Problem Set 9 Solutions

This problem set is due **at 11:59pm** on **Thursday, December 1, 2016.**

EXERCISES (NOT TO BE TURNED IN)

Distributed Algorithms

- Read Lecture 19 Notes
- Read Recitation 11 Notes

Randomized Algorithms

- Do Exercise 7.3-2 in CLRS on page 180.
- Do Exercise 7.4-4 in CLRS on page 185.
- Do Exercise 9.2-2 in CLRS on page 219.
- Do Exercise 1.23 in Mitzenmacher and Upfal on page 19.

Problem 9-1. Distributed Scheduling [50 points]

Suppose we have n processors and $n/2$ threads. Each thread has exactly one job that it wants to run. Your goal is to design a distributed algorithm where each thread chooses a processor to run on.

Consider the following algorithm, which proceeds in rounds. Initially, at time $t = 0$, all threads are active. At each time t, the threads proceed as follows:

- Each active thread chooses a processor independently and uniformly at random.
- For every processor that is picked by only one thread, the thread executes its job on that processor and becomes inactive.
- For every processor that is picked by multiple threads, no jobs are completed and all threads that chose the processor stay active.

This process is repeated by all the active threads until all threads eventually execute their job.

(a) [15 points] Suppose that at a given round, there are δn active threads (where $\delta < 1$). Show that the expected number of active threads in the next round is at most $\delta(1 - e^{-\delta}) \cdot n$.

Note: You may use the fact that $1 - k \leq e^{-k}$ for any k.

Solution:

A given thread is alone if all the $\delta n - 1$ other threads land on different processors - this occurs with probability $(1 - 1/n)^{\delta n - 1}$

Let $m = n - 1$. Working backwards:

$$\begin{aligned}
(1 - \frac{1}{n})^{\delta n - 1} &\geq e^{-\delta} \\
e^{\delta} &\geq (\frac{1}{1 - \frac{1}{n}})^{\delta n - 1} \\
e^{\delta} &\geq (\frac{n}{n-1})^{\delta n - 1} \\
e^{\delta} &\geq (\frac{m+1}{m})^{\delta(m+1) - 1} \\
e^{\delta} &\geq (1 + \frac{1}{m})^{\delta(m+1) - 1} \\
e &\geq (1 + \frac{1}{m})^{(m+1) - \frac{1}{\delta}} \\
e &\geq (1 + \frac{1}{m})^{m + (1 - \frac{1}{\delta})} \\
e &\geq (1 + \frac{1}{m})^{m}
\end{aligned}$$

. The last line is a fact about e.

(b) [10 points] Let m_i denote the number of active threads at the start of round i, and δ_i denote the fraction. That is, $m_i := \delta_i \cdot n$. Using your result from part (a), show that the expected number of threads at the start of round $i+1$ is at most $\delta_i^2 \cdot n$. That is, $\mathbb{E}[m_{i+1}] \leq \delta_i^2 \cdot n$.

Solution: We can define δ_i as the fraction of the n balls left at the start of round i Assume that there are $n/2$ balls at the start of round 1 - this corresponds to a $\delta_1 = 1/2$. By setting the number of balls at the start of round $i+1$, which is $\delta_i n(1 - \frac{1}{e^{\delta_i}})$ to $\delta_{i+1}n$, we get the following equality:

$$\delta_{i+1} = \delta_i(1 - \frac{1}{e^{\delta_i}}) \leq \delta_i(1 - (1 - \delta_i)) = \delta_i^2$$

Hence, there are at most δ_i^2 threads at the start of round $i+1$.

(c) [25 points] The 6.046 professors and TAs claim that they can show a stronger statement, namely that if there are δn threads at the start of a round and $\delta n \geq \sqrt{n}$, then the probability that there are more than $\delta^2 n$ threads active at the end of the round is at most $\frac{1}{n^2}$.

Take this statement at face value, and use it to show that if you start with $n/2$ threads, then after $O(\log \log n)$ rounds, you are left with at most $\sqrt{n}$ threads with probability at least $1 - o(1)$.

Solution: The proof is constructed as follows - first, we assume that if there are δn threads at the start of a round, then there are at most δn^2 threads at the end of a round - and show that in $loglog(n)$ rounds, we have less than $\sqrt(n)$ threads left.

