GUC German University in Cairo

IET Faculty
Dr. Wassim Alexan

COMM1003 Information Theory
Problem Set 1 – Solutions

1.

Prove that $H(X_0 \mid X_n)$ is non–decreasing with n for any Markov chain.

For a Markov chain, by the data processing theorem, we have

$$I(X_0; X_{n-1}) \geq I(X_0; X_n)$$

Therefore,

$$H(X_0) - H(X_0 \mid X_{n-1}) \geq H(X_0) - H(X_0 \mid X_n)$$

or $H(X_0 \mid X_n)$ increases with n.

2.

Let the random variable X have 3 possible outcomes $\{a, b, c\}$. Consider 2 distributions on this random variable

Symbol	$p(x)$	$q(x)$
a	1/2	1/3
b	1/4	1/3
c	1/4	1/3

Calculate $H(p)$, $H(q)$, $D(p \,||\, q)$ and $D(q \,||\, p)$. Verify that in this case $D(p \,||\, q) \neq D(q \,||\, p)$.

$$H(p) = \frac{1}{2}\log 2 + \frac{1}{4}\log 4 + \frac{1}{4}\log 4 = 1.5 \text{ bits}$$

$$H(q) = \frac{1}{3}\log 3 + \frac{1}{3}\log 3 + \frac{1}{3}\log 3 = \log 3 = 1.58496 \text{ bits}$$

$$D(p \,||\, q) = \frac{1}{2}\log\frac{3}{2} + \frac{1}{4}\log\frac{3}{4} + \frac{1}{4}\log\frac{3}{4} = \log(3) - 1.5 = 1.58496 - 1.5 = 0.08496$$

$$D(p \,||\, q) = \frac{1}{2}\log\frac{2}{3} + \frac{1}{3}\log\frac{4}{3} + \frac{1}{3}\log\frac{4}{3} = \frac{5}{3} - \log(3) = 1.66666 - 1.58496 = 0.08170$$

3.

Though, as the in the previous problem, $D(p\,||\,q) \neq D(q\,||\,p)$ in general, there could be distributions for which equality holds. Give an example of two distributions p and q on a binary alphabet such that $D(p\,||\,q) = D(q\,||\,p)$ (other than the trivial case $p = q$).

A simple case for $D((p,\ 1-p)\,||\,(q,\ 1-q)) = D((q,\ 1-q)\,||\,(p,\ 1-p))$, i.e. for

$$p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q} = q \log \frac{q}{p} + (1-q) \log \frac{1-q}{1-p}$$

is when $q = 1 - p$.

4.

Let X, Y and Z be three random variables with a joint probability mass function $p(x, y, z)$. The relative entropy between the joint distribution and the product of the marginals is

$$D(p(x, y, z)\,||\,p(x)\,p(y)\,p(z)) = E\left[\log \frac{p(x, y, z)}{p(x)\,p(y)\,p(z)}\right]$$

Expand this in term of entropies. When is this quantity zero?

$$\begin{aligned} D(p(x, y, z)\,||\,p(x)\,p(y)\,p(z)) &= E\left[\log \frac{p(x, y, z)}{p(x)\,p(y)\,p(z)}\right] \\ &= E[\log p(x, y, z)] - E[\log p(x)] - E[\log p(y)] - E[\log p(z)] \\ &= -H(X, Y, Z) + H(X) + H(Y) + H(Z) \end{aligned}$$

We have $D(p(x, y, z)\,||\,p(x)\,p(y)\,p(z)) = 0$ if and only if $p(x, y, z) = p(x)\,p(y)\,p(z)$ for all (x, y, z), i.e. if X, Y and Z are independent.

5.

Let X and Y be two independent integer–valued random variables. Let X be uniformly distributed over $\{1, 2, \ldots, 8\}$ and let $\Pr\{Y = k\} = 2^{-k}$, $k = 1, 2, 3, \ldots$

(a) Find $H(x)$
(b) Find $H(Y)$
(c) Find $H(X + Y, X - Y)$.

(a) For a uniform distribution, $H(X) = \log m = \log_2 8 = 3$

(b) For a geometric distribution, $H(Y) = \sum_k k\, 2^{-k} = 2$ (see lecture 2)

(c) Since $(X, Y) \to (X + Y, X - Y)$ is a one to one transformation,
$H(X + Y, X - Y) = H(X, Y) = H(x) + H(Y) = 3 + 2 = 5$

6.

Which of the following inequalities are generally ≥, =, ≤?
Label each with ≥, =, ≤.

(a) $H(5X)$ vs. $H(X)$
(b) $H(X_0 \mid X_{-1})$ vs. $H(X_0 \mid X_{-1}, X_1)$
(c) $H(X_1, X_2, \ldots, X_n)$ vs. $H(c(X_1, X_2, \ldots, X_n))$, where $c(X_1, X_2, \ldots, X_n)$ is the Huffman codeword assigned to $(x_1, x_2, \ldots, x_n)$
(d) $H(X, Y)/(H(X)+H(Y))$ vs. 1

(a) $X \to 5X$ is a one to one mapping, and hence $H(X) = H(5X)$

(b) Since conditioning reduces entropy, $H(X_0 \mid X_{-1}) \geq H(X_0 \mid X_{-1}, X_1)$

(c) Source coding removes redundancy and thus a smaller number of bits would be needed to describe the data, thus $H(X_1, X_2, \ldots, X_n) \geq H(c(X_1, X_2, \ldots, X_n))$

(d) $H(X, Y) \leq H(X)+H(Y)$, so $H(X, Y)/(H(X)+H(Y)) \leq 1$

7.

(a) Consider a fair coin flip. What is the mutual information between the top side and the bottom side of the coin?

(b) A 6–sided fair die is rolled. What is the mutual information between the top side and the front side (the side most facing you)?

To prove (a) observe that

$$I(T; B) = H(B) - H(B \mid T) = \log_2 2 = 1$$

To prove (b) note that having observed a side of the cube facing us, F, there are 4 possibilities for the top T, which are equally probable. Thus,

$$\begin{aligned} I(T; F) &= H(T) - H(T \mid F) \\ &= \log_2 6 - \log_2 4 \\ &= \log_2 3 - 1 \end{aligned}$$

Since T has a uniform distribution on $\{1, 2, \ldots, 6\}$.

8.

A computer system can operate in two different modes. Every hour, it remains in the same mode or switches to a different mode according to the transition probability matrix

$$P = \begin{pmatrix} 0.4 & 0.6 \\ 0.6 & 0.4 \end{pmatrix}$$

(a) Compute the 2–step transition probability matrix
(b) If the system is in Mode I at 5:30 PM, what is the probability that it will be in Mode I at 8:30 PM on the same day?

(a)

$$P^{(2)} = P \cdot P = \begin{pmatrix} 0.52 & 0.48 \\ 0.48 & 0.52 \end{pmatrix}$$

(b) There are 3 transitions between 5:30 PM and 8:30 PM, thus, we need to compute $p_{11}^{(3)}$. The 3–step transition probability matrix is

$$P^{(3)} = P^{(2)} \cdot P = \begin{pmatrix} 0.496 & \dots \\ \dots & \dots \end{pmatrix}$$