Fundamentals of Information Theory

Homework 2

Name: Zhou Shouchen
Student ID: 2021533042

Due 23:59 (CST), Oct. 27, 2024

Problem 1

2.16 Bottleneck. Suppose that a (nonstationary) Markov chain starts in one of n states, necks down to $k < n$ states, and then fans back to $m > k$ states. Thus, $X_1 \to X_2 \to X_3$, that is $p(x_1, x_2, x_3) = p(x_1)\, p(x_2 \mid x_1)\, p(x_3 \mid x_2)$, for all $x_1 \in \{1, 2, \ldots, n\}, x_2 \in \{1, 2, \ldots, k\}, x_3 \in \{1, 2, \ldots, m\}$.

(a) Show that the dependence of X_1 and X_3 is limited by the bottleneck by proving that $I(X_1; X_3) \leq \log k$.

(b) Evaluate $I(X_1; X_3)$ for $k = 1$, and conclude that no dependence can survive such a bottleneck.

Solution

(a) Since $X_1 \to X_2 \to X_3$ form the Markov chain, so we can get that:

$$\begin{aligned} I(X_1; X_3) &\leq I(X_1; X_2) && \text{(Data Processing Inequality)} \\ &\leq H(X_2) && \text{(property of mutual information)} \\ &\leq \log k && (|X_2| = k) \quad \square \end{aligned}$$

(b) Since $k = 1$, from (a), we can get that

$$I(X_1; X_3) \leq \log 1 = 0$$

And from the property of mutual information, we can get that

$$I(X_1; X_3) \geq 0$$

So we can conclude that $I(X_1; X_3) = 0$.
Which means that no dependence can survive such a bottleneck.

Problem 2

2.25 Venn diagrams. There isn't really a notion of mutual information common to three random variables. Here is one attempt at a definition: Using Venn diagrams, we can see that the mutual information common to three random variables X, Y, and Z can be defined by

$$I(X;Y;Z) = I(X;Y) - I(X;Y \mid Z)$$

This quantity is symmetric in X, Y, and Z, despite the preceding asymmetric definition. Unfortunately, $I(X;Y;Z)$ is not necessarily nonnegative. Find X, Y, and Z such that $I(X;Y;Z) < 0$, and prove the following two identities:

(a) $I(X;Y;Z) = H(X,Y,Z) - H(X) - H(Y) - H(Z) + I(X;Y) + I(Y;Z) + I(Z;X)$

(b) $I(X;Y;Z) = H(X,Y,Z) - H(X,Y) - H(Y,Z) - H(Z,X) + H(X) + H(Y) + H(Z)$.

The first identity can be understood using the Venn diagram analogy for entropy and mutual information. The second identity follows easily from the first.

Solution

<1> Let $X, Y \overset{i.i.d.}{\sim} Bern\left(\frac{1}{2}\right), Z = X + Y$.

Since $X \perp Y$, so

$$I(X;Y) = 0$$

And

$$I(X;Y|Z) = H(X|Z) - H(X|Y,Z) = H(X|Z) > 0$$

So

$$I(X;Y;Z) = I(X;Y) - I(X;Y|Z) = -I(X;Y|Z) < 0$$

So above all, we have given an example that $I(X;Y;Z) < 0$.

<2> (a)

$$\begin{aligned}
I(X;Y;Z) &= I(X;Y) - I(X;Y|Z) \\
&= I(X;Y) - (I(X;Y,Z) - I(X;Z)) \quad \text{(Chain Rule of Mutual Information)} \\
&= I(X;Y) - (H(X) + H(Y,Z) - H(X,Y,Z)) + I(X;Z) \\
&= H(X,Y,Z) - H(X) - H(Y,Z) + I(X;Y) + I(Z;X) \\
&= H(X,Y,Z) - H(X) - (H(Y) + H(Z) - I(Y;Z)) + I(X;Y) + I(Z;X) \\
&= H(X,Y,Z) - H(X) - H(Y) - H(Z) + I(X;Y) + I(Y;Z) + I(Z;X) \quad \square
\end{aligned}$$

