Problem Set 3 Solutions

Problem 1. **(a)** Instead, let's consider a sequence of k coin flips. We look at the probability of getting at least n heads in that sequence. That is, let

$$Y_i = \begin{cases} 1 & : \text{ if the } i\text{-th coin flip is heads} \\ 0 & : \text{ otherwise} \end{cases}$$

be an indicator random variable. Then we let $Y = \sum_{i=1}^{k} Y_i$.

There is a direct mapping between the problem if you assume the same random flips. That is, consider getting a value of $X_1 = x_1$. We have the same chance of getting $Y_1 = \ldots = Y_{x_1-1} = 0$ and $Y_{x_1} = 1$. Similarly, consider $X_i = x_i$. Then we have $Y_{\sum_{j=1}^{i-1} x_i+1} = \ldots = Y_{\sum_{j=1}^{i} x_i-1} = 0$ and $Y_{\sum_{j=1}^{i} x_i} = 1$ with probability $Pr[X_i = x_i]$. Basically, if some $X_i = x_i$, then we have a mapping to $x_i - 1$ values of j such that $Y_j = 0$ and 1 value of j such that $Y_j = 1$.

We claim that

$$Pr[X = k] = Pr[Y = n \text{ in } k \text{ flips}] \ .$$

This fact should be somewhat obvious given the mapping. Each $X_i = x_i$ corresponds to x_i total flips in the sequence of coin flips. Moreover, each X_i corresponds to exactly one head in the sequence. Thus, if $X = k$, then we have a correspondence to a sequence of k flips with n heads, so $Y = n$ with equal probability.

We can generalize to

$$Pr[X \geq k] = Pr[Y \leq n \text{ in } k \text{ flips}] \ .$$

If we increase the value of X on the LHS, then there are fewer heads in the first k flips. Thus, $Pr[Y \leq n]$ decreases commensurately.

Once we have this reduction, we can use a Chernoff bound, because we have Y_i are indicator random variables. We note that $E[Y_i] = 1/2$ (it's just a coin flip), giving us $\mu_Y = E[Y] = \sum_{i=1}^{k} E[Y_i] = k/2$.

$$\begin{aligned} Pr[Y \leq (1-\delta)\mu_Y \text{ in } k \text{ flips}] &\leq e^{-\delta^2 \mu_Y/2} \\ &= e^{-\delta^2 k/4} \ . \end{aligned} \tag{1}$$

If we set $k = (1+\epsilon)2n$, and $(1-\delta)\mu_Y = n$, we get

$$\begin{aligned} Pr[X \geq (1+\epsilon)\mu_X] &= Pr[Y \leq (1-\delta)\mu_Y \text{ in } (1+\epsilon)\mu_X \text{ flips}] \\ &= Pr[Y \leq n \text{ in } (1+\epsilon)\mu_X \text{ flips}] \\ &\leq e^{-\delta^2(1+\epsilon)n/2} \ . \end{aligned} \tag{2}$$

Then we just need to substitute in for δ. We have $(1-\delta)\mu_Y = n$, so $(1-\delta)(1+\epsilon)n = n$, or $\delta = \epsilon/(1+\epsilon)$, which gives us

$$Pr[X \geq (1+\epsilon)2n] \leq e^{-\epsilon^2 n/(2+2\epsilon)} .$$

(b) As expected, this derivation is very similar to what we did in class.

We start by exponentiating things and applying Markov's inequality to get

$$\begin{aligned} Pr[X > (1+\epsilon)\mu] &= Pr\left[e^{tX} > e^{t(1+\epsilon)\mu}\right] \\ &\leq \frac{E\left[e^{tX}\right]}{e^{t(1+\epsilon)\mu}} . \end{aligned} \tag{3}$$

We take advantage of the independence of our variables to get

$$\begin{aligned} E\left[e^{tX}\right] &= E\left[e^{\sum tX_i}\right] \\ &= E\left[\prod e^{tX_i}\right] \\ &= \prod E\left[e^{tX_i}\right] . \end{aligned} \tag{4}$$

Now, we look at the geometric distribution to solve for $E[e^{tX_i}]$:

$$\begin{aligned} E\left[e^{tX_i}\right] &= \frac{1}{2}e^t + \frac{1}{4}e^{2t} + \frac{1}{8}e^{4t} + \ldots \\ &= \sum_{k=1}^{\infty} \left(\frac{e^t}{2}\right)^k \\ &= \frac{e^t/2}{1-e^t/2} \quad , \text{ if } e^t < 2 \\ &= \frac{e^t}{2-e^t} . \end{aligned} \tag{5}$$

