ADVANCED ALGORITHM DESIGN
Homework 1

September 29, 2024

Problem 1

In class, we saw that, when hashing m items into a hash table of size $\mathcal{O}(m^2)$, the expected number of collisions was < 1. In particular, this meant we could easily find a "perfect" hash function of the table that has no collisions.

Consider the following alternative scheme: build two tables, each of size $\mathcal{O}(m^{1.5})$ and choose a separate hash function for each table independently. To insert an item, hash it to one bucket in each table and insert it only in the emptier bucket (tie-break lexicographically).

(a) Show that, if we're hashing m items, with probability $\frac{1}{2}$, there will be no collisions in either table (a collision occurs when multiple distinct elements are inserted into the same bucket in the same table). You may assume a fully random hash function.

(b) Modify the above scheme to use $\mathcal{O}(\log m)$ tables, each of size $\mathcal{O}(m)$. Prove again that with probability $\frac{1}{2}$, there will be no collisions in any table. Again, you may assume a fully random hash function.

Solution

(a) Let $X = \{x_1, \cdots, x_m\}$ be the elements to be hashed, h, g are the hash functions, and $n = m\sqrt{2m}$ be the size of each hash table. Let E_i be the event that x_i collides. Then $E_i \subseteq F_i$ where F_i is the event that $\exists x_j, x_k$ such that $h(x_i) = h(x_j)$ and $g(x_i) = g(x_k)$. Since g, h are independent, $\mathbb{P}\left[E_i\right] \leq \mathbb{P}\left[\exists x_j \text{ such that } h(x_i) = h(x_j)\right]^2$. Let's take $E_1 \subseteq F_1$ and estimate the latter probability without the square assuming a random hash function. The chance that x_1 collides with a fixed x_j (under h) is $\leq \frac{1}{n}$, so by union bound $\mathbb{P}\left[E_1\right] \leq \left(\frac{m-1}{n}\right)^2 \leq \frac{1}{2m}$. Again by union bound $\mathbb{P}\left[\cup_i E_i\right] \leq \frac{1}{2}$ whence $\mathbb{P}\left[\cap_i E_i^c\right] \geq \frac{1}{2}$. $\cap_i E_i^c$ is precisely the event that there is no collision for x_i for each i.

(b) Now we have m elements in X, $r = \lceil d \ln m \rceil$ tables each of size $n = \lceil cm \rceil$ where $c = e, d = 2$. let the hash functions be $h_1, \cdots, h_r$. Again define E_i as the event that x_i collides and F_i as the event that $\exists x_{j_1}, \cdots, x_{j_r}$ such that $h_k(x_i) = h_k(x_{j_k})$ for each $1 \leq k \leq r$. By a similar argument as the previous part, $\mathbb{P}\left[F_j\right] = \mathbb{P}\left[\exists x_j \text{ s.t. } h_1(x_i) = h_1(x_j)\right]^r$, due to independence. Then

$$\mathbb{P}\left[F_1\right] \leq \left(\frac{m-1}{n}\right)^r = c^{-r}\left(1-\frac{1}{m}\right)^r \leq \frac{1}{c^{d\ln m}} \cdot \exp\left\{-\frac{d\ln m}{m}\right\} = m^{-2} \cdot m^{\frac{2}{m}} \leq \frac{1}{2m}$$

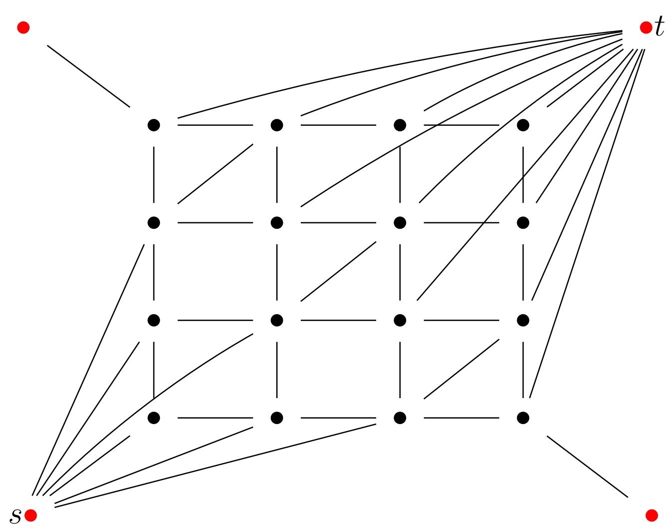
for $m > 100$, because $m^{\frac{2}{m}} \xrightarrow{m\to\infty} 1$ and is eventually decreasing. By union bound and taking complement $\mathbb{P}\left[\cap_i F_i^c\right] = 1 - \mathbb{P}\left[\cup_i F_i\right] \geq 1 - \frac{1}{2} = \frac{1}{2}$.

