# CSIE 7624: Homework 1
# Week 2: Entropy, Relative Entropy and Mutual Information

Due on March 26, 2018

*JL Wu 14:20*

**D06922002 Kuan-Lun Wang**

# Contents

## Problem 1

*Coin flips.* A fair coin is flipped until the first head occurs. Let $X$ denote the number of flips required.

### (a)

Find the entropy $H(X)$ in bits. The following expression may be useful:

$$\sum_{n=0}^{\infty} r^n = \frac{1}{1-r}, \ \sum_{n=0}^{\infty} nr^n = \frac{r}{(1-r)^2}.$$

Recall the definition of this $X$, it is the number of trials required to flip the first successful trial under Bernoulli trials, with success probability $q$; That is, $X$ has a negative binomial distribution, and the p.m.f. of it could be

$$p(x) = \binom{x-1}{1-1}(q)^1(1-q)^{x-1} = q(1-q)^{x-1}$$

for all natural number $x$. By the definition of entropy, we have the entropy of $X$:

$$\begin{aligned} H(X) &= -\sum_{x=1}^{\infty} p(x)\log p(x) \\ &= -\sum_{x=1}^{\infty} q(1-q)^{x-1}\log q(1-q)^{x-1} \\ &= \left[-q(\log q)\sum_{x=1}^{\infty}(1-q)^{x-1}\right] + \left[-q(\log(1-q))\sum_{x=1}^{\infty}(x-1)(1-q)^{x-1}\right] \end{aligned}$$

Introducing $n = x-1$ and $r = 1-q$, we get

$$\begin{aligned} H(X) &= \left[-q(\log q)\sum_{n=0}^{\infty} r^n\right] + \left[-q(\log(1-q))\sum_{n=0}^{\infty} nr^n\right] \\ &= \left[-q(\log q)\frac{1}{1-r}\right] + \left[-q(\log(1-q))\frac{r}{(1-r)^2}\right] \\ &= \left[-q(\log q)\frac{1}{q}\right] + \left[-q(\log(1-q))\frac{1-q}{q^2}\right] \\ &= \frac{-(q\log q + (1-q)\log(1-q))}{q} \\ &= H(q)/q \text{ (bits)}. \end{aligned}$$

Since the coin is fair implies $q = 0.5$, the entropy $H(q)$ is 1 in bits. Hence $H(X)$ is 2 in bits.

(a)

### (b)

A random variable $X$ is drawn according to this distribution. Find an "efficient" sequence of yes-no questions of the form, "Is $X$ contained in the set $S$?" Compare $H(X)$ to the expected number of questions required to determine $X$.

Consider a sequence $S_n$ of yes-no questions: "Is $X = n$?" Clearly, the index of $S_n$ is equal to $X$ if and only if $X = n$. Then, the expected number of $S_n$ required to determine $X$ is equal to $\mathrm{E}[X] = q/(1-q)$. Moreover, we can evaluate the their relative:

$$\begin{cases} H(X) \geq \mathrm{E}[X], & \text{if } q \leq q_0; \\ H(X) < \mathrm{E}[X], & \text{if } q > q_0, \end{cases}$$

where $q_0$ is the root of $H(q)/q - q/(1-q)$, near 0.577886.

## Problem 2

*Minimum entropy.* What is the minimum value of $H(p_1, \ldots, p_n) = H(\vec{p})$ as $\vec{p}$ ranges over the set of $n$-dimensional probability vectors? Find all $\vec{p}$'s that achieve this minimum.

First, we show the minimum value of $H(\vec{p})$ is zero. By the definition of entropy, the minimum value of $H(\vec{p})$ must be not small than 0; that is,

$$H(\vec{p}) = -\sum_{x=1}^{n} p_x \log p_x \geq 0.$$

Assume that the minimum value of $H(\vec{p})$ greater that 0. Take $p_1 = 1$ and $p_2, \ldots, p_n = 0$. By the definition of $0 \log 0$, the entropy in this case is

$$H(\vec{p}) = -(1 \log 1 + 0 \log 0 + \cdots + 0 \log 0) = -(1 \log 1 + (n-1)0) = 0.$$

So, it is a contradiction because that $H(\vec{p})$ must be nonzero. Therefore the minimum value of $H(p_1, \ldots, p_n)$ is 0.

Second, we show $H^{-1}(0) = \{(1, 0, \ldots, 0), (0, 1, 0, \ldots, 0), (0, \ldots, 0, 1)\}$. Denoting $S = \{(1, 0, \ldots, 0), (0, 1, 0, \ldots, 0), (0, \ldots, 0, 1)\}$. Assume that there is a $\vec{p} \notin S$ such that $H(\vec{p}) = 0$; that is, $\vec{p} \in H^{-1}(0) \backslash S$. Then, there is a $x$ such that $p_x \neq 0, 1$. Since $p_x$ must between 0 and 1, $\log p_x$ must be negative and implies

$$H(\vec{p}) \geq -p_x \log p_x > 0.$$

Thus, it is a contradiction because that $H(\vec{p})$ must be 0. So, the $\vec{p}$ achieves minimum value 0 of $H(\vec{p})$ must be $(1, 0, \ldots, 0)$, $(0, 1, 0, \ldots, 0)$, or $(0, \ldots, 0, 1)$.

