CS258 Information Theory Homework 4

Zhou Litao 518030910407 F1803016

March 24, 2020

Exercise 1 (Monotonicity of entropy per element) For a stationary stochastic process $X_1, X_2, \ldots, X_n$, show that

$$\frac{H(X_1, X_2, \ldots, X_n)}{n} \leq \frac{H(X_1, X_2, \ldots, X_{n-1})}{n-1}$$
$$\frac{H(X_1, X_2, \ldots, X_n)}{n} \geq H(X_n | X_{n-1}, \ldots, X_1)$$

Proof.
We first prove the second statement by properties of stationary process.

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

By Equation 1 and entropy equalities we can prove the first statement.

$$\begin{aligned} (n-1)H(X_1, \cdots, X_n) &\leqslant nH(X_1, \cdots, X_n) - H(X_1, \cdots, X_n) \\ &\leqslant n\left[H(X_1, \cdots, X_n) - H(X_n | X_1, \cdots, X_{n-1})\right] \\ &= nH(X_1, \cdots, X_{n-1}) \end{aligned} \tag{2}$$

□

Exercise 2 (Initial conditions) Show, for a Markov chain, that

$$H(X_0 | X_n) \geq H(X_0 | X_{n-1})$$

Thus, initial conditions X_0 become more difficult to recover as the future X_n unfolds.

Proof. For Markov Chain, we have $p(x_0|x_n, x_{n-1}) = p(x_0|x_{n-1})$. Hence

$$\begin{aligned} H(X_0|X_n, X_{n-1}) &= -\sum_{x_n, x_{n-1}} p(x_n, x_{n-1}) \sum_{x_0} p(x_0|x_n, x_{n-1}) \log p(x_0|x_n, x_{n-1}) \\ &= -\sum_{x_{n-1}} \left(\sum_{x_n} p(x_n, x_{n-1}) \right) \sum_{x_0} p(x_0|x_{n-1}) \log p(x_0|x_{n-1}) \\ &= -\sum_{x_{n-1}} p(x_{n-1}) \sum_{x_0} p(x_0|x_{n-1}) \log p(x_0|x_{n-1}) = H(X_0|X_{n-1}) \end{aligned} \tag{3}$$

Since condition reduce entropy, we have $H(X_0|X_{n-1}) = H(X_0|X_n, X_{n-1}) \leq H(X_0|X_n)$ □

Exercise 3 (The past has little to say about the future) For a stationary stochastic process $X_1, X_2, \ldots, X_n, \ldots$, show that

$$\lim_{n\to\infty} \frac{1}{2n} I\left(X_1, X_2, \ldots, X_n; X_{n+1}, X_{n+2}, \ldots, X_{2n}\right) = 0$$

Proof. First Note that

$$I\left(X_1, X_2, \ldots, X_n; X_{n+1}, X_{n+2}, \ldots, X_{2n}\right) = H(X_1, \ldots, X_n) - H(X_{n+1}, \ldots, X_{2n} | X_1, \ldots, X_n) \tag{4}$$

By definition of entropy rate we know that

$$\frac{1}{2}\sum_{i=1}^{n} H(X_1, \ldots, X_n) \to H(\mathcal{X})$$

□

Exercise 4 (Entropy rate) Let $\{X_i\}$ be a discrete stationary stochastic process with entropy rate $H(\mathcal{X})$. Show that

$$\frac{1}{n} H\left(X_n, \ldots, X_1 | X_0, X_{-1}, \ldots, X_{-k}\right) \to H(\mathcal{X})$$

for $k = 1, 2, \ldots$

Proof.

$$H\left(X_n, \ldots, X_1 | X_0, X_{-1}, \ldots, X_{-k}\right) = \sum_{i=1}^{n} H(X_i | X_{i-1}, \ldots, X_{-k}) \tag{5}$$

Note that

$$H'(X) = \lim_{n\to\infty} H(X_n | X_{n-1}, ldots, X_1) = \lim_{n\to\infty} H(X_n | X_{n-1}, \ldots, X_{-k})$$

We already know that $H'(X) \to H(\mathcal{X})$.

