School of Computer Science and Technology
University of Science and Technology of China

Exercise Sheet 2 for
Design and Analysis of Algorithms
Autumn 2022
Solution

Exercise 1 (40 points, graded by Yudong Zhang)

Suppose each CPU can execute at most one process at any time. Consider the following experiment, which proceeds in a sequence of rounds. For the first round, we have n processes, which are assigned independently and uniformly at random to n CPUs. For any $i \geq 1$, after round i, we first find all the processes p such that p has been assigned to a CPU C *by itself*, i.e., p is the *unique* process that has been assigned to C; then we remove all such processes (as they would be executed) in round i. The remaining processes are retained for round $i+1$, in which they are assigned independently and uniformly at random to the n CPUs.

(a) **(20 points)** If there are b processes at the start of a round, what is the expected number of processes at the start of the next round?

(b) **(20 points)** Suppose that every round the number of removed processes was exactly the expected number of removed processes. Show that all the processes will be removed in $O(\log\log n)$ rounds.

Hint: If x_j is the expected number of processes left after j rounds, show and use that $x_{j+1} \leq x_j^2/n$. You can use the fact that $1 - kx \leq (1-x)^k$ for $0 < x < 1$ and $k \leq \frac{1}{x}$.

Solution.

(a) Let Y be the number of CPUs that contain exactly one process after the random assignment in this round. Note that we only need to calculate $b - \mathrm{E}[Y]$.

Let CPU_i denote the i-th CPU, for any $1 \leq i \leq n$. We define X_i as follows:

$$X_i = \begin{cases} 1, & \text{if } CPU_i \text{ contains exactly 1 process} \\ 0, & \text{otherwise} \end{cases}$$

Then, we have

$$\Pr[X_i = 1] = \binom{b}{1} \cdot \frac{1}{n} \cdot \left(1 - \frac{1}{n}\right)^{b-1}$$

Also note that $\mathrm{E}[X_i] = 0 \cdot \Pr[X_i = 0] + 1 \cdot \Pr[X_i = 1] = \Pr[X_i = 1]$ and that $Y = \sum_{i=1}^{n} X_i$. Therefore,

$$\begin{aligned} \mathrm{E}[Y] = \mathrm{E}\left[\sum_{i=1}^{n} X_i\right] &= \sum_{i=1}^{n} \mathrm{E}[X_i] && \text{by linearity of expectation} \\ &= n \cdot b \cdot \frac{1}{n} \cdot \left(1 - \frac{1}{n}\right)^{b-1} \\ &= b\left(1 - \frac{1}{n}\right)^{b-1} \end{aligned}$$

Thus, the expected number of processes at the start of the next round is $b - \mathrm{E}[Y] = b - b\left(1 - \frac{1}{n}\right)^{b-1}$.

(b) First, we will show that $x_{j+1} \le x_j^2/n$. From the previous question, if $x_j = b$, we will have that $x_{j+1} = b - b\left(1 - \frac{1}{n}\right)^{b-1} \le b - b\left(1 - \frac{b-1}{n}\right) \le b^2/n = x_j^2/n$. The first inequality is derived from the fact that $1 - kx \le (1-x)^k$ for $0 < x < 1$ and $k \le 1/x$.

Now, we will show that all the processes will be removed in $O(\log\log n)$ rounds. In the first round, where $x_0 = n$, we have that $x_1 = n - n\left(1 - \frac{1}{n}\right)^{n-1} \le n - n \cdot \frac{1}{e} = n\left(1 - \frac{1}{e}\right)$ where the inequality follows from the fact that $\left(1 - \frac{1}{n}\right)^{n-1} \ge \frac{1}{e}$ for any $n \ge 2$. Letting $\frac{1}{c} = 1 - \frac{1}{e}$, we get $x_1 \le \frac{n}{c}$.

Next, by the inequality $x_{j+1} \le x_j^2/n$, we have that $x_j \le \frac{n}{c^{2^{j-1}}}$ for any $j \ge 1$. (This can be proven by induction: $x_2 \le \frac{x_1^2}{n} \le \frac{n^2}{c^2 n} = \frac{n}{c^2}$, $x_3 \le \frac{x_2^2}{n} \le \frac{n^2}{c^4 n} = \frac{n}{c^4}$, and so on.) We want to find k such that no process left after the k-th round, i.e., $x_k = 0$. This is satisfied when $\frac{n}{c^{2^{k-1}}} < 1$, which leads to $k = \log_2 \log_c n + 1 = O(\log\log n)$.

Exercise 2 **(30 points, graded by Yinhao Dong)**

Suppose you are given a biased coin that has $\Pr[\text{HEADS}] = p \ge a$, for some fixed a, without being given any other information about p.

(a) **(10 points)** Devise a procedure that outputs a value $\tilde{p}$ such that you can guarantee that $\Pr[|p - \tilde{p}| \ge \varepsilon p] \le \delta$, for any choice of the constants $0 < a, \varepsilon, \delta < 1$. (The value $\tilde{p}$ is often called the estimate of p.)

(b) **(20 points)** Let N be the number of times you need to flip the biased coin to obtain the estimate. What is the smallest value of N for which you can still give the above guarantee?

Hint: Flip the coin a few times and consider the fraction of times seeing HEADS.

Solution.