For all values of i, we can use the formula derived in part (b): $log(\delta_{i+1}) = 2log(\delta_i)$

This implies that at some round k, we have

$log(\delta_k) = 2^{k-1}\delta_1 = 2^{k-1}(1 - 1/e)$

We need to find the smallest value of k where $\delta_k n < \sqrt(n)$ which is equivalent to finding the smallest k such that $log(\delta_k) < -.5log(n)$

Using our above expression, this becomes the smallest k such that $log(2^{k-1}(1 - 1/e)) < -.5log(n)$ Taking the logs and solving will give $k = \theta(loglog(n))$

We know that this holds true, if each round satisfies the inequality that if there are δn threads at the start, then there are at most δn^2 threads at the end of the round.

We can define a "bad" round as a round which starts with δn threads and ends with more than δn^2 threads left The probability that a round i, which starts with δn threads is "bad" is:

Pr[At most δn^2 threads left at end of round i] = $1 - \frac{1}{n^2}$

Pr[At least δn^2 threads left at the end of round i] = $\frac{1}{n^2}$
Hence, the probability that a given round i is "bad" is $\frac{1}{n^2}$. We can do a union bound over all $O(loglogn)$ rounds to get that the probability that any round is "bad" is at most:
Pr[any round is "bad"] $\leq loglogn\frac{1}{n^2}$
Pr[any round is "bad"] $\leq \frac{1}{n}$
Hence, with high probability (at least $1 - 1/n$), all our rounds are "good" and in $O(loglogn)$ rounds, there are at most $\sqrt{n}$ threads left.

(d) **(Optional)** Show that if you start with n processors and $\sqrt{n}$ threads, with probability at least $1/2$, all the threads will finish in a single round. Put this together with the previous parts to show that if you start with n processors and $n/2$ threads, all the threads finish in $O(\log \log n)$ rounds with probability at least $1/2 - o(1)$.

Solution: This is the birthday paradox problem! :)

Problem 9-2. Planning Mixers [50 points]

There have recently been reports that our 6.046 students do not really know each other very well. To help facilitate collaboration and friendship between students, our beloved 6.046 professors have decided to throw a series of k separate mixers. They would like each student to attend exactly one of the k mixers and get to know the other students at the event.

There are a total of n students in 6.046 this semester. The friendship status among students can be modeled as an undirected graph $G = (V, E)$ where each vertex $u \in V$ is a student and there is an edge $(u, v) \in E$ if u and v are friends.

In order to encourage students to meet people outside their friend groups, the professors want each student u to attend a mixer on a different day from any of their friends. That is, if u and v are friends, then they must go to two different mixer events.

Let a k-mixer-assignment be an assignment for each student $u \in V$ to one of the mixer days $1, 2, . . k$ so that no two friends attend the same mixer.

(a) [10 points] Suppose that the 6.046 staff somehow discovered the entire friendship graph, G, among all students in the class (perhaps through an overnight session of stalking students on Facebook). Let d be the maximum number of friends that any student has. Give a simple algorithm they can use for constructing a $(d + 1)$-mixer-assignment.

Solution: This is equivalent to a $d + 1$-coloring of G. Consider the greedy algorithm - for each vertex v, assign it the first color from the palette that hasn't been assigned to any of its neighbors.

Before going through with their plan, the professors were notified that it may be frowned upon to stalk their students on Facebook. Unfortunately this means that they can't find out the friendship graph G. Luckily, they come up with the following distributed protocol which allows students to co-operate with their friends and decide on their own $(d+1)$-mixer-assignment.

The algorithm proceeds in rounds, where each round has 2 phases. Initially, all students are undecided. Furthermore, each student u starts out with available choices 1 to $deg(u)+1$.

- In the first phase, each undecided student u tentatively chooses a random number j among their available options, texts all his friends of his chosen number, and receives all his friends' chosen numbers.
- In the second phase, each student u does the following: If none of his friends have chosen his number j, then u decides to attend mixer j and notifies all his friends of this decision. When his friends receive this message, it means they cannot attend mixer j and so they remove j from their future choices.

The protocol continues in multiple rounds until all students have decided on a mixer.