By Cesaro Mean,

$$\frac{1}{n} H\left(X_n, \ldots, X_1 | X_0, X_{-1}, \ldots, X_{-k}\right) \to H(\mathcal{X})$$

□

Exercise 5 (Markov's inequality and Chebyshev's inequality)

1. (Markov's inequality) For any nonnegative random variable X and any $t > 0$, show that

$$\Pr\{X \geq t\} \leq \frac{EX}{t} \tag{6}$$

   Exhibit a random variable that achieves this inequality with equality.

2. (Chebyshev's inequality) Let Y be a random variable with mean μ and variance σ^2. By letting $X = (Y - \mu)^2$, show that for any $\epsilon > 0$

$$\Pr\{|Y - \mu| > \epsilon\} \leq \frac{\sigma^2}{\epsilon^2} \tag{7}$$

3. (Weak law of large numbers) Let $Z_1, Z_2, \ldots, Z_n$ be a sequence of i.i.d. random variables with mean μ and variance σ^2. Let $\bar{Z}_n = \frac{1}{n}\sum_{i=1}^{n} Z_i$ be the sample mean. Show that

$$\Pr\left\{\left|\bar{Z}_n - \mu\right| > \epsilon\right\} \leq \frac{\sigma^2}{n\epsilon^2} \tag{8}$$

   Thus, $\Pr\left\{\left|\bar{Z}_n - \mu\right| > \epsilon\right\} \to 0$ as $n \to \infty$. This is known as the weak law of large numbers.

Proof.

1. Note that for any t,
$$t \cdot 1_{\{X \geq t\}} \leq X \tag{9}$$
By taking expectation at both sides we have
$$E(1_{\{X \geq t\}}) = \Pr(X \geq t) \leq \frac{EX}{t} \tag{10}$$

2. Note $EX = DY = \sigma^2$. By letting $X = (Y - \mu)^2$ and $t = \epsilon^2$ in the Markov inequality, we can derive that
$$\Pr\left\{\left|\bar{Z}_n - \mu\right| > \epsilon\right\} \leq \frac{\sigma^2}{n\epsilon^2} \tag{11}$$

3. Note $E\bar{Z}_n = \mu$ and $D\bar{Z}_n = \frac{1}{n^2}\sum_{i=1}^n DZ_i = \frac{\sigma^2}{n}$. By applying Chebyshev's inequality on $\bar{Z}_n$ we have
$$\Pr\left\{\left|\bar{Z}_n - \mu\right| > \epsilon\right\} \leq \frac{\sigma^2}{n\epsilon^2} \tag{12}$$

□

Exercise 6 (Piece of cake) A cake is sliced roughly in half, the largest piece being chosen each time, the other pieces discarded. We will assume that a random cut creates pieces of proportions
$$P = \begin{cases} \left(\frac{2}{3}, \frac{1}{3}\right) & \text{with probability } \frac{3}{4} \\ \left(\frac{2}{5}, \frac{3}{5}\right) & \text{with probability } \frac{1}{4} \end{cases}$$
Thus, for example, the first cut (and choice of largest piece) may result in a piece of size $\frac{3}{5}$. Cutting and choosing from this piece might reduce it to size $\left(\frac{3}{5}\right)\left(\frac{2}{3}\right)$ at time 2, and so on. How large, to first order in the exponent, is the piece of cake after n cuts?

Solution. Let $C_1, C_2, \ldots, C_n$ denote the choice of each cut. Then after n cuts, the size of the cake $W_n = \prod_{i=1}^n C_i$. By taking the logarithm at the equation, we have $\log W_n = \sum_{i=1}^n \log C_i$. Since C_is are i.i.d., we can apply the law of large numbers as follows.
$$\lim_{n\to\infty} \frac{1}{n} \log W_n = E(\log C) = \frac{3}{4}\log\frac{2}{3} + \frac{1}{4}\log\frac{3}{5} \approx -0.494 \tag{13}$$
Note that the equation above indicates that
$$\frac{1}{n}\log W_n = -0.494 + o(1) W_n \quad = 2^{-0.494t + o(n)} \tag{14}$$
So the first order in the exponent is $\frac{3}{4}\log\frac{2}{3} + \frac{1}{4}\log\frac{3}{5} \approx -0.494$. □