Problem 2

Prove that (the natural variant of) Karger's algorithm does not work for finding the minimum $s-t$ cut in unweighted, undirected graphs. Specifically, design an unweighted, undirected graph G (with no parallel edges), with two nodes s, t, such that repeatedly contracting a random edge that does not contract s and t to the same supernode outputs a minimum $s-t$ cut with probability $2^{-\Omega(n)}$.

Solution

Consider the following graphs parameterized by even k. We draw $4k^2$ nodes $X = (\mathbb{Z} \cap [1, 2k])^2$, add four nodes $(0,0), (0, 2k+1), (2k+1, 0), (2k+1, 2k+1)$, and take $s = (0,0), t = (0, 2k+1, 2k+1)$. So there are $n = 4k^2 + 4$ nodes. We draw edges as follows. For each $(i, j) \in X$ with $i + j \geq 2k + 1$, connect it with t (then t has $k(2k+1)$ neighbors). For each $(i, j) \in X$ with $i + j \leq 2k$, connect it with s (then s has $k(2k-1)$). For each $(i, j) \in X$, connect it with each of its neighbors in X, that is those elements which differ in each coordinate by at most 1 (this gives $8k^2 - 4k$ edges). Connect each $(2k - i, i)$ to $(2k - i + 1, i + 1)$ for each $i = 1, \cdots, 2k - 1$ (this gives another $2k - 1$ edges). Connect $(0, 2k+1)$ to $(1, 2k)$ and connect $(2k+1, 0)$ to $(2k, 1)$ (two more edges). So the total number of edges is $m = 12k^2 - 2k + 1$. The picture for $k = 2$ is as follows.



It's not hard to see from the structure of the graph that the minimum $s-t$ cut value of such a graph is $c = k(2k-1)$ and the unique such cut is $C = (\{s\}, \overline{\{s\}})$. Clearly $m \geq 12k^2 = 6 \cdot 2k^2 \geq 2 \cdot (2k-1) \cdot k = 6c$.

Now we use a crucial observation. Say Karger's (modified) algorithm has performed i contractions where C has survived so far. Here's some terminology. Sequentially while the algorithm contracts edges, we will call an edge *good* if it is adjacent to s, and we will call an edge *valid* if it is not adjacent to both s and t. For example, in the above graph, $\{t, 1\}$ is a valid but not good, $\{s, 1\}$ is both valid and good. If we contract $\{1, t\}$ to form a super-node, so that now s is adjacent to this supernode $1, t$, then the edge $\{s, \{1, t\}\}$ is good but not valid. Say after i steps, conditioned on the fact that C has survived so far, there are z_i *valid* (multi-)edges remaining and k_i out of them are valid and good edges (i.e., adjacent to s but not t), then the probability of survival of C after another *valid* contraction (i.e. contracting a *valid edge*) is $1 - \frac{k_i}{z_i}$. Now we observe that after $i-1$ steps, the contracted edge may or may not make t a neighbor of s. If t becomes a neighbor of s then $(k_i, z_i) = (k_{i-1} - 1, z_{i-1} - 2)$, otherwise $(k_i, z_i) = (k_{i-1}, z_{i-1} - 1)$. In either case $z_i - k_i = z_{i-1} - k_{i-1} - 1$. So the probability that C survives after i^{th} round is $\frac{z_i - k_i}{z_i}$ which is either $\frac{z_{i-1} - k_{i-1} - 1}{z_{i-1} - 1}$ or $\frac{z_{i-1} - k_{i-1} - 1}{z_{i-1} - 2}$. In particular this probability is $\leq \frac{z_{i-1} - k_{i-1} - 1}{z_{i-1} - 2}$. So the probability that C survives after r rounds is $\leq \prod_{i=0}^{r-1} \frac{c-i}{m-2i} \leq \left(\frac{c-r}{m-2r}\right)^r$. Taking $r = \frac{c}{2}$, the probability of survival of C is $\leq \left(\frac{c}{2(m-c)}\right)^{c/2} \leq \left(\frac{c}{10c}\right)^{k^2} = 10^{-k^2} = 2^{-\Omega(n)}$.

Problem 3

In this problem, we investigate whether an algorithm can compute the median of a given set of numbers when it can only access the input set via independent samples.

