Randomness and Computation - Assignment 1

UUN: S1796157

1. (a) We know that Z_i and Z_j are independent fair coin tosses. We know also that $P(Y_p) = P(X_i, X_j)$. Therefore, $P(Y_p)$ can be rewritten as $P(Y_p) = P(X_i)P(X_j)$ for every $p \in P$. Then we can easily show that:

$$P(Y_p = 0) = P(Z_i = 0)P(Z_j = 0) + P(Z_i = 1)P(Z_j = 1) = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

$$P(Y_p = 1) = P(Z_i = 1)P(Z_j = 0) + P(Z_i = 0)P(Z_j = 1) = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

So Y_p represents a fair coin flip.

(b) P contains all the possible combinations of $Z_1, \dots, Z_{2\sqrt{N}}$ subject to the condition: $P : \{i, j : 1 \leq i < j \leq n\}$ where $n = \lceil 2\sqrt{N} \rceil$. The cardinality of P is given by binomial coefficients of the n variables taken in subset of k elements. In our case, $k = 2$:

$$\binom{n}{2} = \frac{n!}{2!(n-2)!} = \frac{n(n-1)(n-2)!}{2(n-2)!} = \frac{n(n-1)}{2} = |\mathcal{P}|$$

Then we need to consider two cases: one in which N is a perfect square, thus leading to n being equal to $2\sqrt{N}$, and when N is not a perfect square, thus leading to n being equal to $\lfloor 2\sqrt{N} \rfloor + 1$.

$$\frac{2\sqrt{N}(2\sqrt{N} - 1)}{2} = 2N - \sqrt{N} > N \qquad \forall\, N > 1 : \text{N is a perfect square}$$

$$\frac{(\lfloor 2\sqrt{N} \rfloor + 1)(\lfloor 2\sqrt{N} \rfloor + 1 - 1)}{2} = 2N + \sqrt{N} > N \qquad \forall\, N > 1 : \text{N is not a perfect square}$$

Therefore, the cardinality of the set $\mathcal{P}$ will be greater than N.

(c) To show that every pair of the Y_p variables satisfies the *pairwise independence*, we must show that $\mathbb{E}[Y_pY_q] = \mathbb{E}[Y_p]\mathbb{E}[Y_q]$. We start our proof considering:

$$\mathbb{E}[Y_pY_q] = \sum_{y_p} \sum_{y_q} y_p \cdot y_q \cdot P(Y_p = y_p, Y_q = y_q)$$

Then, by using the chain rule $P(Y_p = y_p, Y_q = y_q) = P(Y_p = y_p | Y_q = y_q) \cdot (Y_q = y_q)$, we can rewrite the expectation as follow:

$$\mathbb{E}[Y_pY_q] = \sum_{y_p} \sum_{y_q} y_p \cdot y_q \cdot P(Y_p = y_p | Y_q = y_q)P(Y_q = y_q)$$

If Y_q and Y_p are pairwise independent, we know that their expectation has to be equal to:

$$\mathbb{E}[Y_pY_q] = \mathbb{E}[Y_p]\mathbb{E}[Y_q] = \sum_{y_p} \sum_{y_q} y_p \cdot y_q P(Y_p = y_p)P(Y_q = y_q)$$

Therefore, to show that the pairwise relation holds, we must show that:

$$P(Y_p = y_p | Y_q = y_q)P(Y_q = y_q) = P(Y_p = y_p)P(Y_q = y_q) \qquad \forall\, y_p, y_q$$

We start considering two cases: the first one consider when Y_p and Y_q are "made" by not sharing any Z and the second one in which Y_p and Y_q share one of the Z variables.

- **First case**: $Y_p = Z_a \oplus Z_b$ and $Y_q = Z_c \oplus Z_d$

  This is almost trivial, since Y_p and Y_q do not share any Z variables they are independent. Therefore $P(Y_p = y_p|Y_q = y_q)P(Y_q = y_q) = P(Y_P = y_p)P(Y_q = y_q)$ for every y_p and y_q.
- **Second case**: $Y_p = Z_a \oplus Z_b$ and $Y_q = Z_a \oplus Z_c$

  In this case, Y_p and Y_q share one of the Z variables, therefore they are not independent anymore. However, we can show that $P(Y_p = y_p|Y_q = y_q) = P(Y_p = y_p)$.