Exercise 7 (AEP) Let X_i be iid $\sim p(x), x \in \{1, 2, \ldots, m\}$. Let $\mu = EX$ and $H = -\sum p(x)\log p(x)$. Let $A^n = \left\{x^n \in \mathcal{X}^n : | - \frac{1}{n}\log p\left(x^n\right) - \ H| \leq \epsilon\right\}$. Let $B^n = \left\{x^n \in \mathcal{X}^n : \left|\frac{1}{n}\sum_{i=1}^n X_i - \mu\right| \leq \epsilon\right\}$

1. Does $\Pr\left\{X^n \in A^n\right\} \longrightarrow 1$?
2. Does $\Pr\left\{X^n \in A^n \cap B^n\right\} \longrightarrow 1$?
3. Show that $\left|A^n \cap B^n\right| \leq 2^{n(H+\epsilon)}$ for all n
4. Show that $\left|A^n \cap B^n\right| \geq \left(\frac{1}{2}\right) 2^{n(H-\epsilon)}$ for n sufficiently large.

Solution.

1. Yes. By the Large Number Law,

$$-\frac{1}{n}\log p(x^n) = -\frac{1}{n}\sum_{i=1}^{n}\log p(X_i) \to H(X) \quad \text{in probability} \tag{15}$$

, which implies that

$$\Pr\{X^n \in A^n\} = \Pr\left\{|-\frac{1}{n}\log p(x^n) - H| \le \epsilon\right\} \to 1 \tag{16}$$

2. Yes. Part (1) implies that $\lim_{n\to\infty}\Pr\{X^n \in A^n\} = 1$.

   By strong law of large number we have $\lim_{n\to\infty}\Pr\{X^n \in B^n\} = 1$

   For arbitrary $\delta > 0$, there exists N_1, such that $\Pr\{X^n \in A^n\} > 1 - \frac{\delta}{2}$ for all $n > N_1$, and there exists N_2, such that $\Pr\{X^n \in B^n\} > 1 - \frac{\delta}{2}$ for all $n > N_2$. We take $N = \max\{N_1, N_2\}$, for any $n > N$, we have

$$\begin{aligned}\Pr\{X^n \in A^n \cap B^n\} &= \Pr\{X^n \in A^n\} + \Pr\{X^n \in B^n\} - \Pr\{X^n \in A^n \cup B^n\} \\ &> 1 - \frac{\delta}{2} + 1 - \frac{\delta}{2} - 1 = 1 - \delta\end{aligned} \tag{17}$$

   , which indicates that $\Pr\{X^n \in A^n \cap B^n\} \to 1$

3. From $x^n \in A^n$ we know that

$$2^{-n(H+\epsilon)} \le p(x^n) \le 2^{-n(H-\epsilon)} \tag{18}$$

$$\begin{aligned}1 &= \sum_{x^n \in \S^n} p(x) \\ &\ge \sum_{x^n \in A^n \cap B^n} p(x) \\ &\ge 2^{-n(H(X)+\epsilon)}|A^n \cap B^n|\end{aligned} \tag{19}$$

   It follows that $|A^n \cap B^n| \le 2^{-n(H+\epsilon)}$.

4. From $\Pr\{X^n \in A^n \cap B^n\} \to 1$, we take $\delta = \frac{1}{2}$, then there exists sufficiently large n, such that

$$\begin{aligned}\frac{1}{2} &\le \Pr(X^n \in A^n \cap B^n) \\ &\le \sum_{x^n \in A^n \cap B^n} p(x^n) \\ &\le |A^n \cap B^n| 2^{-n(H(X)-\epsilon)}\end{aligned} \tag{20}$$

   It follows that $|A^n \cap B^n| \ge \left(\frac{1}{2}\right) 2^{-n(H-\epsilon)}$.

□

Exercise 8 (Doubly stochastic matrices) An $n \times n$ matrix $P = [P_{ij}]$ is said to be doubly stochastic if $P_{ij} \ge 0$ and $\sum_j P_{ij} = 1$ for all i and $\sum_i P_{ij} = 1$ for all j. An $n \times n$ matrix P is said to be a permutation matrix if it is doubly stochastic and there is precisely one $P_{ij} = 1$ in each row and each column. It can be shown that every doubly stochastic matrix can be written as the convex combination of permutation matrices.

