CS 466/666: Algorithm Design and Analysis **University of Waterloo: Fall 2023**

Homework 2 Solutions

October 22, 2023

Problem 1. We reexamine the balls-and-bins experiment in this question, focusing on aspects other than the maximum load. Suppose we are throwing m balls into n bins where each ball chooses one of the bins independently and uniformly at random.

(*a*) Prove that it is sufficient and necessary for m to be $\Theta(n \log n)$ so that with constant probability, every bin has at least one ball inside it. **(12.5 points)**

Solution. We prove each part separately.

Sufficiency: Let us assume that $m \geqslant 2n \ln n$. Then, for every $i \in [n]$,

$$\Pr(\text{Bin } i \text{ is empty}) = \left(1 - \frac{1}{n}\right)^m \leqslant \exp\left(-\frac{2n \ln n}{n}\right) = n^{-2}.$$

A union bound over all the n bins now implies that in this case, every bin is non-empty with probability at least $1 - 1/n = 1 - o(1)$ (which is larger than a constant > 0).

Necessity: Now assume instead that $m = (1/2) \cdot n \ln n$. For every bin $i \in [n]$, define a random variable $X_i \in \{0, 1\}$ which is 1 iff bin i is empty. Let $X = \sum_{i=1}^{n} X_i$ denote the number of empty bins. We have,

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{i=1}^{n} \mathbb{E}[X_i] && \text{(by linearity of expectation)} \\
&= \sum_{i=1}^{n} \Pr(\text{Bin } i \text{ is empty}) && \text{(as } X_i \text{ is an indicator random variable)} \\
&= \sum_{i=1}^{n} \left(1 - \frac{1}{n}\right)^m && \text{(as the } m \text{ balls are thrown into } n \text{ bins independently and uniformly)} \\
&\geqslant \sum_{i=1}^{n} \exp\left(-\frac{m}{2n}\right) && \text{(as } 1 - x \geqslant \exp(-2x) \text{ for } x \in (0, 1/2)) \\
&= \sum_{i=1}^{n} \exp\left(-(1/2) \ln n\right) && \text{(by the choice of } m = \varepsilon \cdot n \ln n) \\
&= n^{1/2}.
\end{aligned}$$

We also have,

$$\begin{aligned}
\mathrm{Var}[X] &= \sum_{i=1}^{n} \mathrm{Var}[X_i] + \sum_{i \neq j} \mathbb{E}[X_i \cdot X_j] - \mathbb{E}[X_i] \cdot \mathbb{E}[X_j] && \text{(as proven in Proposition 6 of Lecture 3)} \\
&\leqslant \sum_{i=1}^{n} \mathbb{E}[X_i] + \sum_{i \neq j} \mathbb{E}[X_i \cdot X_j] - \mathbb{E}[X_i] \cdot \mathbb{E}[X_j] && \text{(as } X_i\text{'s are indicator random variables)} \\
&= \mathbb{E}[X] + \sum_{i \neq j} \mathbb{E}[X_i \cdot X_j] - \mathbb{E}[X_i] \cdot \mathbb{E}[X_j]. && \text{(by linearity of expectation)}
\end{aligned}$$

We now claim that the second sum above is non-positive. This is because knowing bin i being empty can only lower the chance of bin j becoming empty also, i.e., these variables are *negatively correlated*.

More formally, for every $i \neq j$,

$$\begin{aligned}
&\mathbb{E}[X_i \cdot X_j] - \mathbb{E}[X_i] \cdot \mathbb{E}[X_j] \\
&\quad = \Pr(\text{Bin } i \text{ and Bin } j \text{ are both empty}) - \Pr(\text{Bin } i \text{ is empty}) \cdot \Pr(\text{Bin } j \text{ is empty}) \\
&\quad = \Pr(\text{Bin } i \text{ is empty}) \cdot \Pr(\text{Bin } j \text{ is empty} \mid \text{Bin } i \text{ is empty}) - \Pr(\text{Bin } i \text{ is empty}) \cdot \Pr(\text{Bin } j \text{ is empty}) \\
&\quad < 0,
\end{aligned}$$

because

$$\Pr(\text{Bin } j \text{ is empty} \mid \text{Bin } i \text{ is empty}) < \Pr(\text{Bin } j \text{ is empty}).$$