(a) Let A be an algorithm the input to which is a collection of m independent samples drawn uniformly at random from an arbitrary set $X = \{x_1, \cdots, x_n\}$. Prove that if $m = o(n)$ then A must fail compute the median of X within a multiplicative error of 1.1 with probability at least $1/3$. That is, any algorithm which (possibly randomly) maps $m = o(n)$ samples to a guess at the median is off by a factor of at least 1.1 with probability at least $1/3$.

(b) Say we relax the goal and ask for the algorithm above to output a number y such that at least $n/2 - t$ elements of X exceed y, and $n/2 - t$ numbers are less than y. Prove that if we take $m = \mathcal{O}(n^2 \ln(1/\delta)/t^2)$ samples, and let y denote the median of the m samples, then y has this property with probability at least $1 - \delta$.

Solution

(a) <u>Deterministic:</u>

First assume that we are only looking at a *deterministic* algorithm. Let $f : X^m \to \mathbb{R}$ be this algorithm. Note that the algorithm will output the same value if two inputs differ only by a permutation. Let $\varepsilon = \frac{1}{10}$. Consider the multiset $S = \{n \text{ many } 0's, n \text{ many } 1's\}$. And let the two lists be $S_1 = S \cup \{\varepsilon\}, S_2 = S \cup \{1-\varepsilon\}$. Clearly their medians are $\varepsilon, 1-\varepsilon$ respectively. We sample m numbers from each list and feed it to the algorithm. Say we sample $X_1^{(i)}, \cdots, X_m^{(i)}$ from S_i. Let G_i be the event that $X_j^{(i)} \in S \;\forall\; 1 \le j \le m$. Then for $i = 1, 2$, $\mathbb{P}[G_i] = (\frac{2n}{2n+1})^m = (1 - \frac{1}{2n+1})^m \ge 1 - \frac{m}{2n+1} = 1 - o(1)$ as $n \to \infty$ because $m = o(n)$.

Let A_1 be the event that f on m samples from S_1 gives answer $\in (\varepsilon/1.1, 1.1\varepsilon)$. Let A_2 be the event that f on m samples from S_2 gives answer $\in ((1-\varepsilon)/1.1, 1.1(1-\varepsilon))$ Then $\mathbb{P}[A_i] = \mathbb{P}[A_i \mid G_i]\,\mathbb{P}[G_i] + \mathbb{P}[A_i \mid G_i^c]\,\mathbb{P}[G_i^c]$ for $i = 1, 2$. But $\mathbb{P}[G_1] = \mathbb{P}[G_2]$. So $\mathbb{P}[A_1] + \mathbb{P}[A_2] = \mathbb{P}[A_1 \mid G_1] + \mathbb{P}[A_2 \mid G_2] +$
$\mathbb{P}[G_1^c]\left(\mathbb{P}[A_1 \mid G_1^c] + \mathbb{P}[A_2 \mid G_2^c] - \mathbb{P}[A_1 \mid G_1] - \mathbb{P}[A_2 \mid G_2]\right)$.

We denote $\boldsymbol{X}^{(i)} := (X_1^{(i)}, \cdots, X_m^{(i)})$ and $\boldsymbol{x}$ for a vector in X^m. Now $\mathbb{P}[A_1 \mid G_1] = \sum_{\boldsymbol{x} \in S^m} \mathbb{P}\left[A_1 \mid G_1, \boldsymbol{X}^{(1)} = \boldsymbol{x}\right] \mathbb{P}\left[\boldsymbol{X}^{(1) \mid G_1} = \boldsymbol{x}\right] = \sum_{\boldsymbol{x} \in S^m} \mathbb{P}[f(\boldsymbol{x}) \in (1/1.1, 1.1)\varepsilon]\,\mathbb{P}\left[\boldsymbol{X}^{(1)} = \boldsymbol{x} \mid G_1\right]$.
The sum runs over only S^m because we've already conditioned on G_1. Similarly $\mathbb{P}[A_2 \mid G_2] = \sum_{\boldsymbol{x} \in S^m} \mathbb{P}[f(\boldsymbol{x}) \in (1/1.1, 1.1)(1-\varepsilon)]\,\mathbb{P}\left[\boldsymbol{X}^{(2)} = \boldsymbol{x} \mid G_2\right]$. Trivially for each $\boldsymbol{x} \in S^m$, $\mathbb{P}\left[\boldsymbol{X}^{(1)} = \boldsymbol{x} \mid G_1\right] = \mathbb{P}\left[\boldsymbol{X}^{(2)} = \boldsymbol{x} \mid G_2\right]$. Also for each $\boldsymbol{x} \in S^m$, the events $\{f(\boldsymbol{x}) \in (1/1.1, 1.1)\varepsilon\}, \{f(\boldsymbol{x}) \in (1/1.1, 1.1)(1-\varepsilon)\}$ are disjoint whence $\mathbb{P}[A_1 \mid G_1] +$

$\mathbb{P}[A_2 \mid G_2] \leq 1$. It follows that $\mathbb{P}[A_1] + \mathbb{P}[A_2] \leq 1 + o(1)$. **In short**, what this paragraph is saying is that, conditioned on the fact that all samples are picked only from S, the algorithm's answer can't be correct together for S_1, S_2 which is why the sum of those probabilities is ≤ 1.