<2> (b)

$$\begin{aligned}
I(X;Y;Z) &= I(X;Y) - H(X) - H(Y,Z) + H(X,Y,Z) + I(Z;X) \quad \text{(from (a)'s third line)} \\
&= (H(X) + H(Y) - H(X,Y)) - H(X) - H(Y,Z) + H(X,Y,Z) + (H(X) + H(Z) - H(Z,X)) \\
&= H(X,Y,Z) - H(X,Y) - H(Y,Z) - H(Z,X) + H(X) + H(Y) + H(Z) \quad \square
\end{aligned}$$

Problem 3

2.46 Axiomatic definition of entropy (Difficult). If we assume certain axioms for our measure of information, we will be forced to use a logarithmic measure such as entropy. Shannon used this to justify his initial definition of entropy. In this book we rely more on the other properties of entropy rather than its axiomatic derivation to justify its use. The following problem is considerably more difficult than the other problems in this section.

If a sequence of symmetric functions $H_m(p_1, p_2, \ldots, p_m)$ satisfies the following properties:

- Normalization: $H_2\left(\frac{1}{2}, \frac{1}{2}\right) = 1$,
- Continuity: $H_2(p, 1-p)$ is a continuous function of p,
- Grouping: $H_m(p_1, p_2, \ldots, p_m) = H_{m-1}(p_1 + p_2, p_3, \ldots, p_m) + (p_1 + p_2) H_2\left(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2}\right)$

prove that H_m must be of the form

$$H_m(p_1, p_2, \ldots, p_m) = -\sum_{i=1}^{m} p_i \log p_i, \quad m = 2, 3, \ldots$$

There are various other axiomatic formulations which result in the same definition of entropy. See, for example, the book by Csiszár and Körner [149].

Solution

Notations:

$$f(m) \triangleq H_m\left(\frac{1}{m}, \frac{1}{m}, \ldots, \frac{1}{m}\right)$$
$$S_k \triangleq \sum_{i=1}^{k} p_i \quad k = 1, 2, \ldots, m$$

From the grouping property, we can get its extension:

$$\begin{aligned}
&H_m(p_1, p_2, p_3, \ldots, p_m) \\
&= H_{m-1}(S_2, p_3, \ldots, p_m) + S_2 H_2\left(\frac{p_1}{S_2}, \frac{p_2}{S_2}\right) \\
&= H_{m-2}(S_3, p_4, \ldots, p_m) + S_3 H_2\left(\frac{p_1+p_2}{S_3}, \frac{p_3}{S_3}\right) + S_2 H_2\left(\frac{p_1}{S_2}, \frac{p_2}{S_2}\right) \\
&= \ldots \\
&= H_{m-(k-1)}(S_k, p_{k+1}, \ldots, p_m) + \sum_{i=2}^{k} S_i H_2\left(\frac{S_{i-1}}{S_i}, \frac{p_i}{S_i}\right) \\
&= H_{m-(k-1)}(S_k, p_{k+1}, \ldots, p_m) + S_k H_k\left(\frac{p_1}{S_k}, \frac{p_2}{S_k}, \ldots, \frac{p_k}{S_k}\right)
\end{aligned}$$

The last equality(blue part) can be obtained by expanding $H_k\left(\frac{p_1}{S_k}, \frac{p_2}{S_k}, \ldots, \frac{p_k}{S_k}\right)$.

