Solution to Midterm

Problem 1

(a)

$$H(X|Y) = H(Z+Y|Y) = H(Z|Y)$$

Furthermore, since conditioning decreases entropy,

$$H(Z|Y) \leq H(Z)$$

and thus

$$H(X|Y) \leq H(Z)$$

(b) $H(X|Y) = H(Z)$ if and only if $H(Z|Y) = H(Z)$. That is Z and Y are independent.

(c) The goal is to show that

$$I(U;W) + I(U;T) \leq I(U;V) + I(W;T).$$

By adding the term $I(U;T|W)$ to both sides, it suffices to show that

$$I(U;T|W) + I(U;W) + I(U;T) \leq I(U;V) + I(W;T). + I(U;T|W)$$

By using chain rule, we have that $I(U;T|W) + I(U;W) = I(U;T,W)$ at the left hand side, and $I(U;T|W) + I(W;T) = I(U,W;T)$ at the right hand side. Thus it suffices to show that

$$I(U;T,W) + I(U;T) \leq I(U;V) + I(U,W;T).$$

From the Markov chain $U \leftrightarrow V \leftrightarrow (W,T)$, $I(U;W,T) \leq I(U;V)$.
Furthremore, $I(U;T) \leq I(U,W;T) = I(U;T) + I(W;T|U)$ since $I(W;T|U) \geq 0$. This concludes the solution

Some Remarks:

(1) Note that having the Markov chain $U \leftrightarrow V \leftrightarrow (W,T)$ leads to having the following two Markov chains: $U \leftrightarrow V \leftrightarrow W$ and $U \leftrightarrow V \leftrightarrow T$. Nonetheless, the other way does not necessarily hold; i.e., $U \leftrightarrow V \leftrightarrow (W,T)$ is a stronger Markov chain than $U \leftrightarrow V \leftrightarrow W$ and $U \leftrightarrow V \leftrightarrow T$.

(2) Note that $I(U;W,T)$ is not the same as $I(U,W;T)$. By chain rule we have:

$$\begin{aligned} I(U;W,T) &= I(U;W) + I(U;T|W) \\ &= I(U;T) + I(U;W|T); \end{aligned}$$

and

$$\begin{aligned} I(U,W;T) &= I(T;U) + I(T;W|U) \\ &= I(T;W) + I(T;U|W). \end{aligned}$$

Problem 2

(a) Let $y_n = p(x_1, \cdots, x_n)^{\frac{1}{n}}$. Since $X_1, X_2, \cdots$ is an i.i.d. sequence, we have $p(x_1, \cdots, x_n) = \prod_{i=1}^{n} p(x_i)$ and

$$\begin{aligned}
\log y_n &= \frac{1}{n} p(x_1, \cdots, x_n) \\
&= \frac{1}{n} \log \prod_{i=1}^{n} p(x_i) \\
&= \frac{1}{n} \sum_{i=1}^{n} \log p(x_i) \\
&\overset{\text{in prob.}}{\to} E(\log p(x)) = -H(X),
\end{aligned}$$

where the last statement is due to the fact that the average of n i.i.d. samples of a random variable converges in probability to the expectation of the random variable. As a result, since $\log y_n$ converges in probability to $-H(X)$, y_n itself converges in probability to $2^{-H(X)}$.

(b) If we go along the same lines as part (a), assuming $y_n = (\prod_{i=1}^{n} f(x_i))^{\frac{1}{n}}$ we obtain

$$\log y_n = \frac{1}{n} \sum_{i=1}^{n} \log f(x_i) \to E(\log f(x)).$$

Thus $y_n \to 2^{E(log f(X))}$.

(c) Firstly we have $g''(u) = \frac{1}{n}(\frac{1}{n} - 1)(u)^{\frac{1}{n}-2} \leq 0$. As a result g is a concave function. Thus given a random variable Y, by Jensen's inequality we have

$$E(g(Y)) \leq g(E(Y)).$$

Now if we take $Y = \prod_{i=1}^{n} f(x_i)$, we have

$$\begin{aligned}
E(g(Y)) &= E((\prod_{i=1}^{n} f(x_i))^{\frac{1}{n}}) \\
&\leq g(E(Y)) \\
&= (E(\prod_{i=1}^{n} f(x_i)))^{\frac{1}{n}} \\
&= (\prod_{i=1}^{n} E(f(x_i)))^{\frac{1}{n}} \\
&= E(f(X)).
\end{aligned}$$

Note that this inequality holds for any $n \in \mathbb{N}$ and we have not considered the convergence issues.

