Homework 4

4.1

a.

$$b_i = \sum_{k=1}^{n} p_{ki} \cdot a_k$$

$$\sum_{i=1}^{n} b_i = \sum_{i=1}^{n} \sum_{k=1}^{n} p_{ki} \cdot a_k = \sum_{k=1}^{n} a_k = 1$$

$\therefore$ **b** is a probability vector.

Let

$$f(x) = -x \log x$$

$\because f''(x) = -\frac{1}{x} < 0, x \in [0, 1]$

$\therefore f(x)$ is concave on $[0, 1]$. (deem $-\frac{1}{0} = -\infty < 0$)

Since $\sum_{k=1}^{n} p_{ki} = 1$, by Jensen's inequality:

$$-b_i \log b_i = f(b_i) = f(\sum_{k=1}^{n} p_{ki} \cdot a_k) \geq \sum_{k=1}^{n} p_{ki} f(a_k) = -\sum_{k=1}^{n} p_{ki} a_k \log a_k$$

Then

$$H(b_1, b_2, \ldots, b_n) = -\sum_{i=1}^{n} b_i \log b_i \geq -\sum_{i=1}^{n} \sum_{k=1}^{n} p_{ki} a_k \log a_k = -\sum_{k=1}^{n} a_k \log a_k = H(a_1, a_2, \ldots, a_n)$$

b.

Since $\mu = (\mu_1, \mu_2, \ldots, \mu_n)$ is a stationary distribution, we have

$$\mu_j = \sum_{k=1}^{n} \mu_k \cdot P_{kj}$$

This problem asks us to verify that uniform distribution this property, which is true because:

$$\frac{1}{n} = \mu_j = \frac{1}{n} \sum_{k=1}^{n} = \sum_{k=1}^{n} \mu_k \cdot P_{kj}$$

c.

Since $\mu = (\mu_1, \mu_2, \ldots, \mu_n)$ is a stationary distribution, we have

$$\mu_j = \sum_{k=1}^{n} \mu_k \cdot P_{kj}$$

$\because \mu$ is uniform distribution, we have

$$\mu_j = \frac{1}{n} = \sum_{k=1}^{n} \mu_k \cdot P_{kj} = \frac{1}{n} \sum_{k=1}^{n} P_{kj}$$

which means

$$\sum_{k=1}^{n} P_{kj} = 1$$

And since P is Markov transition matrix, we have

$$\sum_{j=1}^{n} P_{kj} = 1$$

$\therefore P$ is doubly stochastic.

4.3

$$\begin{aligned} H(TX) &= H(TX|T) + I(TX;T) \\ &\geq H(TX|T) \end{aligned}$$

If given T, we can reverse TX since $f(TX) = T^{-1}TX$ is a bijection. Therefore:

$$\begin{aligned} H(TX) &= H(TX|T) + I(TX;T) \\ &\geq H(TX|T) \\ &= H(T^{-1}TX|T) \\ &= H(X|T) \\ &= H(X) \text{ , given T and X are independent} \end{aligned}$$

4.6

a.

$$\begin{aligned} \frac{H\left(X_1, X_2, \ldots, X_n\right)}{n} &= \frac{\sum_{i=1}^{n} H\left(X_i \mid X^{i-1}\right)}{n} \\ &= \frac{H\left(X_n \mid X^{n-1}\right) + \sum_{i=1}^{n-1} H\left(X_i \mid X^{i-1}\right)}{n} \\ &= \frac{H\left(X_n \mid X^{n-1}\right) + H\left(X_1, X_2, \ldots, X_{n-1}\right)}{n} \end{aligned}$$

From the inequalities below

$$\begin{aligned} &H\left(X_{n+1} \mid X_n, \ldots, X_1\right) \\ &\leq H\left(X_{n+1} \mid X_n, \ldots, X_2\right) \\ &= H\left(X_n \mid \boldsymbol{X}_{n-1}, \ldots, \boldsymbol{X}_1\right) \end{aligned}$$

We know that $\forall\ 1 \leq i \leq n, H(X_i|X^{i-1}) \geq H(X_n|X^{n-1})$, which means

$$H(X_n|X^{n-1}) \leq \frac{1}{n-1} \sum_{i=1}^{n-1} H(X_i|X^{i-1}) = \frac{1}{n-1} H(X_1, X_2, \ldots, X_{n-1})$$

Therefore

$$\begin{aligned} \frac{H\left(X_1, X_2, \ldots, X_n\right)}{n} &= \frac{H\left(X_n \mid X^{n-1}\right) + H\left(X_1, X_2, \ldots, X_{n-1}\right)}{n} \\ &\leq \frac{H\left(X_1, X_2, \ldots, X_{n-1}\right)}{n-1} \end{aligned}$$

b.

