Advanced Algorithms 2024 2024

Sample Solutions 00

Lecturer: Johannes L, Maximillian P, Bernhard H *Teaching Assistant: Antti R*

1 Basic Recursions

- $T(n) = 2T(n/3) + \Theta(n)$. Here $a = 2$, $b = 3$ and $\log_b(a) \approx 0.63 < 1$, thus case 3 of the Master theorem applies and the recursion tree is root-heavy. Thus, $T(n) = \Theta(n)$.
- $T(n) = 3T(n/2) + \Theta(n)$. Here $a = 3$, $b = 2$ and $\log_b(a) \approx 1.58 > 1$, thus case 1 of the Master theorem applies and the recursion tree is leaf-heavy. Thus, $T(n) = \Theta(n^{\log_b(a)}) = \Theta(n^{\log_2(3)})$.
- $T(n) = 2T(n/2) + \Theta(n)$. Here $a = 2$, $b = 2$ and $\log_2(2) = 1 = 1$, thus case 2 of the Master theorem applies and the recursion tree is balanced. Thus, $T(n) = \Theta(n^{\log_b(a)} \log^{1+0}(n)) = \Theta(n \log n)$.
- $T(n) = 2T(n/2) + \Theta(\sqrt{n})$. Here $a = 2$, $b = 2$ and $\log_2(2) = 1 > \frac{1}{2}$, thus case 1 of the Master theorem applies and the recursion tree is leaf-heavy. Thus, $T(n) = \Theta(n^{\log_b(a)}) = \Theta(n)$.
- $T(n) = 2T(n/2) + \Theta(n/\log n)$. Here $a = 2$, $b = 2$ and $\log_2(2) = 1 = 1$ but $k = -1$, thus case 2b of the Master theorem applies and the recursion tree is balanced. Thus, $T(n) = \Theta(n^{\log_b(a)} \log\log(n)) = \Theta(n \log\log(n))$.
- $T(n) = T(\sqrt{n}) + 1$. Here the Master theorem does not apply, but letting $T'(k) = T'(k/2) + 1$, we have $T(n) = T'(\log n)$, and for this function, we have $a = 1$, $b = 2$ and $\log_b(a) = 0 = 0$, thus case 2 of the Master theorem applies and the recursion tree is balanced. Thus, $T'(k) = \Theta(k^{\log_b(a)} \log^{1+0}(k)) = \log(k)$, and $T(n) = \log(\log(n))$.

2 Estimation

Assume $\epsilon < \frac{1}{10}$. Suppose we pick k balls (for k divisible by 10) and output YES if at least $k/10$ balls sampled are red, and NO otherwise. Let x_i for $i \in [k] = \{1, \ldots, k\}$ be 1 if the ith sampled ball is red and 0 otherwise. These are independent random variables, each with value 1 with probability equal to the fraction of red balls in the bag. Let $X = \sum_i x_i$.

- If more than a $(1+\epsilon)/10$-fraction of the balls are red, then $\mathbb{E}[X] \geq (1+\epsilon)10k$. By Chernoff, for $\epsilon' = \frac{\epsilon}{1+\epsilon}$, we have

$$\begin{aligned}
P(X \leq \frac{k}{10}) &\leq P\left(X \leq \frac{1}{1+\epsilon}\mathbb{E}[X]\right) \\
&= P\left(X \leq (1-\epsilon')\mathbb{E}[X]\right) \\
&\leq \exp\left(\frac{-\epsilon'^2\mathbb{E}[X]}{2}\right) \\
&\leq \exp\left(\frac{-5\epsilon^2 k}{(1+\epsilon)}\right) \\
&\leq \exp(-\epsilon^2 k)
\end{aligned}$$

- If less than a $(1-\epsilon)/10$-fraction of the balls are red, then $\mathbb{E}[X] \leq (1-\epsilon)10k$. By Chernoff, for $\epsilon' = \frac{\epsilon}{1-\epsilon}$, we have

$$
\begin{aligned}
P(X \geq \frac{k}{10}) &\leq P\left(X \geq \frac{1}{1-\epsilon}\mathbb{E}[X]\right) \\
&= P\left(X \geq (1+\epsilon')\mathbb{E}[X]\right) \\
&\leq \exp\left(\frac{-\epsilon'^2\mathbb{E}[X]}{3}\right) \\
&\leq \exp\left(\frac{-3\epsilon^2 k}{(1+\epsilon)}\right) \\
&\leq \exp(-\epsilon^2 k)
\end{aligned}
$$

Thus, selecting $k = \frac{\ln(1/\delta)}{\epsilon^2}$, the tester is correct with probability $1-\delta$ as long as the fraction of red balls is at least $(1+\epsilon)/10$ or at most $(1-\epsilon)/10$.

