Information-Theoretic Methods in Computer Science (CS1.502)

Spring 2023, IIIT Hyderabad
Assignment 1

Abhinav S Menon

Problem 1

Part (a)

Let us define a new r.v. $Z = g(X, Y)$. We then need to show that

$$\mathbb{E}[Z]_{p_Z} = \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} g(x, y) P_{XY}(x, y).$$

We have

$$\mathbb{E}[Z] = \sum_{z \in \mathcal{Z}} z p(z)$$

where each z in the summation on the RHS can be expressed as $g(x, y)$ for some x, y. Making this substitution, we get

$$\begin{aligned}
\mathbb{E}[Z] &= \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} g(x, y) p(Z = z) \\
&= \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} g(x, y) p(X = x, Y = y) \\
&= \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} g(x, y) P_{XY}(x, y)
\end{aligned}$$

Part (b)

Consider the special case $g(x, y) = x + y$ in the above expression.

$$\mathbb{E}[X + Y] = \sum_{x,y} (x + y) P_{XY}(x, y)$$

Splitting the summation into two and the addition into two terms, we get

$$\begin{aligned}
\mathbb{E}[X+Y] &= \sum_x \sum_y x P_{XY}(x,y) + y P_{XY}(x,y) \\
&= \sum_x \sum_y x P_{XY}(x,y) + \sum_y \sum_x y P_{XY}(x,y) \\
&= \sum_x x \sum_y P_{XY}(x,y) + \sum_y y \sum_x P_{XY}(x,y) \\
&= \sum_x x P_X(x) + \sum_y y P_Y(y) \\
&= \mathbb{E}[X] + \mathbb{E}[Y].
\end{aligned}$$

Problem 2

Part (a)

First, consider the forward implication; if $H(Y \mid X) = 0$, then each x has exactly one value of y for which $p(x,y) = 0$.
We can express $H(Y \mid X)$ as

$$\sum_x p(x) \sum_y p(y \mid X = x) \log \frac{1}{p(y \mid X = x)}.$$

We know that each term of the x-summation is non-negative (since all probabilities are between 0 and 1); therefore, if the sum is 0, then each term is 0.
Now, for any x, we have that

$$\sum_y p(y \mid X = x) \log \frac{1}{p(y \mid X = x)},$$

which means that either $p(y \mid X = x)$ is 0 or $p(y \mid X = x)$ is $\frac{1}{1} = 1$.
Since $\sum_y p(y \mid X = x) = 1$, we know that the latter can be true for exactly one value of y. For this value, $p(x,y) = p(x)p(y \mid X = x) = p(x) > 0$.
Therefore, for each x, there is exactly one value of y for which $p(x,y) > 0$.

The backward implication is straightforward. Let there be only one possible value of y for which $p(x) > 0$.
Then consider $p(Y \mid X)$; for a certain x, we have that $\sum_y p(y \mid X = x) = \sum_y \frac{p(x,y)}{p(x)} = 1$. If only one term of this sum is positive, it must be one; therefore, $p(Y \mid X = x)$ has its entire mass concentrated at one point. Therefore it has entropy 0, for all x.
Therefore, $H(Y \mid X) = \sum_x p(x) H(Y \mid X = x)$ must also be 0.

Part (b)

Consider the joint entropy $H(X, g(X))$. We can write this as

$$H(X) + H(g(X) \mid X)$$

or as

$$H(g(X)) + H(X \mid g(X)).$$

From the equality of these expressions, we get that

$$H(X) - H(g(X)) = H(X \mid g(X)) - H(g(X) \mid X).$$

As proved above, however, the second term on the RHS is zero, and the first term (by the properties of entropy) is nonnegative. Therefore,

$$\begin{aligned} H(X) - H(g(X)) &= H(X \mid g(X)) - H(g(X) \mid X) \\ &= H(X \mid g(X)) \\ &\geq 0, \end{aligned}$$

QED.

Part (c)

Let $f(t) = t \log t$. Then

$$f'(t) = t\left(\frac{1}{t}\right) + 1 \cdot \log t = 1 + \log t$$

and

$$f''(t) = 0 + \frac{1}{t} = \frac{1}{t} > 0.$$

This proves that f is strictly convex.