$$\begin{aligned}
P(Y_p = 1) &= \sum P(Y_p = 1|Y_q = 1)P(Y_q = y_q) \\
&= \sum P(Z_a \oplus Z_b = 1|Z_a \oplus Z_c = y_q) \\
&= \sum P(Z_a \oplus Z_b = 1|Z_a \oplus Z_c = y_q) \\
&= \sum P(Z_a \oplus Z_b = 1|Z_a \oplus Z_c = 1) + P(Z_a \oplus Z_b = 1|Z_a \oplus Z_c = 0) \\
&= \frac{1}{2} \cdot \frac{1}{2} + \frac{1/2}{1/2} \cdot \frac{1}{2}\frac{1}{2} = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}
\end{aligned}$$

The same is true for $P(Y_p = 0) = 1 - P(Y_p = 1)$, therefore $P(Y_p = y_p|Y_q = y_q) = P(Y_p = y_p)$. Hence, we can conclude or proof by rewriting the expecation as:

$$\begin{aligned}
\mathbb{E}[Y_pY_q] &= \sum_{y_p} \sum_{y_q} y_p \cdot y_q P(Y_p = y_p|Y_q = y_q)P(Y_q = y_q) \\
&= \sum_{y_p} \sum_{y_q} y_p \cdot y_q P(Y_p = y_p)P(Y_q = y_q) \\
&= \sum_{y_p} y_p \cdot P(Y_p = y_p) \sum_{y_q} y_q \cdot P(Y_q = y_q) \\
&= \mathbb{E}[Y_p]\mathbb{E}[Y_q]
\end{aligned}$$

(d) $P(\bigcap_{i=1}^{2\sqrt{n}} z_i) \neq \prod_{i=1}^{2\sqrt{n}} P(z_i)$ Let us have $X_1 = Z_a \oplus Z_b$, $X_2 = Z_d \oplus Z_c$ and $X_3 = Z_a \oplus Z_c$. We can show that: $P(X_1, X_2, X_3) \neq P(X_1)P(X_2)P(X_3)$ because knowing X_1 and X_2 gives us information about X_3. This can be shown to hold for $|P| \geq 3$.

(e)

$$\mathbb{E}[Y] = \mathbb{E}\left[\sum_{i=0}^{m} Y_i\right] = \sum_{i=0}^{m} \mathbb{E}[Y_i] = \sum_{i=0}^{m} \frac{1}{2} = \frac{m}{2}$$

$$\text{we know that} \quad m = 2N - \sqrt{N} \quad \text{hence}$$

$$\mathbb{E}[Y] = \frac{2N - \sqrt{N}}{2} = N - \frac{\sqrt{N}}{2}$$

(f)

$$Var[Y] = Var\left[\sum_{i=0}^{m} Y_i\right] = \sum_{i=1}^{k} Var[Y_i]$$

$$Var[Y_i] = \mathbb{E}[(Y - \mathbb{E}[Y])^2] = \mathbb{E}[Y_i^2] - \mathbb{E}[Y_i]^2 = \sum_{i=0}^{k} \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$$

$$Var[Y] = \sum_{i=1}^{m} Var[Y_i] = \sum_{i=1}^{m} \frac{1}{4} = \frac{m}{4}$$

$$\text{we know that} \quad m = 2N - \sqrt{N} \quad \text{hence}$$

$$Var(Y) = \frac{2N - \sqrt{N}}{4}$$

(g) We want to find an upper bound for $P(|Y - \mathbb{E}[Y]|) \geq n$, using the Chebyshev's inequality. We can proceed as follow:

$$P(|Y - \mathbb{E}[Y]| \geq n) \leq \frac{Var[Y]}{n^2} = \frac{n}{4} \cdot \frac{1}{n} = \frac{1}{4n}$$

2. Considering the coupon collector problem, the child has only $\frac{n}{2}$ spaces in his book that has to be filled with player's stickers (which remains n). Therefore, by starting from the Coupon Collector analysis on slide 15 of the Lecture 5 we can infer that:

$$\mathbb{E}[X] = \sum_{i=1}^{\frac{n}{2}} \frac{n}{n-(i-1)} = n \sum_{i=1}^{n/2} \frac{1}{n-(i-1)} = n \sum_{i=1}^{\frac{n}{2}+1} \frac{1}{i}$$

$$t = \frac{n}{2} + 1 \qquad \int_{x=1}^{t} \frac{1}{x} < \sum_{i=1}^{t} \frac{1}{i} \quad \text{and} \quad \sum_{i=2}^{t} \frac{1}{i} < \int_{x=1}^{t} \frac{1}{x} \quad \text{hence} \quad \ln(t) < \sum_{i=1}^{t} \frac{1}{i} \leq \ln(t) + 1$$

therefore, we can conclude that $E[X] \sim n \cdot ln(t) = n \cdot ln(\frac{n}{2} + 1)$