COMP 4804 Assignment 2

Braeden Hall (101143403)

October 31, 2021

Question 1.

1. Interval w contains $\frac{n}{4}$ points.

$$\begin{aligned} Pr(\text{"}x \text{ captures } w\text{"}) &= Pr(\text{"}x \text{ captures } w\text{"} | x \in w) \cdot Pr(x \in w) + Pr(\text{"}x \text{ captures } w\text{"} | x \notin w) \cdot Pr(x \notin w) \\ &= 1 \cdot Pr(x \in w) + 0 \cdot Pr(x \notin w) = Pr(x \in w) \\ &= \frac{\frac{n}{4}}{n} = \frac{1}{4} \end{aligned}$$

   Therefore, the probability that x captures w is $\frac{1}{4}$.

2. Each interval contains $\frac{n}{4}$ points. So we can calculate the number of intervals as

$$n - \left(\frac{n}{4} - 1\right) = \frac{4n}{4} - \frac{n-4}{4} = \frac{3n+4}{4}$$

   Define $X =$"number of intervals captured by k random points"
   For $\forall w \in \{1, 2, \ldots, \frac{3n+4}{4}\}$ define:

$$X_w = \begin{cases} 1 & \text{if interval w is captured by one or more of the } k \text{ points} \\ 0 & \text{otherwise} \end{cases}$$

$$X = \sum_{w=1}^{\frac{3n+4}{4}} X_w \Rightarrow E(X) = \sum_{w=1}^{\frac{3n+4}{4}} E(X_w) = \sum_{w=1}^{\frac{3n+4}{4}} Pr(X_w = 1)$$

$$Pr(X_w = 1) = 1 - Pr(X_w = 0) = 1 - Pr\left(\bigcap_{i=1}^{k} p_i \notin w\right) = 1 - \left(\frac{3}{4}\right)^k = \frac{4^k - 3^k}{4^k}$$

$$E(X) = \sum_{w=1}^{\frac{3n+4}{4}} Pr(X_w = 1) = \sum_{w=1}^{\frac{3n+4}{4}} \frac{4^k - 3^k}{4^k} = \frac{3n+4}{4}\left(\frac{4^k - 3^k}{4^k}\right) = \frac{(3n+4)(4^k - 3^k)}{4^{k+1}}$$

   Therefore, the expected number of captured by k random points is $\frac{(3n+4)(4^k-3^k)}{4^{k+1}}$.

3. We can use Markov's inequality to find a k such that, with high probability, all intervals in I are captured.

   Define $X =$"number of intervals captured by k random points"
   Note that $E(X) = \frac{(3n+4)(4^k-3^k)}{4^{k+1}}$, as calculated in part 2.

$$Pr\left(X \geq \frac{3n+4}{4}\right) \leq \frac{(3n+4)(4^k-3^k)}{4^{k+1}} \cdot \frac{4}{3n+4} = \frac{4^k-3^k}{4^k}$$

$$Pr\left(X < \frac{3n+4}{4}\right) \geq 1 - \frac{4^k-3^k}{4^k} = \frac{4^k-4^k+3^k}{4^k} = \frac{3^k}{4^k}$$

   Take $k = \ln n$
   Them the probability that the random sample does not capture every interval in I is:

$$\geq \frac{3^k}{4^k} = \frac{3^{\ln n}}{4^{\ln n}} = \frac{n^{\ln 3}}{n^{\ln 4}}$$

   .
   Note that $\frac{n^{\ln 3}}{n^{\ln 4}} \to 0$ as $n \to \infty$. Therefore, with high probability, random samples of size $k = \ln n = o(n)$ will capture all intervals of I.

Question 2.

Omitted.

Question 3.

1. For $\forall v \in V$, let $col(v)$ denote the colour of vertex v.
   An edge $e = (u, v)$ is bad $\iff col(u) = col(v)$.

$$Pr(col(u) = col(v)) = \frac{k}{k^2} = \frac{1}{k}$$

   Therefore, an edge is bad with probability $\frac{1}{k}$.

2. Define $X =$"number of bad edges".
   For $\forall e \in E$ define:

$$X_e = \begin{cases} 1 & \text{if edge } e \text{ is bad} \\ 0 & \text{otherwise} \end{cases}$$

$$X = \sum_{e \in E} X_e \Rightarrow E(X) = \sum_{e \in E} E(X_e) = \sum_{e \in E} Pr(X_e = 1) = \sum_{e \in E} \frac{1}{k} = \frac{m}{k}$$

   Therefore, the expected number of bad edges is $\frac{m}{k}$. By a symmetric argument, the expected number of good edges is $m\left(1 - \frac{1}{k}\right) = m - \frac{m}{k}$.