So, WLOG, $\mathbb{P}[A_1^c] \geq \frac{1}{2} - o(1)$, that is, the algorithm fails to estimate (within multiplicative error 1.1) the median of S_1 with at least, say, $\frac{1}{3}$ probability.

Randomized:

Now we argue why there cannot be a *randomized* algorithm for the given task. Suppose there is some randomized algorithm f which performs the given task with high probability. Since f is randomized, for every input $\boldsymbol{x} \in X^m$, it will pick one of many *deterministic* algorithms $f_1, \cdots, f_k$, with respective probabilities $p_1, \cdots, p_k$, and output $f_i(\boldsymbol{x})$ with probability p_i. In the deterministic case, we proved that for any deterministic algorithm the probability of existence of a list in which the algorithm fails to give an estimate within the given error on $m = o(n)$ samples at least for $1/3$. Let E be the event that f fails to give a good estimate via m samples from X. Note that each f_i is deterministic, whence $\mathbb{P}[E \mid f = f_i] \geq 1/3$. Consequently $\mathbb{P}[E] = \sum_{i=1}^k \mathbb{P}[E \mid f = f_i] p_i \geq \frac{1}{3}\sum_{i=1}^k p_i = \frac{1}{3}$.

(b) We assume the ratio of t to n is small, say $t < n/10$. We sample numbers $Z_1, \cdots, Z_m$, where $m = \frac{12n^2 \ln(1/\delta)}{t^2}$ and $\delta < \frac{1}{2}$, from X. Let M be the true median of X and Z be the median of the samples. Let $a, b \in X$ be such that $\frac{n}{2} - t$ elements in X are $< a$ and $\frac{n}{2} - t$ elements in X are $> b$. Let $L = X \cap [0, a), R = X \cap (b, 1]$. We want that Z lies in $L \cup R$ with low probability. Consider the indicators $I_i = \begin{cases} 1 & \text{if } Z_i \in L \\ 0 & \text{otherwise} \end{cases}$.
Then $\mathbb{E}[I_i] \leq \frac{1}{2} - \frac{t}{n}$. The random variable $X_L = \sum_{i=1}^m I_i$ counts the number of samples in L, and $\mathbb{E}[X_L] \leq m\left(\frac{1}{2} - \frac{t}{n}\right)$. Note that the I_i are all iid. By Chernoff with $\varepsilon = \frac{t}{n}$, $\mathbb{P}\left[X_L \geq \frac{m}{2}\right] \leq \mathbb{P}[X_L > (1+\varepsilon)\mathbb{E}[X_L]] \leq 2\exp\left\{-\frac{\varepsilon^2 m}{3(1+\varepsilon)}\right\} \leq \exp\left\{-\frac{t^2 m}{6n^2}\right\} = \delta^2 < \frac{\delta}{2}$. By a similar argument, if we define X_R in a similar way as above, we will have $\mathbb{P}\left[X_L \geq \frac{m}{2}\right] \leq \frac{\delta}{2}$. So the probability that more than $\frac{m}{2}$ sampled elements come from L and more than $\frac{m}{2}$ sampled elements come from R is at least $1 - \delta$ (use union bound for $\{X_L > m/2\} \cup \{X_R > m/2\}$). Hence, the chance that the sample median lies in $X \setminus (L \cup R)$ is $\geq 1 - \delta$.

Problem 4

A cut is said to be a $B-$*approximate* min cut if the number of edges in it is at most B times that of the minimum cut. Show that all undirected graphs have at most $(2n)^{2B}$ cuts that are $B-$approximate.

