

Name: _____

Entry number: _____

There are 3 questions for a total of 30 points.

- (10) 1. n balls are thrown randomly into n bins. Show that the probability that the number of empty bins is more than $((3/4) \cdot n)$ is exponentially small in n .

Solution: Let X_i denote the bin number in which the i^{th} balls is thrown. Let $f(X)$ denote the number of empty bins when the assignment is denoted by $X = (X_1, \dots, X_n)$. Let $Z_i = E[f(X)|X_1, \dots, X_i]$. Note that $Z_0, \dots, Z_n$ is a martingale. Moreover, $Z_0, \dots, Z_n$ is 1-Lipschitz. Applying Azuma-Hoeffding, we get that $\Pr[Z_n - Z_0 \geq \lambda] = \Pr[f(X) - E[f(X)] \geq \lambda] \leq e^{-\frac{\lambda^2}{2n}}$. Note that the expected number of empty bins, denoted by $E[f(X)] \leq n/e$. Now we have

$$\begin{aligned} \Pr[f(X) \geq (3/4) \cdot n] &= \Pr[(f(X) - n/e) \geq (3/4 - 1/e) \cdot n] \\ &\leq \Pr[f(X) - E[f(X)] \geq (3/4 - 1/e) \cdot n] \\ &\leq e^{-\frac{(3/4 - 1/e)^2 n}{2}} \end{aligned}$$

- (10) 2. Let $\{0, 1, \dots, 2^n - 1\}$ be the universe of elements to be hashed into a table of size 2^m . Elements of the universe and table entries are represented as bit vectors of size n and m respectively. Consider the following family of hash functions:

$$\mathcal{H} = \{h_M | M \text{ is a randomly chosen } 0/1 \text{ matrix of size } m \times n \text{ and } h_M(x) = M \cdot x \pmod{2}\}$$

Show that $\mathcal{H}$ is a 2-universal family of hash function.

Solution: For any $x \neq y$, let $u = h_M(x)$ and $v = h_M(y)$. Let the i^{th} bit of a vector p be denoted by p_i . We will show that $\forall i, \Pr_M[u = v] = 1/2$. From this it follows that $\Pr_M[u = v] = 1/2^m$ showing that $\mathcal{H}$ is 2-universal.

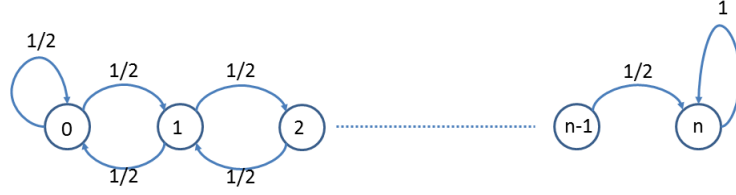
Let M denote the subset of indices on which x and y do not match. Note that since $x \neq y$, M is not the empty set. Let $M' \subseteq M$ be the subset of those indices i such that $x_i = 0$. Let $M'' = M \setminus M'$. Now we will analyse the probability that $u_i = v_i$.

$$\begin{aligned} \Pr_M[u_i = v_i] &= \Pr_{r \leftarrow \{0,1\}^n} \left[\sum_i (x_i \cdot r_i) \pmod{2} = \sum_i (y_i \cdot r_i) \pmod{2} \right] \\ &= \Pr_{r \leftarrow \{0,1\}^{|M|}} \left[\sum_{i \in M} (x_i \cdot r_i) \pmod{2} = \sum_{i \in M} (y_i \cdot r_i) \pmod{2} \right] \\ &= \Pr_{r \leftarrow \{0,1\}^{|M|}} \left[\sum_{i \in M''} (x_i \cdot r_i) \pmod{2} = \sum_{i \in M'} (y_i \cdot r_i) \pmod{2} \right] \\ &= 1/2 \end{aligned}$$

The last equality is due to the fact that for any fixed bit b and any $s \neq 0$,

$$\Pr_{r \leftarrow \{0,1\}^s} \left[\sum_i r_i \pmod{2} = b \right] = \frac{1}{2^s} \left(\binom{s}{0} + \binom{s}{2} + \dots \right) = \frac{1}{2^s} \left(\binom{s}{1} + \binom{s}{3} + \dots \right) = 1/2.$$

- (10) 3. Consider the following markov chain, that is very similar to the Markov Chain we studied when talking about the 2-SAT problem. Compute the expected time to reach the state n starting from any state i .

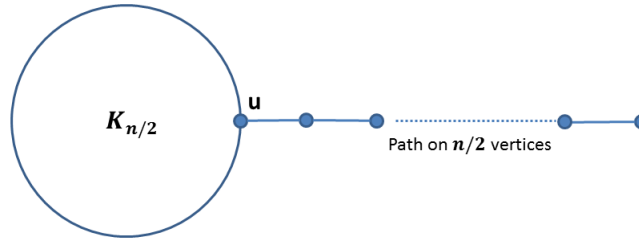


Solution: Let h_j denote the expected time to reach the state n starting at the state j . We have the following recursive relations for h_i 's.

$$\begin{aligned} h_0 &= \frac{1}{2} \cdot (1 + h_0) + \frac{1}{2} \cdot (1 + h_1) \\ \forall i, 1 \leq i \leq n-1, \quad h_i &= \frac{1}{2} \cdot (1 + h_{i-1}) + \frac{1}{2} \cdot (1 + h_{i+1}) \\ h_n &= 0. \end{aligned}$$

Solving the above equations we get $\forall j, 1 \leq j \leq n-1, h_{j-1} = h_j + 2j$. This gives $h_j = 2n + 2(n-1) + \dots + 2(j+1) = n \cdot (n+1) - j \cdot (j+1)$.

Extra credit question (5 points) Consider the lollipop graph graph shown below consisting of a clique on $n/2$ vertices connected to a path on $n/2$ vertices. Argue that the expected cover time of a random walk starting at u is $\theta(n^3)$.



Solution: From the discussion in the class, we know that the expected cover time is $O(n^3)$. We will show that the expected cover time starting at vertex u is $\Omega(n^3)$. Let us label the vertices on the path $P_1, P_2, \dots, P_{n/2}$ from left to right. Let h_j denote the expected time to reach the state $P_{n/2}$ (the tail of the lollipop) starting at state P_j . We know that the expected cover time starting at state P_1 is at least h_1 . So a lower bound on h_1 will give us a lower bound on the cover time. Starting at P_1 , the probability that in the next step the random walk enters the clique is $\frac{n/2-1}{n/2}$. Once this happens, the expected number of steps in which the random walk comes back to state P_1 is $(n/2-1)$. So we may write

$$h_1 = \frac{n/2-1}{n/2} \cdot (1 + (n/2-1) + h_1) + \frac{1}{n/2} \cdot (1 + h_2) \quad (1)$$

We may also write

$$\forall i, 2 \leq i \leq n/2-1, h_i = 1/2 \cdot (1 + h_{i-1}) + 1/2 \cdot (1 + h_{i+1}) \quad (2)$$

Solving (1) and (2), we get that $h_{j-1} = h_j + 2(j-2) + \Omega(n^2)$. This gives us that $h_1 = \Omega(n^3)$.