3. i) Note: each edge is bad with probability $\frac{1}{k}$, independent of any other edges.

$$Pr(\text{"vertex } v \text{ is dead"}) = Pr\left(\bigcap_{(u,v)\in E} \text{edge } (u,v) \text{ is bad}\right) = \prod_{i=1}^{6} \frac{1}{k} = \left(\frac{1}{k}\right)^6 = \frac{1}{k^6}$$

ii) Define $X =$"number of dead vertices"
For $\forall v \in V$ define:

$$X_v = \begin{cases} 1 & \text{if vertex } v \text{ is dead} \\ 0 & \text{otherwise} \end{cases}$$

$$X = \sum_{v\in V} X_v \Rightarrow E(X) = \sum_{v\in V} E(X_v) = \sum_{v\in V} Pr(X_v = 1) = \sum_{v\in V} \frac{1}{k^6} = \frac{n}{k^6}$$

4. Let c be some constant. We want $E(X) < \frac{n}{c}$

$$E(X) < \frac{n}{c} \Rightarrow \frac{n}{k^6} < \frac{n}{c} \Rightarrow n < \frac{k^6 n}{c} \Rightarrow k^6 > \frac{cn}{n} \Rightarrow k > c^{\frac{1}{6}}$$

a) k must be greater that $10^{\frac{1}{6}}$, i.e $k > 1$.

b) k must be greater that $100^{\frac{1}{6}}$. i.e. $k > 2$.

c) k must be greater that $1000^{\frac{1}{6}}$, i.e $k > 3$.

Question 4.

1. We eat both candies $\iff$ the two candies are of different types.

$$\begin{aligned} Pr(\text{"both candies are different types"}) &= 1 - Pr(\text{"both candies are the same time"}) \\ &= 1 - Pr(\text{"both candies are lime"} \cup \text{"both candies are lemon"}) \\ &= 1 - \big(Pr(\text{"both candies are lime"}) + Pr(\text{"both candies are lime"})\big) \end{aligned}$$

$$\begin{aligned} Pr(\text{"both candies are lime"}) &= \frac{\binom{\frac{n}{2}}{2}}{\binom{n}{2}} = \frac{\frac{n}{2}\left(\frac{n}{2}-1\right)}{2} \cdot \frac{2}{n(n-1)} = \frac{\left(\frac{n}{2}-1\right)}{2(n-1)} = \frac{n-2}{4(n-1)} \\ &= Pr(\text{"both candies are lemon"}) \end{aligned}$$

$$\begin{aligned} Pr(\text{"both candies are different types"}) &= 1 - \big(Pr(\text{"both candies are lime"}) + Pr(\text{"both candies are lime"})\big) \\ &= 1 - 2\left(\frac{n-2}{4(n-1)}\right) = 1 - \frac{n-2}{2(n-1)} = \frac{n}{2(n-1)} \end{aligned}$$

Therefore, the probability we eat both candies is $\frac{n}{2(n-1)}$.

2. Define X ="number of times we repeat the experiment until we eat 2 candies".
   Note that X is a geometric random variable where the probability of success is $p = \frac{n}{2(n-1)}$. As such,

$$E(X) = \frac{1}{p} = \frac{1}{\frac{n}{2(n-1)}} = \frac{2(n-1)}{n}$$

   Therefore, the expected number of times we must repeat the experiment until we eat 2 candies is $\frac{2(n-1)}{n}$.

3.

$$Pr(\text{"both candies are lemon"}) = \frac{\binom{n_1}{2}}{\binom{n_1+n_2}{2}} = \frac{n_1(n_1-1)}{2} \cdot \frac{2}{(n_1+n_2)((n_1+n_2-1)} = \frac{n_1(n_1-1)}{(n_1+n_2)(n_1+n_2-1)}$$

   By a symmetric argument,

$$Pr(\text{"both candies are lime"}) = \frac{n_2(n_2-1)}{(n_1+n_2)(n_1+n_2-1)}$$

$$\begin{aligned} Pr(\text{"both candies are different types"}) &= 1 - \big(Pr(\text{"both candies are lime"}) + Pr(\text{"both candies are lime"})\big) \\ &= 1 - \left(\frac{n_1(n_1-1)}{(n_1+n_2)(n_1+n_2-1)} + \frac{n_2(n_2-1)}{(n_1+n_2)(n_1+n_2-1)}\right) \\ &= 1 - \left(\frac{n_1(n_1-1)+n_2(n_2-1)}{(n_1+n_2)(n_1+n_2-1)}\right) \end{aligned}$$

   Therefore, the probability we eat both candies is $1 - \left(\frac{n_1(n_1-1)+n_2(n_2-1)}{(n_1+n_2)(n_1+n_2-1)}\right)$.