Solution

We argue as the hint. Let $m = \lceil 2B \rceil$. We stop Karger's algorithm when m supernodes remain, and output a random cut among them. Say c is the actual mincut value. Consider a particular $B-$approximate mincut $(S, \overline{S})$ and we compute its probability of **survival** (when stopped at m supernodes). From the analysis given in notes, this probability is at least $\prod_{i=1}^{n-m} \left(1 - \frac{Bc}{(n-i+1)c/2}\right)$. This gives $\prod_{i=1}^{n-m} \left(1 - \frac{Bc}{(n-i+1)c/2}\right) = \frac{n-2B}{n} \cdot \frac{n-1-2B}{n-1} \cdots \frac{m+1-2B}{m+1} = \frac{(m!) \prod_{i=m+1}^{n} (i-2B)}{n!} \geq \frac{m!}{n^{2B}}$. The probability that it is output by the algorithm, in the final step, is $\frac{2}{2^m-2}$, because there are a total of $2^m - 2$ nontrivial choices of vertex subsets and feasible outputs are $(S, \overline{S}), (\overline{S}, S)$. The chance that $(S, \overline{S})$ is output by the algorithm is $\geq \frac{2m!}{(2^m-2)n^{2B}} \geq \frac{m!}{(2^{m-1}-1)n^{2B}} \geq \frac{1}{2^{2B}n^{2B}}$.

Let $C_1, \cdots, C_k$ (where $C_i = (X_i, \overline{X_i})$) be the distinct $B-$approximate mincuts in the given graph, and let E_i be the event that C_i is output by the algorithm. By the above calculation $\mathbb{P}[E_i] \geq (2n)^{-2B}$. So $1 \geq \mathbb{P}[\cup_i E_i] = \sum_i \mathbb{P}[E_i] \geq k(2n)^{-2B}$ whence $k \leq (2n)^{2B}$.

Problem 5

Consider the following process for matching n jobs to n processors. In each step, every job picks a processor at random. The jobs that have no contention on the processors they picked get executed, and all the other jobs back off and then try again. Jobs only take one round of time to execute, so in every round all the processors are available. Show that all the jobs finish executing after $\mathcal{O}(\log\log n)$ steps, with high probability.

Solution

Claim 1
If instead we had m jobs and n processors, the expected number of completed jobs in the first round is $m\left(1-\frac{1}{n}\right)^{m-1}$.

Proof. Let I_i be the random variable that takes value 1 if i^{th} job is completed, and 0 otherwise. They have the same density. A particular job is completed in a round iff it is matched to a processor distinct from the processors that the other jobs get. The chance that the i^{th} job is assigned to a different processor than that of a given job, say j, is exactly $1-\frac{1}{n}$. By independence, $\mathbb{E}\left[I_i\right]=\mathbb{P}\left[I_i=1\right]=\left(1-\frac{1}{n}\right)^{m-1}$. Conclude by linearlity of expectation. Just to use later, we will keep in mind that $\text{Var}\left[I_i\right]=\left(1-\frac{1}{n}\right)^{m-1}\left(1-\left(1-\frac{1}{n}\right)^{m-1}\right)$. ■

Corollary 2
Say m jobs remain at the beginning of some round. Then the expected number of jobs remaining after one more round is $m\left[1-\left(1-\frac{1}{n}\right)^{m-1}\right]$.

Claim 3
Say $m\leq\sqrt{n}$ jobs remain at the beginning of some round. Then, with high probability, all jobs are completed.

Proof. Let X be the random variable denoting the number of remaining jobs after the round starting with m jobs. By corollary 2, $\mathbb{E}\left[X\right]=m\left[1-\left(1-\frac{1}{n}\right)^{m-1}\right]$. From claim 1, taking $J_i=1-I_i$ we observe that $X=\sum J_i$. Independence of J_i's implies that $\text{Var}\left[X\right]=\sum\text{Var}\left[J_i\right]=m\cdot\text{Var}\left[J_1\right]=m\left(1-\frac{1}{n}\right)^{m-1}\left(1-\left(1-\frac{1}{n}\right)^{m-1}\right)$. Chebyshev gives $\mathbb{P}\left[X\geq 1\right]\leq$

$$\mathbb{P}\left[X\geq\mathbb{E}\left[X\right]+\frac{1}{\sqrt{rn}}\right]\leq\frac{m}{n}\underbrace{\left(1-\frac{1}{n}\right)^{m-1}}_{\leq e^{-\frac{m-1}{n}}}\left(1-\underbrace{\left(1-\frac{1}{n}\right)^{m-1}}_{\geq 1-\frac{m-1}{n}}\right)\leq\frac{m(m-1)}{n^2}\cdot e^{-\frac{m-1}{n}}\leq\frac{m^2}{n^2}\cdot 1\leq\frac{1}{n}.$$