And using this, we can get that

$$\begin{aligned}
f(mn) &= H_{mn}\left(\frac{1}{mn}, \frac{1}{mn}, \ldots, \frac{1}{mn}\right) \\
&= H_{mn-(n-1)}\left(S_n, \frac{1}{mn}, \ldots, \frac{1}{mn}\right) + S_n H_n\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right) \\
&= H_{mn-2(n-1)}\left(S_n, S_n, \frac{1}{mn}, \ldots, \frac{1}{mn}\right) + 2S_n H_n\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right) \\
&= \ldots \\
&= H_{mn-m(n-1)}\left(S_n, S_n, \ldots, S_n\right) + mS_n H_n\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right) \\
&= H_m\left(\frac{1}{m}, \ldots, \frac{1}{m}\right) + m\frac{1}{m}H_n\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right) \\
&= f(m) + f(n)
\end{aligned}$$

And since we have the Continuity property, i.e. $H_2(p, 1-p)$ is a continuous function of p, from the property and provement of Cauchy function, we could get that

$$f(m) = \log_a m$$

And from the Normalization property, we have

$$f(2) = 1$$

So we can get that $a = 2$.

So above all, we have proved that

$$f(m) = H_m\left(\frac{1}{m}, \frac{1}{m}, \ldots, \frac{1}{m}\right) = \log_2 m$$

And then prove $H_2(p, 1-p) = -p\log_2 p - (1-p)\log_2(1-p)$:

1. When p is rational, suppose $p = \frac{r}{s}$, where r, s are integers, $s > 1, 0 \le r \le s$, $\gcd(r, s) = 1$. Then from the Grouping property and its extension, we have:

$$\begin{aligned}
f(s) = H_s\left(\frac{1}{s}, \ldots, \frac{1}{s}\right) &= H_s\left(\underbrace{\frac{1}{s}, \ldots, \frac{1}{s}}_{r}, \underbrace{\frac{1}{s}, \ldots, \frac{1}{s}}_{s-r}\right) \\
&= H_{s-(r-1)}\left(\frac{r}{s}, \underbrace{\frac{1}{s}, \ldots, \frac{1}{s}}_{s-r}\right) + \frac{r}{s}H_r\left(\frac{1}{r}, \ldots, \frac{1}{r}\right) \\
&= H_{s-(r-1)-(s-1)}\left(\frac{r}{s}, \frac{s-r}{s}\right) + \frac{s-r}{s}H_{s-r}\left(\frac{1}{s-r}, \ldots, \frac{1}{s-r}\right) + \frac{r}{s}H_r\left(\frac{1}{r}, \ldots, \frac{1}{r}\right) \\
&= H_2(\frac{r}{s}, \frac{s-r}{s}) + \frac{s-r}{s}f(s-r) + \frac{r}{s}f(r)
\end{aligned}$$

And since $p = \frac{r}{s}$, so we can get that

$$\begin{aligned} H_2(p, 1-p) &= H_2(\frac{r}{s}, \frac{s-r}{s}) \\ &= f(s) - \frac{s-r}{s} f(s-r) - \frac{r}{s} f(r) \\ &= \log_2 s - (1-p)\log_2 (s(1-p)) - p\log_2(sp) \\ &= -p\log_2 p - (1-p)\log_2(1-p) \end{aligned}$$

2. When p is irrational, since we have the Continuity property, so we can also get that

$$H_2(p, 1-p) = -p\log_2 p - (1-p)\log_2(1-p)$$

So above all, we have prove that $\forall p \in [0, 1]$, we have

$$H_2(p, 1-p) = -p\log_2 p - (1-p)\log_2(1-p)$$

Then we use the induction to prove that, suppose that

$$H_k(p_1, p_2, \ldots, p_k) = -\sum_{i=1}^{k} p_i \log p_i, \forall k = 1, 2, \ldots, m$$