Substituting back in our formula for $E[e^{tX}]$, we get

$$\begin{aligned} E\left[e^{tX}\right] &= \prod E\left[e^{tX_i}\right] \\ &= \left(\frac{e^t}{2-e^t}\right)^n \\ &= \left(\frac{e^t}{2-e^t}\right)^{\mu/2} . \end{aligned} \tag{6}$$

So, for our overall bound, we have

$$
\begin{aligned}
Pr[X > (1+\epsilon)\mu] &= Pr\left[e^{tX} > e^{t(1+\epsilon)\mu}\right] \\
&\leq \left(\frac{e^t}{2-e^t}\right)^{\mu/2}\left(\frac{1}{e^{t(1+\epsilon)\mu}}\right) \\
&= \left(\frac{e^t}{2-e^t}\right)^{\mu/2}\left(\frac{1}{e^{2t(1+\epsilon)}}\right)^{\mu/2} \\
&= \left(\frac{1}{2-e^t}\right)^{\mu/2}\left(\frac{1}{e^{t+2t\epsilon)}}\right)^{\mu/2} \\
&= \left(\frac{1}{(2-e^t)(e^{t(1+2\epsilon)})}\right)^{\mu/2}. \qquad (7)
\end{aligned}
$$

This inequality holds for all t, so we want to pick t as to minimize $Pr[X > (1+\epsilon)\mu]$. This probability is minimized when $e^t = (1+2\epsilon)/(1+\epsilon)$.
Plugging back in, we get

$$
\begin{aligned}
Pr[X > (1+\epsilon)\mu] &\leq \left(\frac{1}{(2-e^t)(e^{t(1+2\epsilon)})}\right)^{\mu/2} \\
&= \left(\frac{1}{(2-e^t)(e^t)^{1+2\epsilon}}\right)^{\mu/2} \\
&= \left(\frac{1}{\left(2-\frac{1+2\epsilon}{1+\epsilon}\right)\left(\frac{1+2\epsilon}{1+\epsilon}\right)^{1+2\epsilon}}\right)^{\mu/2} \\
&= \left(\frac{1+\epsilon}{\left(\frac{1+2\epsilon}{1+\epsilon}\right)^{1+2\epsilon}}\right)^{\mu/2} \\
&= \left((1+\epsilon)\left(\frac{1+\epsilon}{1+2\epsilon}\right)^{1+2\epsilon}\right)^{\mu/2} \\
&= \left((1+\epsilon)\left(1-\frac{\epsilon}{1+2\epsilon}\right)^{1+2\epsilon}\right)^{\mu/2} \\
&\leq \left((1+\epsilon)e^{-\epsilon}\right)^{\mu/2} \\
&= \left(\frac{1+\epsilon}{e^{\epsilon}}\right)^{n}. \qquad (8)
\end{aligned}
$$

Problem 2. Let's consider an item x in a recursive call on n elements. We call a pivoting round **good** for x if x ends up in a subproblem of size at most $3n/4$. Naturally, each time x ends up in such a subproblem the problem size reduces by a factor of $4/3$, so x can be in at most $\log_{4/3} n$ such problems.

We can think of x as belonging to at most $\log_{4/3} n$ **segments**, where a segment is some number of bad rounds followed by a good round. We let X_i be the value of the i-th segment for x. That is, if $X_i = k$, then after the $(i-1)$-st good round for x, we have $k-1$ bad rounds followed by the i-th good round. Wow, would you look at that? The X_i follow a geometric distribution as in the previous problem.[1] Naturally we let $X = \sum_{i=\leq \lg n} X_i$.

So now we just apply the bound from the previous problem. We can use the part (a) bound from the previous problem, to get

$$Pr[X \geq (1+\epsilon) 2 \log_{4/3} n] \leq e^{-\epsilon^2 \log_{4/3} n/(2+2\epsilon)} .$$

For our purposes, we can set $\epsilon = 5$ to get

$$Pr[X \geq 12 \log_{4/3} n] \leq e^{-25 \log_{4/3} n/12} \leq e^{-2 \log_{4/3} n} \leq e^{-2 \lg n} = 1/n^2 .$$

Notice that increasing ϵ increases the power of n, so this is really a high probability bound.