Yet another way to see the above more directly is that

$$\begin{aligned}
\mathbb{E}[X_i \cdot X_j] &= \left(1 - \frac{2}{n}\right)^m \\
\mathbb{E}[X_i] \cdot \mathbb{E}[X_j] &= \left(1 - \frac{1}{n}\right)^{2m},
\end{aligned}$$

and the former is smaller than the latter because $(1-2x) < (1-x)^2$ for all $x \neq 0$.

All in all, this implies that $\mathrm{Var}[X] \leqslant \mathbb{E}[X]$. We now have,

$$\begin{aligned}
\Pr(X = 0) &\leqslant \Pr(|X - \mathbb{E}[X]| \geqslant \mathbb{E}[X]) \leqslant \frac{\mathrm{Var}[X]}{\mathbb{E}[X]^2} && \text{(by Chebyshev's inequality)} \\
&\leqslant \frac{1}{\mathbb{E}[X]} && \text{(by the bound on the variance of } X\text{)} \\
&= \frac{1}{n^{1/2}}.
\end{aligned}$$

Thus, in this case, the probability that every bin is non-empty is at most $1/n^{1/2} = o(1)$ (which is less than any constant > 0).

To conclude, $m = \Theta(n \log n)$ is the tight bound on this problem.

(*b*) Find the sufficient and necessary asymptotic value for m to be, so that with constant probability, at least one bin has two or more balls inside it. **(12.5 points)**

Solution. The first part of the argument is the same for both cases. For every pair of balls $i \neq j \in [m]$, we define a random variable $Y_{ij} \in \{0, 1\}$ which is 1 iff both balls i and j are assigned to the same bin. Define $Y = \sum_{i \neq j} Y_{ij}$. Note that there is at least one bin with two or more balls inside it if and only if $Y \neq 0$. Moreover,

$$\begin{aligned}
\mathbb{E}[Y] &= \sum_{i \neq j} \mathbb{E}[Y_{ij}] && \text{(by linearity of expectation)} \\
&= \sum_{i \neq j} \Pr(\text{Ball } i \text{ and Ball } j \text{ are assigned to the same bin}) \\
&= \sum_{i \neq j} \frac{1}{n} && \text{(as each ball is choosing its destination randomly and independently)} \\
&= \frac{\binom{m}{2}}{n}. && \text{(as we are picking two different balls out of } m\text{)}
\end{aligned}$$

Sufficiency: Suppose $m \geqslant 2\sqrt{n}+1$. We have the variables $Y_{ij}, Y_{k\ell}$ are *pairwise* independently, because knowing Balls i and j are assigned to the same bin, still leaves the choice of at least k or ℓ entirely

independent (at most one of them can be equal to i and j, otherwise, we are not considering two different pairs). Thus,

$$\begin{aligned}\operatorname{Var}[X] &= \operatorname{Var}\left[\sum_{i\neq j} Y_{ij}\right] = \sum_{i\neq j}\operatorname{Var}[Y_{ij}] \\ &\text{(as proven in Lecture 3, variance of sum of pairwise-independent variables is sum of variances)} \\ &\leqslant \sum_{i\neq j}\mathbb{E}[Y_{ij}] \qquad\qquad (\text{as } \operatorname{Var}[Y_{ij}] \leqslant \mathbb{E}\left[Y_{ij}^2\right] = \mathbb{E}[Y_{ij}] \text{ as } Y_{ij}\in\{0,1\}) \\ &= \mathbb{E}[Y].\end{aligned}$$

Hence, by applying Chebyshev's inequality,

$$\Pr(Y=0) \leqslant \Pr(|Y-\mathbb{E}[Y]| \geqslant \mathbb{E}[Y]) \leqslant \frac{\operatorname{Var}[Y]}{\mathbb{E}[Y]^2} \leqslant \frac{1}{\mathbb{E}[Y]} = \frac{n}{\binom{m}{2}} = \frac{2n}{m\cdot(m-1)} < \frac{2n}{2\sqrt{n}\cdot 2\sqrt{n}} = 1/2.$$

Thus, in this case, with probability at least $1/2 > 0$, there is a bin that contains at least two balls.