As mentioned in a. ,

$$H(X_n|X^{n-1}) \leq \frac{1}{n}\sum_{i=1}^{n} H(X_i|X^{i-1}) = \frac{1}{n}H(X_1, X_2, \ldots, X_n)$$

4.7

a.

To calculate the stationary distribution, let

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

We have

$$\mu_1 = \frac{p_{10}}{p_{10}+p_{01}}$$
$$\mu_2 = \frac{p_{01}}{p_{10}+p_{01}}$$

The entropy rate:

$$H(\mathcal{X}) = H(X_2|X_1) = \mu_1 H(p_{01}) + \mu_2 H(p_{10}) = \frac{p_{10}H(p_{01}) + p_{01}H(p_{10})}{p_{10}+p_{01}}$$

b.

Since $H(p)$ is concave,

$$\frac{p_{10}H(p_{01}) + p_{01}H(p_{10})}{p_{10}+p_{01}} \leq H(\frac{2p_{10}p_{01}}{p_{01}+p_{10}}) \leq H(\frac{1}{2})$$

According to the equality-hold condition of Jensen's inequality, the equality is reached if and only if $p_{10} = p_{01} = \frac{1}{2}$.

c.

Let $p_{01} = p, p_{10} = 1$

$$H(\mathcal{X}) = H(X_2|X_1) = \frac{p_{10}H(p_{01}) + p_{01}H(p_{10})}{p_{10}+p_{01}} = \frac{H(p)}{1+p}$$

d.

Let $f(p) = \frac{H(p)}{1+p}$

$$f'(p) = \frac{2\log(1-p) - \log p}{(1+p)^2}$$

Make $f'(p) \geq 0$, we have $p \geq \frac{\sqrt{5}+3}{2}$(excluded) or $p \leq \frac{3-\sqrt{5}}{2}$

The maximum entropy rate is $H(\frac{\sqrt{5}+3}{2}) \approx 0.694 bits$

e.

We might as well assume $\mu = (\mu_1, \mu_2)$ respectively denotes two states: 0 and 1.

Then from the transition matrix, we notice that if the current state is 1, then the next state must be 0 while there is no constraint when the current state is 0.

From this observation, we could divide the sequence into two cases:

1. Started with 1, then the next state is 0, and the following $t-2$ states are not fixed.
2. Started with 0, then the next $t-1$ states are not fixed.

Then we have

$$N(t) = N(t-1) + N(t-2), t \geq 3$$

For base cases, we have $N(1) = 2, N(2) = 3$

This form implies that $\{N(t)\}$ is Fibonacci sequence, which have the expression :

$$N(t) = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^{t+1} - \left(\frac{1-\sqrt{5}}{2}\right)^{t+1}\right]$$

Since $|\frac{1-\sqrt{5}}{2}| < 1$, we have

$$\lim_{t\to\infty} \frac{1}{t}\log N(t) = \lim_{t\to\infty}[\frac{t+1}{t}\log(\frac{1+\sqrt{5}}{2}) + \frac{1}{t}\log\frac{1}{\sqrt{5}}] \approx \log\frac{1}{\sqrt{5}} \approx 0.694 bits$$

Since there are $N(t)$ cases of sequence $\{X_i\}_{i=1}^t$, the upper bound of $H(X_1, X_2, \ldots, X_t)$ is $\log N(t)$. And the entropy rate of Markov chain is $\lim_{t\to\infty} \frac{1}{t}H(X_1, X_2, \ldots, X_t)$, thus $H_0 = \lim_{t\to\infty} \frac{1}{t}\log N(t)$ is an upper bound on the entropy rate of Markov chain.

Actually, we see the upper bound is reached in part (d).

4.10

a.

1. If $i, j \in \{1, 2, \ldots, n-1\}$, it has been pointed out that X_i, X_j are i.i.d. random variables, thus independent.
2. If $i \in \{1, 2, \ldots, n-1\}, j = n$.

   We want to prove $p(\sum_{i=1}^n X_i \text{ is odd}) = p(\sum_{i=1}^n X_i \text{ is even}) = \frac{1}{2}$ by induction.

   Assume that it's true for $k-1$, then

   $$P(\sum_{i=1}^k X_i \text{ is odd}) = p(\sum_{i=1}^{k-1} X_i \text{ is odd})p(X_k = 0) + p(\sum_{i=1}^{k-1} X_i \text{ is even})p(X_k = 1) = \frac{1}{2}$$

   Therefore,

   $$p(X_n = 1) = p(X_n = 0) = \frac{1}{2}$$

   WOLG, we just focus on the case below:

   $$p(X_i = 1, X_n = 1) = p(X_i = 1, \sum_{k=1, k\neq i}^n X_k \text{ is even}) = \frac{1}{2}\cdot\frac{1}{2} = p(X_i = 1)p(X_n = 1)$$

Other cases are completely similar, which proves that X_n, X_i are independent.

b.