(b) [15 points] Prove that the above distributed protocol outputs a valid $(d+1)$-mixer-assignment. That is, prove that after all students have made a decision, (1) each student will be assigned to a mixer from 1 to $d+1$, and (2) no two friends will be assigned to the same mixer.

Solution: Once again, we use the equivalent terminology of assigning $d+1$ colors.

Lemma: When the algorithm terminates, only $d+1$ colors are used.

For any vertex, u, we provide a palette of colors from 1 to $deg(u)+1$. Since $deg(u) \leq d$ for all u, this implies that each vertex is given one of the $d+1$ colors.

Lemma: The algorithm produces a valid vertex coloring.

For a given vertex u, u chooses a final color, only if no other neighbor has picked the same color in that round.

Furthermore since the final color is chosen from u's palette, which does not include any final colors from it's neighbors, this is a valid color for u to pick - no neighbors will already have this color.

After u chooses its color, all neighbors remove that color from their palette - so none of u's neighbors will pick that color in the future.

(c) [25 points] Prove that with probability at least $1-\frac{1}{n}$, all students will have decided on a mixer after $O(d\log n)$ rounds of the given protocol.
Hint: First, for any student v, show that he will decide on a mixer after $O(\log n)$ rounds with high probability.

Solution: Consider a specific vertex u. The probability that u decides on a mixer at any given round, is exactly the probability that u picks a j that is different from all of it's currently undecided neighbors. This can easily be lower bounded by $\geq \frac{1}{d+1}$ since in the worst case, u has $d+1$ undecided neighbors.

Therefore, the probability that u does not decide in any given round is $\leq 1 - \frac{1}{d+1}$. Let X_u be an indicator variable so that $X_u = 1$ if u has not decided on a mixer after $2(d+1)\log n = O(d \log n)$ rounds. Then we have that,

$$\begin{aligned} Pr[X_u = 1] &\leq \left(1 - \frac{1}{d+1}\right)^{2(d+1)\log n} && (1) \\ &\leq e^{-2\log n} && (2) \\ &= \frac{1}{n^2} && (3) \end{aligned}$$

We can now bound the probability that all students have decided after $2(d+1)\log n = O(d \log n)$ rounds as,

$$\begin{aligned} Pr\left[\sum_u X_u = 0\right] &= 1 - Pr\left[\sum_u X_u > 0\right] \\ &\geq 1 - \sum_u Pr[X_u = 1] && \text{by union bound} \\ &\geq 1 - n(\frac{1}{n^2}) \\ &\geq 1 - \frac{1}{n} \end{aligned}$$

Therefore the number of rounds is $O(d \log n)$ with high probability as desired.

Problem 9-3. Amplifying Probabilities [50 points]

Bob is an electrical engineer at Cisco, working on routers. For one of his assignments, he wants to estimate the number of distinct packets streaming through a router. Let d be the exact number of distinct packets. Using the algorithm discussed in lecture for estimating the number of distinct elements in a stream, he can compute an estimate X for d such that $\frac{d}{5} \leq X \leq 5d$ with probability at least $1 - 2/5 = 0.6$.

Unfortunately, a correctness probability of 0.6 is not good enough for Bob. He wants to get an estimate Y such that $\frac{d}{5} \leq Y \leq 5d$ with probability at least $1 - \delta$. To do this, Bob simultaneously computes $k = O(\log(1/\delta))$ independent estimates $X_1, X_2, .., X_k$ of d using the above algorithm so that for each X_i, he has that $\frac{d}{5} \leq X_i \leq 5d$ with probability at least 0.6.

(a) [15 points] Suppose Bob knows nothing about the distribution of the X_i except that they will be within the given range with probability at least 0.6. His first strategy is to take the average of his k measurements and set $Y = \frac{1}{k}\sum_{i=1}^{k} X_i$. Here you will prove that Y does not necessarily give a good estimate for d.

Specifically, for $k = 2$ construct random variables X_1 and X_2, such that X_1 and X_2 satisfy the above requirement, but $Y = (X_1 + X_2)/2$ will be outside the range with probability strictly greater than $1/2$.

Solution: Consider the following random variables X_1, X_2, where $X_1 = d$ with probability 0.6 and $X_1 = 10^9 d$ with probability 0.4. X_2 has the same distribution as X_1.

