ECE 231A HW 2

Lawrence Liu

October 22, 2022

Problem 1

(a)

No this is not necessarily true. Consider the following example: X_1 is mutually independent fo X_2, but conditional on X_3 they are not independent. Therefore $I(X_1; X_2) = 0$ but $I(X_1; X_2|X_3) > 0$. Therefore $I(X_1; X_2; X_3) < 0$ in this case.

(b)

We have that

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

Since that

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

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

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

Therefore we have that

$$I(X_1;X_2;X_3) = I(X_1;X_2) - I(X_1;X_2|X_3)$$

$$I(X_1;X_2;X_3) = I(X_2;X_3) - I(X_2;X_3|X_1)$$

$$I(X_1;X_2;X_3) = I(X_1;X_3) - I(X_1;X_3|X_2)$$

Since $I(X_1;X_2) \geq 0$, $I(X_2;X_3) \geq 0$, and $I(X_1;X_3) \geq 0$, we have that

$$I(X_1;X_2;X_3) \geq -I(X_1;X_2|X_3)$$

$$I(X_1;X_2;X_3) \geq -I(X_2;X_3|X_1)$$

$$I(X_1;X_2;X_3) \geq -I(X_1;X_3|X_2)$$

Therefore we have that

$$I(X_1;X_2;X_3) \geq -\min(I(X_1;X_2|X_3), I(X_2;X_3|X_1), I(X_1;X_3|X_2))$$

(c)

Once again from

$$I(X_1;X_2;X_3) = I(X_1;X_2) - I(X_1;X_2|X_3)$$

$$I(X_1;X_2;X_3) = I(X_2;X_3) - I(X_2;X_3|X_1)$$

$$I(X_1;X_2;X_3) = I(X_1;X_3) - I(X_1;X_3|X_2)$$

Since $I(X_1;X_2|X_3) \geq 0$, $I(X_2;X_3|X_1) \geq 0$, and $I(X_1;X_3|X_2) \geq 0$, we have that

$$I(X_1;X_2;X_3) \leq I(X_1;X_2)$$
$$I(X_1;X_2;X_3) \leq I(X_2;X_3)$$
$$I(X_1;X_2;X_3) \leq I(X_1;X_3)$$

Therefore we have that

$$I(X_1;X_2;X_3) \leq \min(I(X_1;X_2), I(X_2;X_3), I(X_1;X_3))$$

Problem 2

We have that

$$I(X;Y|U) = H(X|U) - H(X|Y,U)$$

since X and U are indepndent we have

$$I(X;Y|U) = H(X) - H(X|Y,U)$$

Since $I(X;Y,U)$ we get:

$$I(X;Y,U) = I(X;Y|U)$$

Since $H(X|Y,Z) = 0$, $H(X|U,Y,Z) = 0$ since conditioning reduces probability but entropy cannot be negative. Then we have

$$\begin{aligned}
H(X|Y,U) &= H(X,Z|Y,U) - H(Z|X,Y,U) \\
&\leq H(X,Z|Y,U) \\
&= H(X,Z,Y,U) - H(Y,U) \\
&= H(X|U,Y,Z) + H(U,Y,Z) - H(Y,U) \\
&= H(U,Y,Z) - H(Y,U) \\
&= H(Z|U,Y) \\
&\leq H(Z|U)
\end{aligned}$$

Therefore we have that

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

Therefore

$$H(X) - H(X|Y,U) \geq H(X) - H(Z|U)$$

Therefore we have that

$$I(X;Y,U) = I(X;Y|U) \geq H(X) - H(Z|U)$$

Problem 3

(a)

Let the increase in Alice's score after the ith round be represented by the random variable Z_A^i we have that

$$Z_A^i = \begin{cases} 4 & \text{w.p. } 1/15 \\ 7 & \text{w.p. } 5/15 \\ 0 & \text{w.p. } 9/15 \end{cases}$$

And the increase in Bob's score after the ith round be represented by the random variable Z_B^i we have that

$$Z_B^i = \begin{cases} 3 & \text{w.p. } 1/15 \\ 5 & \text{w.p. } 2/15 \\ 6 & \text{w.p. } 6/15 \\ 0 & \text{w.p. } 6/15 \end{cases}$$

Then we have that

$$S_A^n = \sum_{i=1}^{n} Z_A^i$$

and

$$S_B^n = \sum_{i=1}^{n} Z_B^i$$

As $n \to \infty$ we have that

$$\lim_{n\to\infty} S_A^n = \lim_{n\to\infty} \sum_{i=1}^{n} Z_A^i = E[Z_A^i] = \boxed{2.6}$$

and

$$\lim_{n\to\infty} S_B^n = \lim_{n\to\infty} \sum_{i=1}^{n} Z_B^i = E[Z_B^i] = \boxed{3.2666}$$