## Problem 3

*Entropy of functions of a random variable.* Let $X$ be a discrete random variable. Show that the entropy of

 Problem 3 continued on next page...

a function of $X$ is less than or equal to the entropy of $X$ by justifying the following steps:

$$\begin{aligned} H(X, g(X)) &\overset{(a)}{=} H(X) + H(g(X)|X) \\ &\overset{(b)}{=} H(X), \\ H(X, g(X)) &\overset{(c)}{=} H(g(X)) + H(X|g(X)) \\ &\overset{(d)}{\geq} H(g(X)). \end{aligned}$$

Thus, $H(g(X)) \leq H(X)$.

First, we show the steps (a) and (b). Take $X_1 = g(X)$ and $X_2 = X$. By the definition of entropy, we have the following equation:

$$H(X, g(X)) = H(X_2, X_1) = H(X_1, X_2).$$

By the entropy chain rule (Theorem 2.5.1), we have the following equation:

$$H(X_1, X_2) = H(g(X)|X) + H(X) \overset{(a)}{=} H(X) + H(g(X)|X).$$

Denote the p.m.f. of $(X_1, X_2)$ by $p(x_1, x_2)$. Assume that there is a $x_2 \in \mathcal{X}_2$, where $\mathcal{X}_2$ is the alphabet of $X_2$, such that $p(x_1, x_2) \neq 0, 1$ for some $x_1 \in \mathcal{X}_1 \backslash \{g(x_2)\}$, where $\mathcal{X}_1$ is the alphabet of $X_1$. Since $X_1 = g(X) = g(X_2)$, $p(g(x_2), x_2)$ must be 1. Clearly, it is a contradiction because that $p(x_1, x_2)$ must be not a zero or one. That is, we have:

$$\begin{aligned} H(g(X)|X = x) &= -\sum_{x_1 \in \mathcal{X}_1} p(x_1, x_2) \log p(x_1, x_2) \\ &= -p(g(x_2), x_2) \log p(g(x_2), x_2) - \sum_{\substack{x_1 \in \mathcal{X}_1 \\ x_1 \neq g(x_2)}} p(x_1, x_2) \log p(x_1, x_2) \\ &= -1 \log 1 - (|\mathcal{X}_1| - 1) 0 \log 0 \\ &= 0 \end{aligned}$$

for $x \in \mathcal{X}_2$, and the step (b) is true.
Similarly, we take $X_1^* = X$ and $X_2^* = g(X)$, and we have:

$$H(X, g(X)) = H(X_1^*, X_2^*) \overset{(c)}{=} H(g(X)) + H(X|g(X)).$$

By the definition of entropy, $H(X|g(X))$ must be greater than 0. So, the step (d) is true. Therefore the proof is true; i.e., $H(g(X)) \leq H(X, g(X)) = H(X)$.

**Definition 1.** A function from $A$ to $B$ is a triple of object $\langle f, A, B \rangle$, where $A$ and $B$ are classes and $f$ is a subclass of $A \times B$ with the following properties:

1. $\forall x \in A$, $\exists y \in B$ such that $(x, y) \in f$;

2. If $(x, y_1) \in f$ and $(x, y_2) \in f$, then $y_1 = y_2$.

## Problem 4

*Zero conditional entropy.* Show that if $H(Y|X) = 0$, then $Y$ is a function of $X$ [i.e., for all $x$ with $p(x) > 0$, there is only one possible value of $y$ with $p(x, y) > 0$].

Denote the alphabets of $X$ and $Y$ by $\mathcal{X}$ and $\mathcal{Y}$, respectively, and denote the p.m.f. of $(X, Y)$ by $p(x, y)$. By the definition of entropy, we have:

$$H(Y|X) = \sum_{x\in\mathcal{X}} p_X(x)H(Y|X = x),$$

where $p_X(x) = \sum_{y\in\mathcal{Y}} p(x, y)$. Assume that $Y$ is not a function of $X$. We consider the case: there is a $x_0 \in \mathcal{X}$ such that $p(x, y) \neq 0, 1$ for all $y \in \mathcal{Y}$. Clearly, for these $x_0$, $p(x, y) > 0$ implies

$$H(Y|X = x_0) = -\sum_{y\in\mathcal{Y}} \frac{p(x_0, y)}{p(x_0)} \log \frac{p(x_0, y)}{p(x_0)} > 0$$

and

$$H(Y|X) = \sum_{x\in\mathcal{X}} p_X(x)H(Y|X = x) > p_X(x_0)H(Y|X = x_0) > 0.$$

So, it is a contradiction, in this case, because that $H(Y|X)$ must be zero; that is, for all $x \in \mathcal{X}$, there is a $y \in \mathcal{Y}$ such that $p(x, y) = 0$ or 1. By the definition of p.m.f., for all $x \in \mathcal{X}$, there is an unique $y \in \mathcal{Y}$ such that $p(x, y) = 1$ ($\because 0 \leq p_X(x) = \sum_{y\in\mathcal{Y}} p(x, y) \leq 1$). Clearly, it is a contradiction because that $H(Y|X) = 0$. Therefore $Y$ must be a function of $X$.