Problem 3

(a) We look at two different solutions:

Solution (1)

$$
\begin{aligned}
H(X|Y) &= H(X,Y) - H(Y) \\
&\stackrel{(1)}{=} H(X) - H(Y)
\end{aligned}
$$

where (1) follows since $Y = f(X)$.
To caculate $H(X|Y) = H(X) - H(Y)$ we need to find $p(X = i) \forall i \in \{2, \cdots, 12\}$ and $p(Y = A), P(Y = B)$, and $P(Y = C)$:

$$
\begin{aligned}
&H(X) = -\sum_{i=2}^{12} p(X = i) \log p(X = i), \\
&H(Y) = -p(Y = A) \log p(Y = A) - p(Y = B) \log p(Y = B) - p(Y = C) \log p(Y = C).
\end{aligned}
$$

Since the two dice are fair, each outcome (i, j) occurs with probability $\frac{1}{36}$. This way,

$$
\begin{aligned}
&p(X = 2) = Pr\{(1,1)\} = \frac{1}{36}, \\
&p(X = 3) = Pr\{(1,2),(2,1)\} = \frac{2}{36}, \\
&p(X = 4) = Pr\{(1,3),(3,1),(2,2)\} = \frac{3}{36}, \\
&p(X = 5) = Pr\{(1,4),(4,1),(3,2),(2,3)\} = \frac{4}{36}, \\
&p(X = 6) = Pr\{(1,5),(5,1),(2,4),(4,2),(3,3)\} = \frac{5}{36}, \\
&p(X = 7) = Pr\{(1,6),(6,1),(2,5),(5,2),(3,4),(4,3)\} = \frac{6}{36}, \\
&p(X = 8) = Pr\{(2,6),(6,2),(3,5),(5,3),(4,4)\} = \frac{5}{36}, \\
&p(X = 9) = Pr\{(3,6),(6,3),(4,5),(5,4)\} = \frac{4}{36}, \\
&p(X = 10) = Pr\{(4,6),(6,4),(5,5)\} = \frac{3}{36}, \\
&p(X = 11) = Pr\{(5,6),(6,5)\} = \frac{2}{36}, \\
&p(X = 12) = Pr\{(6,6)\} = \frac{1}{36}.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&p(Y = A) = p(X = 2) + p(X = 12) = \frac{2}{36}, \\
&p(Y = B) = p(X = 3) + p(X = 11) = \frac{4}{36}, \\
&p(Y = C) = 1 - p(Y = A) - p(Y = B) = \frac{30}{36}.
\end{aligned}
$$

From equations (1) and (1), $H(X)$ and $H(Y)$ are calculated to be $H(X) = 3.7052$ and $H(Y) = 0.8031$ Thus $H(X|Y) = 3.7052 - 0.8031 = 2.9021$.

Solution (2) One should note that

$$
H(X|Y) = p(Y = A)H(X|Y = A)p(Y = B)H(X|Y = B) + p(Y = C)H(X|Y = C).
$$

and furthermore,

$$H(X|Y=S) = -\sum_{i=2}^{12} p_{X|Y}(X=i|Y=S)\log p_{X|Y}(X=i|Y=S),\ \forall S \in \{A,B,C\}.$$

$p(Y=A) = \frac{2}{36}$, $p(Y=B) = \frac{4}{36}$, and $p(Y=C) = \frac{30}{36}$ are found as above. To calculate $H(X|Y=S)$, we need $p_{X|Y}(X=i|Y=S)\ \forall i \in \{2,\cdots,12\}$:

$$p_{X|Y}(X=i|Y=A) = \begin{cases} 0 & \text{if } i \notin \{2,12\} \\ \frac{1}{2} & \text{if } i \in \{2,12\} \end{cases}$$
$$p_{X|Y}(X=i|Y=B) = \begin{cases} 0 & \text{if } i \notin \{3,11\} \\ \frac{1}{2} & \text{if } i \in \{3,11\} \end{cases}$$
$$p_{X|Y}(X=i|Y=C) = \begin{cases} 0 & \text{if } i \notin \{4\cdots,10\} \\ \frac{p(X=i)}{p(Y=C)} & \text{if } i \in \{4,10\} \end{cases}$$