(b)

We will need to make $\alpha > 7$ since we need to increase the expected value, therefore the probabilites for Z_A would not change, however instead of being 7 with the probability of 5/15 we would have α with the probability of 5/15. Likewise Z_B would not change. Therefore we would have that our new

$$E[Z_A^i] = \frac{4}{15} + \frac{5}{15}\alpha$$

To make this greater than 3.2666 we would have that $\boxed{\alpha > 9}$,

Problem 4

(a)

We have that $H(\hat{p}(Y_m)|Y_m) = 0$ and $H(\hat{p}(Y_m)|Y_m, p) = 0$, therefore we have that $I(\hat{p}(Y_m); p|Y_m) = 0$

$$\begin{aligned} I(p; Y_m|\hat{p}) + I(\hat{p}(Y_m); p) &= I(\hat{p}(Y_m); p, Y_m) = I(p; \hat{p}(Y_m)|Y_m) + I(p; Y_m) \\ I(\hat{p}(Y_m); p) + I(\hat{p}(Y_m); p) &= I(p; Y_m) \\ I(\hat{p}(Y_m); p) &\leq I(p; Y_m) \end{aligned}$$

With the last line coming from the fact that $I(\hat{p}(Y_m); p) \geq 0$

(b)

If $\sum_{i=1}^n X_i^m \neq k$ then $\hat{p}(Y_m) \neq \frac{k}{n}$ therefore we must have $P[Y_m|\hat{p}(Y_m) = \frac{k}{n}] = 0$ for all $\sum_{i=1}^n X_i^m \neq k$. If $\sum_{i=1}^n X_i^m = k$ then we have $\binom{n}{k}$ possible Y_m, since X_i^m is iid, each of these are equally probable therefor we have

$$P[Y_m|\hat{p}(Y_m) = \frac{k}{n}] = \begin{cases} \frac{1}{\binom{n}{k}} & \text{if } \sum_{i=1}^n X_i^m = k \\ 0 & \text{otherwise} \end{cases}$$

(c)

Let $\hat{p} = \hat{p}(Y_m)$ First let us prove that $P(Y_m|\hat{p}, p) = P(Y_m|\hat{p})$.

$$P(Y_m|\hat{p}, p) = P(Y_m|\hat{p})$$
$$P(Y_m, \hat{p}, p) = P(Y_m|\hat{p})p(\hat{p}|p)p(p)$$

For all the cases of Y_m such that $\sum_{i=1}^n X_i^m \neq k$ we have that $P(Y_m, \hat{p}, p) = 0$ and $P(Y_m|\hat{p}) = 0$, therefore we will only need to Consider the cases where $\sum_{i=1}^n X_i^m = k$, for these cases we have

$$P(Y_m, \hat{p}, p) = \frac{1}{\binom{n}{k}} p(\hat{p}|p)p(p)$$

Since X_i^m is iid we have that $p(\hat{p}|p) = \binom{n}{k} p(Y_m|p)$, therefore Therefore we have that

$$\begin{aligned} P(Y_m, \hat{p}, p) &= p(Y_m|p)p(p) \\ &= P(Y_m, p) \end{aligned}$$

This is true since $\hat{p}$ is a function of Y_m. Therefore we have that, p, $\hat{p}$ and Y_m form the following Markov chain:

$$p \to \hat{p} \to Y_m$$

Therefore from the data processing inequality we have

$$I(\hat{p}(Y_m); p) \geq I(p; Y_m)$$

therefore from this and problem 4(a) we have that

$$I(\hat{p}(Y_m); p|Y_m) = I(p; Y_m)$$

(d)

From part (c) we have that $I(p; \hat{p}(Y_i)) = I(p; Y_i)$, therefore we have that we just need to prove

$$I(Y_i; p) \geq I(Y_j; p)$$

If we prove this inequality for $j = i+1$, we have proved it for all $j > i$. Since $H(Y_{i+1}|Y_i, p) = H(Y_{i+1}|Y_i)$, therefore we have that $I(Y_{i+1}; p|Y_i) = 0$ therefore we have

$$\begin{aligned} I(Y_i; p|Y_{i+1}) + I(Y_{i+1}; p) &= I(Y_i; p, Y_{i+1}) = I(Y_{i+1}; p|Y_i) + I(Y_i; p) \\ I(Y_i; p|Y_{i+1}) + I(Y_{i+1}; p) &= I(Y_i; p) \\ I(Y_{i+1}; p) &\leq I(Y_i; p) \end{aligned}$$

With the last line coming from the fact that $I(Y_i; p|Y_{i+1}) \geq 0$ Therefore we have that

$$I(Y_i; p) \geq I(Y_j; p)$$

for all $j > i$.

