Information Theory — Assignment I

1. Let X, Y and random variables from $\Omega \longrightarrow (0, 1)$ such that for each $\omega \in \Omega$, $X(\omega) \geq Y(\omega)$. Show that $E(X) \geq E(Y)$. Can we conclude that $var(X) \geq var(Y)$?

2. Estimate the mean and variance of the negative binomial distribution.

3. Consider the following experiment. A fair coin is tossed first. If the outcome is a head, then an unfair dice with probabilities of $1, 2, 3, 4, 5$ and 6 respectively $\frac{1}{2}\frac{1}{4}\frac{1}{8}\frac{1}{16}\frac{1}{32}\frac{1}{32}$ is thrown. If the outcome of the coin toss is a tail, a fair dice is thrown. Let X denote a random variable indicating the value of the dice and Y denote a random variable which is 1 if the coin gave heads and 0 when the coin gave tails.

    1. Write down the sample space and probabilities of each point in the sample space for the experiment. Are the distributions of X and Y independent?
    2. Find $E(X|Y = 0)$ and $E(X|Y = 1)$. Hence find $E(X)$.
    3. What is $Pr(Y = 1|X \in \{2, 4, 6\})$?
    4. Find $H(X)$, $H(X|Y)$, $H(Y|X)$ and $I(X, Y)$.

4. The binary symmetric channel comprises of a transmitter emitting a signal $X \in 0, 1$ with $Pr(X = 1) = t$ and $Pr(X = 0) = 1 - t$. The signal passes through a communication channel that adds a **noise** $Z \in 0, 1$ as follows: $Pr(Z = 1) = \epsilon$, $Pr(Z = 0) = 1 - \epsilon$. At the receiving end of the communication channel $Y = X \oplus Z$ is received (XOR of the channel output and the source output). Use the abbreviation $h(x)$ for the expression $x \log \frac{1}{x} + (1 - x) \log \frac{1}{(1-x)}$. Assuming $t = \frac{1}{2}$, find the entropy $H(Y)$ of the receiver and the mutual information $I(X, Y)$ between the source and the receiver.

5. Consider a **complete graph** with n vertices. A particle starts at vertex number 1 and randomly walks to one of its neighbours. In the next step, it picks another neighbour at random and moves to that vertex and so on. In this problem, we estimate the expected number of steps necessary to visit every vertex at least once.

    Let random variables X_2,X_3... X_n be defined in the following way. X_i denotes the number of steps required for the count of visited vertices to become i **after** the $i - 1^{th}$ vertex was visited. For instance, X_2 denotes the number of steps the particle needs to visit two vertices. X_2 is always 1 because the start vertex is already visited before the first jump and the first jump itself goes to a new vertex making visit count 2. However X_3 need not be two because there is a probability $(\frac{1}{n})$ that the second jump is back to the start vertex.

    1. Show that each X_i has a geometric distribution. Find $E(X_i)$ for each i.
    2. Define the random variable $X = X_2 + X_3 + ... + X_n$. Estimate $E(X)$. What is this value indicating?

6. Consider a Hash table with m slots with simple uniform hashing. (Recall your data structures lessons!). Suppose n items are hashed with **chaining** into the hash table. What is the expected number of insertions needed to ensure that every slot has at least one item hashed into it? (Do you see the relationship with the previous question?).

7. Consider a graph with n vertices and m edges. Suppose each vertex in the graph is independently coloured with either RED or BLUE colour. For each edge $e \in E(G)$, define the random variable X_e having value 1 if the two end points of e have different colours whereas X_e takes value 0 if the end points have the same colour.

    1. Find $E(X_e)$.
    2. Define the random variable $X = \Sigma_{e \in E} X_e$. This gives the total number of edges with differently coloured end points.
    3. Show that $E(X) = \frac{m}{2}$.

4. Conclude from the above that a randomly chosen **cut** in a graph has expected size $\frac{m}{2}$ edges. Hence conclude that every graph must have at least one cut of size $\frac{m}{2}$

8. The following is a randomized algorithm for finding a large independent set in a graph. The questions below develop the analysis of the algorithm. Recall that an independent set in a graph $G(V, E)$ is a collection of vertices S such that there is no edge between any of the vertices in S. We construct a random independent set and prove that it must have a certain size on the average. □

   Let G be a graph with n vertices and m edges. First take a random permutation of $\{1, 2, .., n\}$ and assign each vertex with a unique number. Select to the set S those vertices whose number is **larger than all its neighbours**.

   1. Show that the set S selected this way is indeed an independent set.
   2. For each vertex v, let X_v be a random variable taking value 1 if v is selected in S. Show that $Pr(X_v = 1) = \frac{1}{(d_v+1)}$ where d_v is the degree of v.
   3. Let $X = \Sigma_{v \in V} X_v$. Clearly X_v denote the total number of vertices selected in S. Hence find the expected number of vertices in S.
   4. Let d denote the **average** degree of G. Show that $E(X) \geq \frac{n}{d+1}$. (From the expression in the previous question use the fact that $f(x) = \frac{1}{(1+x)}$ is convex when $x > -1$).