-First Colloquium in Information Theory (TOR III)-
05. May 2022

1. (20p) A gambler has a fair coin and a two-headed coin in his pocket. He selects one of the coins at random.
   (a) When he flips it, it shows heads. What is the probability that it is the fair coin?
   (b) Suppose that he flips the same coin a second time and, again, it shows heads. Now what is the probability that it is the fair coin?
   (c) Suppose that he flips the same coin n times and it is heads every time. What is the probability that it is the fair coin?
   (d) How much information, in bits, is gained by learning that the coin was indeed the fair coin in (a), how much in (b), and how much in (c)?
   (e) Suppose that he flips the same coin an $(n+1)$-th time and it shows tails. Now what is the probability that it is the fair coin? What is the information content of this event?

2. (20p) Suppose that X is a random variable whose entropy $H(X)$ is 8 bits. Suppose that $Y(X)$ is a random variable which is a deterministic function of X that takes on a different value for each value of X.
   (a) What then is $H(Y)$, the entropy of Y?
   (b) What is $H(Y|X)$, the conditional entropy of Y given X?
   (c) What is $H(X|Y)$, the conditional entropy of X given Y?
   (d) What is $H(X,Y)$, the joint entropy of X and Y?
   (e) Suppose now that the deterministic function $Y(X)$ is not invertible; in other words, different values of X may correspond to the same value of $Y(X)$. In that case, what could you say about $H(Y)$?
   (f) In that case, what can you say about $H(X|Y)$?

3. (20p)
   (a) Let the join probability mass function $p(x,y)$ for two random variables X and Y be given as:

	X=0	X=1
Y=0	1/4	1/3
Y=1	1/3	1/12

   Find: $H(X)$, $H(Y)$, $H(X|Y)$, $H(Y|X)$, $H(X,Y)$ and $I(X;Y)$.

   (b) Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed discrete random variables. Let
   $$C_n(t) = \{(x_1, \ldots, x_n) : P(X_1 = x_1, \ldots, X_n = x_n) \geq 2^{-nt}\}.$$
      i. Show that $|C_n(t)| \leq 2^{nt}$.
      ii. For what values of t does $P(C_n(t)) \to 1$ when $n \to +\infty$?

4. (20p)
   (a) Using Lempel–Ziv–Welch algorithm compress the following text: ”bababcccbaaaac” with the binary initial dictionary:

a	00
b	01
c	10

(b) The source of information A generates the symbols $A = \{a, b, c\}$ with the probabilities $P_A = \{0.5, 0.3, 0.2\}$. Using arithmetic coding encode the message "cabb".

5. (20 p) In several proofs we have used the so-called relative distance (Kullback-Lebler distance) between two probability distributions $p(x)$ (true one) and $q(x)$ (estimated one). This distance was defined as $D(p||q) = \sum_{x \in \mathcal{X}} p(x) \log_2 \frac{p(x)}{q(x)}$.

   (a) Using Jensen's inequality which claims that $E[f(X)] \geq f(E[X])$ (valid for convex functions f) prove that $D(p||q) \geq 0$. Hint: Consider $-\log_2$ as function f.

   (b) Show that the mutual information $I(X;Y)$ can be expressed as $I(X;Y) = H(X)+H(Y)-H(X,Y)$ and therefore by definition of the (joint) entropy can be expressed as

$$I(X;Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{X,Y}(x,y) \log_2 \frac{p_{X,Y}(x,y)}{p_X(x) p_Y(y)}.$$

   (c) Use the above parts to show that $I(X;Y) \geq 0$.