Problem 1. A random variable X has three values, $\{0, 1, 2\}$. It is known that the probabilities $P(X = 2) = 1/2$ and $P(X = 1) = p$. Determine the values of p that maximize $H(X)$. (*Hint:* consider grouping).

Let the groups be $\{2\}$ and $\{0,1\}$.

$$H(X) = H(\underbrace{P(X\in\{2\})}_{=1/2}, \underbrace{P(X\in\{0,1\})}_{=1/2})$$

$$+ \underbrace{P(X\in\{2\})}_{=1/2} H(X \mid X\in\{2\})$$ X is determined by $X\in\{2\}$

$$+ \underbrace{P(X\in\{0,1\})}_{=1/2} H(X \mid X\in\{0,1\})$$

If $X\in\{0,1\}$, $P(X=1)=2p$, $P(X=0)=1-2p$

$$= H\left(\tfrac{1}{2},\tfrac{1}{2}\right) + \tfrac{1}{2}\cdot 0 + \tfrac{1}{2} H(2p, 1-2p)$$

$$= H\left(\tfrac{1}{2},\tfrac{1}{2}\right) + \tfrac{1}{2} H(2p, 1-2p).$$

This is maximal when $H(2p, 1-2p)$ is max.

$H(q, 1-q)$ is max when $q = 1/2$, so $2p = 1/2$,

so $p = 1/4$.

Max if $P(X=1)=1/4$ $P(X=0)=1/4$

(and $P(X=2)=1/2$).

Problem 2. Construct a Huffman code for the symbols $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ with probabilities

