Information Theory
Practice Set 1

Lakshmi Prasad Natarajan

Solutions are not to be returned

Reading Exercise

Example 2.2.1 from Cover and Thomas.

Practice Set

1. *Joint entropy of independent random variables.*
   Recall that the random variables $X_1, \ldots, X_n$ are independent if $p(x_1, \ldots, x_n) = \prod_{i=1}^n p(x_i)$.

   (a) Argue that if $X_1, \ldots, X_n$ are independent then

   $$H(X_n|X_1, \ldots, X_{n-1}) = H(X_n).$$

   (b) Show that if $X_1, \ldots, X_n$ are independent then

   $$H(X_1, \ldots, X_n) = H(X_1) + \cdots + H(X_n). \quad (1)$$

2. We showed in the class that the entropy of any random variable is greater than on equal to zero. What about the conditional entropy $H(Y|X)$? Can you argue that $H(Y|X) \geq 0$ for any jointly distributed random variables (X, Y) ?

3. Suppose a random variable Y is a function of another random variable X. That is, the value of Y is completely (deterministically) based on the value of X.

   (a) For each choice of $x \in \mathcal{X}$, what is the value of $H(Y|X = x)$?

   (b) Is the value of $H(Y|X)$ equal to zero ?

   (c) Can you now argue that $H(X) = H(Y) + H(X|Y)$?
   `Hint`: Use chain rule on $H(X, Y)$ in two different ways.

   (d) Now use the fact that conditional entropy $H(X|Y) \geq 0$ to conclude that $H(X) \geq H(Y)$.

   `Remark`: This proves that the entropy of a function of a random variable is less than or equal to the entropy of the random variable itself. That is, if g is any function then $H(g(X)) \leq H(X)$. In other words, performing deterministic computations on a random variable can not increase the entropy.

4. Can you design examples of jointly distributed random variables X, Y (can you create a table of joint probability mass functions as done in Lecture 3) such that

   (a) $H(Y|X) = 0$.

   (b) $H(Y|X) = H(Y)$.

   (c) $H(Y|X = x) > H(Y)$ for at least one choice of $x \in \mathcal{X}$.

5. *Entropy of Geometric Distribution.*
   From Example 1.2 of Information Theory: From Coding to Learning *by Polyanskiy and Wu.*
   Let $\mathcal{X} = \{1, 2, 3, \ldots\}$ and α be such that $0 < \alpha < 1$. Let X be a random variable with distribution

   $$P[X = x] \quad = \quad \alpha\,(1 - \alpha)^{x-1}, x \in \mathcal{X}.$$

   Compute the entropy of X. You can express this using the binary entropy function
   $h_2(p) = -p \log p - (1 - p) \log(1 - p)$.

6. *Condensation test for convergence.*

   We will use a result from real analysis, called *condensation test*, to verify if some series converge. The context of the condensation test is as follows:

   Let $\{a_n : n \geq 1\}$ be a non-increasing sequence of non-negative real numbers, i.e., $a_n \geq 0$ and $a_n \geq a_{n+1}$ for all n. Let the sequence $\{b_n\}$ be defined as follows

   $$b_n = 2^n a_{2^n},$$

   that is, the n^{th} element of sequence b is equal to the 2^nth element of sequence of a multiplied by 2^n.

   For instance, if $a_n = \frac{1}{\log_2 n}$, $n \geq 2$, then $b_n = \frac{2^n}{n}$, $n \geq 2$.

   Theorem (Condensation Test). $\sum_{n=1}^{\infty} a_n$ converges if and only if $\sum_{n=1}^{\infty} b_n$ converges.

   That is, $\sum_{n=1}^{\infty} a_n < \infty$ if and only if $\sum_{n=1}^{\infty} b_n < \infty$.

   Use the condensation test to check if the following series $\sum_n a_n$ converge.

   (a) $a_n = \dfrac{1}{n}$

   (b) $a_n = \dfrac{1}{n^2}$

   (c) $a_n = \dfrac{1}{n \log n}$, $n \geq 2$

   (d) $a_n = \dfrac{1}{n \left(\log n\right)^2}$, $n \geq 2$

7. *Infinite Entropy.*
   From Example 1.3 of Information Theory: From Coding to Learning *by Polyanskiy and Wu.*

   (a) Show that there exists an $\alpha \in \mathbb{R}$ such that $\displaystyle\sum_{n=2}^{\infty} \frac{\alpha}{n \left(\log n\right)^2} = 1$.

   (b) With α as above, consider a random variable X with $\mathcal{X} = \{2, 3, 4, \dots\}$ and

   $$P[X = n] = \frac{\alpha}{n \left(\log n\right)^2}.$$

   Argue that $H(X)$ is not finite.