UC Berkeley
Department of Electrical Engineering and Computer Sciences

EECS 126: Probability and Random Processes

Discussion 10
Fall 2021

1. **Entropy Warmup** Entropy seems like a very weird concept. We'll walk through a few examples together to build up your intuition. Let's say that we have a random variable X that can take on values from $\{\text{lecture}, \text{midterm}, \text{pop quiz}\}$. (Don't worry, we don't actually have pop quizzes in this class). Each day you go to class, you observe a random value of X, which is determined according to the distribution P_X. If $P(\text{lecture}) = 0.85$, $P(\text{midterm}) = 0.1$, and $P(\text{pop quiz}) = 0.05$, we'd mainly see lectures, occasionally have a midterm, and have a pop quiz very rarely. We can describe how "interesting" it is to see a particular $X = x$ with the notion of the "surprise," which is a function $S(x) = \log_2 \frac{1}{P_X(x)}$. This function is large for low-probability events and small for high-probability events.

   (a) For the probabilities above, calculate $S(\text{lecture})$, $S(\text{midterm})$, and $S(\text{pop quiz})$.

   (b) If $P(\text{lecture}) = \frac{1}{3}$, $P(\text{midterm}) = \frac{1}{3}$, and $P(\text{pop quiz}) = \frac{1}{3}$, calculate the surprises again. Given that $\log_2 \frac{1}{0.85} = 0.234$, $\log_2 \frac{1}{1/3} = 1.58$, $\log_2 \frac{1}{0.1} = 3.32$, and $\log_2 \frac{1}{0.05} = 4.32$, do the relative magnitudes of the values in (a) and (b) make sense intuitively?

   (c) The entropy is the *expected surprise*. Formally,

   $$H(X) = \sum_x P_X(x) S(x) = \sum_x P_X(x) \log_2 \frac{1}{P_X(x)}$$

   We will follow the convention that, if for a particular x, $P_X(x) = 0$, then $P_X(x) \log_2 \frac{1}{P_X(x)} = 0$. Calculate the entropy for the original probability values $(0.85, 0.1, 0.05)$, the entropy of the uniform distribution $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, and the entropy of the deterministic RV with distribution $(1, 0, 0)$.

   (d) Do the entropies in part (c) make sense to you?

2. **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:

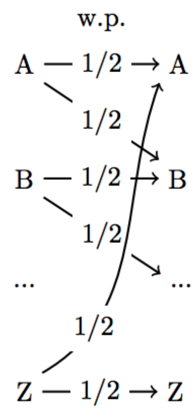
   $$\begin{aligned} H(X \mid Y) &= \sum_{y \in \mathcal{Y}} p_Y(y) H(X \mid Y = y) \\ &= \sum_{y \in \mathcal{Y}} p_Y(y) \sum_{x \in \mathcal{X}} p_{X|Y}(x \mid y) \log_2 \frac{1}{p_{X|Y}(x \mid y)} \end{aligned}$$

   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.

3. **Huffman Questions**

Consider a set of n objects. Let $X_i = 1$ or 0 accordingly as the i-th object is good or defective. Let $X_1, X_2, \ldots, X_n$ be independent with $P(X_i = 1) = p_i$; and $p_1 > p_2 > \cdots > p_n > 1/2$. We are asked to determine the set of all defective objects. Any yes-no question you can think of is admissible.

(a) Propose an algorithm based on Huffman coding in order to identify all defective objects.

(b) Suppose the worst case scenario happens and we have to ask the maximimum number of questions. What (in words) is the last question we should ask? And what two sets are we distinguishing with this question?