Now, we can write $H(X \mid Y)$ as

$$\sum_y p(y) \sum_x p(x \mid Y = y) \log \frac{1}{p(x \mid Y = y)}.$$

By the convexity of f, we have

$$\sum \lambda_i f(t) \geq f\left(\sum \lambda_i t\right).$$

Therefore,

$$\begin{aligned}
-H(X \mid Y) &= \sum_y p(y) \sum_x p(x \mid Y=y) \log p(x \mid Y=y) \\
&= \sum_x \sum_y p(y) \left[p(x \mid Y=y) \log p(x \mid Y=y) \right] \\
&\geq \sum_x \left(\sum_y p(y) p(x \mid Y=y) \right) \log \left(\sum_y p(y) p(x \mid Y=y) \right) \\
&= \sum_x \left(\sum_y p(x,y) \right) \log \left(\sum_y p(x,y) \right) \\
&= \sum_x p(x) \log p(x) \\
&= -H(X),
\end{aligned}$$

which implies that $H(X \mid Y) \leq H(X)$, QED.

If X and Y are independent, then we have

$$\begin{aligned}
H(X \mid Y) &= \sum_{x,y} p(x,y) \log \frac{1}{p(x \mid Y=y)} \\
&= \sum_{x,y} p(x)p(y) \log \frac{1}{p(x)} \\
&= \sum_x p(x) \log \frac{1}{p(x)} \sum_y p(y) \\
&= \sum_x p(x) \log \frac{1}{p(x)} \\
&= H(X).
\end{aligned}$$

Conversely, if $H(X \mid Y) = H(X)$, then we have

$$\begin{aligned}
H(X \mid Y) - H(X) &= \sum_x p(x) \log \frac{1}{p(x)} - \sum_{x,y} p(x,y) \log \frac{1}{p(x \mid Y=y)} \\
&= \sum_{x,y} p(x,y) \log \frac{1}{p(x)} - \sum_{x,y} p(x,y) \log \frac{1}{p(x \mid Y=y)} \\
&= \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \\
&= 0.
\end{aligned}$$

which means that for every x, y, either $p(x,y) = 0$ or $p(x,y) = p(x)p(y)$, so X and Y are independent.

Thus, X and Y are independent iff $H(X \mid Y) = H(X)$, QED.

Part (d)

Consider two binary r.vs. X and Y. Let their PMF be given by

$$p(x,y) = \begin{cases} 0.3 & x=0, y=0 \\ 0.3 & x=0, y=1 \\ 0.3 & x=1, y=0 \\ 0.1 & x=1, y=1 \end{cases}.$$

Then we have

$$H(X \mid Y=0) = 0.5 \log \frac{1}{0.5} + 0.5 \log \frac{1}{0.5} = 1,$$

while

$$H(X) = 0.6 \log \frac{1}{0.6} + 0.4 \log \frac{1}{0.4} \approx 0.495.$$

Thus, $H(X \mid Y=0) < H(X)$ for this distribution, as required.

Problem 3

Part (a)

We need to show that $I(X;Y \mid Z) \le I(X;Y)$.

First, consider the quantity $p(z \mid x, y)$.

$$\begin{aligned} p(Z \mid X, Y) &= \frac{p(Z, X, Y)}{p(X, Y)} \\ &= \frac{p(X, Z \mid Y) p(Y)}{p(X \mid Y) p(Y)} \\ &= \frac{p(X \mid Y) p(Z \mid Y)}{p(p \mid Y)} \\ &= p(Z \mid Y), \end{aligned}$$

due to the Markov property $p(X, Z \mid Y) = p(X \mid Y) p(Z \mid Y)$.
Since these probability distributions are equal, their entropies also must be:

$$H(Z \mid X, Y) = H(Z \mid Y).$$

Now, we can write $I(X;Y) - I(X;Y \mid Z)$ as

$$\begin{aligned} I(X;Y) - I(X;Y \mid Z) &= (H(X) - H(X \mid Y)) - (H(X \mid Z) - H(X \mid Y, Z)) \\ &= (H(X) - H(X \mid Z)) - (H(X \mid Y) - H(X \mid Y, Z)) \\ &= I(X;Z) - I(X;Z \mid Y) \\ &= I(Z;X) - I(Z;X \mid Y). \end{aligned}$$

The second term on the RHS is equal to

$$I(Z; X \mid Y) = H(Z \mid Y) - H(Z \mid Y, X)$$

which, as we proved above, is zero.

Therefore

$$I(X; Y) - I(X; Y \mid Z) = I(Z; X) \geq 0,$$

by the nonnegativity of mutual information.