## Problem 5

*Entropy of a disjoint mixture.* Let $X_1$ and $X_2$ be discrete random variables drawn according to probability mass functions $p_1(\cdot)$ and $p_2(\cdot)$ over the respective alphabets $\mathcal{X}_1 = \{1, 2, \ldots, m\}$ and $\mathcal{X}_2 = \{m + 1, \ldots, n\}$. Let

$$X = \begin{cases} X_1 & \text{with probability } \alpha, \\ X_2 & \text{with probability } 1 - \alpha. \end{cases}$$

### (a)

Find $H(X)$ in terms of $H(X_1)$, $H(X_2)$, and $\alpha$.


Problem 5 [(a)] continued on next page. . .

By the definition of entropy, we have:

$$\begin{aligned}
H(X) &= -\sum_{x_1\in\mathcal{X}_1} \alpha p_1(x_1)\log \alpha p_1(x_1) - \sum_{x_2\in\mathcal{X}_2} (1-\alpha)p_2(x_2)\log(1-\alpha)p_2(x_2) \\
&= \alpha\left(-\sum_{x_1\in\mathcal{X}_1} p_1(x_1)\log p_1(x_1) - \log\alpha \sum_{x_1\in\mathcal{X}_1} p_1(x_1)\right) \\
&\quad +(1-\alpha)\left(-\sum_{x_2\in\mathcal{X}_2} p_2(x_2)\log p_2(x_2) - \log(1-\alpha) \sum_{x_2\in\mathcal{X}_2} p_2(x_2)\right) \\
&= \alpha(H(X_1)-\log\alpha) + (1-\alpha)(H(X_2)-\log(1-\alpha)) \\
&= \alpha H(X_1) + (1-\alpha)H(X_2) + (-\alpha\log\alpha - (1-\alpha)\log(1-\alpha)) \\
&= \alpha H(X_1) + (1-\alpha)H(X_2) + H(\alpha).
\end{aligned}$$

(a)

## (b)

Maximize over $\alpha$ to show that $2^{H(X)} \le 2^{H(X_1)} + 2^{H(X_2)}$ and interpret using the notion that $2^{H(X)}$ is the effective alphabet size.

By the above answer of this problem (a), we have:

$$\frac{\partial H(X)}{\partial \alpha} = H(X_1) - H(X_2) + \log\frac{1-\alpha}{\alpha}$$

implies that

$$\left.\frac{\partial H(X)}{\partial \alpha}\right|_{\alpha=\alpha_0} = 0 \text{ and } \left.\frac{\partial^2 H(X)}{\partial \alpha^2}\right|_{\alpha=\alpha_0} = -\frac{1}{\alpha_0(1-\alpha_0)} = -\frac{\left(2^{H(X_1)}+2^{H(X_2)}\right)^2}{2^{H(X_1)+H(X_2)}} < 0,$$

where

$$\alpha_0 = \frac{1}{1+2^{-H(X_1)+H(X_2)}}$$

which confirms that $H(X)$ and $\alpha_0$ are indeed the maximum value and the maximum point; that is,

(b)

$$\begin{aligned}
H(X) &\le \max_{\alpha\in[0,1]} H(X) \\
&= \frac{H(X_1)}{1+2^{-H(X_1)+H(X_2)}} + \frac{2^{-H(X_1)+H(X_2)}H(X_2)}{1+2^{-H(X_1)+H(X_2)}} \\
&\quad -\frac{1}{1+2^{-H(X_1)+H(X_2)}}\log\frac{1}{1+2^{-H(X_1)+H(X_2)}} \\
&\quad -\frac{2^{-H(X_1)+H(X_2)}}{1+2^{-H(X_1)+H(X_2)}}\log\frac{2^{-H(X_1)+H(X_2)}}{1+2^{-H(X_1)+H(X_2)}} \\
&= H(X_1) + \log\left(1+2^{-H(X_1)+H(X_2)}\right).
\end{aligned}$$

Therefore $2^{H(X)} \le 2^{H(X_1)} + 2^{H(X_2)}$. By the solution of Problem 1, $2^{H(X)}$ is the effective alphabet size.