(e)

From part (d) we have that the Y_i's form a markov chain thus we have that

$$Y_i \to Y_{i+1} \to Y_{i+2} \to \cdots Y_{m-1} \to Y_m$$

Since $\hat{p}(Y_m)$ is a function of Y_m we have that

$$Y_i \to Y_{i+1} \to \cdots Y_{m-1} \to Y_m \to \hat{p}(Y_m)$$

Forms a markov chain as well. We have

$$\begin{aligned} I(\hat{p}(Y_m); Y_{m-1}|Y_i) &\geq I(\hat{p}(Y_m); Y_{m-1}|Y_j) \\ H(\hat{p}(Y_m)|Y_i) - H(\hat{p}(Y_m)|Y_i, Y_{m-1}) &\geq H(\hat{p}(Y_m)|Y_j) - H(\hat{p}(Y_m)|Y_j, Y_{m-1}) \\ H(\hat{p}(Y_m)|Y_i) - H(\hat{p}(Y_m)|Y_{m-1}) &\geq H(\hat{p}(Y_m)|Y_j) - H(\hat{p}(Y_m)|Y_{m-1}) \\ H(\hat{p}(Y_m)|Y_i) &\geq H(\hat{p}(Y_m)|Y_j) \end{aligned}$$

We can see that this is intuitively true, since Y_1 is "further" from Y_m and $\hat{p}(Y_m)$ than Y_j is, therefore it has less "determination" on Y_m and $\hat{p}(Y_m)$.

Problem 5

(a)

In order for $\mu = [\mu_1, \mu_2]^T$ to be a stationary distribution for a Markov chain we must have $\mu^T \Pi = \mu^T$, therefore we have the following two series of equations

$$\frac{1}{4}\mu_1 + \frac{2}{3}\mu_2 = \mu_1$$
$$\frac{3}{4}\mu_1 + \frac{1}{3}\mu_2 = \mu_2$$

Solving these we get, and applying the condition that $\mu_1 + \mu_2 = 1$, we get $\mu_1 = \frac{\frac{2}{3}}{\frac{3}{4}+\frac{2}{3}} = \boxed{0.470}$ and $\mu_2 = \frac{\frac{3}{4}}{\frac{3}{4}+\frac{2}{3}} = \boxed{0.529}$ Therefore the entropy of this $H(X) = -0.470\log_2(0.470) - 0.529\log_2(0.529) = \boxed{0.9975 \text{ shannons}}$

(b)

We have that

$$\begin{aligned}\mathcal{H}(x) &= -\sum_i \mu_i \sum_j P_{ij}\log_2(P_{ij}) \\ &= -0.470\left(\frac{1}{4}\log_2\left(\frac{1}{4}\right) + \frac{3}{4}\log_2\left(\frac{3}{4}\right)\right) - 0.529\left(\frac{2}{3}\log_2\left(\frac{2}{3}\right) + \frac{1}{3}\log_2\left(\frac{1}{3}\right)\right) \\ &= \boxed{0.8679345589507923}\end{aligned}$$

This is less than $H(X)$ since $\mathcal{H}(x) = \lim_{n\to\infty} H(X_n|X_{n-1},\ldots,X_1)$, since conditioning reduces entropy, thus we have that $H(X_n|X_{n-1},\ldots,X_1) < H(X)$.

(c)

Since we can determine X_n given $Y_n, ..., Y_1$ and since Y_n is a deterministic function we have

$$H(Y_n|Y_{n-1},\ldots,Y_1) = H(X_n|X_{n-1},...,X_1)$$

And therefore we have that $\mathcal{H}(Y) = \mathcal{H}(X) = \boxed{0.86793455589507923}$

(d)

We have that the probability of X_n being different from X_{n-1} is just

$$P(Y_n = 1) = \frac{3}{4}\mu_1 + \frac{2}{3}\mu_2 = 0.705$$

Therefore we have

$$H(Y_n) = -P(Y_n = 1)\log_2(P(Y_n = 1)) - (1 - P(Y_n = 1))\log_2(1 - P(Y_n = 1)) = \boxed{0.8739}$$