Then for $k = m + 1$, we have

$$\begin{aligned} H_k(p_1, p_2, \ldots, p_k) &= H_{m+1}(p_1, p_2, \ldots, p_{m+1}) \\ &= H_m(p_1 + p_2, p_3, \ldots, p_{m+1}) + (p_1 + p_2) H_2\left(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2}\right) \\ &= \left(-(p_1 + p_2)\log(p_1 + p_2) - \sum_{i=3}^{m+1} p_i \log p_i\right) - (p_1 + p_2)\left(\frac{p_1}{p_1 + p_2}\log\frac{p_1}{p_1 + p_2} + \frac{p_2}{p_1 + p_2}\log\frac{p_2}{p_1 + p_2}\right) \\ &= -(p_1 + p_2)\log(p_1 + p_2) - \sum_{i=3}^{m+1} p_i \log p_i - p_1 \log p_1 - p_2 \log p_2 + (p_1 + p_2)(\frac{p_1}{p_1 + p_2} + \frac{p_2}{p_1 + p_2})\log(p_1 + p_2) \\ &= -\sum_{i=1}^{m+1} p_i \log p_i \\ &= -\sum_{i=1}^{k} p_i \log p_i \end{aligned}$$

So we have proved that $\forall m$, we have

$$H_m(p_1, \ldots, p_m) = -\sum_{i=1}^{m} p_i \log p_i.$$

So above all, we have proved that [1]

$$H_m(p_1, p_2, \ldots, p_m) = -\sum_{i=1}^{m} p_i \log p_i, \quad m = 2, 3, \ldots$$

[1] The proof above has referenced from A Rényi.Wahrscheinlichkeitsrechnung, mit einem Anhang über Informationstheorie. Veb Deutscher Verlag der Wissenschaften, Berlin, 1962.

Problem 4

4.33 Chain inequality. Let $X_1 \to X_2 \to X_3 \to X_4$ form a Markov chain. Show that

$$I(X_1; X_3) + I(X_2; X_4) \leq I(X_1; X_4) + I(X_2; X_3)$$

Solution

From the chain rule of mutual information, we have

$$\begin{aligned}
I(X_1, X_2; X_3) &= I(X_1; X_3) + I(X_2; X_3|X_1) \\
&= I(X_2; X_3) + I(X_1; X_3|X_2) \\
I(X_1, X_2; X_4) &= I(X_1; X_4) + I(X_2; X_4|X_1) \\
&= I(X_2; X_4) + I(X_1; X_4|X_2) \\
I(X_3, X_4; X_2|X_1) &= I(X_3; X_2|X_1) + I(X_4; X_2|X_1, X_3) \\
&= I(X_4; X_2|X_1) + I(X_3; X_2|X_1, X_4)
\end{aligned}$$

From the property of Markov chain, we have

$$\begin{aligned}
I(X_1; X_3|X_2) &= 0 \\
I(X_1; X_4|X_2) &= 0 \\
I(X_2; X_4|X_1, X_3) &= 0
\end{aligned}$$

So we have

$$I(X_1; X_3) - I(X_2; X_3) = -I(X_2; X_3|X_1) \tag{1}$$

$$I(X_2; X_4) - I(X_1; X_4) = I(X_2; X_4|X_1) \tag{2}$$

$$I(X_3; X_2|X_1) - I(X_4; X_2|X_1) = I(X_2; X_3|X_1, X_4) \tag{3}$$

So we have

$$\begin{aligned}
& I(X_1; X_3) + I(X_2; X_4) - I(X_1; X_4) - I(X_2; X_3) \\
&= [I(X_1; X_3) - I(X_2; X_3)] + [I(X_2; X_4) - I(X_1; X_4)] \\
&= I(X_2; X_4|X_1) - I(X_2; X_3|X_1) && \text{(from (1) and (2))} \\
&= -I(X_2; X_3|X_1, X_4) && \text{(from (3))} \\
&\leq 0 && \text{(property of mutual information)}
\end{aligned}$$

So above all, we have proved that

$$I(X_1; X_3) + I(X_2; X_4) \leq I(X_1; X_4) + I(X_2; X_3)$$

Problem 5

5.4 Huffman coding. Consider the random variable

$$X = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \\ 0.49 & 0.26 & 0.12 & 0.04 & 0.04 & 0.03 & 0.02 \end{pmatrix}.$$