Necessity: Now suppose instead $m = \varepsilon\cdot\sqrt{n}$ for some $\varepsilon \to 0$ in the limit. Then, by Markov's inequality,

$$\Pr(Y \geqslant 1) \leqslant \mathbb{E}[Y] = \frac{\binom{m}{2}}{n} \leqslant \frac{m^2}{n} = \varepsilon^2.$$

Thus, as $\varepsilon \to 0$, this probability also tends to zero (and hence is less than any constant > 0).

To conclude, $m = \Theta(\sqrt{n})$ is the tight bound on this problem.

Problem 2. For any $\varepsilon \in (0, 1/4)$, a $(1\pm\varepsilon)$-**cut sparsifier** of an undirected unweighted graph $G = (V, E)$ is a *weighted* spanning subgraph $H = (V, E_H)$ of G with weights $w : E_H \to \mathbb{R}^+$ such that for *every* cut $S \subseteq V$,

$$(1-\varepsilon)\cdot|\delta_G(S)| \leqslant w(\delta_H(S)) \leqslant (1+\varepsilon)\cdot|\delta_G(S)|;$$

here, $\delta_G(S)$ is the set of edges in the cut S in G and $w(\delta_H(S))$ denotes the *weight* of the cut S in H. In other words, the weight of every cut in H is a $(1\pm\varepsilon)$-approximation of the size of the same cut in G.

We are generally interested in constructing cut sparsifiers that are *sparse*, i.e., has few edges (otherwise, we could have taken $H = G$ with weight 1 everywhere). In this question, we see a simple (but not that efficient) way of constructing a cut sparsifier using random sampling.

Let λ denote the minimum cut value of G. Suppose we sample each edge in G with probability

$$p := \frac{100\log n}{\varepsilon^2\cdot\lambda},$$

to obtain the graph H and set the weights of all sampled edges to be $1/p$. Prove that with high probability, H is a $(1\pm\varepsilon)$-cut sparsifier of G with

$$O(\frac{m\log n}{\varepsilon^2\cdot\lambda})$$

many edges. Note that for "large" values of λ, this approach indeed sparsifies the graph. **(25 points)**

Note: There is a fundamental graph structural result that is crucial for solving this problem: in any graph $G = (V, E)$ with minimum cut λ, and for any $\alpha \geqslant 1$, the number of cuts with size at most $\alpha\cdot\lambda$ is $O(n^{2\alpha})$. This is a generalization of the bound we proved earlier in the course that the number of minimum cuts is $O(n^2)$ (the aforementioned result extends this from exact minimum cuts to approximate ones). You can prove this generalization along the same lines of the proof for exact minimum cuts, but for this question, you can just directly use this fact without a proof.

Solution. Fix any cut $S \subseteq V$ in G and assume $|\delta_G(S)| = \alpha \cdot \lambda$ for some (real-valued) $\alpha \geqslant 1$. We have that

$$w(\delta_H(S)) = \frac{1}{p} \cdot \delta_H(S),$$

by the choice of weights in H. Thus, to prove $w(\delta_H(S))$ is a $(1 \pm \varepsilon)$-approximation of $\delta_G(S)$, we only need to prove $\delta_H(S)$ is a $(1 \pm \varepsilon)$-approximation of $p \cdot \delta_G(S)$.