# Problem 6

*Example of joint entropy.* Let $p(x,y)$ be given by

| $X \backslash Y$ | 0 | 1 |
|---|---|---|
| 0 | $\frac{1}{3}$ | $\frac{1}{3}$ |
| 1 | 0 | $\frac{1}{3}$ |

Find:

## (a)

$H(X)$, $H(Y)$.

Denote the alphabets of $X$ and $Y$ by $\mathcal{X}$ and $\mathcal{Y}$, respectively, and denote the p.m.f. of $(X, Y)$ by $p(x, y)$. By the above table, the p.m.f. $p_X$ of $X$ must be

$$p_X(x) = \begin{cases} p(0,0) + p(0,1), & x = 0; \\ p(1,0) + p(1,1), & x = 1 \end{cases}$$
$$= \begin{cases} \frac{1}{3} + \frac{1}{3}, & x = 0; \\ 0 + \frac{1}{3}, & x = 1 \end{cases}$$
$$= \begin{cases} \frac{2}{3}, & x = 0; \\ \frac{1}{3}, & x = 1. \end{cases}$$

Then, by the definition of entropy, we have:

$$H(X) = -\sum_{x \in \mathcal{X}} p_X(x) \log p_X(x) = -\tfrac{2}{3} \log \tfrac{2}{3} - \tfrac{1}{3} \log \tfrac{1}{3} = -\frac{2}{3} + \log 3.$$

Similarly, we have the p.m.f. $p_Y$ of $X$:

$$p_Y(y) = \begin{cases} p(0,0) + p(1,0), & y = 0; \\ p(0,1) + p(1,1), & y = 1 \end{cases}$$
$$= \begin{cases} \frac{1}{3}, & y = 0; \\ \frac{2}{3}, & y = 1, \end{cases}$$

and the entropy of $Y$:

$$H(Y) = H(\tfrac{1}{3}) = H(X) = -\frac{2}{3} + \log 3.$$

## (b)

$H(X|Y)$, $H(Y|X)$.


Problem 6 [(b)] continued on next page...

Using the notation in this problem (a), by the definition of entropy, we have:

$$
\begin{aligned}
H(X|Y=y) &= -\sum_{x\in\mathcal{X}} \frac{p(x,y)}{p_Y(y)} \log \frac{p(x,y)}{p_Y(y)} \\
&= \begin{cases} -\frac{p(0,0)}{p_Y(0)} \log \frac{p(0,0)}{p_Y(0)} - \frac{p(1,0)}{p_Y(0)} \log \frac{p(1,0)}{p_Y(0)}, & y=0; \\ -\frac{p(0,1)}{p_Y(1)} \log \frac{p(0,1)}{p_Y(1)} - \frac{p(1,1)}{p_Y(1)} \log \frac{p(1,1)}{p_Y(1)}, & y=1 \end{cases} \\
&= \begin{cases} -\frac{1/3}{1/3} \log \frac{1/3}{1/3} - \frac{0}{1/3} \log \frac{0}{1/3}, & y=0; \\ -\frac{1/3}{2/3} \log \frac{1/3}{2/3} - \frac{1/3}{2/3} \log \frac{1/3}{2/3}, & y=1 \end{cases} \\
&= \begin{cases} 0, & y=0; \\ 1, & y=1, \end{cases}
\end{aligned}
$$

and

$$
H(X|Y) = \sum_{y\in\mathcal{Y}} p_Y(y) H(X|Y=y) = \frac{1}{3} H(X|Y=0) + \frac{2}{3} H(X|Y=1) = \frac{1}{3}\cdot 0 + \frac{2}{3}\cdot 1 = \frac{2}{3}.
$$

(b)

Similarly, we have:

$$
\begin{aligned}
H(Y|X=x) &= -\sum_{y\in\mathcal{Y}} \frac{p(x,y)}{p_X(x)} \log \frac{p(x,y)}{p_X(x)} \\
&= \begin{cases} -\frac{p(0,0)}{p_X(0)} \log \frac{p(0,0)}{p_X(0)} - \frac{p(0,1)}{p_X(0)} \log \frac{p(0,1)}{p_X(0)}, & x=0; \\ -\frac{p(1,0)}{p_X(1)} \log \frac{p(1,0)}{p_X(1)} - \frac{p(1,1)}{p_X(1)} \log \frac{p(1,1)}{p_X(1)}, & x=1 \end{cases} \\
&= \begin{cases} -\frac{1/3}{2/3} \log \frac{1/3}{2/3} - \frac{1/3}{2/3} \log \frac{1/3}{2/3}, & x=0; \\ -\frac{0}{1/3} \log \frac{0}{1/3} - \frac{1/3}{1/3} \log \frac{1/3}{1/3}, & x=1 \end{cases} \\
&= \begin{cases} 1, & x=0; \\ 0, & x=1, \end{cases}
\end{aligned}
$$

and

$$
H(Y|X) = \sum_{x\in\mathcal{X}} p_X(x) H(Y|X=x) = \frac{2}{3} H(Y|X=0) + \frac{1}{3} H(Y|X=1) = \frac{2}{3}\cdot 1 + \frac{1}{3}\cdot 0 = \frac{2}{3}.
$$

## (c)

$H(X,Y)$.

