UC Berkeley
Department of Electrical Engineering and Computer Sciences

EECS 126: Probability and Random Processes

Discussion 07
Spring 2023

1. **Entropy of a Sum**

   Let X_1, X_2 be i.i.d. Bernoulli($\frac{1}{2}$). Calculate $H(X_1+X_2)$ and show that $H(X_1+X_2) \geq H(X_1)$. Does this make intuitive sense?

2. **Mutual Information and Channel Coding**

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

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

where $H(X \mid Y)$ is the *conditional entropy* of X given Y,

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

Conditional entropy can be interpreted as the average amount of uncertainty remaining in the random variable X after observing Y. Then, mutual information is the amount of information about X gained by observing Y.

Now, the channel coding theorem says that the capacity of a channel with input X and output Y is the maximal possible amount of mutual information between them:

$$C = \max_{p_X} I(X;Y) = \max_{p_X} H(X) - H(X \mid Y).$$

a. Let X be the roll of a fair die and $Y = \mathbb{1}_{X \geq 5}$. What is $H(X \mid Y)$?

b. Suppose the channel is a noiseless binary channel, i.e. $X \in \{0,1\}$ and $Y = X$. Use the theorem above to find its capacity C.

c. Consider a binary erasure channel with probability of erasure p. Use the theorem above to find C. *Hint*: To find the optimal p_X, it is helpful to let $p_X(1) = \mathbb{P}(X = 1) = \alpha$.

3. **Binary Coding**

A system has 6 possible configurations $[1, 2, 3, 4, 5, 6]$. It takes on each configuration i with probability p_i, where

$$[p_1, p_2, p_3, p_4, p_5, p_6] = \left[\frac{1}{2}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16}\right].$$

We want to *encode* the configurations, i.e. assign a binary string *codeword* γ_i to each configuration i, such that no codeword is a prefix of another codeword. Let ℓ_i be the length of the codeword γ_i, and let $L = \sum_{i=1}^{6} p_i \ell_i$ be the expected codeword length. Come up with a code for which L equals the entropy of the distribution above. (This code will in fact *minimize* L.)

Hint: Consider organizing your codewords in a *trie*, a binary tree in which each codeword corresponds to the path from the root to a leaf. For example, the codeword 011 would be represented as the leaf `root.left.right.right`.