ITCT Homework 1

- r07944007 網媒碩一 林良翰

1. [Prob. 2.1.]

a.

- Let r.v. X be the number of flips required, thus we have $\{x, p(x)\}, \forall x \in X$.
- Find the formula of $P(x)$, and let r be the probability of flipping a head.

$$\begin{array}{llll} x=1 & , & P(x=1) & = & r \\ x=2 & , & P(x=2) & = & (1-r)\,r \\ x=3 & , & P(x=3) & = & (1-r)^2 r \\ \vdots & & & \vdots & \\ x=i & , & P(x=i) & = & (1-r)^{n-1} r \end{array}$$

- Find the entropy $H(X)$

$$\begin{aligned} H(X) &= -\sum_{x\in X} P(x)\log P(x) \\ &= -\sum_{i=1}^{\infty} (1-r)^{i-1} r \log (1-r)^{i-1} r \\ &= -\sum_{n=0}^{\infty} (1-r)^{n} r \log (1-r)^{n} r \\ &= -\left[\sum_{n=0}^{\infty} n\,(1-r)^{n} r \log (1-r) + \sum_{n=0}^{\infty} (1-r)^{n} r \log r\right] \\ &= -\left[\frac{1-r}{[1-(1-r)]^2} r \log (1-r) + \frac{1}{1-(1-r)} r \log r\right] \\ &= -\left[\frac{1-r}{r}\log (1-r) + \log r\right] \end{aligned}$$

- $\because$ it is a fair coin $\Rightarrow r = \frac{1}{2}$

  $H(x) = -\left[\log\left(\frac{1}{2}\right) + \log\left(\frac{1}{2}\right)\right] = 2$

b.

- Let r.v. Y be the number of questions required, thus we have $\{y, p(y)\}, \forall y \in Y$.
- List the questions

  1st question ($y=1$): Is $x=1$?

  2nd question ($y=2$): If not, is $x=2$?

  3rd question ($y=3$): If not, is $x=3$?

  ...

n-th question ($y=n$): If not, is $x=n$?

$\Rightarrow p(y)=p(x)$

- The entropy of Y is same as X

  $H(Y)=H(X)=2$

2. [Prob. 2.3.]

- $H(p)$ will reach its minimum when only one element of p is non-zero, with total n possible cases.
- All possible p that make $H(p)$ achieve minimum

$$p=\left[p_1,\overbrace{0,\ldots,0}^{n-1}\right],p_1>0$$

$$p=\left[0,p_2,\overbrace{0,\ldots,0}^{n-2}\right],p_2>0$$

$$\vdots$$

$$p=\left[\overbrace{0,\ldots,0}^{n-1},p_n\right],p_n>0$$

3. [Prob. 2.4.]

- (a) uses the chain rule of entropy.
- (b) $\because$ the relation of x and $g(x)$ is clearly defined by function $g\Rightarrow p(g(x)|x)=1$

  $H(g(X)|X)=\sum\limits_{x\in X}\sum\limits_{g(x)\in g(X)}p(g(x),x)\log p(g(x)|x)=\sum\limits_{x\in X}\sum\limits_{g(x)\in g(X)}p(g(x),x)\log 1=0$
- (c) also uses the chain rule of entropy.
- (d) $\because$ entropy always larger or equals to zero: $H(X|g(X))\geq 0$

4. [Prob. 2.5.]

- Assume $p(x)>0,\forall x\in X$
- $H(Y|X)=\sum\limits_{y\in Y}\sum\limits_{x\in X}p(y,x)\log p(y|x)=0$

  We have two cases
  - Case 1 : $p(y,x)=0,\forall x\in X,\forall y\in Y$

    $\because p(x)>0$, we could only assure that $p(y,x)\geq 0$, so this case couldn't work.
  - Case 2 : $\log p(y|x)=0,\forall x\in X,\forall y\in Y$

    $\forall x\in X,\forall y\in Y,\because \log p(y|x)=0$

    $\Rightarrow p(y|x)=1$

$\Rightarrow p(y, x) = p(x)$

$\Rightarrow y$ is a function of x

$\Rightarrow$ r.v. Y is a function of r.v. X

- So if $H(Y|X) = 0$, then Y is a function of X

5. [Prob. 2.10.]

a.