Thus,

$$\begin{aligned} H(X|Y) &= \frac{2}{36}\left(-\frac{1}{2}\log(2) - \frac{1}{2}\log(2)\right) + \frac{4}{36}\left(-\frac{1}{2}\log(2) - \frac{1}{2}\log(2)\right) + \\ &\quad \frac{30}{36}\left(-2*\frac{3}{30}\log\left(\frac{30}{3}\right) - 2*\frac{4}{30}\log\left(\frac{30}{4}\right) - 2*\frac{5}{30}\log\left(\frac{30}{5}\right) - \right. \\ &\quad \left.\frac{6}{30}\log\left(\frac{30}{6}\right)\right) \\ &= 2.9 \end{aligned}$$

(b)

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) \\ &= H(Y) - H(Y|X) \end{aligned}$$

Since $H(X)$ is fixed, and since $H(Y|X) = 0$, maximizing $I(X;Y)$ is just equivalent to maximizing $H(Y)$, or minimizing $H(X|Y)$. In order to maximize $H(Y)$, we know that the maximum is achieved for uniform distribution on Y; i.e., $Pr(Y=A) = Pr(Y=B) = Pr(Y=C) = \frac{1}{3}$ and one such configuration is given by:

$$\begin{cases} A & \text{if } i \in \{2,5,6,11\} \\ B & \text{if } i \in \{3,7,10,12\} \\ C & \text{if } i \in \{4,8,9\} \end{cases}$$

In this case, $I(X;Y) = \log(3)$

Some Remarks:

Although building a ternary Huffman tree on X, divides the X outcomes to three groups of roughly the same probability, in words, in the first layer of the tree, this is not the optimal method to construct Y. Huffman procedure is proved optimal in terms of average codeword-length, or average depth of tree and this is not what we intend to optimize here. you should always be careful when talking about optimality of Huffman procedure. It always matters what quantity you are optimizing.

(c)

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) \\ &= H(Y) - H(Y|X) \end{aligned}$$

since $H(X)$ is fixed, maximizing $I(X;Y)$ is equivalent to minimizing $H(X|Y)$. Since $H(X|Y) \geq 0$, choosing for example

$$\begin{cases} A & \text{if } i \in \{2\} \\ B & \text{if } i \in \{8\} \\ C & \text{if } i \in \{3,4,5,6,7,9,10,11,12\} \end{cases}$$

makes $H(X|Y) = 0$ and achieves the minimum possible $H(X|Y)$. One notes that with this random variable Y, there is a one to one correspondence between X and Y and both $H(X|Y) = 0$ and $H(Y|X) = 0$. In this case $I(X;Y) = H(X) = H(p)$, where we have assumed p to be the probability of $X = 2$ and $1 - p$ the probability of $X = 8$.

Problem 4

The set of outcome probabilities are $p_1 = \frac{1}{12}$, $p_2 = \frac{1}{9}$, $p_3 = \frac{1}{18}$, $p_4 = \frac{1}{6}$, $p_5 = \frac{1}{12}$, $p_6 = \frac{1}{2}$, where $p_i = \mathbb{P}[X = i]$.