In this case, if we take averages, we will only get a good approximation if $X_1 = X_2 = d$. This happens with probability $0.6^2 = 0.36$. Thus, with probability 0.72, our estimate Y will be outside the range.

(b) [35 points] Suppose instead that Bob sets Y to be the **median** of $X_1, X_2, .. X_k$. Prove that for a choice of $k = O(\log(1/\delta))$, $\frac{d}{5} \leq Y \leq 5d$ with probability at least $1 - \delta$.

Hint: Think about how you can use Chernoff bound.

Solution: We want to know the probability that Y is outside the range $\frac{d}{5} \leq Y \leq 5d$. Since Y is the median of $X_1, X_2, .., X_k$, we know that Y must be in the desired range if a majority of the X_i are inside the range (note that this is not an if and only if relation). In other words, it is true that,

$$Pr\left[\frac{d}{5} \leq Y \leq 5d\right] \geq Pr\left[\text{at least half of } X_i \text{ have } \frac{d}{5} \leq X_i \leq 5d\right]$$

Let T_i be an indicator random variable such that $T_i = 1$ if $\frac{d}{5} \leq X_i \leq 5d$ and 0 otherwise. Note that $Pr[T_i = 1] = \gamma$ where $\gamma \geq 0.6$ and it follows that $E[T_i] = \gamma$. Note furthermore that $E[\sum_{i=1}^{k} T_i] = \gamma k$ by linearity of expectation.

We therefore have that,

$$
\begin{aligned}
Pr\left[\frac{d}{5} \le Y \le 5d\right] &\ge Pr\left[\sum_{i=1}^{k} T_i \ge \frac{k}{2}\right] \\
&= 1 - Pr\left[\sum_{i=1}^{k} T_i < \frac{k}{2}\right]
\end{aligned}
$$

We now apply Chernoff bound to show $Pr\left[\sum_{i=1}^{k} T_i < \frac{k}{2}\right] \le \delta$.

$$
\begin{aligned}
Pr\left[\sum_{i=1}^{k} T_i < \frac{k}{2}\right] &= Pr\left[\sum_{i=1}^{k} T_i < (1-\beta)\gamma k\right] && \text{for } \beta = \frac{2\gamma - 1}{2\gamma} \\
&\le e^{-\beta^2 \gamma k/2} \\
&= e^{-(2\gamma-1)^2 k/(8\gamma)} \\
&\le e^{-(2\gamma-1)^2 k/8} \\
&\le \delta && \text{for } k = \frac{8}{(2\gamma-1)^2} \log\frac{1}{\delta} = O\left(\log\frac{1}{\delta}\right)
\end{aligned}
$$

Hence with probability at least $1-\delta$, Y is in the desired range.

Problem 9-4. Tricky Select [50 points]

Recall that we have a deterministic median finding algorithm which runs in $O(n)$ time. In this problem, we are interested in finding an estimate for the median of a set of n numbers, in sub-linear time.

Specifically, we are given a set of n distinct elements S, and two parameters $\epsilon, \delta < \frac{1}{2}$, and we want to find a number $x \in S$, such that $(\frac{1}{2} - \epsilon)n \le rank(x) \le (\frac{1}{2} + \epsilon)n$ with probability at least $1 - \delta$.

Consider the following algorithm for computing x:

- Create a set T by sampling $m = O(\frac{1}{\epsilon^2}\log(\frac{1}{\delta}))$ elements uniformly at random from S. Hence $|T| = m$.
- Let x be the median of T, which can be found by running the deterministic median finding algorithm on T. Return x.

Notice that by setting $\delta \le \frac{1}{n}$, $m = O(\frac{1}{\epsilon^2}\log n)$ and the guarantee holds with probability at least $1 - \frac{1}{n}$.

(a) [35 points] Prove the claim that with probability at least $1-\delta$, the rank of x is in $[(\frac{1}{2}-\epsilon)n, (\frac{1}{2}+\epsilon)n]$.

Hint: Notice that in order to have $rank(x) > r$ for some $1 \le r \le n$, we must have picked at most $\frac{m}{2}$ elements from S that have rank less than r.

Hint: Think about how you can use Chernoff to bound the probability that this happens.

