ldots, x_j\}$. Thus, x_i will lie in the subtree to its left.

(d) Compute $\mathbb{E}[D_i] = \sum_{j \neq i} \mathbb{E}[D_{ij}]$, taking care to consider both $j < i$ and $j > i$.

Solution sketch:

$$\begin{aligned}
\mathbb{E}[D_i] &= \mathbb{E}\left[\sum_{j<i} \frac{1}{i-j+1} + \sum_{j>i} \frac{1}{j-i+1}\right] \\
&\leq \sum_{j=2}^{i} \frac{1}{i} + \sum_{j=2}^{n-i+1} \frac{1}{j} \\
&\leq H_i + H_{n-i+1} = O(\log n)
\end{aligned}$$

2. n people queue up to attend a movie which has n seats. However, the first person has lost his ticket and sits in one of the empty seats uniformly at random. Subsequently, each person (and no one else has lost a ticket) sits either in his assigned seat, or if that seat is already taken sits in an empty seat uniformly at random. What is the expected number of people *not* sitting in their correct seats?

Solution sketch: A crucial observation is that if person j finds that his seat is already occupied, then the unoccupied seats at that point are $\{1, j+1, j+2, \ldots, n\}$. Why? Consider the sitting process. Suppose person 1 sits in seat $i_1 > 1$. Now persons $2, \ldots, i_1 - 1$ sit in their correct seats. Person i_1 finds his seat taken but seats $\{1\} \cup \{i_1 + 1, \ldots, n\}$ free. The proof follows by induction.

Let p_i be the probability that the i'th person sits in the wrong seat. We want to compute $\sum_i p_i$. Clearly, $p_1 = 1 - \frac{1}{n}$. For $i > 1$ and $j < i$, let $p_{i,j}$ be the probability that person j is sitting in seat i. We see that $p_{i,1} = \frac{1}{n}$. For $j > 1$, note that person j sits in seat i exactly if j is sitting in the wrong seat and he chooses to sit in seat i. Now, from the above observation, we know that if j sees his seat is taken, seats $\{1\} \cup \{j+1, \ldots, n\}$ will be free. So, if person j is sitting in the wrong seat, he sits in i with probability $1/(n-j+1)$. Therefore, for $j > 1$:

$$p_{i,j} = \frac{p_j}{n-j+1}.$$

So, for $i > 1$

$$p_i = \frac{1}{n} + \sum_{j=2}^{i-1} \frac{p_j}{n-j+1}.$$

Solving the recurrence, we get $p_i = 1/(n-i+2)$ for $i > 1$. This gives us: .

$$\sum_{i=1}^{n} p_i = 1 - \frac{1}{n} + \sum_{i=2}^{n} \frac{1}{n-i+2} = 1 + \sum_{i=3}^{n} \frac{1}{n-i+2} = \sum_{i=1}^{n-1} \frac{1}{i} = \Theta(\log n)$$

3. Given a permutation π of $\{1, 2, \ldots, n\}$, let $L(\pi)$ denote the length of the longest increasing subsequence in π. Note that a subsequence may not be contiguous; for instance in the permutation $\pi = (1, 6, 4, 5, 2, 7, 3)$ for $n = 7$, the longest subsequence is $(1, 4, 5, 7)$ and so $L(\pi) = 4$. In this exercise, you need to prove $\mathbf{E}[L(\pi)] = \Theta(\sqrt{n})$ where the expectation is over a random permutation of $\{1, \ldots, n\}$.

(a) Prove that $\mathbf{E}[L(\pi)] = O(\sqrt{n})$. **Hint:** Use the following fact: for any non-negative integer random variable Z, $\mathbb{E}[Z] = \sum_{z \geq 0} \Pr(Z \geq z)$. Now, for a fixed k, calculate an upper bound on the probability that $L(\pi) \geq k$.

Solution sketch: For any fixed k-tuple $(i_1, \ldots, i_k)$, the probability that $\pi(i_1) < \pi(i_2) < \cdots < \pi(i_k)$ is exactly $\frac{1}{k!}$ because any of the possible orderings are equally likely. Hence by the union bound:

$$\Pr[L(\pi) \geq k] \leq \frac{\binom{n}{k}}{k!} \leq e^{2k} \frac{n^k}{k^{2k}}$$

using the inequalities $\binom{n}{k} \leq (en/k)^k$ and $k! \geq (k/e)^k$. So:

$$\begin{aligned}
\mathbf{E}[L(\pi)] &= \sum_{k \geq 0} \Pr[L(\pi) \geq k] \\
&= \sum_{k=1}^{10\sqrt{n}} \Pr[L(\pi) \geq k] + \sum_{k > 10\sqrt{n}} \Pr[L(\pi) \geq k] \\
&\leq 10\sqrt{n} + \sum_{k > 10\sqrt{n}} \left(\frac{e^2 n}{k^2} \right)^k \\
&\leq 10\sqrt{n} + \sum_{k > 1} 2^{-k} = O(\sqrt{n})
\end{aligned}$$

where in the last line, we substituted $e^2 n / k^2 < 1/2$ for $k > 10\sqrt{n}$.

(b) Prove that $\mathbf{E}[L(\pi)] = \Omega(\sqrt{n})$. **Hint:** Assume n is a perfect square. For $i = 1, \ldots, \sqrt{n}$, define the indicator random variable X_i which takes the value 1 if and only if some entry in $(i-1)\sqrt{n} + 1 \leq j \leq i\sqrt{n}$ satisfies $(i-1)\sqrt{n} + 1 \leq \pi(j) \leq i\sqrt{n}$. Can you relate these $\sqrt{n}$ variables with $L(\pi)$?

Solution sketch: If each event X_i holds, then clearly there's an increasing subsequence of length $\sqrt{n}$. Note that $X_i = 0$ when all the entries in the interval $[(i-1)\sqrt{n} + 1, i\sqrt{n}]$ are from the $n - \sqrt{n}$ numbers outside that interval. So:

$$\Pr[X_i = 0] = \frac{\binom{n - \sqrt{n}}{\sqrt{n}}}{\binom{n}{\sqrt{n}}} = \prod_{i=0}^{\sqrt{n}-1} \frac{n - \sqrt{n} - i}{n - i} \leq \left(1 - \frac{1}{\sqrt{n}}\right)^{\sqrt{n}} \leq e^{-1}$$

Then, $\mathbb{E}[X_i] \geq 1 - e^{-1}$ and hence, $\mathbb{E}[L(\pi)] \geq (1 - e^{-1})\sqrt{n}$. The calculation for non-perfect squares is similar.