Symbol	Probability
0	~~0.1~~
1	~~0.2~~
2	~~0.05~~
3	~~0.05~~
4	~~0.12~~

Symbol	Probability
5	~~0.08~~
6	~~0.1~~
7	~~0.1~~
8	~~0.03~~
9	~~0.17~~

~~0.13~~ ~~0.18~~ ~~0.2~~ ~~0.25~~ ~~0.35~~ 0.4 0.6

$Z_1 = \{3 \text{ or } 8\}$ $P(Z_1) = 0.05 + 0.03 = 0.08$ ✓

$Z_2 = \{Z_1 \text{ or } 2\}$ $P(Z_2) = 0.08 + 0.05 = 0.13$

$Z_3 = \{5 \text{ or } 6\}$ $P(Z_3) = 0.08 + 0.1 = 0.18$ ✓

$Z_4 = \{0 \text{ or } 7\}$ $P(Z_4) = 0.1 + 0.1 = 0.2$ ✓

$Z_5 = \{4 \text{ or } Z_2\}$ $P(Z_5) = 0.12 + 0.13 = 0.25$

$Z_6 = \{9 \text{ or } Z_3\}$ $P(Z_6) = 0.18 + 0.17 = 0.35$

$Z_7 = \{1 \text{ or } Z_4\}$ $P(Z_7) = 0.2 + 0.2 = 0.4$

$Z_8 = \{Z_5 \text{ or } Z_6\}$ $P(Z_8) = 0.6$

$0 \Rightarrow 10$

$Z_8 \Rightarrow 0$

$Z_7 \Rightarrow 10$

$Z_6 \Rightarrow 110$

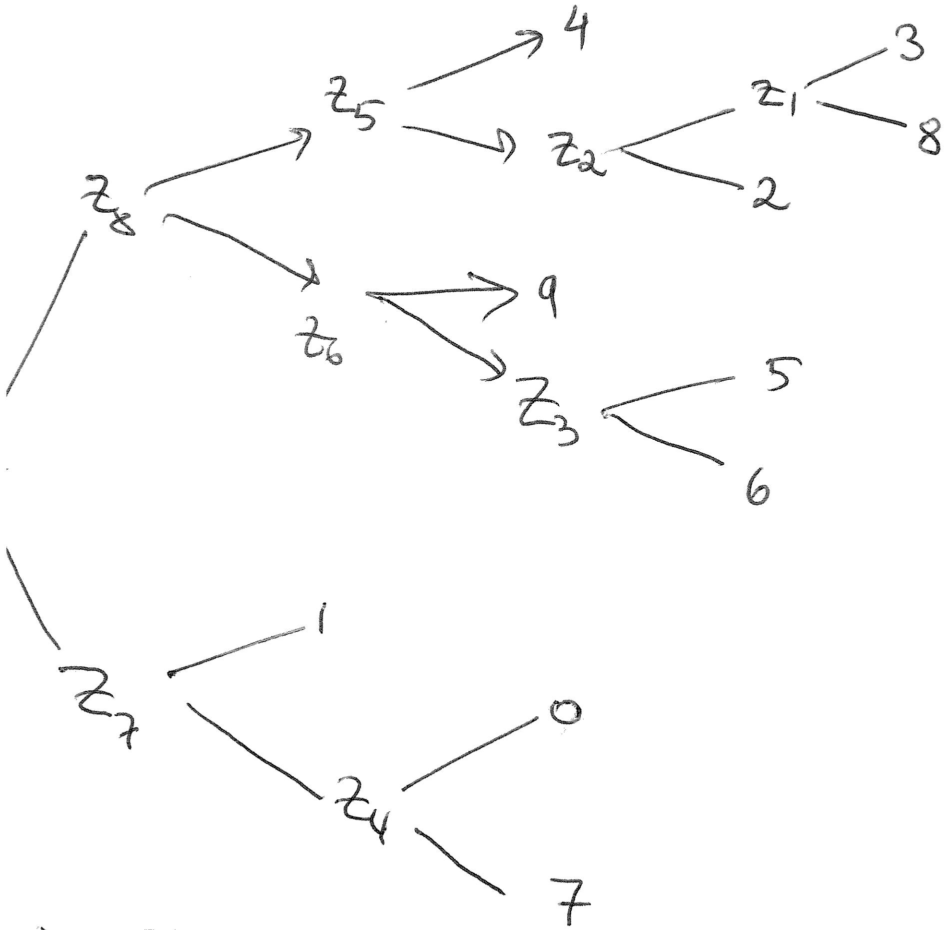
$Z_5 \Rightarrow \cancel{0}110$

$Z_4 \Rightarrow 111\cancel{0}0$

$Z_3 \Rightarrow \cancel{0}11110$

$Z_2 \Rightarrow 111111$

$Z_1 \Rightarrow 1111111$



So eg:

$0 \longmapsto 110$

$1 \longmapsto 10$

$2 \longmapsto 0011$

$3 \longmapsto 00100$

$4 \longmapsto 000$

$5 \longmapsto 0110$

$6 \longmapsto 0111$

$7 \longmapsto 111$

$8 \longmapsto 00101$

$9 \longmapsto 010$

Problem 3. A diamond is hidden in one of five identical boxes, numbered 1 through 5. You are allowed to select any number of these 5 boxes and then receive an answer as to whether the diamond is in one of the boxes you selected. (For example, if you selected boxes #1, #3 and #4, then you would receive a "Yes" answer if the diamond is in any of these three boxes and a "No" answer if the diamond is in box #2 or #5; of course in neither case would you be told the exact location of the diamond).

Prove using information theory that you cannot always determine the exact location of the diamond by asking two questions, and devise a way by which you can determine the location by asking three questions.

What is your lower bound in the case of N boxes?

Let $X =$ # box with diamond.
Assume that X is random with all values eq. likely.
Then $H(X) = \log 5 > 2$.

Hence by the discussion on p. 39, 40 ("yes/no answers") we know that no algorithm can even on average determine the value of X in 2 steps (& hence can't do it always, either).

So min # questions ≥ 3.

Can do this with $n=3$:

- "Is $X \in \{1,2\}$?"
  - yes → {Is $X = 1$?}
    - Yes → $X = 1$
    - No → $X = 2$
  - No → {Is $X \in \{3,4\}$}
    - Yes → Is $X = 3$?
      - Yes → $X = 3$
      - No → $X = 4$
    - No → $X = 5$.

In general, can't do this with less than $\log N$ questions.