The first inequality is true since $\mathbb{E}\left[X\right]+\frac{1}{\sqrt{n}}\leq\frac{m-1}{n}+\frac{1}{\sqrt{n}}=\frac{m-1+\sqrt{n}}{n}\overset{(*)}{\leq}\frac{2\sqrt{n}-1}{n}\leq 1$. $(*)$ is true as $n-2\sqrt{n}+1=(\sqrt{n}-1)^2\geq 1\implies 1\geq\frac{2\sqrt{n}-1}{n}$. So $\mathbb{P}\left[X=0\right]=1-\mathbb{P}\left[X\geq 1\right]\geq 1-\frac{1}{n}$. ■

Claim 4
If $m = \alpha n$ jobs remain at the beginning of some round (this is the conditioning on a particular round), with $\frac{1}{\sqrt{n}} < \alpha < 1$ (so that $\sqrt{n} < m < n$), then after one round, with probability $1 - \frac{1}{n}$, at most $3\alpha^2 n/2$ jobs remain.

Proof. We denote $[t] \coloneqq \{1, \cdots, t\}$ for a positive integer t.
Let P_i (for $1 \le i \le m$) be the random variable, taking values in $\{1, \cdots, n\}$, which denotes the processor assigned to job i. Denote the tuples $\boldsymbol{Q}_i = (P_1, \cdots, P_i)$ and for simplicity $\boldsymbol{Q} = \boldsymbol{Q}_m$. Note that P_i are iid $\sim$ Uniform$(1, n)$. Define a function $f : [n]^m \to \mathbb{R}$ given by $f(x_1, \cdots, x_m) = \#\{i \in [m] \mid \exists j \in [m] \smallsetminus \{i\} \text{ for which } x_i = x_j\}$. That is, $f(x_1, \cdots, x_m)$ counts the number of repeated elements. For example if $m = 5$ then $f(1, 1, 2, 3, 7) = 2$ and $f(6, 2, 3, 4, 8) = 0$. So $f(\mathbb{Q})$ is the number of uncompleted jobs. We compute $\mu \coloneqq \mathbb{E}\left[f(\boldsymbol{Q})\right] = m\left(1 - \left(1 - \frac{1}{n}\right)^{m-1}\right) \simeq m \cdot \frac{m-1}{n} \simeq \alpha^2 n$. In fact, $\alpha^2 n$ is only slightly more that $\mu = m\left(1 - \left(1 - \frac{1}{n}\right)^{m-1}\right)$ because $\left(1 - \frac{1}{n}\right)^{m-1} \ge 1 - \frac{m-1}{n}$ and the difference is $\simeq \frac{\mu}{\alpha n}$.

Define random variables $R_i \coloneqq \mathbb{E}\left[f(\boldsymbol{Q}) \mid \boldsymbol{Q}_i\right]$ and $R_0 \coloneqq \mathbb{E}\left[f(\boldsymbol{Q})\right]$. Note $R_m = \mathbb{E}\left[f(\boldsymbol{Q}) \mid \boldsymbol{Q}\right] = f(\boldsymbol{Q})$. Using the moment-method e^{tX} along with Markov inequality gives

$$\mathbb{P}\left[f(\boldsymbol{Q}) - \mu \ge \varepsilon\right] \le e^{-t\varepsilon}\mathbb{E}\left[e^{t(f(\boldsymbol{Q})-\mu)}\right].$$

To bound the RHS, let's first note that $f(\boldsymbol{Q}) - \mu = R_m - R_0 = \sum_{i=1}^{m}(R_i - R_{i-1})$. Recall the property of expectation that $\mathbb{E}\left[h(X, Y)\right] = \mathbb{E}_X\left[\mathbb{E}_Y\left(h(X, Y) \mid X\right)\right]$ where we will take $h = \exp$. We have the following calculation (the last line in the following is due to independence of P_i's)

$$\begin{aligned}
\mathbb{E}\left[e^{t(f(\boldsymbol{Q})-\mu)}\right] &= \mathbb{E}\left[\exp\left\{t\sum_{i=1}^{m}(R_i - R_{i-1})\right\}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\exp\left\{t\sum_{i=1}^{m}(R_i - R_{i-1})\right\} \;\middle|\; \boldsymbol{Q}_{m-1}\right]\right] \\
&= \mathbb{E}\left[\exp\left\{t\sum_{i=1}^{m-1}(R_i - R_{i-1})\right\}\mathbb{E}\left[\exp\left\{t\left(R_m - R_{m-1}\right)\right\} \mid \boldsymbol{Q}_{m-1}\right]\right] \\
&\le \alpha_m \mathbb{E}\left[\exp\left\{t\sum_{i=1}^{m-1}(R_i - R_{i-1})\right\}\right] \le \cdots \le \prod_{i=1}^{m}\alpha_i
\end{aligned}$$

