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

Exercise Sheet 2 for
Design and Analysis of Algorithms
Autumn 2022

Due 18 Oct 2022 at 23:59

Exercise 1

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) 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) 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) $Pr(\text{the ith bin has exactly 1 ball}) = \binom{b}{1}\frac{1}{n}(1-\frac{1}{n})^{(b-1)} = \frac{b}{n}(1-\frac{1}{n})^{(b-1)}$
$E(\text{numbers of 1 ball bins}) = \sum_{i=1}^{n} Pr(\text{the ith bin has exactly 1 ball}) = b(1-\frac{1}{n})^{(b-1)}$
so, the expected number of processes at the start of the next round is $b - E(\text{numbers of 1 ball bins}) = b - b(1-\frac{1}{n})^{(b-1)}$

(b) If x_j is the expected number of processes left after j rounds, $x_{j+1} = x_j(1-(1-\frac{1}{n})^{x_j-1}) \leq \frac{x_j(x_j-1)}{n} \leq \frac{x_j^2}{n}$
$x_1 = n(1-(1-\frac{1}{n})^{n-1}) = n(1-\frac{1}{e})$, when $n \to \infty$. Set $k = (1-\frac{1}{e})^{-1} > 1$ which is a small constant.
So, $x_1 = \frac{n}{k}$, $x_2 \leq x_1^2/n = n^2/(nk^2) = n/k^2$, $x_3 = n/k^4$, ... , $x_(j+1) = n/k^{2^j} = 1$, which suppose on the $j+1$ ground, all the process are removed.
$\therefore n = k^{2^j}$, $j = log_2 log_k n$, $\therefore j+1 = O(loglogn)$.

Exercise 2

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

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

(b) 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:

Algorithm 1: COINSALG

Initialization: $X \leftarrow 0$; $N \leftarrow$ a selected number;
for $i \leftarrow 1$ to N do **do**
  Flip the biased coin;
  if seeing HEADS **then**
    $X \leftarrow X + 1$;
  end if
end for
$\tilde{p} \leftarrow \frac{X}{N}$;
return $\tilde{p}$

(a) As the Algorithm 1, I devise a procedure that return a value $\tilde{p}$ that can guarantee that $\Pr[|p - \tilde{p}| \geq \varepsilon p] \leq \delta$, for any choice of the constants $0 < a, \varepsilon, \delta < 1$, and I will prove it as bellow.

(b) Set $X_i = 1$ if and only if at the ith flip time, seeing the HEADS, otherwise $X_i = 0$. $X = \sum_{i=1}^{N} X_i$, E(number of all HEADS in N times) $= EX = NEX_i = Np$, $\tilde{p} = \frac{X}{N}$
$Pr(|p - \tilde{p}| \geq \varepsilon p) = Pr(|p - \frac{X}{N}| \geq \varepsilon p) = Pr(|X - Np| \geq \varepsilon Np) = Pr(|X - EX| \geq \varepsilon EX)$, $\because X_i \in [0, 1]$, we can use the Chernoff bound, $\therefore Pr(|X - EX| \geq \varepsilon EX) \leq 2exp(\frac{-EX\varepsilon^2}{3}) \leq \delta$, $\therefore 2exp(\frac{-Np\varepsilon^2}{3}) \leq \delta$, and we can solve the N.
$N \geq \frac{-3ln\delta}{2p\varepsilon^2}$, $\because p \geq a$, we get the smallest N value that can still give the above guarantee is $N = \frac{-3ln\delta}{2a\varepsilon^2}$

Exercise 3

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 give a 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 $1/(2c)$.

Solution:

$$\forall x, x' \in S \; \forall y, y' \in Y \left(x \neq x' \Rightarrow \Pr_h[h(x) = y \wedge h(x') = y'] = \frac{1}{|Y|^2} = \frac{1}{c^2 s^4} \right).$$

Pr(the i,jth distinct elements of S hash to the kth location) $= \binom{s}{2}\binom{|Y|}{1}\frac{1}{|Y|^2} = \frac{s(s-1)cs^2}{2c^2s^4} \leq \frac{1}{2c}$.