1. Let $\mathbf{a}^t = (a_1, a_2, \ldots, a_n), a_i \ge 0, \sum a_i = 1$, be a probability vector. Let $\mathbf{b} = \mathbf{a}P$, where P is doubly stochastic. Show that $\mathbf{b}$ is a probability vector and that $H(b_1, b_2, \ldots, b_n) \ge H(a_1, a_2, \ldots, a_n)$. Thus, stochastic mixing increases entropy.

2. Show that a stationary distribution μ for a doubly stochastic matrix P is the uniform distribution.

3. Conversely, prove that if the uniform distribution is a stationary distribution for a Markov transition matrix P, then P is doubly stochastic.

Proof.

1. From $\mathbf{b} = \mathbf{a}P$ we know that

$$b_j = \sum_{i=1}^{m} a_i p_{ij} \tag{21}$$

Then we have.

$$\begin{aligned} H(\mathbf{b}) - H(\mathbf{a}) &= -\sum_{j=1}^{m} \left(\sum_{i=1}^{m} a_i p_{ij} \right) \log b_j + \sum_{i=1}^{m} a_i \log a_i \\ &= -\sum_{i=1}^{m} a_i \left(\sum_{j=1}^{m} p_{ij} \log b_j \right) + \sum_{i=1}^{m} a_i \left(\sum_{j=1}^{m} p_{ij} \log a_i \right) \\ &= \sum_{i=1}^{m} \sum_{j=1}^{m} a_i p_{ij} \log \frac{a_i}{b_j} \\ &= \sum_{i=1}^{m} \sum_{j=1}^{m} b_j p_{ij} \left(\frac{a_i}{b_j} \log \frac{a_i}{b_j} \right) \\ &\geq \left(\sum_{i=1}^{m} \sum_{j=1}^{m} b_j p_{ij} \frac{a_i}{b_j} \right) \log \left(\sum_{i=1}^{m} \sum_{j=1}^{m} b_j p_{ij} \frac{a_i}{b_j} \right) \\ &= \left(\sum_{i=1}^{m} \sum_{j=1}^{m} p_{ij} a_i \right) \log \left(\sum_{i=1}^{m} \sum_{j=1}^{m} p_{ij} a_i \right) = 1 \cdot \log 1 = 0 \end{aligned} \tag{22}$$

2. By condition we have $\mu_i = \frac{1}{m}$ for any i. Since for any j,

$$\sum_{i=1}^{m} \mu_i p_{ij} = \frac{1}{m} \sum_{i=1}^{m} p_{ij} = \frac{1}{m} = \mu_i \tag{23}$$

We have that $\mu P = \mu$. The uniform distribution is a stationary distribution for a doubly stochastic matrix.

3. From $\mu P = \mu$ and $\mu = \frac{1}{m}$, we know that

$$\sum_{i=1}^{m} \frac{1}{m} p_{ij} = \frac{1}{m} \tag{24}$$

holds for any j, which implies

$$\sum_{i=1}^{m} p_{ij} = 1 \quad \text{for any} j \tag{25}$$

Then P is doubly stochastic.

□

Exercise 9 (Shuffes increase entropy) Argue that for any distribution on shuffles T and any distribution on card positions X that

$$\begin{aligned} H(TX) &\geq H(TX|T) \\ &= H\left(T^{-1}TX|T\right) \\ &= H(X|T) \\ &= H(X) \end{aligned}$$

if X and T are independent.

Proof. The first line holds because condition reduces entropy. The second line holds since T^{-1} can be given by the condition T. The last line holds if X and T are independent, which finishes the proof. □

Exercise 10 (Entropy rates of Markov chains)

1. Find the entropy rate of the two-state Markov chain with transition matrix

$$P = \begin{bmatrix} 1 - p_{01} & p_{01} \\ p_{10} & 1 - p_{10} \end{bmatrix}$$

2. What values of p_{01}, p_{10} maximize the entropy rate?

3. Find the entropy rate of the two-state Markov chain with transition matrix

$$P = \begin{bmatrix} 1 - p & p \\ 1 & 0 \end{bmatrix}$$

4. Find the maximum value of the entropy rate of the Markov chain of part (c). We expect that the maximizing value of p should be less than $\frac{1}{2}$, since the 0 state permits more information to be generated than the 1 state.