- By the definition of entropy

$$H(X_1) = -\sum_{x \in X_1} p_1(x) \log p_1(x)$$
$$H(X_2) = -\sum_{x \in X_2} p_2(x) \log p_2(x)$$

- Let $p(\cdot)$ be the probability mass function of r.v. X

$$p(x) = \begin{cases} \alpha p_1(x), & \forall x \in X_1 \\ (1-\alpha) p_2(x), & \forall x \in X_2 \end{cases}$$

- Again, by the definition of entropy

$$H(X) = -\sum_{x \in X_1 \cup X_2} p(x) \log p(x)$$
$$\because X_1 \cap X_2 = \phi$$
$$\Rightarrow -\sum_{x \in X_1 \cup X_2} p(x) \log p(x) = -\sum_{x \in X_1} p(x) \log p(x) - \sum_{x \in X_2} p(x) \log p(x)$$
$$= -\sum_{x \in X_1} \alpha p_1(x) \log \alpha p_1(x) - \sum_{x \in X_2} (1-\alpha) p_2(x) \log (1-\alpha) p_2(x)$$
$$= \alpha H(X_1) + (1-\alpha) H(X_2) - \log \alpha^{\alpha} (1-\alpha)^{(1-\alpha)}$$

b.

- Let $H(X) = h,\ H(X_1) = h_1,\ H(X_2) = h_2,\ H(A) = \log \alpha^{\alpha}(1-\alpha)^{(1-\alpha)} = h_{\alpha}$
- $\because h$ is a concave function, thus it has maximum value. We conduct first derivative test to find it

$$\begin{aligned} \frac{dh}{d\alpha} &= \frac{d(\alpha h_1 + (1-\alpha) h_2 + h_\alpha)}{d\alpha} \\ &= h_1 - h_2 + \frac{d \log \alpha^{\alpha}(1-\alpha)^{(1-\alpha)}}{d\alpha} \\ &= h_1 - h_2 + \log \frac{1-\alpha}{\alpha} \overset{set}{=} 0 \end{aligned}$$

Let $\alpha^* = \frac{1}{1+2^{-h_1+h_2}}$ such that h reaches its maximum value

- Now we have the upper bound of h

$$\begin{aligned} h &\leq \alpha^* h_1 + (1-\alpha^*) h_2 - \log \alpha^{*\alpha^*} (1-\alpha^*)^{(1-\alpha^*)} \\ &= \alpha^* h_1 + (1-\alpha^*) h_2 - \alpha^* \log \alpha^* - (1-\alpha^*) \log (1-\alpha^*) \end{aligned}$$

Let $\beta = 2^{-h_1+h_2}$

$$
\begin{aligned}
&\Rightarrow \alpha^* = \tfrac{1}{1+2^{-h_1+h_2}} = \tfrac{1}{1+\beta} \\
h \quad &\leq \quad \tfrac{1}{1+\beta}h_1 + \left(1 - \tfrac{1}{1+\beta}\right)h_2 - \tfrac{1}{1+\beta}\log\tfrac{1}{1+\beta} - \left(1 - \tfrac{1}{1+\beta}\right)\log\left(1 - \tfrac{1}{1+\beta}\right) \\
&= \quad \tfrac{1}{1+\beta}(h_1 - h_2) + h_2 + \log(1+\beta) - \tfrac{\beta}{1+\beta}(-h_1 + h_2) \\
&= \quad h_1 + \log(1+\beta)
\end{aligned}
$$

- Since n^x is strictly increasing when $n > 1,\ \forall x \in \mathbb{R}$

$$
\begin{aligned}
2^h \quad &\leq \quad 2^{h_1 + \log(1+\beta)} \\
&= \quad 2^{h_1} \times 2^{\log(1+\beta)} \\
&= \quad 2^{h_1}(1+\beta) = 2^{h_1} + \beta 2^{h_1} \\
&= \quad 2^{h_1} + 2^{-h_1+h_2+h_1} \\
&= \quad 2^{h_1} + 2^{h_2}
\end{aligned}
$$

- Finally,

$$
2^{H(X)} \quad \leq \quad 2^{H(X_1)} + 2^{H(X_2)}
$$

6. [Prob. 2.12.]

a.

- $H(X) = -\sum\limits_{x \in X} p(x)\log p(x) = -p(X=0)\log p(X=0) - p(X=1)\log p(X=1) = -\frac{2}{3} + \log 3$
- $H(Y) = -\sum\limits_{y \in Y} p(y)\log p(y) = -p(Y=0)\log p(Y=0) - p(Y=1)\log p(Y=1) = -\frac{2}{3} + \log 3$

b.

- $H(X|Y) = -\sum\limits_{x \in X}\sum\limits_{y \in Y} p(x,y)\log p(x|y) = \frac{2}{3}$
- $H(Y|X) = -\sum\limits_{x \in Y}\sum\limits_{y \in X} p(y,x)\log p(y|x) = \frac{2}{3}$

c.