Problem 6 [(c)] continued on next page. . .

By the entropy chain rule, we have:

$$H(X,Y) = H(X|Y) + H(Y) = \frac{2}{3} + \log 3.$$

### (d)

$H(Y) - H(Y|X)$.

By the solution of this problem (a) and (b), we have:

$$H(Y) - H(Y|X) = \log 3 - \frac{2}{3}.$$

### (e)

$I(X;Y)$.

By the mutual information and entropy theorem (Theorem 2.4.1), we have:

$$I(X;Y) = H(Y) - H(Y|X) = \log 3 - \frac{2}{3}.$$

### (f)

Draw a Venn diagram for the quantities in parts (a) through (e).

See the Figure 1.

## Problem 7

*Bottleneck.* Suppose that a (non-stationary) 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, $p(x_1, x_2, x_3) = p(x_1)\ p(x_2|x_1)\ p(x_3|x_2)$, for all $x_1 \in \{1, 2, \dots, n\}$, $x_2 \in \{1, 2, \dots, k\}$, $x_3 \in \{1, 2, \dots, m\}$.

### (a)

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

By the data-processing inequality, Theorem 2.8.1, and Theorem 2.6.4, we have the following inequality:

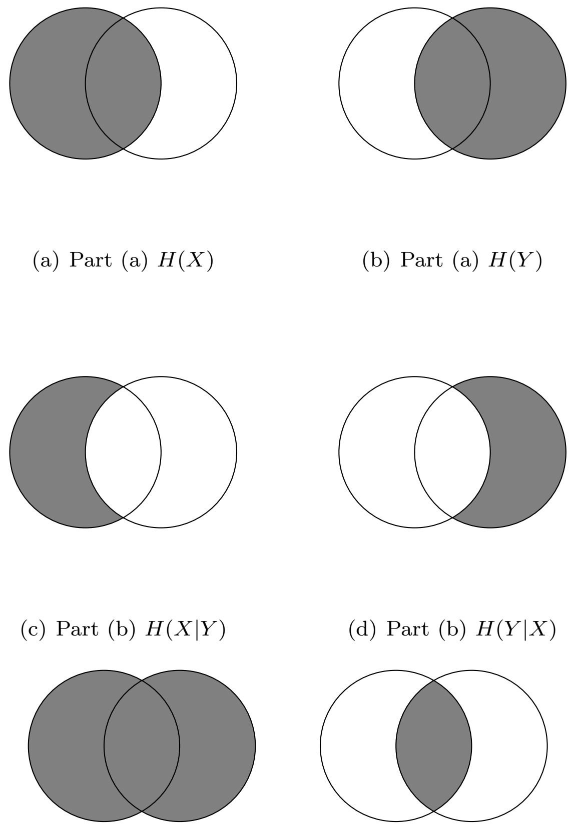
$$I(X_1; X_3) \le I(X_1; X_2) = H(X_2) - H(X_2|X_1) \le H(X_2) \le \log k;$$

that is, the size of the bottleneck limits the dependence between $X_1$ and $X_3$.

### (b)

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

 Problem 7 [(b)] continued on next page...



(a) Part (a) $H(X)$

(b) Part (a) $H(Y)$

(c) Part (b) $H(X|Y)$

(d) Part (b) $H(Y|X)$

(e) Part (c) $H(X,Y)$

(f) Part (d) $H(Y)-H(Y|X)$ and Part (e) $I(X;Y)$

Figure 1: The Venn diagram for the quantities in parts (a) through (e)

By the above inequality, the mutual information $I(X_1;X_3)$ must be zero whenever $k=1$. By the definition of mutual information, two variables must be independent if has zero mutual information. (b)

## Problem 8

*World Series.* The World Series is a seven-game series that terminates as soon as either team wins four games. Let $X$ be the random variable that represents the outcome of a World Series between teams $A$ and $B$; possible values of $X$ are AAAA, BABABAB, and BBBAAAA. Let $Y$ be the number of games played, which ranges from 4 to 7. Assuming that A and B are equally matched and that the games are independent, calculate $H(X)$, $H(Y)$, $H(Y|X)$, and $H(X|Y)$.