where α_i is an upper bound for $\mathbb{E}\left[\exp\left\{t\left(R_i - R_{i-1}\right)\right\} \mid \boldsymbol{Q}_{i-1}\right]$. Let's try to get α_i's. First note that if the jobs-to-processor assignment were to be changed for just one job (that is change just one P_i) then the number of collisions changes (increases or decreases) at most by 5 (5 is just something arbitrarily big). More formally, $|f(b, x_2 \cdots, x_m) - f(a, x_2 \cdots, x_n)| \le 5$ for any $x_2, \cdots, x_n, b, a \in [n]$ and this is true for all coordinates i, not just $i = 1$. It's also true that

$$\begin{aligned}\mathbb{E}\left[R_i - R_{i-1} \;\mid\; \boldsymbol{Q}_{i-1}\right] &= \mathbb{E}\left[\mathbb{E}\left[f(\boldsymbol{Q}) \;\mid\; \boldsymbol{Q}_i\right] - \mathbb{E}\left[f(\boldsymbol{Q}) \;\mid\; \boldsymbol{Q}_{i-1}\right] \;\mid\; \boldsymbol{Q}_{i-1}\right] \\ &= \mathbb{E}_{P_i,\cdots,P_m}\Big(\mathbb{E}\left[f(\boldsymbol{Q}) \;\mid\; \boldsymbol{Q}_i\right] - \mathbb{E}\left[f(\boldsymbol{Q}) \;\mid\; \boldsymbol{Q}_{i-1}\right]\Big) \\ &= \mathbb{E}_{P_i,\cdots,P_m}\Big(\mathbb{E}\left[f(\boldsymbol{Q}) \;\mid\; \boldsymbol{Q}_i\right]\Big) - \mathbb{E}_{P_i,\cdots,P_m}\Big(\mathbb{E}\left[f(\boldsymbol{Q}) \;\mid\; \boldsymbol{Q}_{i-1}\right]\Big) \\ &= \mathbb{E}_{P_i,\cdots,P_m}\Big(\mathbb{E}\left[f(\boldsymbol{Q}) \;\mid\; \boldsymbol{Q}_{i-1}\right]\Big) - \mathbb{E}_{P_i,\cdots,P_m}\Big(\mathbb{E}\left[f(\boldsymbol{Q}) \;\mid\; \boldsymbol{Q}_{i-1}\right]\Big) = 0\end{aligned}$$

By Hoeffding's lemma (`https://en.wikipedia.org/wiki/Hoeffding%27s_lemma`), $\mathbb{E}\left[\exp\left\{t\left(R_i - R_{i-1}\right)\right\} \;\mid\; \boldsymbol{Q}_{i-1}\right] \le \exp\left\{25t^2/8\right\}$ (the difference in supremum and infimum of the required random variables is at most 5 because changing the allocation of one job changes at most 5 in the number of uncompleted jobs.) Combining the above we get

$$\mathbb{P}\left[f(\boldsymbol{Q}) - \mu \ge \varepsilon\right] \le e^{-t\varepsilon}\prod_{i=1}^{m} e^{\frac{25t^2}{8}} = \exp\left\{-t\varepsilon + \frac{ct^2m}{2}\right\}$$

where $c = \frac{25}{16}$. This is true for all t. Optimizing over t, the best RHS is $\exp\left\{-\frac{\varepsilon^2}{2mc}\right\}$. Picking $\varepsilon = \sqrt{2cm\ln n}$ and noting that $\mu/2 \simeq \alpha^2 n/2 \ge \sqrt{2cm\ln n}$ (because $\ln n = o(n)$) we get

$$\mathbb{P}\left[\#\text{unfinished jobs} \ge 2\alpha^2 n\right] \le \mathbb{P}\left[\#\text{unfinished jobs} \ge \mu + \sqrt{2cm\ln n}\right] \le \frac{1}{n}.$$

■

The expected number of remaining jobs after just first round is $n\left(1 - \left(1 - \frac{1}{n}\right)^{n-1}\right) \simeq n\left(1 - \frac{1}{e}\right)$. It is easy to note that $\frac{3}{2}\alpha^2 = K\alpha^2 < \alpha$ for $\alpha_0 = 1 - \frac{1}{e}$. Due to the above proof, this event also have high chance of occurrence, say with probability $\ge 1 - \frac{1}{n}$.