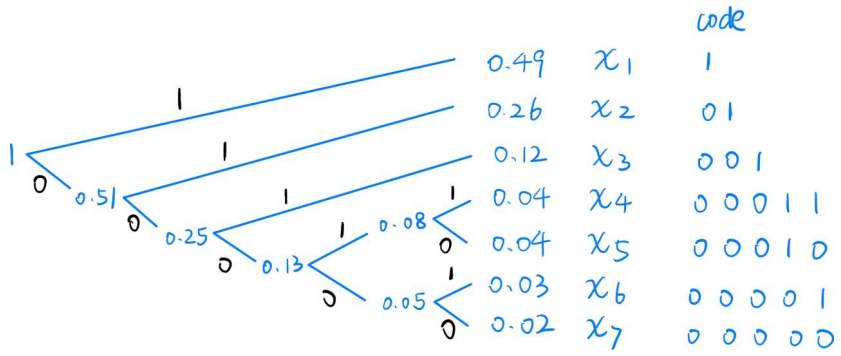
(a) Find a binary Huffman code for X.

(b) Find the expected code length for this encoding.

(c) Find a ternary Huffman code for X.

Solution

(a) The binary Huffman tree and the corresponding binary Huffman code for X are shown below.



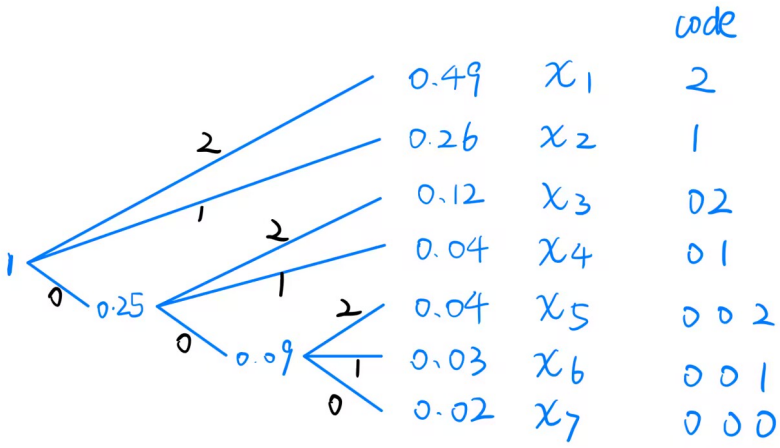
(b) The length of the binary Huffman code for x_i is $l(x_i)$:

x_i	x_1	x_2	x_3	x_4	x_5	x_6	x_7
$l(x_i)$	1	2	3	5	5	5	5

So the expected code length for this encoding is

$$\bar{L} = \sum_{i=1}^{t} p(x_i) l(x_i) = 2.02 \text{ bits}$$

(c) The ternary Huffman tree and the corresponding ternary Huffman code for X are shown below.



Problem 6

5.6 Bad codes. Which of these codes cannot be Huffman codes for any probability assignment?

(a) $\{0, 10, 11\}$

(b) $\{00, 01, 10, 110\}$

(c) $\{01, 10\}$

Solution

(a) The code could be a Huffman code. For example

$$p(x_1) = 0.7, p(x_2) = 0.2, p(x_3) = 0.1$$

Then the Huffman code could be:

$$c(x_1) = 0, c(x_2) = 10, c(x_3) = 11$$

(b) The code is a bad code. The codeword with longest length should not be single, at least 2 variables' codewords should be the longest, otherwise, the longest codeword could be shorten.

i.e. The codeword 110 could be shorten to 11.

(c) The two codewords' first bit and the second bit are all different, so we could shorten the codewords with their first bit, and the property would remain, i.e. it is still a prefix-free code.

i.e. The codewords $\{01, 10\}$ could be shorten to $\{0, 1\}$.