UC Berkeley
Department of Electrical Engineering and Computer Sciences

EECS 126: Probability and Random Processes

Discussion 6

Spring 2021

1. **Mutual Information and Noisy Typewriter**

The **mutual information** of X and Y is defined as

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

Here, $H(X \mid Y)$ denotes the **conditional entropy** of X given Y, which is defined as:

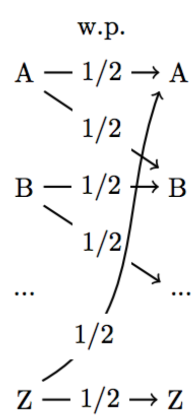
$$\begin{aligned} H(X \mid Y) &= \sum_{y \in \mathcal{Y}} p_Y(y) H(X \mid Y = y) \\ &= \sum_{y \in \mathcal{Y}} p_Y(y) \sum_{x \in \mathcal{X}} p_{X|Y}(x \mid y) \log_2 \frac{1}{p_{X|Y}(x \mid y)} \\ &= - \sum_{y \in \mathcal{Y}, x \in \mathcal{X}} p_{X,Y}(x,y) \log_2 \frac{p_{X,Y}(x,y)}{p_Y(y)} \end{aligned}$$

The interpretation of conditional entropy is the average amount of uncertainty remaining in the random variable X after observing Y. The interpretation of mutual information is therefore the amount of information about X gained by observing Y.

(a) Show that $H(X,Y) = H(Y) + H(X \mid Y) = H(X) + H(Y \mid X)$. This is often called the **Chain Rule**. Interpret this rule.

(b) Show that $I(X;Y) = H(X) + H(Y) - H(X,Y)$. Note that this shows that $I(X;Y) = I(Y;X)$, i.e., mutual information is symmetric.

(c) Consider the noisy typewriter.



Each symbol gets sent to one of the adjacent symbols with probability $1/2$. Let X be the input to the noisy typewriter, and let Y be the output (X is a random variable that takes values in the English alphabet). What is the distribution of X that maximizes $I(X;Y)$?

Note

It turns out that $I(X;Y) \geq 0$ with equality if and only if X and Y are independent. The mutual information is an important quantity for channel coding.

2. **Entropy of a Sum**

(a) Let X_1, X_2 be i.i.d. Bernoulli(1/2) (fair coin flips). Calculate $H(X_1 + X_2)$ and show that $H(X_1 + X_2) \geq H(X_1)$.

(b) It turns out that in general adding independent random variables increases entropy. To prove this, first let's prove that conditioning decreases entropy. Let X and Y be two (not necessarily i.i.d) random variables, prove that $H(X|Y) \leq H(X)$

Hint: Use the the fact that $I(X;Y) \geq 0$ for any r.v.s X and Y.

(c) Now suppose X_1, X_2 are two independent (but not necessarily identically distributed) random variables. Prove that $H(X_1 + X_2) \geq H(X_1)$.