3 Selection

Repeat the following: sample a uniformly random index and compare it to every element. If less than k elements are smaller, delete those elements and subtract from k however many there were. If at least k were smaller or equal, delete all greater elements. Once at most 4 elements remain, sort them using a constant number of comparisons and return the kth element.

To analyze the algorithm, consider the number of elements deleted in an iteration that started with m elements. During this iteration, $m-1$ comparisons are made, and if the sampled pivot is not in the smallest or largest $\lfloor m/4 \rfloor$ elements, at least $\lfloor m/4 \rfloor \geq m/8$ elements are deleted. This occurs with probability at least $\frac{1}{2}$.

Thus, denoting by $T(n)$ the expected number of comparisons to find the kth element for some k out of n elements, we have the recursion $T(n) \leq n + \frac{1}{2}T(\frac{7}{8}n) + \frac{1}{2}T(n)$, thus $T(n) \leq 2n + T(\frac{7}{8}n)$ (with the base case $T(n) = O(1)$ for $n \leq 4$. Thus, by the master theorem, we have $T(n) = O(n)$.

4 Quicksort

The quicksort algorithm is the following: sample a uniformly random pivot, compare it to every element, recursively sort the lesser and greater elements, and return the sorted lesser elements concatenated with the pivot concatenated with the sorted greater elements.

In a call to Quicksort, say the sampled pivot is the kth element in the sorted order.

- If $k = i$ or $k = j$, the ith and jth element in the sorted order are compared.
- If $\min(i,j) < k < \max(i,j)$, the ith and jth element in the sorted order are not compared, and they do not appear in the same recursive call, thus cannot possibly be compared.
- If $k < \min(i,j)$ or $k > \max(i,j)$, the ith and jth element in the sorted order are not compared, and they appear in the same recursive call (thus might be compared in the recursive call).

Thus, the probability the ith and jth element $(i < j)$ in the sorted order are compared is exactly $\frac{1}{j-i}$. Thus, the expected number of comparisons is $\sum_{m=1}^{n-1} \frac{n-m}{m} \leq nH_n \leq 2n \ln n$, where $H_n = \sum_{m=1}^{n} \frac{1}{m}$ is the nth harmonic number.

5 Finding a Common Friend

After the mth day, both Alice and Bob have m friends. Order the friends of Alice arbitrarily. Since these sets of friends are independent, the probability the ith friend of Alice is a friend of Bob's given that no lower-index friend of Alice is a friend of Bob's is $\frac{m}{n-(i-1)} \geq \frac{m}{n}$. Thus, the probability none of Alice's friends is a friend of Bob's is at most $(1 - \frac{m}{n})^m \leq e^{-m^2/n}$, thus if $m = \sqrt{n}$, Alice and Bob share a friend with at least a constant probability. For a lower bound, similarly for $i \leq m < \frac{n}{2}$, we have $\frac{m}{n-(i-1)} \leq \frac{2m}{n}$, thus the probability none of Alice's first $m \leq \frac{n}{4}$ friends is a friend of Bob's is at least $(1 - \frac{2m}{n})^m \geq e^{-4m^2/n}$. Thus if $m = \sqrt{n}$, Alice and Bob share a friend with at most a constant probability. Thus, the expected time until Alice and Bob share a friend is $\Theta(\sqrt{n})$.

6 Rumor Spreading

Suppose $m \leq \frac{n}{2}$ people know the rumor before the gossip on some specific day occurs. Order the people who know the rumor arbitrarily. Let x_i be a random variable with value 1 if the ith person who knew the rumor gossiped to a person who didn't know the rumor and wasn't gossiped to by a person with lower index who already knew the rumor, and value 0 otherwise. Then, $\mathbb{E}[x_i] \geq \frac{n-m-(i-1)}{n}$, thus the expected number of people who know the rumor after the day $\mathbb{E}[\sum_{i\in[m]} x_i]$ is by linearity of expectation at least $\sum_{i\in[m]} \mathbb{E}[x_i] \geq m + \frac{m}{4}$. Thus, by Markov the probability at least $m(1 + \frac{1}{8})$ people know the rumor is at least $\frac{1}{2}$. After this occurs $8\lceil\log n\rceil$ times, at least half the people know the rumor. By Chernoff, this occurs $8\lceil\log n\rceil$ times during the first $\Theta(\log n)$ days with high probability.