(a) Flip the coin N times, for any $N \ge \frac{1-a}{\varepsilon^2 \delta a}$ (or $N \ge \frac{3}{\varepsilon^2 a} \ln \frac{2}{\delta}$). Let #HEADS denote the number of HEADS during all these N flips, and output $\tilde{p} = \frac{\#\text{HEADS}}{N}$.

(b) Now we show how to obtain the above lower bound for N which guarantees that $\Pr[|p - \tilde{p}| \ge \varepsilon p] \le \delta$, for any choice of the constants $0 < a, \varepsilon, \delta < 1$.

We define N indicator random variables $X_1, \cdots, X_N$ where

$$X_i = \begin{cases} 1, & \text{if we get HEADS for the } i\text{-th flip} \\ 0, & \text{otherwise} \end{cases}$$

Let $Y = \sum_{i=1}^N X_i$, we have that $\tilde{p} = Y/N$. Notice that Y is a binomial distribution, so $\mathrm{E}[Y] = Np$ and $\sigma_Y = \sqrt{Np(1-p)}$.

Using Chebyshev's inequality, we get

$$\begin{aligned} \Pr[|\tilde{p} - p| \ge \varepsilon p] &= \Pr[|Y/N - \mathrm{E}[Y/N]| \ge \varepsilon \mathrm{E}[Y/N]] \\ &= \Pr[|Y - \mathrm{E}[Y]| \ge \varepsilon \mathrm{E}[Y]] \\ &\le \frac{\sigma_Y^2}{(\varepsilon \mathrm{E}[Y])^2} \\ &= \frac{Np(1-p)}{(\varepsilon Np)^2} \\ &= \frac{1-p}{\varepsilon^2 Np} \end{aligned}$$

Note that $p \ge a$. Then by setting $N \ge \frac{1-a}{\varepsilon^2 \delta a}$, we will have that $\Pr[|\tilde{p} - p| \ge \varepsilon p] \le \frac{1-p}{\varepsilon^2 Np} \le \frac{\varepsilon^2 \delta a}{1-a} \cdot \frac{1-p}{\varepsilon^2 p} \le \delta$.

Or

Using Chernoff bound, we get

$$\begin{aligned}\Pr[|\tilde{p}-p| \geq \varepsilon p] &= \Pr[|Y - \mathrm{E}[Y]| \geq \varepsilon \mathrm{E}[Y]] \\ &\leq 2\exp\left(\frac{-\varepsilon^2 \mathrm{E}[Y]}{3}\right) \\ &= 2\exp\left(-\frac{\varepsilon^2 N p}{3}\right)\end{aligned}$$

By Noting that $p \geq a$, we can choose $N \geq \frac{3}{\varepsilon^2 a} \ln \frac{2}{\delta}$ so that the above probability will be at most δ.

Exercise 3 **(30 points, graded by Di Wu)**

Let X and Y be finite sets and let Y^X denote the set of all functions from X to Y. We will think of these functions as "hash" functions. A family $\mathcal{H} \subseteq Y^X$ is said to be strongly 2-universal if the following property holds, with $h \in \mathcal{H}$ picked uniformly at random:

$$\forall x, x' \in X \; \forall y, y' \in Y \left(x \neq x' \Rightarrow \Pr_h[h(x) = y \wedge h(x') = y'] = \frac{1}{|Y|^2} \right).$$

We are given a stream $\mathcal{S}$ of elements of X, and suppose that $\mathcal{S}$ contains at most s distinct elements. Let $\mathcal{H} \subseteq Y^X$ be a strongly 2-universal hash family with $|Y| = cs^2$ for some constant $c > 0$. Suppose we use a random function $h \in \mathcal{H}$ to hash.
Prove that the probability of a collision (i.e., the event that two distinct elements of $\mathcal{S}$ hash to the same location) is at most $1/(2c)$.

Proof. We use h chosen uniformly at random from $\mathcal{H}$ to hash the stream $\mathcal{S}$. Let the random variable C be the total number of collisions between any pair of distinct numbers in the stream. We are asked to show that $\Pr[C \geq 1] \leq 1/(2c)$. For any pair of distinct numbers x and x', let $\chi_{\{h(x)=h(x')\}} = 1$ if $h(x) = h(x')$, and $\chi_{\{h(x)=h(x')\}} = 0$ otherwise. Then the total number of pairwise collisions can be written as $C = \sum_{x \neq x'} \chi_{\{h(x)=h(x')\}}$, and by linearity of expectation,

$$\mathrm{E}[C] = \mathrm{E}\left[\sum_{x \neq x'} \chi_{\{h(x)=h(x')\}}\right] = \sum_{x \neq x'} \mathrm{E}\left[\chi_{\{h(x)=h(x')\}}\right].$$

From the definition of strong 2-universality follows with $y = y'$, that for $x \neq x'$, $\mathrm{E}\left[\chi_{\{h(x)=h(x')\}}\right] = \Pr[h(x) = h(x')] \leq \sum_y \Pr\left[(h(x) = y) \wedge (h(x') = y)\right] = \sum_y 1/|Y|^2 = 1/|Y| = 1/cs^2$. There are at most $\binom{s}{2}$ distinct pairs in the stream $\mathcal{S}$, therefore

$$\mathrm{E}[C] \leq \frac{\binom{s}{2}}{cs^2} \leq \frac{\frac{1}{2}s^2}{cs^2} = \frac{1}{2c}$$

Using Markov's inequality, we obtain

$$\Pr[C \geq 1] \leq \frac{\mathrm{E}[C]}{1} = \frac{1}{2c}.$$

□