Let $X_1, \ldots, X_{\alpha\lambda}$ be a random variables in $\{0, 1\}$ for edges of $\delta_G(S)$ where $X_i = 1$ iff i-th edge of the cut is sampled in H. We have,

$$\mathbb{E}[\delta_H(S)] = \mathbb{E}[X] = \sum_{i=1}^{\alpha\lambda} \mathbb{E}[X_i] = p \cdot \alpha \cdot \lambda = p \cdot \delta_G(S).$$

Thus, we only need to prove X is also concentrated around its expectation. Since X is a sum of 0/1 independent random variables, we can apply Chernoff bound and have,

$$\Pr(|X - \mathbb{E}[X]| > \varepsilon \cdot \mathbb{E}[X]) \leqslant 2 \cdot \exp\left(-\frac{\varepsilon^2 \cdot \mathbb{E}[X]}{3}\right) \leqslant 2 \cdot \exp\left(-\frac{\varepsilon^2 \cdot \frac{100 \log n}{\varepsilon^2 \cdot \lambda} \cdot \alpha\lambda}{3}\right) < n^{-10\alpha}.$$

As the total number of cuts of size $\alpha \cdot \lambda$ is at most $O(n^{2\alpha})$ (by the note in the question statement), by union bound, the probability that even one cut of size $\alpha \cdot \lambda$ is not a $(1 \pm \varepsilon)$-approximation in H is at most $O(n^{-8\alpha}) < n^{-7}$. As there can be at most n^2 different values for α (as the size of the cut is a number between 1 to $O(n^2)$), we obtain that probability at least $1 - n^{-5}$, H is a cut sparsifier of G.

As for the number of edges, denoted by Y, since we sample each edge independently with probability p we have,

$$\mathbb{E}[Y] = p \cdot m = \frac{m \cdot 100 \log n}{\varepsilon^2 \cdot \lambda}.$$

Moreover, as Y is a sum of 0/1 independent random variables (one for choice of each edge being sampled or not), by Chernoff bound,

$$\Pr(|Y - \mathbb{E}[Y]| \geqslant 2 \cdot \mathbb{E}[Y]) \leqslant \exp(-2\,\mathbb{E}[Y]) \ll 1/\text{poly}(n).$$

Thus, we also obtain the bound on the number of edges with high probability.

Problem 3. In this question, we prove two other simple results from *random graph theory* (for a slightly different family of random graphs than the ones studied in the lecture). Let $G = (V, E)$ be a graph on n vertices obtained by adding an edge between each pair of vertices independently and with probability

$$p := \frac{100 \log n}{n}.$$

(*a*) Prove that with high probability G is connected. **(10 points)**

Solution. For G to not be connected, we need to have at least one cut $(S, V \setminus S)$ with no edges inside it. Moreover, without loss of generality we can assume $|S| \leqslant n/2$ (as the cut is symmetric in undirected graphs and we can always focus on the smaller side of the cut). Thus, we will focus on proving that no such cut exists with high probability.

Fix any set $S \subseteq V$ with $1 \leqslant |S| \leqslant n/2$. We have,

$$\begin{aligned}
\Pr(\text{cut } S \text{ has no edges in } G) &= \prod_{u \in S, v \in V \setminus S} (1-p) \\
&\quad \text{(none of the pairs of vertices between } S \text{ and } V \setminus S \text{ can become an edge in } G) \\
&= (1-p)^{|S| \cdot (n-|S|)} \\
&\leqslant \exp\left(-p \cdot |S| \cdot \frac{n}{2}\right) \qquad \text{(as } 1-x \leqslant e^{-x} \text{ and } n-|S| \geqslant n/2) \\
&= \exp\left(-\frac{100 \log n}{n} \cdot |S| \cdot \frac{n}{2}\right) \qquad \text{(by the choice of } p) \\
&\leqslant n^{-50|S|}.
\end{aligned}$$

We thus have,

$$\begin{aligned}
\Pr(\text{there is a cut } S \text{ with no edges in } G) &\leqslant \sum_{s=1}^{n/2} \sum_{S \subseteq V : |S| = s} n^{-50s} \\
&\quad \text{(by union bound and the calculation above)} \\
&\leqslant n \cdot \binom{n}{2} \cdot n^{-50s} < n^{-25},
\end{aligned}$$

concluding the proof.