5. Let $N(t)$ be the number of allowable state sequences of length t for the Markov chain of part (c). Find $N(t)$ and calculate

$$H_0 = \lim_{t \to \infty} \frac{1}{t} \log N(t)$$

[Hint: Find a linear recurrence that expresses $N(t)$ in terms of $N(t-1)$ and $N(t-2)$. Why is H_0 an upper bound on the entropy rate of the Markov chain? Compare H_0 with the maximum entropy found in part (d).]

Solution.

1. We first calculate the stationary distribution μ.

$$\begin{cases} \mu P = \mu \\ \mu \mathbf{1}^T = 1 \end{cases} \Rightarrow \mu = \left[\frac{p_{10}}{p_{01} + p_{10}}, \frac{p_{01}}{p_{01} + p_{10}} \right] \tag{26}$$

   By the entropy rate of Markov Chain,

$$\begin{aligned} H(\mathcal{X}) &= \sum_{i=1,2} \mu_i \left(\sum_{j=1,2} -p_{ij} \log p_{ij} \right) \\ &= \frac{p_{10} H(p_{01}) + p_{01} H(p_{10})}{p_{01} + p_{10}} \end{aligned} \tag{27}$$

2. Note that the entropy is upper bounded by its alphabet, since

$$\frac{1}{n} H\left(X_1, X_2, \ldots, X_n\right) \leq \frac{1}{n} \log |\mathcal{X}|^n = \log |\mathcal{X}| \tag{28}$$

   Hence the maximal entropy rate for this problem is $\log 2 = 1$. This can be obtained when $p_{01} = p_{10} = \frac{1}{2}$, where

$$\begin{aligned} H(\mathcal{X}) &= \frac{p_{10} H(p_{01}) + p_{01} H(p_{10})}{p_{01} + p_{10}} \\ &= \frac{1}{2} + \frac{1}{2} = 1 \end{aligned} \tag{29}$$

   Therefore, $p_{01} = p_{10} = \frac{1}{2}$ maximize the entropy rate.

3. We first calculate the stationary distribution μ.

$$\begin{cases} \mu P = \mu \\ \mu \mathbf{1}^T = 1 \end{cases} \Rightarrow \mu = \left[\frac{1}{p+1}, \frac{p}{p+1} \right] \tag{30}$$

By the entropy rate of Markov Chain,

$$\begin{aligned} H(\mathcal{X}) &= \sum_{i=1,2} \mu_i \left(\sum_{j=1,2} -p_{ij} \log p_{ij} \right) \\ &= \frac{-p \log p - (1-p) \log(1-p)}{p+1} \end{aligned} \tag{31}$$

4. We take the derivative of $H(\mathcal{X})$.

$$\frac{dH(\mathcal{X})}{dp} = \frac{\log(1-p) - \log p + \log(1-p)}{(p+1)^2} := 0 \quad \Rightarrow \quad p = \frac{3-\sqrt{5}}{2} \tag{32}$$

The maximal entropy rate is $\log \frac{1+\sqrt{5}}{2} \approx 0.6942$.

5. The transition matrix implies that there is no possibility that the last state is 1 with the last but one state to be 1. We can calculate the $N(t)$ recursively. If the last state is 0, the previous state possibilities add up to $N(t-1)$. If the last state is 1, however, the last but one state can only be 0. Then all the previous state possibilities will add up to $N(t-2)$. Further more, we can manually check the initial length that $N(1) = 2$, $N(2) = 3$. Now we have

$$N(t) = N(t-1) + N(t-2) \tag{33}$$

By solving the characteristic equation we know that $N(t)$ must be in the form like

$$N(t) = C_1 (\frac{1+\sqrt{5}}{2})^n + C_2 (\frac{1-\sqrt{5}}{2})^t \tag{34}$$

Then

$$\begin{aligned} H_0 &= \lim_{t\to\infty} \frac{1}{t} \log N(t) \\ &= \lim_{t\to\infty} \frac{1}{t} \log \left(C_1 (\frac{1+\sqrt{5}}{2})^t + C_2 (\frac{1-\sqrt{5}}{2})^t \right) \\ &= \lim_{t\to\infty} \frac{1}{t} \log C_1 (\frac{1+\sqrt{5}}{2})^t \\ &= \lim_{t\to\infty} \frac{1}{t} t \log \frac{1+\sqrt{5}}{2} = \log \frac{1+\sqrt{5}}{2} \end{aligned} \tag{35}$$