(a) For the entropy $H(X)$ we have

$$\begin{aligned}
H(X) &= -\sum_{i=1}^{6} p_i \log_2 p_i \\
&= \frac{1}{12}\log_2 12 + \frac{1}{9}\log_2 9 + \frac{1}{18}\log_2 18 + \frac{1}{6}\log_2 6 + \frac{1}{12}\log_2 12 + \frac{1}{2}\log_2 2 \\
&= \frac{1}{6}\log_2(3 \times 2^2) + \frac{1}{9}\log_2(3^2) + \frac{1}{18}\log_2(2 \times 3^2) + \frac{1}{6}\log_2(2 \times 3) + \frac{1}{2} \\
&= \frac{1}{6}(2 + \log_2 3) + \frac{2}{9}\log_2 3 + \frac{1}{18}(2\log_2 3 + \log_2 2) + \frac{1}{6}(\log_2 2 + \log_2 3) + \frac{1}{2} \\
&= (\frac{1}{3} + \frac{1}{18} + \frac{1}{6} + \frac{1}{2}) + (\frac{1}{6} + \frac{2}{9} + \frac{1}{9} + \frac{1}{6})\log_2 3 \\
&= \frac{19}{18} + \frac{2}{3}\log_2 3 \quad \text{bits} \\
&\approx 2.11 \quad \text{bits.}
\end{aligned}$$

(b) For the equivalence of source coding problem and 20 questions problem you may refer to the lectures or you can find the explanation in the book *Elements of Information Theory*, second edition by Cover and Thomas on page 120.

You can see the Huffman tree for this set of probabilities in Figure 1. To find the best strategy to ask questions, we start from the root of the tree and ask the question wheather the dice outcome belong to the right branch of the tree or not. Depending on the answer to the this question we go to the right or left branch and continue to ask the same question until we reach a leaf of the tree (the outcome of the dice). So the sequence of questions whould be

$$S_1 = \{6\} \rightarrow \begin{cases} \text{Yes:} \quad X = 6 \\ \text{No:} \quad S_2 = \{2,5\} \rightarrow \begin{cases} \text{Yes:} \quad S_3 = \{2\} \rightarrow \begin{cases} \text{Yes:} & X = 2 \\ \text{No:} & X = 5 \end{cases} \\ \text{No:} \quad S_3 = \{4\} \rightarrow \begin{cases} \text{Yes:} \quad X = 4 \\ \text{No:} \quad S_4 = \{1\} \rightarrow \begin{cases} \text{Yes:} & X = 1 \\ \text{No:} & X = 3 \end{cases} \end{cases} \end{cases} \end{cases}$$

Then for the average number of questions we can write

$$
\begin{aligned}
\overline{L} &= \sum_{i=1}^{6} p_i l_i \\
&= 1/2 \times 1 + (1/6 + 1/9 + 1/12) \times 3 + (1/12 + 1/18) \times 4 \\
&= 77/36 \quad \text{bits} \\
&= 2.14 \quad \text{bits.}
\end{aligned}
$$

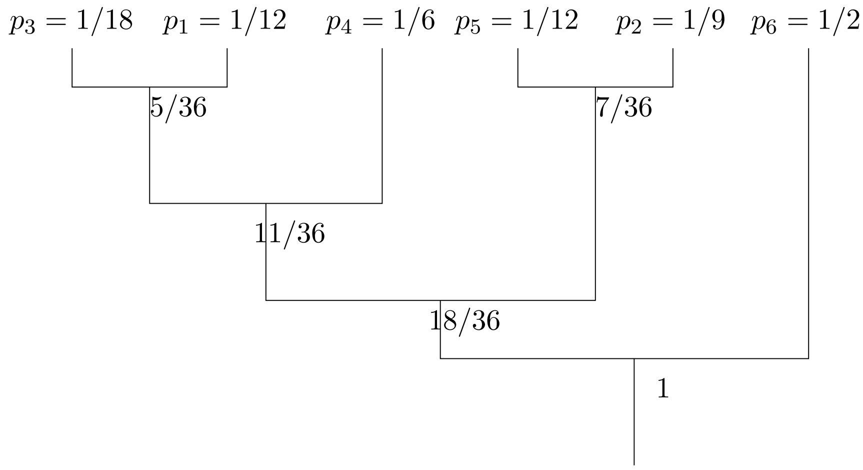

Figure 1: The Huffman tree.

(c) In this part we toss the same dice until the first 6 appears. The random variable Y is the number of required tosses until the first 6 appears. For example if the outcome of tossing is $1, 4, 2, 4, 6$, then $Y = 5$.