Since X_i, X_j are independent, we have
$H(X_i, X_j) = H(X_i) + H(X_j) = 2H((X_i)) = 2bits$

c.

Since X_n is the function of $X_1, X_2, \ldots, X_{n-1}$, we have:

$$\begin{aligned} H(X_1, X_2, \ldots, X_n) &= \sum_{i=1}^{n} H(X_i|X_1, X_2, \ldots, X_{i-1}) \\ &= H(X_n|X_1, X_2, \ldots, X_{n-1}) + \sum_{i=1}^{n-1} H(X_i) \\ &= (n-1)H(X_1) \neq nH(X_1) \end{aligned}$$

4.12

a.

Since this is a Markov chain, we have

$$\begin{aligned} H(X_1, X_2, \ldots, X_n) &= H(X_1) + \sum_{i=2}^{n} H(X_i|X_1, X_2, \ldots, X_{i-1}) \\ &= \sum_{i=1}^{n} H(X_i|X_{i-1}) \\ &= H(X_1) + (n-1)H(p) \end{aligned}$$

Since the first step is equally like to be positive or negative, we have $H(X_1) = 1bits$

Therefore,

$$H(X_1, X_2, \ldots, X_n) = H(X_1) + (n-1)H(p) \approx 1 + 0.469(n-1) \quad (bits)$$

b.

Entropy rate is given by

$$\lim_{n\to\infty} \frac{1}{n} H(X_1, X_2, \ldots, X_n) = H(p) \approx 0.469$$

c.

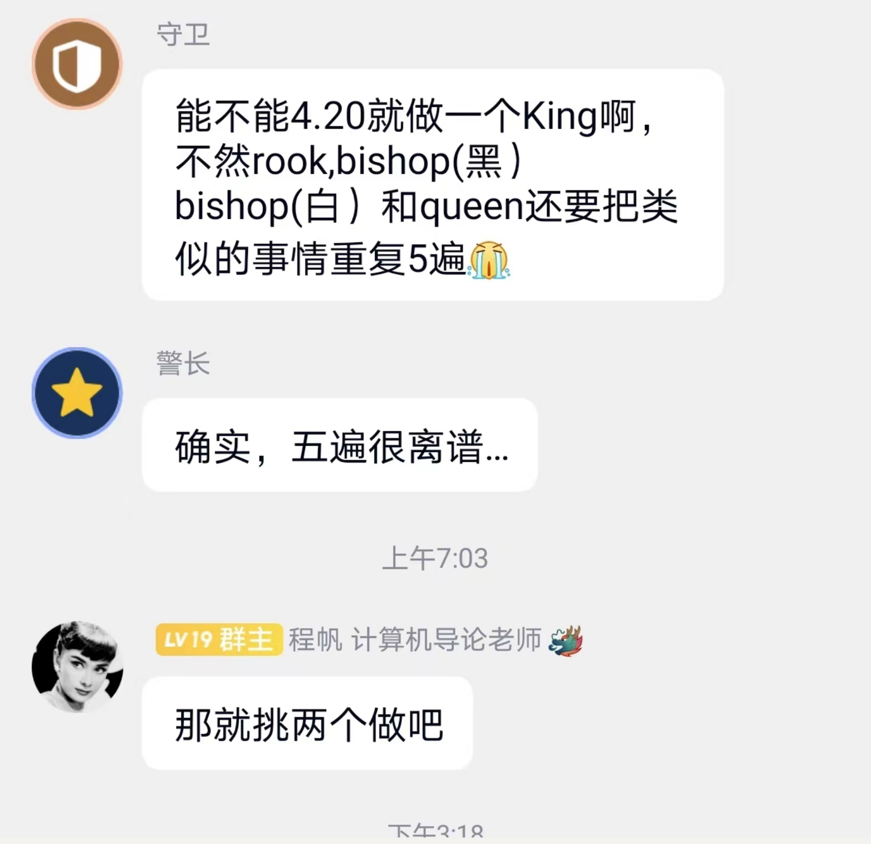
Let $q = 1 - p = 0.9$

$$\begin{aligned} E(S) &= p\sum_{k=0}^{\infty}(k+2)q^{k+2} \\ &= p\lim_{k\to\infty}[\frac{q(1-q^k)}{(1-q)^2} + \frac{2}{1-q} - \frac{(k+2)q^{k+1}}{1-k}] \\ &= 11 \end{aligned}$$

4.15

$$
\begin{aligned}
&\lim_{n\to\infty} \frac{1}{n} H(X_n, \dots, X_1 | X_0, X_{-1}, \dots, X_{-k}) \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} H(X_i | X_{i-1}, X_{i-2}, \dots, X_{-k}) \\
&= \lim_{n\to\infty} H(X_n | X_{n-1}, X_{n-2}, \dots, X_{-k}) \\
&= H(\mathcal{X})
\end{aligned}
$$

