UC Berkeley
Department of Electrical Engineering and Computer Sciences

EECS 126: Probability and Random Processes

Discussion 6
Spring 2020

---

1. **Gambling Game**

   Let's play a game. You stake a positive initial amount of money w_0. You toss a fair coin. If it comes up heads you earn an amount equal to three times your stake, so you quadruple your wealth. If it comes up tails you lose everything. There is one requirement though, you are not allowed to quit and have to keep playing, by staking all your available wealth, over and over again.

   Let W_n be a random variable which is equal to your wealth after n plays.

   (a) Find $\mathbb{E}[W_n]$ and show that $\lim_{n\to\infty} \mathbb{E}[W_n] = \infty$.

   (b) Since $\lim_{n\to\infty} \mathbb{E}[W_n] = \infty$, this game sounds like a good deal! But wait a moment!! Where does the sequence of random variables $\{W_n\}_{n\geq 0}$ converges in probability to?

2. **Entropy of a Sum**

   Let X_1, X_2 be i.i.d. Bernoulli(1/2) (fair coin flips). Calculate $H(X_1 + X_2)$ and show that $H(X_1 + X_2) \geq H(X_1)$. In fact it is generally true that adding independent random variables increases entropy.

   Note: It is known that the Gaussian distribution maximizes entropy given a constraint on the variance. Therefore, one intuitive interpretation of the CLT is that convolving independent random variables tends to increase uncertainty until the sum approaches the distribution which "maximizes uncertainty", the Gaussian distribution. Proving the CLT along these lines is far from easy, however.

3. 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:

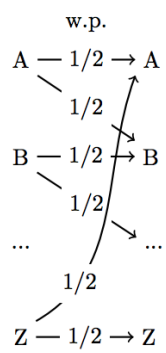
$$\begin{aligned} H(X \mid Y) &= \sum_{y \in \mathcal{Y}} p_Y(y) H(X \mid Y = y) \\ &= \sum_{y \in \mathcal{Y}} p_Y(y) \sum_{x \in \mathcal{X}} p_{X|Y}(x \mid y) \log_2 \frac{1}{p_{X|Y}(x \mid y)} \end{aligned}$$

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.