Now we want to find the probability $\mathbb{P}[Y = k]$. We know that the probability of observing a 6 is $1/2$ and the probability of not observing 6 is also $1/2$. Because the trials are independent we have

$$
\begin{aligned}
\mathbb{P}[Y = k] &= \mathbb{P}[\text{Probability of not observing 6 for the first } k-1 \text{ tossing}] \\
&\qquad \times \mathbb{P}[\text{Probability of observing 6 at the } k\text{th tossing}] \\
&= \left(\frac{1}{2}\right)^{k-1} \times \left(\frac{1}{2}\right) \\
&= \left(\frac{1}{2}\right)^{k}.
\end{aligned}
$$

(d) For the entropy of Y we can wirte

$$\begin{aligned}
H(Y) &= -\sum_{k=1}^{\infty} \mathbb{P}[Y=k] \cdot \log_2 \mathbb{P}[Y=k] \\
&= \sum_{k=1}^{\infty} \left(\frac{1}{2}\right)^k \cdot \log_2 2^k \\
&= \sum_{k=1}^{\infty} k \cdot \left(\frac{1}{2}\right)^k \\
&= \left.\frac{r}{(1-r)^2}\right|_{r=1/2} \\
&= 2 \quad \text{bits.}
\end{aligned}$$

(e) Again we should use the huffman procedure to find the optimal sequence of "Yes-No" questions. The Huffman tree for the random variable Y is depicted in Figure 2. So

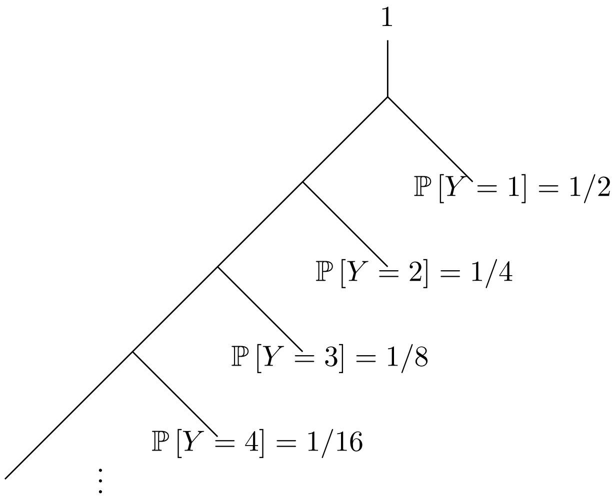

Figure 2: The Huffman tree for the second experiment.

according to the Figure 2 the sequence of questions are $S_1 = \{1\}$, $S_2 = \{2\}$, *etc.* or equivalently "Is $Y = 1$?", "Is $Y = 2$?", *etc.*

(f) With the set of questions designed in part (e), the probability of finding the value of Y after asking k questions is equal to

$$\mathbb{P}[\text{Finding the answer after asking } k \text{ questions}] = \mathbb{P}[Y=k] = (1/2)^k,$$

so for the average number of required questions need to be asked we have

$$\mathbb{E}\text{Number of required questions} = \sum_{k=1}^{\infty} k \times (1/2)^k = H(Y).$$

In this case the average number of required questions is the same as the entropy of Y. Remember that the length of codewords in the Huffman procedure are $\lceil -\log_2 \mathbb{P}[Y=k] \rceil$ where in this case there is no need to apply the function $\lceil \cdot \rceil$ because the probabilities are so that $-\log_2 \mathbb{P}[Y=k]$ are integers themselves.

Problem 5

(a) We can easily observe that the two output bits x_{2i} and x_{2i-1} only depend on the previous two bits x_{2i-2} and x_{2i-3} so formally we have

$$\mathbb{P}[x_{2i}, x_{2i-1} | x_{2i-2}, \ldots, x_1] = \mathbb{P}[x_{2i}, x_{2i-1} | x_{2i-2}, x_{2i-3}].$$