4. Define X ="number of times we must repeat the experiment until the bag is empty".
   For $\forall i \in \{\frac{n}{2}, \frac{n-2}{2}, \ldots, 1\}$, define:
   X_i ="number of times we must repeat the experiment until we eat 2 candies when there are $2i$ candies in the bag".

$$E(X_i) = \frac{2(2i-1)}{2i} = \frac{2i-1}{i}$$

$$\begin{aligned} X = \sum_{i=1}^{\frac{n}{2}} X_i \Rightarrow E(X) = \sum_{i=1}^{\frac{n}{2}} E(X_i) = \sum_{i=1}^{\frac{n}{2}} \left((2i-1) \cdot \frac{1}{i}\right) &= H_{\frac{n}{2}} \sum_{i=1}^{\frac{n}{2}} 2i - 1 \\ &\leq 2H_{\frac{n}{2}} \sum_{i=1}^{\frac{n}{2}} i \leq 2H_{\frac{n}{2}} \cdot \frac{\frac{n}{2}\left(\frac{n}{2}+1\right)}{2} \leq H_{\frac{n}{2}} \cdot \left(\frac{n(n+2)}{4}\right) \end{aligned}$$

   Therefore the expected number of times we must repeat the experiment until the bag is empty is less than or equal to $H_{\frac{n}{2}} \cdot \left(\frac{n(n+2)}{4}\right)$.

5. Omitted.

Question 5.

1. Suppose there exists a deterministic algorithm that can find the element x in an n element array in less than n comparisons. This means that the algorithm can always find x after checking $\leq n-1$ indices in the array. Assume, without lack of generality, that the algorithm checks the first $n-1$ elements in the array. It is possible for an adversary to structure the array so that x is the n^{th} element, then the algorithm will not find x. Thus, we have a contradiction and such an algorithm does not exist.

2. Algorithm:

Algorithm 1 $findElem(A, x, rb)$:

```
if rb = 0 then
    start ← 1
    end ← ⌈n/2⌉
else
    start ← ⌈n/2⌉ + 1
    end ← n
end if

for i = start to end do
    if A[i] = x then
        return i
    end if
end for

return −1
```

In this algorithm, in the worst case (when it returns -1), will perform $\lceil \frac{n}{2} \rceil \leq \frac{n+1}{2}$ comparisons. The algorithm will return the index of x in the array if it is found in the half of the array that the algorithm searches.

Question 6.

1. Define $X =$"number of right empty points in P"
For $\forall i \in \{1, \dots, n\}$, define:

$$X_i = \begin{cases} 1 & \text{if point } p_i \text{ is right empty} \\ 0 & \text{otherwise} \end{cases}$$

$$X = \sum_{i=1}^{n} X_i \Rightarrow E(X) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} Pr(X_i = 1)$$

Note: point p_i is right empty $\iff \forall p_j \in P \setminus \{p_i\}$, $x_j < x_i$ or $y_j < y_i$

$$\begin{aligned}
Pr(X_i = 1) &= Pr\Bigg(\bigcap_{p_j \in P \setminus \{p_i\}} (x_j < x_i \cup y_j < y_i) \Bigg) = \prod_{p_j \in P \setminus \{p_i\}} Pr(x_j < x_i \cup y_j < y_i) \\
&= \prod_{p_j \in P \setminus \{p_i\}} \Big(Pr(x_j < x_i) + Pr(y_j < y_i) - Pr(x_j < x_i \cap y_j < y_i) \Big) \\
&= \prod_{p_j \in P \setminus \{p_i\}} \left(\frac{1}{2} + \frac{1}{2} - \left(\frac{1}{2} \cdot \frac{1}{2} \right) \right) = \prod_{p_j \in P \setminus \{p_i\}} \frac{3}{4} = \left(\frac{3}{4} \right)^{n-1}
\end{aligned}$$

$$E(X) = \sum_{i=1}^{n} Pr(X_i = 1) = \sum_{i=1}^{n} \left(\frac{3}{4} \right)^{n-1} = n \left(\frac{3}{4} \right)^{n-1} = \frac{n \cdot 3^{n-1}}{4^{n-1}}$$

Showing $E(X)$ is $O(\log n)$:

$$\frac{n \cdot 3^{n-1}}{4^{n-1}} \leq c \cdot \log n \Rightarrow \frac{n \cdot 3^{n-1}}{4^{n-1} \log n} \leq c$$

$$\lim_{n \to \infty} \frac{n \cdot 3^{n-1}}{4^{n-1} \log n} = 0$$

Therefore, the expected number of right empty points is $O(\log n)$.

2. Omitted.