Solution: Let x be the median of T and $rank(x)$ be the rank of x in S. Then by definition have that,

$$\begin{aligned}
&Pr\left[\left(\frac{1}{2}-\epsilon\right)n \le rank(x) \le \left(\frac{1}{2}+\epsilon\right)n\right] \\
&= 1 - Pr\left[\left(rank(x) < \left(\frac{1}{2}-\epsilon\right)n\right) \text{ OR } \left(rank(x) > \left(\frac{1}{2}+\epsilon\right)n\right)\right] \\
&\ge 1 - Pr\left[rank(x) < \left(\frac{1}{2}-\epsilon\right)n\right] - Pr\left[rank(x) > \left(\frac{1}{2}+\epsilon\right)n\right]
\end{aligned}$$

Where the last inequality holds by the union bound. We will now show the following,

$$Pr\left[rank(x) > \left(\frac{1}{2}+\epsilon\right)n\right] \le \frac{\delta}{2}. \tag{4}$$

Notice that since the probability distribution of T is symmetric around the median of S, this would imply that

$$Pr\left[rank(x) < \left(\frac{1}{2}-\epsilon\right)n\right] = Pr\left[rank(x) > \left(\frac{1}{2}+\epsilon\right)n\right] \le \frac{\delta}{2}.$$

Hence, proving (4) implies that $rank(x)$ will be within the desired range with probability at least $1-\delta$.

To prove (4) we start by making an observation that in order for $rank(x) > \left(\frac{1}{2}+\epsilon\right)n$, we must have picked at most $\frac{m}{2}$ elements from S that have rank less than $\left(\frac{1}{2}+\epsilon\right)n$.

Let T_i be an indicator random variable such that $T_i = 1$ if the i-th element picked had rank less than $\left(\frac{1}{2}+\epsilon\right)n$ and 0 otherwise. Then $Pr[T_i = 1] = \left(\frac{1}{2}+\epsilon\right)$. We then let $Y = \sum_{i=1}^{m} T_i$. We can compute the expectation of Y as,

$$\mathbb{E}[Y] = \mathbb{E}\left[\sum_{i=1}^{m} T_i\right] = \sum_{i=1}^{m} \mathbb{E}[T_i] = m\left(\frac{1}{2}+\epsilon\right).$$

We can now use Chernoff to upper bound the probability that at most $\frac{m}{2}$ elements that have rank less than $\left(\frac{1}{2}+\epsilon\right)n$.

$$\begin{aligned}
Pr\left[Y \leq \frac{m}{2}\right] &= Pr\left[Y \leq (1-\beta)m\left(\frac{1}{2}+\epsilon\right)\right] && \text{where } \beta = \frac{2\epsilon}{1+2\epsilon} \\
&\leq e^{-\beta^2 \mathbb{E}[Y]/2} \\
&= e^{-\epsilon^2 m/(1+2\epsilon)} \\
&\leq e^{-\epsilon^2 m/2} \\
&\leq e^{-\log\frac{2}{\delta}} && \text{for } m \geq \frac{2}{\epsilon^2}\log\frac{2}{\delta} \\
&\leq \frac{\delta}{2}
\end{aligned}$$

Hence we have proven (4), as desired, since

$$Pr\left[rank(x) > \left(\frac{1}{2}+\epsilon\right)n\right] = Pr\left[Y \leq \frac{m}{2}\right] \leq \frac{\delta}{2}.$$

(b) [15 points] Generalize the algorithm from part (a) to compute the approximate rank k element in S with probability $1-\delta$. That is, the algorithm should output a number x such that $k - \epsilon n \leq rank(x) \leq k + \epsilon n$.

Hint: Instead of returning the median element in T, think of how you can scale k down to find the corresponding rank in T.

Solution: Our algorithm is similar to part a. Instead of the median, let x be the rank $\frac{km}{n}$ element in Y. We run our deterministic linear time select algorithm on T with to find and return x.

We will now show that with probability at least δ, $k - \epsilon n \leq rank(x) \leq k + \epsilon n$. Note that when $k = \frac{n}{2}$, this is exactly the problem in part (a), hence this is a generalization.