Then for the probabilities we can write

$$\begin{aligned}
\mathbb{P}[x_{2i-1} = 0, x_{2i} = 1 | x_{2i-3} = 1, x_{2i-2} = 0] = 1/2, \\
\mathbb{P}[x_{2i-1} = 1, x_{2i} = 0 | x_{2i-3} = 1, x_{2i-2} = 0] = 1/2, \\
\mathbb{P}[x_{2i-1} = 0, x_{2i} = 1 | x_{2i-3} = 0, x_{2i-2} = 1] = 1/2, \\
\mathbb{P}[x_{2i-1} = 0, x_{2i} = 0 | x_{2i-3} = 0, x_{2i-2} = 1] = 1/2, \\
\mathbb{P}[x_{2i-1} = 1, x_{2i} = 0 | x_{2i-3} = 0, x_{2i-2} = 0] = 1/2.
\end{aligned}$$

(b) We can group the outputs of the FSM two by two. Then from part (a) we know that the sequence $\{(X_{2i-1}, X_{2i})\}$ form a Markov chain. For the state of the new Markov chain we define $Z_i \triangleq (X_{2i-1}, X_{2i})$. Then we have Figure 3 for the transition graph of this Markov chain by using results from part (a).

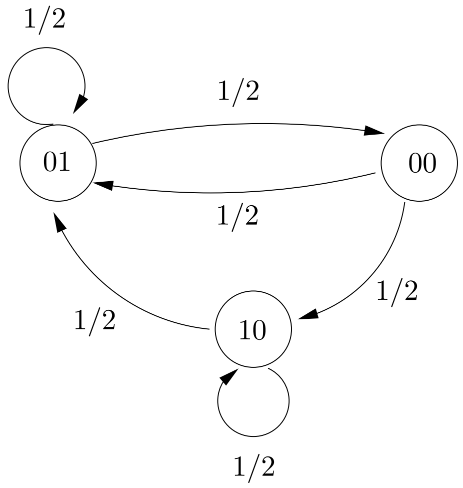

Figure 3: The transition graph of the Markov chain for problem 5, part (b).

(c) For the entropy rate of the source, $\mathcal{H}(X)$, by definition we have

$$\begin{aligned}
\mathcal{H}(X) &\triangleq \lim_{n\to\infty} \frac{1}{2n} H(X_1, \ldots, X_{2n}) \\
&= \lim_{n\to\infty} \frac{1}{2n} H(Z_1, \ldots, Z_n) \\
&= \frac{1}{2} \lim_{n\to\infty} \frac{1}{n} H(Z_1, \ldots, Z_n) \\
&= \frac{1}{2} \mathcal{H}(Z),
\end{aligned}$$

so the entropy rate of the source is half of the entropy rate of the markov chain defined in part (b).

(d) From parts (b) and (c) we know that to find the entropy rate of the source it is sufficient to find the entropy rate of the Markov chain defined in part (b), see Figure 3.

For the transition matrix of the markov chain $\{Z_i\}$ we can write

$$P = \begin{array}{c c} & \begin{array}{ccc} 00 & 01 & 10 \end{array} \\ \begin{array}{c} 00 \\ 01 \\ 10 \end{array} & \left[\begin{array}{ccc} 0 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0 \\ 0 & 0.5 & 0.5 \end{array}\right] \end{array},$$

where $P_{ij} = \mathbb{P}[Z_n = j | Z_{n-1} = i]$ and $i, j \in \{00, 01, 01\}$.

To find the entropy rate of this Markov chain whave to find its stationary distribution so we have to solve the following linear equation

$$\mu \cdot P = \mu,$$

constraint to $\sum_i \mu_i = 1$. It can be easily shown that the solution to this equation is as follows

$$\mu_{00} = \mu_{10} = 1/4, \quad \mu_{01} = 1/2.$$

Then for the entropy rate of the Markov chain $\{Z_i\}$ we have

$$\begin{aligned}
\mathcal{H}(Z) &= H(Z_2|Z_1) \\
&= -\sum_{i,j\in\{00,10,01\}} \mu_i P_{ij} \log_2 P_{ij} \\
&= -1/4(1/2 \log_2 1/2 + 1/2 \log_2 1/2) - 1/2(1/2 \log_2 1/2 + 1/2 \log_2 1/2) \\
&\quad - 1/4(1/2 \log_2 1/2 + 1/2 \log_2 1/2) \\
&= 1.
\end{aligned}$$

So from part (c) for the entropy rate of the source we have

$$\mathcal{H}(X) = \frac{1}{2}\mathcal{H}(Z) = 1/2.$$