Introduction to Information Theory, Fall 2019

Practice problem set #6

You do **not** have to hand in these exercises, they are for your practice only.

1. **Markov chains and data processing** Suppose we are given three (correlated) random variable X, Y and Z. Then we can always write

$$P(x, y, z) = P(x)P(y|x)P(z|x, y).$$

If we can actually write

$$P(x, y, z) = P(x)P(y|x)P(z|y)$$

then we say that $X \to Y \to Z$ forms a *Markov chain*, which means essentially that Y depends on X, and Z depends on Y but not on X.

(a) Show that $X \to Y \to Z$ is a Markov chain if and only if $P(x, z|y) = P(x|y)P(z|y)$. Argue that if $X \to Y \to Z$ is a Markov chain, then $Z \to Y \to X$ is also a Markov chain.

(b) Show that

$$H(Z|X, Y) \leqslant H(Z|Y)$$

with equality if and only if $X \to Y \to Z$ is a Markov chain.

(c) Prove the *Data Processing Inequality*: if $X \to Y \to Z$ is a Markov chain then

$$I(X : Y) \geqslant I(X : Z).$$

Explain what this tells you about data processing.