Now, assume that at least half the people know the rumor before a certain day. Fix a person who does not know the rumor, and consider the probability that after i further days, they still do not know the rumor. This probability is at most $\left((1 - \frac{1}{n})^{n/2}\right)^i \leq e^{-i/2}$. Thus, with high probability, after $O(\log n)$ further days, every person knows the rumor with high probability.

Since both phases take at most $O(\log n)$ days with high probability, everyone knows the rumor with high probability in $O(\log n)$ days. Since the number of people who know the rumor can at most double each day, the minimum number of days until every person knows the rumor is $\Omega(\log n)$, the expected time until everyone knows the rumor is $\Theta(\log n)$.

7 Independent Set

7.1 Part 1

Let S be a randomly sampled subset of vertices containing each vertex independently with probability $p = \frac{1}{d}$. The induced graph $G[S]$ has in expectation $\mathbb{E}[E[S]] = |E|p^2 = \frac{n}{2d}$ edges and $\mathbb{E}[|S|] = np = \frac{n}{d}$ vertices.

For the sampled set S, let $S' \subseteq S$ be a subset where for every edge in $G[S]$, one of its endpoint vertices is removed. Thus, S' is an independent set, and we have $|S'| \geq |S| - |E[S]|$. Thus, using linearity of expectation, $\mathbb{E}[|S'|] \geq \mathbb{E}[|S|] - \mathbb{E}[|E[S]|] = \frac{n}{d} - \frac{n}{2d} = \frac{n}{2d}$.

As we have described a random process that produces an independent set with expected size $\frac{n}{2d}$, there in particular must exist an independent set of size at least $\frac{n}{2d}$.

7.2 Part 2

Let $v_1, v_2, \ldots, v_n$ be a uniformly random ordering of the vertices. Construct the independent set greedily: for $i = 1, 2, \ldots, n$, if v_i is not adjacent to a vertex in the independent set, add it to the independent set.

Constructing the independent set this way, a sufficient condition for a vertex v to appear in the independent set is that all of its neighbours appear after it in the ordering; this occurs with probability $\frac{1}{\deg(v)+1}$. Thus, the expected number of vertices in the independent set is at least $\sum_v \frac{1}{\deg(v)+1}$. The function $f(x) = \frac{1}{1+x}$ is convex, thus $\sum_v f(\deg(v)) \geq nf(\overline{d}) = \frac{n}{\overline{d}+1}$, as desired.

As we have described a random process that produces an independent set with expected size $\frac{n}{\overline{d}+1}$, there in particular must exist an independent set of size at least $\frac{n}{\overline{d}+1}$.

8 Balanced Colouring

Assign to every vertex a random colour, and let x_b be a random variable that equals 1 if the element $b \in B$ is assigned the colour red and 0 if it is assigned the colour blue. For a subset S_i, let $X_i := \sum_{b \in S_i} x_b$. We have $\mathbb{E}[X_i] = |S_i|/2$, and the number of red and blue elements in a set i differ by at most $2T$ if and only if $|X_i - \mathbb{E}[X_i]| \leq T$.

Fix an arbitrary constant $c \geq 1$, and let $T = \sqrt{3c \cdot n \log m}$. By Chernoff, for any $0 < \epsilon < 1$, we have

$$P(|X_i - \mathbb{E}[X_i]| \geq \epsilon \mathbb{E}[X_i]) \leq 2 \exp\left(\frac{-\epsilon^2 \mathbb{E}[X_i]}{3}\right)$$

If a set $_i$ has size at most $2T$, the number of red and blue elements in that set differ by at most $2T$ with probability 1. If the set S_i has size greater than $2T$, for $\epsilon = \frac{T}{\mathbb{E}[X_i]}$,

$$P(|X_i - \mathbb{E}[X_i]| \geq T) \leq 2 \exp\left(-\frac{cn \log m}{\mathbb{E}[X_i]}\right) \leq 2m^{-c}$$

Union bounding over each of the m sets, the number of red and blue elements in every set differs by at most $2T = O(\sqrt{n \log m}$ with probability at least $2m^{-(c-1)}$. As c was an arbitrary constant, this holds with high probability.