Alright, so we have a high probability bound that it takes $O(\lg n)$ rounds before x is in a subproblem of size 1. But there are n different elements x that we can be talking about, so we take the union bound, giving us probability at most $1/n$ (or $1/n^{c-1}$ if we adjust the constant) that any element has not been reduced to a subproblem of size 1. Thus, with high probability, all elements are completed by $O(\lg n)$ steps (i.e., this many levels of recursion). At each level of recursion, we do at most n comparisons (counting all the subproblems), so with high probability, the total number of comparisons (or work) is $O(n \lg n)$.

Problem 3. **(a)** Let's consider just the node with an address of all 0s, denoted by $0^{(n)}$. We argue that there are $\Omega(\sqrt{N})$ packets routed through this node, so the total routing must take $\Omega(\sqrt{N})$ steps.

Consider all packets coming from $a_i \circ 0^{(n/2)}$, where $0^{(n/2)}$ is $n/2$ 0s. There are $2^{n/2}$ such packets, because a_i is $n/2$ bits long. The bit fixing strategy corrects each bit in order. Thus, a_i is corrected to $0^{(n/2)}$ before the second half of the string is touched. Therefore, each of these packets goes through $O^{(n)}$. Again, there are $2^{n/2} = 2^{\lg N/2} = N^{1/2}$ such packets, so we're done.

(b) Again, we argue that a lot of packets will go through $0^{(n)}$ with high probability. Specifically, we argue that there are at least $2^{\Omega(n)}$ such packets with high probability..

We again consider a subset of packets starting from $a_i \circ 0^{(n/2)}$. We let $S = \left\{a_i \circ 0^{(n/2)} \mid a_i \text{ contains } k \text{ 1s}\right\}$. For every $s_i \in S$, we have an indicator random

[1] The mean is at most 2. Having a smaller mean only helps (stochastic domination), so let's just argue that the mean is at most 2. i.e., with probability at least 1/2, a round is good. Let x_i be the element of rank i. Well, with probability 1/2, we pick a pivot between $x_{n/4}$ and $x_{3n/4}$. For any pivot within this range, both subproblems are smaller than $3n/4$. Thus, x is in a subproblem of size at most $3n/4$, and x has a good round. We can also have a good round for x for some other partition choices, but it doesn't matter, we just need to show that the probability is at least 1/2.

variable X_i, with

$$X_i = \begin{cases} 1 & : \text{ if } s_i \text{ goes through } 0^{(n)} \\ 0 & : \text{ otherwise} \end{cases}$$

Similarly, $X = \sum X_i$.
Note that $|S| = \binom{n/2}{k}$, because we are just choosing k of $n/2$ locations to be 1s.
So now let's look at $E[X_i] = Pr[s_i \text{ goes through } 0^{(n)}]$. The packet from s_i goes through $0^{(n)}$ if we choose to fix the k bits in the first half of the string to 0s before fixing the corresponding k bits in the second half to 1s. Since there are $2k$ bits to choose from, and we need to choose k of them first, we have

$$E[X_i] = \frac{1}{\binom{2k}{k}} \geq \left(\frac{k}{2ek}\right)^k = \left(\frac{1}{2e}\right)^k .$$

Thus, we have

$$\begin{aligned} E[X] &= E\left[\sum_{i=1}^{\binom{n/2}{k}} X_i\right] \\ &= \sum_{i=1}^{\binom{n/2}{k}} E[X_i] \\ &= \binom{n/2}{k} E[X_i] \\ &\geq \binom{n/2}{k} \left(\frac{1}{2e}\right)^k \\ &\geq \left(\frac{n}{2k}\right)^k \left(\frac{1}{2e}\right)^k \\ &= \left(\frac{n}{4ek}\right)^k . \end{aligned} \tag{9}$$

So now we just choose $k = n/(8e)$. Then we have $E[X] \geq 2^{n/(8e)}$. Next, we apply the Chernoff bound to get

$$\begin{aligned} Pr[X < (1-\epsilon)E[X]] &\leq e^{-\epsilon^2 E[X]/2} \\ &\leq e^{-\epsilon^2 2^{n/(8e)}/2} \\ &= e^{-\epsilon^2 2^{n/(8e)-1}} . \end{aligned} \tag{10}$$

Suppose we choose something simple, like $\epsilon = 1/2$. Then we have $Pr[X < 1/2E[X]] \leq e^{-2^{n/(8e)-3}} = e^{-N^{1/(8e)}/8}$, which is exponentially small in N. Note that $E[X] \geq 2^{\Omega(n)}$ from above, so we have

$$Pr[X < 1/2E[X]] \leq Pr[X < 2^{\Omega(n)}] \leq e^{-N^{1/(8e)}/8} .$$

So with high probability, we have $2^{\Omega(n)}$ packets passing through $0^{(n)}$, and the total routing time must be at least $2^{\Omega(n)}$.