First, we calculate $H(Y)$. Denote the alphabets of $X$ and $Y$ by $\mathcal{X}$ and $\mathcal{Y}$, respectively. By the definition of $Y$, the p.m.f. of it could be

$$p_Y(y) = 2\binom{y-1}{4-1}\left(\frac{1}{2}\right)^4\left(1-\frac{1}{2}\right)^{y-4} = \frac{\binom{y-1}{3}}{2^{y-1}}$$

Problem 8 continued on next page. . .

for all $y = 4, \dots, 7$. Then, the entropy of $Y$ is

$$H(Y) = -\sum_{y\in\mathcal{Y}} p_y(y) \log p_Y(y) = \frac{27 - 5\log 5}{8}.$$

Second, we calculate $H(X|Y)$. For the World Series with $y$ games, the p.m.f. of $X$ could be

$$\frac{p(x,y)}{p_Y(y)} = \frac{1}{2\binom{y-1}{3}}$$

for each possible $x$, say, the possible alphabet is $\mathcal{X}_y$ ($|\mathcal{X}_y| = 2\binom{y-1}{3}$). So, the entropy of $X|Y = y$ is

$$H(X|Y = y) = -\sum_{x\in\mathcal{X}_y} \frac{1}{2\binom{y-1}{3}} \log \frac{1}{2\binom{y-1}{3}} = 1 + \log \binom{y-1}{3},$$

and the entropy of $X|Y$ is

$$H(X|Y) = \sum_{y\in\mathcal{Y}} p_Y(y) H(X|Y = y) = \sum_{y=4}^{7} \frac{\binom{y-1}{3}}{2^{y-1}} \left(1 + \log \binom{y-1}{3}\right) = \frac{39 + 10\log 5}{16}.$$

Third, we calculate $H(Y|X)$. Since $Y$ is the length of $X$, $H(Y|X)$ must be zero, by the Problem 4. Fourth, we calculate $H(X)$. By the chain rule, Theorem 2.2.1, we have:

$$H(X) = H(X,Y) - H(Y|X) = H(Y) + H(X|Y) - H(Y|X) = \frac{93}{16}.$$

Therefore the calculate is completed.

## Problem 9

*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

$$I(X;Y;Z) = I(X;Y) - I(X;Y|Z).$$

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

Consider $X$ and $Y$ are i.i.d. uniform from $\{0,1\}$, and $Z = X + Y$ is a function of $(X,Y)$. Clearly, $H(X) = H(Y) = 1$ and $H(X,Y) = 2$ because $H(1/2) = 1$ and $H((1/2)^2) = H(1/4) = 2$; that is, $I(X;Y) = H(X) + H(Y) - H(X,Y) = 0$. By the definitions of conditional mutual information and chain

 Problem 9 continued on next page...

rule, Theorem 2.2.1, we have:

$$\begin{aligned} I(X;Y|Z) &= H(X|Z) - H(X|Y,Z) \\ &= H(X,Z) - H(Z) - H(X,Y,Z) + H(Y) + H(Z|Y) \\ &= H(X,Z) - H(Z) - H(X,Y,Z) + H(Y) + H(Y,Z) - H(Y) \\ &= -H(X,Y,Z) + H(X,Z) + H(Y,Z) - H(Z) \\ &= -H(X,Y) + H(X,Z) + H(Y,Z) - H(Z). \end{aligned}$$

Since $H(X,Y) = 2$, $H(X,Z) = H(Y,Z) = H(1/4) = 2$, and $H(Z) = H(1/4, 1/2, 1/4) = 3/2$, the value of $I(X;Y|Z)$ is 1/2. Therefore, for this case, $I(X;Y;Z) = I(X;Y) - I(X;Y|Z) = -1/2 < 0$.

## (a)

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

By information chain rule, Theorem 2.5.2, we have:

$$\begin{aligned} I(X;Y;Z) &= I(X;Y) - I(X;Y|Z) \\ &= I(X;Y) - I(X,Z;Y) + I(Z;Y) \qquad (1) \\ &= I(X;Y) - I(X,Z;Y) + I(Y;Z). \end{aligned}$$

Since $I(X,Z;Y) = H(X,Z) + H(Y) - H(X,Y,Z)$ and the equation (2.41) in the textbook, we can rewrite the above equation (1) as:

$$\begin{aligned} I(X;Y;Z) &= H(X,Y,Z) - H(Y) + I(X;Y) + I(Y;Z) - H(X,Z) \\ &= H(X,Y,Z) - H(X) - H(Y) - H(Z) + I(X;Y) + I(Y;Z) + I(Z;X). \end{aligned} \qquad (2)$$

(a)

## (b)

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

By the equation (2.41) in the textbook, we have:

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

and

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

(b)

Then, the equation (2) can be written as:

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

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

## Problem 10

*Inequalities.* Let $X$, $Y$, and $Z$ be joint random variables. Prove the following inequalities and find conditions for equality.

### (a)

$H(X,Y|Z) \geq H(X|Z)$.