Formally we showed that if M_i is the random variable denoting the number of jobs remaining at the end of i^{th} round, then $\mathbb{P}\left[M_i \le \frac{1}{K}(K\alpha_0)^{2^{i-1}} n\right] \ge (1 - \frac{1}{n})^i$.

Let T denote the (least) number of steps for which $M_T \le \sqrt{n}$. By our probability bounds $\mathbb{P}\left[T \le \log_2 \log_{2\alpha_0} n\right] \ge (1 - \frac{1}{n})^{\log_2 \log_{2\alpha_0} n} \ge 1 - \frac{\log_2 \log_{2\alpha_0} n}{n} = 1 - o(1)$. Again, with probability $\ge 1 - \frac{1}{n}$, the remaining $M_T < \sqrt{n}$ jobs are completed in one more round. So $\mathbb{P}\left[\text{ all jobs completed in } \le \mathcal{O}(\log\log n) \text{ rounds }\right] \ge 1 - \frac{\mathcal{O}(\log\log n)}{n}$.

Problem 6

Consider the following random process: there are $n+1$ coupons $\{0, \cdots, n\}$. Each step, you draw a uniformly random coupon independently with replacement, and you repeat this until you have drawn all coupons in $\{1, \cdots, n\}$ (that is, you may terminate without ever drawing 0). Prove that, with probability at least $1 - \mathcal{O}(1/n)$, you draw the 0 coupon at most $\mathcal{O}(\log n)$ times.

Solution

Denote $[n] = \{1, \cdots, n\}$. Let X_i be the random variable denoting the number of steps to observe a new coupon in $[n]$ if i distinct coupons in $[n]$ have already been observed. Then the number of steps required to terminate is $X = \sum_{i=0}^{n-1} X_i$. X_i follows a geometric distribution with parameter $\frac{n-i}{n+1}$, so $\mathbb{E}[X_i] = \frac{n+1}{n-i}$. This means $\mathbb{E}[X] = (n+1)H_n$ where H_n is the harmonic sum $\sum_{i=1}^{n} \frac{1}{i}$. Using Chebyshev with $\varepsilon = (n+1)H_n$ we can get $\mathbb{P}[X > 2(n+1)H_n] \leq \frac{1}{\ln n}$ because $\mathbb{E}[X] = (n+1)H_n, \mathrm{Var}[X] = (n+1)((n+1)H_n - n)$.

Let Y denote the number of 0 coupons collected. Take $t = \lfloor 2(n+1)H_n \rfloor$. Let Y_i be the indicator denoting that 0 was picked in the i^{th} step, and $\mathbf{1}_{i \geq X}$ be the indicator denoting $i \leq X$. Then $Y = \sum_{i=1}^{\infty} Y_i \mathbf{1}_{i \leq X}$. Since X (and thus $\mathbf{1}_{i \leq X}$) and all Y_i have finite expectations and they are independent, we can write $\mathbb{E}[Y] = \sum_{i=1}^{\infty} \mathbb{E}[Y_i \mathbf{1}_{i \leq X}] = \mathbb{E}[Y_1] \sum_{i=1}^{\infty} \mathbb{E}[\mathbf{1}_{i \leq X}] = \sum_{i=1}^{\infty} \mathbb{E}[Y_i] \mathbb{E}[\mathbf{1}_{i \leq X}] = \mathbb{E}[Y_1] \cdot \mathbb{E}[X] = \frac{(n+1)H_n}{n+1} = H_n$.[1] With n large and $\varepsilon = 2$, Chernoff gives (for small constants c, c') $\mathbb{P}[Y > 10 \ln n] \leq \mathbb{P}\left[\sum_{i=1}^{t} Y_i > 10 \ln n\right] \mathbb{P}[X \leq t] + \frac{c}{\ln n} = \mathbb{P}\left[\sum_{i=1}^{t} Y_i > (1+\varepsilon) 2H_n\right] + \frac{c'}{\ln n} \leq e^{-2H_n} + \frac{c'}{\ln n} \leq \mathcal{O}\left(\frac{1}{n}\right)$. So $\mathbb{P}[Y \leq 10 \ln n] \geq 1 - \mathcal{O}\left(\frac{1}{n}\right)$.

[1] Another way to see this: $X - Y$ is the random variable denoting the number of steps to terminate if 0 was never collected. By a similar argument as before (each parameter just becomes $\frac{n-i}{n}$, so the individual expectations are $\frac{n}{n-i}$ for $0 \leq i < n$), the expected number of steps for such operation is nH_n. So $\mathbb{E}[Y] = H_n$.