Problem 4. **(a)** We start with the fact

$$Pr[k \text{ balls in bin } 1] = \binom{n}{k}\left(\frac{1}{n}\right)^k\left(1-\frac{1}{n}\right)^{n-k} .$$

I don't feel the need to argue this probability is correct, because we did this in class. Anyway, we just continue from here:

$$\begin{aligned}
Pr[k \text{ balls in bin } 1] &= \binom{n}{k}\left(\frac{1}{n}\right)^k\left(1-\frac{1}{n}\right)^{n-k} \\
&\geq \left(\frac{n}{k}\right)^k\left(\frac{1}{n}\right)^k\left(1-\frac{1}{n}\right)^{n-k} \\
&= \left(\frac{1}{k}\right)^k\left(1-\frac{1}{n}\right)^{n-k} \\
&\geq \left(\frac{1}{k}\right)^k\left(\frac{1}{2e}\right) \quad , \text{ for } n \geq 2 \\
&= \frac{1}{2e}\left(\frac{1}{k}\right)^k . \qquad (11)
\end{aligned}$$

Now, we just set $k = c\lg n/\lg\lg n$, giving us

$$\begin{aligned}
Pr[c\lg n/\lg\lg n \text{ balls in bin } 1] &\geq \frac{1}{2e}\left(\frac{1}{c\lg n/\lg\lg n}\right)^{c\lg n/\lg\lg n} \\
&= \frac{1}{2e}\left(\frac{\lg\lg n}{c\lg n}\right)^{c\lg n/\lg\lg n} \\
&\geq \left(\frac{1}{c\lg n}\right)^{c\lg n/\lg\lg n} \quad , \text{ for } n \geq 4 \\
&= \left(\frac{1}{c2^{\lg\lg n}}\right)^{c\lg n/\lg\lg n} \\
&= \frac{1}{c2^{\lg\lg n\cdot(c\lg n/\lg\lg n)}} \\
&= \frac{1}{c2^{c\lg n}} \\
&= \frac{1}{cn^c} \\
&= \Omega(n^{-c}) . \qquad (12)
\end{aligned}$$

Setting $c = 1/2$, we get $Pr[\lg n/2\lg\lg n \text{ balls in bin } 1] \geq \Omega(1/\sqrt{n})$.

(b) Let us first argue that conditioning on a bin not having k balls only increases the probability that the next bin does have k balls. We use induction on number of bins we are conditioning on. Let B_i be the event that bin i has at least k balls. The base case is as follows, for $i > 1$:

$$\begin{aligned} Pr[B_i] &= Pr[B_i|B_1] \cdot Pr[B_1] + Pr[B_i|\neg B_1] \cdot Pr[\neg B_1] \\ &\leq Pr[B_i|\neg B_1] \cdot Pr[B_1] + Pr[B_i|\neg B_1] \cdot Pr[\neg B_1] \qquad (13) \\ &= Pr[B_i|\neg B_1](Pr[B_1] + Pr[\neg B_1]) \\ &= Pr[B_i|\neg B_1] \ . \qquad (14) \end{aligned}$$

We notice that B_i is more likely if B_1 does not have k balls, because then there are more balls that can be in B_i.

So now we assume that it works condition on up to k events, and we condition on the next one. Note that we are solving for every event B_i, with $i > k+1$. We have

$$\begin{aligned} Pr[B_i] &\leq Pr[B_i|\neg B_1 \wedge \neg B_2 \wedge \ldots \wedge \neg B_k] \\ &= Pr[B_i|\neg B_1 \wedge \ldots \wedge \neg B_k \wedge B_{k+1}] \cdot Pr[B_{k+1}] \\ &\quad + Pr[B_i|\neg B_1 \wedge \ldots \wedge \neg B_k \wedge \neg B_{k+1}] \cdot Pr[\neg B_{k+1}] \\ &\leq Pr[B_i|\neg B_1 \wedge \ldots \wedge \neg B_k \wedge \neg B_{k+1}] \end{aligned} \qquad (15)$$

The argument is the same as in the base case.