Denote the alphabets of $X$, $Y$, and $Z$, by $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$, respectively, and denote the p.m.f. of $(X,Y,Z)$ and $Z$ by $p(x,y,z)$, $p_Z(z)$. Give $z_0 \in \mathcal{Z}$. By the definition of entropy, we have:

$$\begin{aligned}
H(X|Z=z_0) &= -\sum_{x\in\mathcal{X}} \frac{\sum_{y\in\mathcal{Y}} p(x,y,z)}{p_Z(z_0)} \log \frac{\sum_{y\in\mathcal{Y}} p(x,y,z)}{p_Z(z_0)} \\
&\leq -\sum_{x\in\mathcal{X}} \frac{\sum_{y\in\mathcal{Y}} p(x,y,z)}{p_Z(z_0)} \log \frac{p(x,y,z)}{p_Z(z_0)} \\
&\leq -\sum_{x\in\mathcal{X}} \frac{p(x,y,z)}{p_Z(z_0)} \log \frac{p(x,y,z)}{p_Z(z_0)} \\
&= H(X,Y|Z=z_0).
\end{aligned}$$

So, we have the following inequality:

$$H(X|Z) = \sum_{z\in\mathcal{Z}} p_Z(z) H(X|Z=z_0) \leq \sum_{z\in\mathcal{Z}} p_Z(z) H(X,Y|Z=z_0) = H(X,Y|Z).$$

Moreover, the condition for equality is

$$\sum_{y\in\mathcal{Y}} p(x,y,z) = p(x,y,z),\ \forall x \in \mathcal{X},\ \forall z \in \mathcal{Z}.$$

That is, the inequality with equality when $Y$ is a function of $X$ and $Z$.

### (b)

$I(X,Y;Z) \geq I(X;Z)$.


Problem 10 [(b)] continued on next page...

By the equation (2.41) in the textbook and chain rule, Theorem 2.2.1, we have:

$$\begin{aligned}
I(X,Y;Z) &= H(X,Y) + H(Z) - H(X,Y,Z) \\
&= H(X,Y) - H(X,Y|Z) \\
&\geq H(X,Y) - H(X|Z) \\
&= H(X) + H(Y|X) - H(X|Z) \\
&\geq H(X) - H(X|Z) \\
&= I(X;Z).
\end{aligned}$$

Clearly, the condition for equality is $Y$ and $Z$ are conditionally independent given $X$.

(b)

### (c)

$H(X,Y,Z) - H(X,Y) \leq H(X,Z) - H(X)$.

By chain rule, Theorem 2.2.1, and conditioning reduces entropy, Theorem 2.6.5, we have:

$$H(X,Y,Z) - H(X,Y) = H(X,Z) + H(Y|X,Z) - H(X) - H(Y|X).$$

and

$$H(Y|X,Z) \leq H(Y|X),$$

respectively. Thus, the following inequality is true:

$$H(X,Y,Z) - H(X,Y) \leq H(X,Z) - H(X).$$

Moreover, the condition for equality is $Y$ and $Z$ are conditionally independent given $X$.

(c)

### (d)

$I(X;Z|Y) \geq I(Z;Y|X) - I(Z;Y) + I(X;Z)$.

By information chain rule, we have:

$$\begin{aligned}
I(X;Z|Y) &= I(X,Y;Z) - I(Y;Z) \\
&= I(X;Z) + I(Y;Z|X) - I(Y;Z) \\
&= I(Z;Y|X) - I(Z;Y) + I(X;Z).
\end{aligned}$$

(d)

## Problem 11

*Fano.* We are given the following joint distribution on $(X,Y)$:

Problem 11 continued on next page...

| $X \backslash Y$ | $a$ | $b$ | $c$ |
|---|---|---|---|
| 1 | $\frac{1}{6}$ | $\frac{1}{12}$ | $\frac{1}{12}$ |
| 2 | $\frac{1}{12}$ | $\frac{1}{6}$ | $\frac{1}{12}$ |
| 3 | $\frac{1}{12}$ | $\frac{1}{12}$ | $\frac{1}{6}$ |

Let $\hat{X}(Y)$ be an estimator for $X$ (based on $Y$) and let $P_e = \Pr\{\hat{X}(Y) \neq X\}$.

## (a)

Find the minimum probability of error estimator $\hat{X}(Y)$ and the associated $P_e$.

Take an estimator

$$\hat{X}^*(Y) = \begin{cases} 1, & \text{if } Y = a; \\ 2, & \text{if } Y = b; \\ 3, & \text{if } Y = c. \end{cases}$$

Assume the estimator $\hat{X}^*(Y)$ without the minimum probability of error. Then, there is an estimator $\hat{X}_0(Y)$ such that $\Pr\{\hat{X}_0 \neq X\} < \Pr\{\hat{X}^* \neq X\}$; that is, there exits a $y = a$, $b$, or $c$ such that $\hat{X}_0(y) \neq \hat{X}^*(y)$ implies $\Pr\{\hat{X}^0 \neq X|Y = y\} < \Pr\{\hat{X}^* \neq X|Y = y\}$. Without loss of generality, we consider $y = a$, we have

$$\Pr\{\hat{X}_0 \neq X|Y = a\} = \frac{1/6 + 1/12}{1/6 + 1/12 + 1/12} = \frac{3}{4}$$

and

$$\Pr\{\hat{X}^* \neq X|Y = a\} = \frac{1/12 + 1/12}{1/6 + 1/12 + 1/12} = \frac{1}{2}.$$

Clearly, it is a contradiction because $3/4 \not\leq 1/2$. Therefore the estimator $\hat{X}^*(Y)$ has minimum error probability. Moreover, the probability of error estimator $\hat{X}(Y)$ is $1/2$.