4.20



King

The stationary distribution μ_i is E_i/E, where E_i is the number of adjacent points that can be reached by King, E is the sum up of E_i. There are several types of E_i:

- $E_i = 3, i \in \{1, 3, 7, 9\}$
- $E_i = 5, i \in \{2, 4, 6, 8\}$
- $E_i = 8, i = 5$

Then $E = 40$. Since it's of same probabilty for the king to choose a possible direction, we have

- $H(X_2|X_1 = x) = \log 3$ bits, $x \in \{1, 3, 7, 9\}$
- $H(X_2|X_1 = x) = \log 5$ bits, $x \in \{2, 4, 6, 8\}$
- $H(X_2|X_1 = x) = \log 8 = 3$ bits, $x = 5$

$$
\begin{aligned}
H(\mathcal{X}) = H(X_2|X_1) &= \sum p(x) H(X_2|X_1 = x) \\
&= \sum \mu H(X_2|X_1 = x) \\
&= (4 \cdot \frac{3}{40} \log 3 + 4 \cdot \frac{5}{40} \log 5 + \frac{8}{40} \cdot 3) \text{ bits} \\
&\approx 2.24 \text{ bits}
\end{aligned}
$$

Queen

- $E_i = 6, i \in \{1,3,7,9,2,4,6,8\}$
- $E_i = 8, i = 5$

Then $E = 56$. Since it's of same probabilty for the queen to choose a possible direction, we have

- $H(X_2|X_1 = x) = \log 6$ bits, $x \in \{1,3,7,9,2,4,6,8\}$
- $H(X_2|X_1 = x) = \log 8 = 3$ bits, $x = 5$

$$\begin{aligned}
H(\mathcal{X}) = H(X_2|X_1) &= \sum p(x) H(X_2|X_1 = x) \\
&= \sum \mu H(X_2|X_1 = x) \\
&= (8 \cdot \frac{6}{56} \log 6 + \frac{8}{56} \cdot 3) \text{ bits} \\
&\approx 2.644 \text{ bits}
\end{aligned}$$

4.33

$I(X_1; X_4) + I(X_2; X_3) - I(X_1; X_3) - I(X_2; X_4)$

$$\begin{aligned}
&= H(X_1) - H(X_1|X_4) + H(X_2) - H(X_2|X_3) - H(X_1) + H(X_1|X_3) - H(X_2) + H(X_2|X_4) \\
&= -H(X_1|X_4) - H(X_2|X_3) + H(X_1|X_3) + H(X_2|X_4) \\
&= H(X_1, X_2|X_3) - H(X_2|X_1, X_3) - (H(X_1, X_2|X_3) - H(X_1|X_2, X_3)) \\
&+ H(X_1, X_2|X_4) - H(X_1|X_2, X_4) - (H(X_1, X_2|X_4) - H(X_2|X_1, X_4)) \\
&= H(X_1|X_2, X_3) - H(X_2|X_1, X_3) + H(X_2|X_1, X_4) - H(X_1|X_2, X_4)
\end{aligned}$$

Since this is a markov chain, we have $H(X_1|X_2, X_3) = H(X_1|X_2, X_4)$, as the information of X_1 only relates to X_2.

Therefore:

$I(X_1; X_4) + I(X_2; X_3) - I(X_1; X_3) - I(X_2; X_4)$

$$\begin{aligned}
&= H(X_1|X_2, X_3) - H(X_2|X_1, X_3) + H(X_2|X_1, X_4) - H(X_1|X_2, X_4) \\
&= -H(X_2|X_1, X_3) + H(X_2|X_1, X_4) \\
&= -H(X_2|X_1, X_3, X_4) + H(X_2|X_1, X_4) \\
&= I(X_2; X_3|X_1, X_4) \geq 0
\end{aligned}$$

4.34

From data process inequality, we know that $I(X;Y) \geq I(X;Z,W)$, therefore

$I(X;Y) + I(Z;W) - I(X;Z) - I(X;W)$

$$\begin{aligned}
&\geq I(X;Z,W) + I(Z;W) - I(X;Z) - I(X;W) \\
&= H(Z,W) - H(Z,W|X) + H(Z) - H(Z|W) - H(X) + H(X|Z) - H(X) + H(X|W) \\
&= H(Z,W) - H(Z,W,X) + H(X) + H(Z) + H(W) - H(Z,W) - H(X) - H(Z) \\
&+ H(X,Z) - H(X) - H(W) + H(X,W) \\
&= -H(Z,W,X) + H(X,Z) - H(X) + H(X,W) \\
&= H(W|X) - H(W|X,Z) \\
&= I(W;Z|X) \geq 0
\end{aligned}$$