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

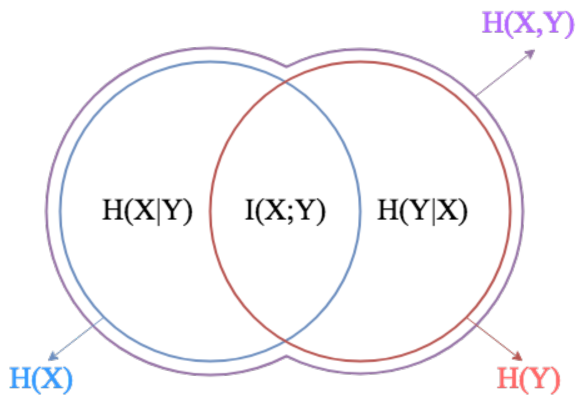
d.

- $H(Y) - H(Y|X) = -\frac{2}{3} + \log 3 - \frac{2}{3} = -\frac{4}{3} + \log 3$

e.

- $I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = -\frac{4}{3} + \log 3$

f.



7. [Prob. 2.16.]

a.

- By data process inequality

  If $X \to Y \to Z$, then $I(X;Z) \le I(X;Y)$

  Thus,

  $$\begin{aligned} I(X_1;X_3) &\le (X_1;X_2) \\ &= H(X_2) - H(X_2|X_1) \\ &\le H(X_2) \\ &\le \log|X_2| \\ &= \log k \end{aligned}$$

b.

- By a., if $k = 1$, then

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

  Because mutual information between two r.v. is always non-negative, therefore

  $I(X_1;X_3) = 0 \Rightarrow X_1$ and X_3 are independent.

8. [Prob. 2.18.]

- List all possibilities of X and Y
  - If $y = 4$, x has 2 possible cases with probablility $\left(\frac{1}{2}\right)^4$, $P(Y=4) = \frac{1}{8}$
  - If $y = 5$, x has 8 possible cases with probablility $\left(\frac{1}{2}\right)^5$, $P(Y=5) = \frac{1}{4}$
  - If $y = 6$, x has 20 possible cases with probablility $\left(\frac{1}{2}\right)^6$, $P(Y=6) = \frac{5}{16}$
  - If $y = 7$, x has 40 possible cases with probablility $\left(\frac{1}{2}\right)^7$, $P(Y=7) = \frac{5}{16}$
- $H(X) = -\sum_{x\in X} P(x)\log P(x) = -\left(\frac{2}{2^4}\log\frac{1}{2^4} + \frac{8}{2^5}\log\frac{1}{2^5} + \frac{20}{2^6}\log\frac{1}{2^6} + \frac{40}{2^7}\log\frac{1}{2^7}\right) = 5.8125$
- $H(Y) = -\sum_{y\in Y} P(y)\log P(y) = -\left(\frac{1}{8}\log\frac{1}{8} + \frac{1}{4}\log\frac{1}{4} + \frac{5}{16}\log\frac{5}{16} + \frac{5}{16}\log\frac{5}{16}\right) \approx 1.924$

- $\because Y$ is a deterministic function of X, $H\left(Y|X\right)=0$
- $\because H\left(X\right)+H\left(Y|X\right)=H\left(X,Y\right)=H\left(Y\right)+H\left(X|Y\right)$
  $\Rightarrow H\left(X|Y\right)=H\left(X\right)+H\left(Y|X\right)-H\left(Y\right)=3.889$

9. [Prob. 2.25.]

a.

- Using definition and venn diagram

$$\begin{aligned}
& I\left(X;Y;Z\right)\\
\overset{def.}{=}\ & I\left(X;Y\right)-I\left(X;Y|Z\right)\\
\overset{chain\ rule}{=}\ & I\left(X;Y\right)-\left[I\left(X;Y,Z\right)-I\left(X;Z\right)\right]\\
\overset{venn.}{=}\ & I\left(X;Y\right)+I\left(X;Z\right)-\left[H\left(X\right)+H\left(Y,Z\right)-H\left(X,Y,Z\right)\right]\\
\overset{def.}{=}\ & I\left(X;Y\right)+I\left(X;Z\right)-\left[H\left(X\right)+H\left(Y,Z\right)-H\left(X,Y,Z\right)\right]\\
\overset{def.}{=}\ & I\left(X;Y\right)+I\left(X;Z\right)-\left[H\left(X\right)+H\left(Y\right)+H\left(Z\right)-I\left(Y;Z\right)-H\left(X,Y,Z\right)\right]\\
=\ & I\left(X;Y\right)+I\left(X;Z\right)+I\left(Y;Z\right)-H\left(X\right)-H\left(Y\right)-H\left(Z\right)+H\left(X,Y,Z\right)
\end{aligned}$$

b.