(*b*) Prove that with high probability the *chromatic number* of G is $\Omega(\frac{\log n}{\log \log n})$. Recall that the chromatic number is the minimum number of colors we can assign to the vertices so that no edge receives the same color on both its endpoints. **(15 points)**

Hint: Try to upper bound the size of the largest *independent set* in G first and then deterministically relate that to the chromatic number.

Solution. Any k-coloring of a graph G implies that G also has an independent set of size at least n/k. This is because each color-class is an independent set and by applying pigeonhole principle. Thus, we are going to prove that the largest independent set in G is of size

$$s = \frac{n \cdot \log \log n}{\log n},$$

with high probability, which immediately implies the given lower bound on the chromatic number.

Let $S \subseteq V$ be any set of s vertices. We have,

$$\begin{aligned}
\Pr(S \text{ is an independent set in } G) &= \prod_{u,v \in S} (1-p) \\
&\quad \text{(none of the pairs of vertices inside } S \text{ can become an edge in } G) \\
&= (1-p)^{\binom{s}{2}} \\
&\leqslant \exp\left(-p \cdot \frac{s^2}{4}\right) \qquad \text{(as } 1-x \leqslant e^{-x} \text{ for } x \in (0,1)) \\
&= \exp\left(-\frac{25 \log n}{n} \cdot \left(\frac{n \cdot \log \log n}{\log n}\right)^2\right) \\
&= \exp\left(-\frac{25n \cdot (\log \log n)^2}{\log n}\right).
\end{aligned}$$

We can now union bound over all sets S of s vertices and have,

$$\begin{aligned}
\Pr\left(G \text{ has an independent set of size } s\right) &\leqslant \binom{n}{s} \cdot \exp\left(-\frac{25n \cdot (\log\log n)^2}{\log n}\right) \\
&\leqslant \left(\frac{e \cdot n}{s}\right)^s \cdot \exp\left(-\frac{25n \cdot (\log\log n)^2}{\log n}\right) \quad (\text{as } \tbinom{a}{b} \leqslant ((e \cdot a)/b)^b) \\
&\leqslant (e \cdot \log n)^s \cdot \exp\left(-\frac{25n \cdot (\log\log n)^2}{\log n}\right) \\
&\leqslant \exp\left(2 \cdot \frac{n \log\log n}{\log n} \cdot \log\log n - \frac{25n \cdot (\log\log n)^2}{\log n}\right) \\
&\leqslant \exp\left(-\frac{10n \cdot (\log\log n)^2}{\log n}\right) \ll 1/\text{poly}(n).
\end{aligned}$$

If G has no independent set of size s, it cannot have an independent set of size $\geqslant s$ either, finalizing the proof.

Problem 4. In this question, we design another simple algorithm for MSTs with runtime better than the classical algorithms (although not as good as the advanced ones we studied). Recall the following two facts:

- Each round of Boruvka's algorithm takes $O(m)$ time and reduces the number of vertices by at least a half.
- Prim's algorithm can be implemented in $O(m + n \log n)$ time using Fibonacci heaps.

Combine these two algorithms in a careful way to obtain an $O(m \log\log n)$ time algorithm for MSTs.

(25 points)

Solution. The algorithm is quite simple.

- Run $k = \log\log n$ rounds of Boruvka's algorithm first which takes $O(mk) = O(m \log\log n)$ time and reduces the number of vertices from n to at most $n/2^k = n/\log n$ (as proven in the class, each round of Boruvka's algorithm reduces the number of vertices by half at least).
- Then, run Prim's algorithm on this new graph using Fibonacci heaps which takes $O(m' + n' \log(n'))$ time where m', n' are the number of edges and vertices in this graph. Since,

$$m' \leqslant m \qquad n' \leqslant \frac{n}{\log n} \qquad \log(n') \leqslant \log(n),$$

  we get that the runtime is $O(m + n)$ in this step.

Hence, the total runtime is $O(m \log\log n)$ as desired. The proof of correctness is simply by the correctness of Boruvka's and Prim's algorithms, concluding the proof.