Similar to part (a), we have that,

$$\begin{aligned}
&Pr\left[k - \epsilon n \leq rank(x) \leq k + \epsilon n\right] \\
&= 1 - Pr\left[(rank(x) < k - \epsilon n) \text{ OR } (rank(x) > k + \epsilon n)\right] \\
&\geq 1 - Pr\left[rank(x) < k - \epsilon n\right] - Pr\left[rank(x) > k + \epsilon n\right]
\end{aligned}$$

Once again, we will show that $Pr\left[rank(x) > k + \epsilon n\right] \leq \frac{\delta}{2}$ and $Pr\left[rank(x) < k - \epsilon n\right] \leq \frac{\delta}{2}$. For $k \geq n/2$, it would be sufficient to show the first inequality since the second inequality is upper bounded by the first. We can then appeal to symmetry for the $k \leq n/2$ case. However, here we will compute both bounds explicitly without making assumptions on n.

We make the same observation as part (a) that in order for $rank(x) > k + \epsilon n$, we must have picked at most $\frac{km}{n}$ elements that have rank less than $k + \epsilon n$.

Let T_i be an indicator random variables such that $T_i = 1$ if the i-th element picked had rank less than $k + \epsilon n$. Then $Pr[T_i = 1] = \frac{k}{n} + \epsilon$. And let $Y = \sum_{i=1}^{m} T_i$ be the total number of such elements. We can compute the expectation of Y as follows.

$$\begin{aligned}
\mathbb{E}[Y] &= \mathbb{E}\left[\sum_{i=1}^{m} T_i\right] \\
&= \sum_{i=1}^{m} \mathbb{E}[T_i] \\
&= m\left(\frac{k}{n} + \epsilon\right)
\end{aligned}$$

We can now apply Chernoff bound to upper bound the probability that at most $\frac{km}{n}$ elements have rank less than $k + \epsilon n$.

$$\begin{aligned}
Pr\left[Y \leq \frac{km}{n}\right] &= Pr\left[Y \leq (1-\beta)m\left(\frac{k}{n} + \epsilon\right)\right] && \text{for } \beta = \frac{\epsilon n}{k + \epsilon n} \\
&\leq e^{-\beta^2 \mathbb{E}[Y]/2} \\
&= e^{-\epsilon^2 m / \left(2\left(\frac{k}{n} + \epsilon\right)\right)} \\
&\leq e^{-\epsilon^2 m/4} \\
&\leq \frac{\delta}{2} && \text{for } m \geq \frac{4}{\epsilon^2} \log \frac{2}{\delta}
\end{aligned}$$

We follow a similar derivation to show $Pr\left[rank(x) < k - \epsilon n\right] \leq \frac{2}{\delta}$.

We make the observation that in order to have $rank(x) < k - \epsilon n$ we must have picked at most $m - \frac{km}{n}$ elements that have rank greater than $k - \epsilon n$.

Let T_i be an indicator random variable such that $T_i = 1$ if the it element has rank greater than $k - \epsilon n$ and 0 otherwise. Then $Pr[T_i = 1] = 1 - \frac{k}{n} + \epsilon$. We then let $Y = \sum_{i=1}^{m} T_i$. We can go through the same derivation as before to get $\mathbb{E}[Y] = m\left(1 - \frac{k}{n} + \epsilon\right)$.

We can now bound the probability that we pick at most $m - \frac{km}{n}$ elements with rank greater than $k - \epsilon n$ as follows.

$$\begin{aligned} Pr\left[Y \leq m - \frac{km}{n}\right] &= Pr\left[Y \leq (1-\beta)m\left(1 - \frac{k}{n} + \epsilon\right)\right] \quad \text{for } \beta = \frac{n\epsilon}{n\epsilon + n - k} \\ &\leq e^{-\beta^2 \mathbb{E}[Y]/2} \\ &= e^{\epsilon^2 m / \left(2\left(\epsilon + 1 - \frac{k}{n}\right)\right)} \\ &\leq e^{\epsilon^2 m/4} \\ &\leq \frac{\delta}{2} \qquad \text{for } m \geq \frac{4}{\epsilon^2} \log \frac{2}{\delta} \end{aligned}$$

Therefore, by setting $m = \frac{4}{\epsilon^2} \log \frac{2}{\delta}$, we get the desired bound for $rank(x)$ with probability at least $1 - \delta$.