## (b)

Evaluate Fano's inequality for this problem and compare.

Recall the error probability, in our estimator, is $P_e = 1/2$, and denote the alphabets of $X$ and $Y$ by $\mathcal{X}$ and $\mathcal{Y}$, respectively. Since

$$H(P_e) = -\sum_{y\in\mathcal{Y}} \frac{1}{2} \log \frac{1}{2} = 1$$

and

$$P_e \log |\mathcal{Y}| = \log 3,$$

we have the first value in Fano's inequality $H(P_e) + P_e \log |\mathcal{Y}| = 5$. Since the estimator $\hat{X}^*(Y)$ is an one-to-one and onto function, the following equation shows "=" in Fano's inequality:

$$\begin{aligned} H(X|\hat{X}^*(Y)) &= H(X,\hat{X}^*(Y)) - H(\hat{X}^*(Y)) \\ &\quad \text{(Chain rule, Theorem 2.2.1)} \\ &= H(X,Y) - H(Y) \\ &\quad \text{(Problem 3)} \\ &= H(X|Y) \\ &\quad \text{(Chain rule, Theorem 2.2.1).} \end{aligned}$$

Denote the the p.m.f. of $(X,Y)$ and $Y$ by $p(x,y)$ and $p_Y(y)$, respectively, and note

$$p_Y(y) = \frac{1}{6} + \frac{1}{12} + \frac{1}{12} = \frac{1}{3}$$

for all $y \in \mathcal{Y}$. Thus, the value of $H(X|\hat{X}^*(Y))$ and $H(X|Y)$ is (b)

$$\begin{aligned} H(X|Y) &= \sum_{y\in\mathcal{Y}} p_Y(y) H(X|Y=y) \\ &= \sum_{y\in\mathcal{Y}} \frac{1}{3}\left(-\sum_{x\in\mathcal{X}} \frac{p(x,y)}{p_Y(y)} \log \frac{p(x,y)}{p_Y(y)}\right) \\ &= 3\cdot\frac{1}{3}\left(-\frac{1/6}{1/3}\log\frac{1/6}{1/3} - \frac{1/12}{1/3}\log\frac{1/12}{1/3} - \frac{1/12}{1/3}\log\frac{1/12}{1/3}\right) \\ &= \frac{3}{2}. \end{aligned}$$

Therefore we have the following inequality for this problem:

$$\begin{array}{ccccc} H(P_e) + P_e \log|\mathcal{Y}| & \geq & H(X|\hat{X}^*(Y)) & \geq & H(X|Y); \\ 1 + \log 3 & \geq & \frac{3}{2} & \geq & \frac{3}{2}. \end{array}$$

# Problem 12

*Relative entropy is not symmetric.* Let the random variable $X$ have three possible outcomes $\{a, b, c\}$. Consider two distributions on this random variable:

| Symbol | $p(x)$ | $q(x)$ |
|---|---|---|
| $a$ | $\frac{1}{2}$ | $\frac{1}{3}$ |
| $b$ | $\frac{1}{4}$ | $\frac{1}{3}$ |
| $c$ | $\frac{1}{4}$ | $\frac{1}{3}$ |

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

Problem 12 continued on next page...

By the definition of entropy, we have:

$$H(p) = -\left(\frac{1}{2}\log\frac{1}{2} + \frac{1}{4}\log\frac{1}{4} + \frac{1}{4}\log\frac{1}{4}\right) = \frac{3}{2}$$

and

$$H(q) = -\left(\frac{1}{3}\log\frac{1}{3} + \frac{1}{3}\log\frac{1}{3} + \frac{1}{3}\log\frac{1}{3}\right) = \log 3.$$

By the definition of relative entropy, we have:

$$D(p||q) = \frac{1}{2}\log\frac{1/2}{1/3} + \frac{1}{4}\log\frac{1/4}{1/3} + \frac{1}{4}\log\frac{1/4}{1/3} = -\frac{3}{2} + \log 3$$

and

$$D(q||p) = \frac{1}{3}\log\frac{1/3}{1/2} + \frac{1}{3}\log\frac{1/3}{1/4} + \frac{1}{3}\log\frac{1/3}{1/4} = \frac{5}{3} - \log 3.$$

Moreover, we verify $D(p||q) \neq D(q||p)$ because that

$$D(p||q) = \frac{3}{2} + \log 3 \neq \frac{5}{3} - \log 3 = D(q||p).$$