Problem Set 2
Information Measures

For the Exercise Sessions on Sept 17 and Oct 1 — **Due: Tue, October 7, 10am, on Moodle**

1 Problems for Class

We will solve these problems together in class on Tuesday, Sept 16, and on Tuesday, Sept 23.

Problem 1: Entropy and pairwise independence

Consider three binary random variables X, Y, Z. Each of the three random variables is uniformly distributed, but they are not independent. However, we know that they are *pairwise* independent. That is, X and Y are independent, and X and Z are independent, and Y and Z are independent.

(a) What is $H(X, Y)$?

(b) Give a lower bound to the value of $H(X, Y, Z)$.

(c) Give an example that achieves this bound.

Solution 1. *(a)* Since X, Y, Z are pairwise independent fair flips, $H(X) = H(Y) = H(Z) = 1$, and $H(Y|X) = H(Y)$. Therefore, using the chain rule for entropy, $H(X, Y) = H(X) + H(Y|X) = H(X) + H(Y) = 2$.

(b) Using the chain rule for entropy, we can write: $H(X, Y, Z) = H(X, Y) + H(Z|X, Y) \geq H(X, Y) = 2$, which holds since (conditional) entropy is non-negative, thus $H(Z|X, Y) \geq 0$.

(c) Let $Z = X + Y \mod 2$, then $H(Z|X, Y) = 0$ and $H(X, Y, Z) = H(X, Y)$.

Problem 2: Conditional KL divergence

We saw in class that a *probability kernel* $P_{Y|X} : \mathcal{X} \to \mathcal{Y}$ is a matrix $P_{Y|X} = P_{Y|X}(y|x) : x \in \mathcal{X}, y \in \mathcal{Y}$ such that $P_{Y|X}(y|x) \geq 0$, and for each $x \in \mathcal{X}, \sum_y P_{Y|X}(y|x) = 1$. Let $P_X \in \Pi(\mathcal{X})$ be a probability distribution on $\mathcal{X}$. We define the *conditional KL divergence* between two probability kernels $P_{Y|X} : \mathcal{X} \to \mathcal{Y}$ and $Q_{Y|X} : \mathcal{X} \to \mathcal{Y}$ given P_X to be

$$D(P_{Y|X} \| Q_{Y|X} | P_X) \triangleq \sum_{x \in \mathcal{X}} P_X(x) D(P_{Y|X}(\cdot|x) \| Q_{Y|X}(\cdot|x))$$

where for every x, $D(P_{Y|X}(\cdot|x) \| Q_{Y|X}(\cdot|x))$ is the standard KL divergence between the two distributions $P_{Y|X}(\cdot|x)$ and $Q_{Y|X}(\cdot|x)$ over $\mathcal{Y}$.

(a) *(Chain rule of the KL divergence)* Show that

$$D(P_{X,Y} \| Q_{X,Y}) = D(P_X \| Q_X) + D(P_{Y|X} \| Q_{Y|X} | P_X)$$

where $P_{X,Y}$ and $Q_{X,Y}$ are two joint distributions on $\mathcal{X} \times \mathcal{Y}$ such that $P_{X,Y}(x, y) = P_X(x) P_{Y|X}(y|x)$ and $Q_{X,Y}(x, y) = Q_X(x) Q_{Y|X}(y|x)$.

(b) Using (a), show that

$$D(P_{Y|X} \| Q_{Y|X} | P_X) = D(P_{X,Y} \| Q_{X,Y})$$

where $P_{X,Y}(x,y) = P_X(x) P_{Y|X}(y|x)$ and $Q_{X,Y}(x,y) = P_X(x) Q_{Y|X}(y|x)$.

(c) *(Conditioning increases divergence)* Using (b) and the Data Processing Inequality seen in class, show that

$$D(P_Y \| Q_Y) \leq D(P_{Y|X} \| Q_{Y|X} | P_X)$$

where $P_Y(y) = \sum_{x \in \mathcal{X}} P_X(x) P_{Y|X}(y|x)$ and $Q_Y(y) = \sum_{x \in \mathcal{X}} P_X(x) Q_{Y|X}(y|x)$.

Solution 2. *(a)*

$$\begin{aligned}
D(P_{XY} \| Q_{XY}) &= \sum_{x,y} P_{XY}(x,y) \log \frac{P_{XY}(x,y)}{Q_{XY}(x,y)} \\
&= \sum_{x,y} P_X(x) P_{Y|X}(y|x) \log \frac{P_X(x) P_{Y|X}(y|x)}{Q_X(x) Q_{Y|X}(y|x)} \\
&= \sum_{x,y} P_X(x) P_{Y|X}(y|x) \log \frac{P_X(x)}{Q_X(x)} + \sum_{x,y} P_X(x) P_{Y|X}(y|x) \log \frac{P_{Y|X}(y|x)}{Q_{Y|X}(y|x)} \\
&= D(P_X \| Q_X) + \sum_x P_X(x) D(P_{Y|X}(\cdot|x) \| Q_{Y|X}(\cdot|x)) = D(P_X \| Q_X) + D(P_{Y|X} \| Q_{Y|X} | P_X).
\end{aligned}$$

(b)

$$D(P_{XY} \| Q_{XY}) = D(P_X \| P_X) + D(P_{Y|X} \| Q_{Y|X} | P_X) = D(P_{Y|X} \| Q_{Y|X} | P_X).$$

(c) Define the kernel

$$W(\tilde{y}|x,y) = \begin{cases} 1, & \text{if } \tilde{y} = y, \\ 0, & \text{otherwise.} \end{cases}$$

Then we have $P_{\tilde{Y}}(\tilde{y}) = \sum_{x,y} P_{XY}(x,y) W(\tilde{y}|x,y) = P_Y(\tilde{y})$ and $Q_{\tilde{Y}}(\tilde{y}) = \sum_{x,y} Q_{XY}(x,y) W(\tilde{y}|x,y) = Q_Y(\tilde{y})$. Hence, we have

$$D(P_{Y|X} \| Q_{Y|X} | P_X) = D(P_{XY} \| Q_{XY}) \geq D(P_{\tilde{Y}} \| Q_{\tilde{Y}}) = D(P_Y \| Q_Y).$$

where the equality follows from part (b) and the inequality follows from DPI.

2 The Homework

The next three problems are the core of this homework. You work out solutions and turn them in. Problem 3 is suitable for Sept 17. Problems 4 and 5 are suitable for Oct 1.

Problem 3: Entropy and Geometry

Suppose X, Y and Z are random variables.

(a) Show that $H(X) + H(Y) + H(Z) \geq \frac{1}{2}\big[H(X,Y) + H(Y,Z) + H(Z,X)\big]$.

(b) Show that $H(X,Y) + H(Y,Z) \geq H(X,Y,Z) + H(Y)$.

(c) Show that

$$2\big[H(X,Y) + H(Y,Z) + H(Z,X)\big] \geq 3H(X,Y,Z) + H(X) + H(Y) + H(Z).$$