Part (b)

First, consider $I(X; Y \mid Z) = H(X \mid Z) - H(X \mid Y, Z)$.
For $H(X \mid Z)$, if $z \in \{0, 2\}$, then $H(X \mid Z = z)$ becomes 0 (probability mass concentrated at a point).
If $z = 1$, then $p(X \mid Z)$ is uniform, so $H(X \mid Z) = 1$. This event has probability 0.5, so the total entropy is $H(X \mid Z) = 0.25 \cdot 0 + 0.25 \cdot 0 + 0.5 \cdot 1 = 0.5$. For $H(X \mid Y, Z)$, we can always uniquely identify X given Y and Z (by the relation $X = Z - Y$). This means that $H(X \mid Y, Z) = 0$.
Therefore $I(X; Y \mid Z) = 0.5 - 0 = 0.5$.

Next, consider $I(X; Y) = H(X) - H(X \mid Y)$.
Since X and Y are independent, $H(X \mid Y) = H(X)$. THerefore $I(X; Y) = 0$.

This proves that $I(X; Y \mid Z) > I(X; Y)$, QED.

Part (c)

We have two Markov chains $X \to Y \to Z$ and $X \to Z \to Y$.

The data processing inequality on the first chain tells us that

$$I(X; Y) \geq I(X; Z),$$

while on the second one it tells us that

$$I(X; Z) \geq I(X; Y).$$

Combining these two statements we see that $I(X; Y) = I(X; Z)$, QED.

To prove the data processing inequality for an arbitrary chain $A \to B \to C$, note that

$$I(A; B, C) = I(A; C) + I(A; B \mid C) = I(A; B) + I(A; C \mid B),$$

by the chain rule of mutual information. However, since $I(A; B \mid C) = 0$ and $I(A; C \mid B) \geq 0$, we get $I(A; C) \geq I(A; B)$.

Problem 4

Part (a)

We will proceed by induction on n. For the base case, consider $n = 2$.

FIrst, note that

$$\begin{aligned} p(x_1, x_2 \mid y) &= \frac{p(x_1, x_2, y)}{p(y)} \\ &= \frac{p(x_2 \mid x_1, y)p(x_1, y)}{p(y)} \\ &= p(x_2 \mid x_1, y)p(x_1 \mid y). \end{aligned}$$

Using this, we can show that

$$\begin{aligned} H(X_1, X_2 \mid Y) &= \sum_{x_1, x_2, y} p(x_1, x_2, y) \log \frac{1}{p(x_1, x_2 \mid y)} \\ &= \sum_{x_1, x_2, y} p(x_1, x_2, y) \left[\log \frac{1}{p(x_2 \mid x_1, y)} + \log \frac{1}{p(x_1 \mid y)}\right] \\ &= \sum_{x_1, x_2, y} p(x_1, x_2, y) \log \frac{1}{p(x_2 \mid x_1, y)} \\ &+ \sum_{x_1, y} \log \frac{1}{p(x_1 \mid y)} \sum_{x_2} p(x_1, x_2, y) \\ &= H(X_2 \mid X_1, Y) + \sum_{x_1, y} p(x_1, y) \log \frac{1}{p(x_1 \mid y)} \\ &= H(H_2 \mid X_1, Y) + H(X_1 \mid Y). \end{aligned}$$

This proves our base case.

Now, assume $H(X_1, \ldots, X_{n-1} \mid Y) = \sum_{i=1}^{n-1} H(X_i \mid X_1, \ldots, X_{i-1} \mid Y)$.
Let $(X_1, \ldots, X_{n-1}) = \mathbf{X}$. Then,

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

QED.

Part (b)

We know that

$$I(X;Y) = H(X) - H(X \mid Y)$$

and that

$$H(X_1, \ldots, X_n) = \sum_{i=1}^{n} H(X_i \mid X_1, \ldots, X_{i-1}).$$

Using these two identities, we have

$$\begin{aligned}
I(X_1, \ldots, X_n; Y) &= H(X_1, \ldots, X_n) - H(X_1, \ldots, X_n \mid Y) \\
&= \sum_{i=1}^{n} H(X_i \mid X_1, \ldots, X_{i-1}) - \sum_{i=1}^{n} H(X_i \mid X_1, \ldots, X_{i-1}, Y) \\
&= \sum_{i=1}^{n} H(X_i \mid X_1, \ldots, X_{i-1}) - H(X_i \mid X_1, \ldots, X_{i-1}, Y) \\
&= \sum_{i=1}^{n} I(X_i; Y \mid X_1, \ldots, X_{i-1}),
\end{aligned}$$

QED.