We find that H_0 is the upper bound of the entropy rate, and can be obtained with the maximum entropy found in pard (d). This is because by the property of entropy,

$$H(X_1, X_2, \ldots, X_n) \leq \log |X_1, X_2, \ldots, X_n| = N(t) \tag{36}$$

□

Exercise 11 (Maximal entropy graphs) Consider a random walk on a connected graph with four edges.

1. Which graph has the highest entropy rate?
2. Which graph has the lowest?

Solution. In a random walk, the next vertex will be arbitrarily chosen from the adjancent vertices of the current vertex. That is to say, all the edges are given the same weight.

We can formulate this problem as follows. By the inclusion-exclusion principle, there are at most 5 vertices in the graph given four edges. Their adjancent relation can be represented in a 5×5 0-1 matrix$W$, where $W_{ij} = 0$ or 1 and $W_{ij} = W_{ji}$.

Then the transition probability matrix P and the stationary distributionμ can be calculated.

$$P_{ij} = \frac{W_{ij}}{\sum_k W_{ik}} \tag{37}$$

$$\mu_i = \frac{W_i}{2W} \tag{38}$$

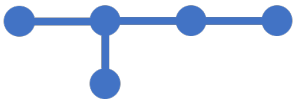
where $W_i = \sum_j w_{ij}$ and $W = \sum_i \frac{W_i}{2}$

Then the entropy rate can be calculated as

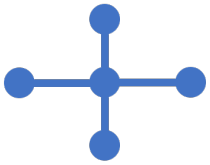
$$\begin{aligned} H(\mathcal{X}) &= H\left(X_2|X_1\right) \\ &= H(X_2, X_1) - H(X_1) \\ &= H(\frac{1}{8}, \frac{1}{8}, \ldots, \frac{1}{8}) - H(\mu) \\ &= 3 - H(\mu) \end{aligned} \tag{39}$$

For 4 edges, there are five possible graphs, as has been shown in Figure 1.

$$W_1 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \quad P_1 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \quad \mu_1^T = \begin{bmatrix} 1/8 \\ 1/4 \\ 1/4 \\ 1/4 \\ 1/8 \end{bmatrix}$$

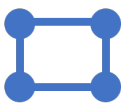


$$W_2 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix} \quad P_2 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1/3 & 0 & 1/3 & 0 & 1/3 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1 & 0 & 1/2 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \quad \mu_2^T = \begin{bmatrix} 1/8 \\ 3/8 \\ 1/4 \\ 1/8 \\ 1/8 \end{bmatrix}$$



$$W_3 = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} \quad P_3 = \begin{bmatrix} 0 & 1/4 & 1/4 & 1/4 & 1/4 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \mu_3^T = \begin{bmatrix} 1/2 \\ 1/8 \\ 1/8 \\ 1/8 \\ 1/8 \end{bmatrix}$$

$$W_4 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} \quad P_4 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 \\ 0 & 1/2 & 0 & 1/2 \\ 0 & 1/2 & 1/2 & 0 \end{bmatrix} \quad \mu_4^T = \begin{bmatrix} 1/8 \\ 3/8 \\ 1/4 \\ 1/4 \end{bmatrix}$$



$$W_5 = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} \quad P_5 = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 \\ 1/2 & 0 & 0 & 1/2 \\ 1/2 & 0 & 0 & 1/2 \\ 0 & 1/2 & 1/2 & 0 \end{bmatrix} \quad \mu_5^T = \begin{bmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix}$$

Figure 1: Five Possible Graphs for Four Edges

Their corresponding entropy rates are 0.75, 0.84436, 1, 1.09436 and 1.

1. The fourth graph has the largest entropy rate.
2. The first graph has the lowest entropy rate.

□