Thus, we have concluded that conditioning on bins not having k balls increases the chances that the next bin does. Specifically, the induction ends at proving

$$Pr[B_i] \leq Pr[B_i|\neg B_1 \wedge \ldots \wedge \neg B_{i-1}] \ .$$

Conversely, we have

$$Pr[\neg B_i] \geq Pr[\neg B_i|\neg B_1 \wedge \ldots \wedge \neg B_{i-1}] \ ,$$

because this is exactly $1 - Pr[B_i]$.

So now let's solve the real problem, with $k = \lg n/2\lg\lg n$. From part (a), we have

$$Pr[B_i] = Pr[\text{Bin } i \text{ has at least } \lg n/2\lg\lg n \text{ balls}] \geq \frac{1}{2\sqrt{n}} \ .$$

Thus, we have

$$Pr[\neg B_i] = Pr[\text{Bin } i \text{ has at most } \lg n/2\lg\lg n \text{ balls}] \leq 1 - \frac{1}{2\sqrt{n}} \ .$$

So now we just solve for all bins having at most this many balls:

$$\begin{aligned}
&Pr[\text{all bins have} \leq \lg n/2 \lg\lg n \text{ balls}] \\
&= Pr[\neg B_1] \cdot Pr[\neg B_2 | \neg B_1] \cdots Pr[\neg B_n | \neg B_1 \wedge \ldots \wedge \neg B_{n-1}] \\
&\leq Pr[\neg B_1] \cdot Pr[\neg B_2] \cdots Pr[\neg B_n] \\
&\leq \left(1 - \frac{1}{2\sqrt{n}}\right)^n \\
&\leq e^{-\left(\frac{1}{2\sqrt{n}}\right)n} \\
&= e^{-\sqrt{n}/2} \; .
\end{aligned}$$

So the probability is exponentially small that all bins have have fewer than $\lg n/2 \lg\lg n$ balls. Therefore, we conclude that with high probability, some bin has $\Omega(\lg n/\lg\lg n)$ balls.

Alternative solution: we can show $Pr[\neg B_i] \geq Pr[\neg B_i | \neg B_1 \wedge \ldots \wedge \neg B_{i-1}]$ via a more formal proof. First, by Bayes, the above inequality is equivalent to $Pr[\neg B_1 \wedge \ldots \wedge \neg B_{i-1}] \geq Pr[\neg B_1 \wedge \ldots \wedge \neg B_{i-1} | \neg B_i]$. Similar to the base case of the above solution, it suffices to show

$$Pr[\neg B_1 \wedge \ldots \wedge \neg B_{i-1} | B_i] \geq Pr[\neg B_1 \wedge \ldots \wedge \neg B_{i-1} | \neg B_i].$$

We consider the probability $Pr[\neg B_1 \wedge \ldots \wedge \neg B_{i-1} |$ bin i has x balls$]$ for any integer x, which we will denote by $f(x)$. When bin i has x balls, the rest bins have $n - x$ balls in total. Thus, conditioned on bin i having x balls, the distribution of balls in other bins behave as if we put $n - x$ balls randomly into $n - 1$ bins. The probability that $\neg B_1 \wedge \ldots \wedge \neg B_{i-1}$ happens is obviously smaller when $n - x$ is larger. Hence, $f(x)$ is nondecreasing when x increases. We can write

$$Pr[\neg B_1 \wedge \ldots \wedge \neg B_{i-1} | B_i] = \sum_{x \geq k} \frac{Pr[\text{bin } i \text{ has } x \text{ balls}]}{Pr[B_i]} f(x).$$

and similarly,

$$Pr[\neg B_1 \wedge \ldots \wedge \neg B_{i-1} | \neg B_i] = \sum_{x < k} \frac{Pr[\text{bin } i \text{ has } x \text{ balls}]}{Pr[\neg B_i]} f(x).$$

We notice that both $Pr[\neg B_1 \wedge \ldots \wedge \neg B_{i-1} | B_i]$ and $Pr[\neg B_1 \wedge \ldots \wedge \neg B_{i-1} | \neg B_i]$ are weighted averages of $f(x)$, but the x values in the expression of $Pr[\neg B_1 \wedge \ldots \wedge \neg B_{i-1} | B_i]$ are larger than those in the expression of $Pr[\neg B_1 \wedge \ldots \wedge \neg B_{i-1} | \neg B_i]$. Since $f(x)$ is nondecreasing, we conclude that $Pr[\neg B_1 \wedge \ldots \wedge \neg B_{i-1} | B_i] \geq Pr[\neg B_1 \wedge \ldots \wedge \neg B_{i-1} | \neg B_i]$.