- Using the properties

$$\begin{cases}
I\left(X;Y\right)=H\left(X\right)+H\left(Y\right)-H\left(X,Y\right)\\
I\left(Y;Z\right)=H\left(Y\right)+H\left(Z\right)-H\left(Y,Z\right)\\
I\left(Z;X\right)=H\left(Z\right)+H\left(X\right)-H\left(Z,X\right)
\end{cases}$$

  We can get

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

- By a. and b., we can find that $I\left(X,Y,Z\right)$ is symmetric and not necessary nonnegative.

10. [Prob. 2.29.]

a.

- By the definition of probability

  $p\left(x|z\right)=\frac{p(x,z)}{p(z)}\Rightarrow p\left(z\right)=\frac{p(x,z)}{p(x|z)}$

  $p\left(x,y|z\right)=\frac{p(x,y,z)}{p(z)}=\frac{p(x,y,z)}{p(x,z)}p\left(x|z\right)=p\left(y|x,z\right)p\left(x|z\right)$
- By the definition of conditional entropy

$$
\begin{aligned}
H(X,Y|Z) &= -\sum_{x\in X}\sum_{y\in Y}\sum_{z\in Z} p(x,y,z)\log p(x,y|z) \\
&= -\sum_{x\in X}\sum_{y\in Y}\sum_{z\in Z} p(x,y,z)\log p(y|x,z)\,p(x|z) \\
&= -\sum_{x\in X}\sum_{y\in Y}\sum_{z\in Z} p(x,y,z)\log p(y|x,z) - \sum_{x\in X}\sum_{y\in Y}\sum_{z\in Z} p(x,y,z)\log p(x|z) \\
&= -\sum_{x\in X}\sum_{y\in Y}\sum_{z\in Z} p(x,y,z)\log p(y|x,z) - \sum_{x\in X}\sum_{z\in Z} p(x,z)\log p(x|z) \\
&= H(Y|X,Z) + H(X|Z) \ \geq H(X|Z)
\end{aligned}
$$

- The equality holds when $H(Y|X,Z) = 0$

  (When Y is the function of X and Z)

b.

- By the chain rule of mutual information

  $I(X,Y;Z) = I(X;Z) + I(Y;Z|X) \geq I(X;Z)$
- The equality holds when $I(Y;Z|X) = 0$

  (When Y and Z are conditionally independent given X)

c.

- By the chain rule for entropy and definition of conditional mutual information

  $H(X,Y,Z) - H(X,Y) = H(Z|X,Y)$
  $= \quad H(Z|X) - I(Z;Y|X)$
  $\geq \quad H(Z|X)$
  $= \quad H(Z,X) - H(X)$
- The equality holds when $I(Z;Y|X) = 0$

  (When Z and Y are conditionally independent given X)

d.

- By the chain rule of mutual information

  $I(X_1, \ldots X_n; Y) = \sum_{i=1}^{n} I(X_i; Y|X_{i-1}, \ldots, X_1)$

  $\Rightarrow I(X,Y;Z) = I(X;Z) + I(Y;Z|X) = I(Y;Z) + I(X;Z|Y)$

  $\Rightarrow I(X;Z|Y) = I(X;Z) + I(Y;Z|X) - I(Y;Z)$
- The equality always hold.

11. [Prob. 2.32.]

a.

- The minimum probability of error estimator

$$\hat{X}(y) = \begin{cases} 1, & y = a \\ 2, & y = b \\ 3, & y = c \end{cases}$$

- The associated P_e

$$\begin{aligned} & P_e \\ = \quad & P\left(\hat{X}(Y) \neq X\right) \\ = \quad & P\left(\hat{X}(Y = a) \neq X\right) + P\left(\hat{X}(Y = b) \neq X\right) + P\left(\hat{X}(Y = c) \neq X\right) \\ = \quad & \tfrac{1}{6} + \tfrac{1}{6} + \tfrac{1}{6} \\ = \quad & \tfrac{1}{2} \end{aligned}$$

b.

- Fano's inequality

$$H(P_e) + P_e \log(|X| - 1) \geq H(X|Y)$$

- By weakened Fano's inequality

$$P_e \geq \frac{H(X|Y) - 1}{\log(|X|)}$$

$$\begin{aligned} & H(X|Y) \\ = \quad & \sum_{y=Y} P(Y = y) H(X|Y = y) \\ = \quad & -\sum_{y \in Y} P(Y = y) \sum_{x \in X} P(X = x|Y = y) \log P(X = x|Y = y) \\ = \quad & -\tfrac{1}{3}\left(\tfrac{1}{2}\log\tfrac{1}{2} + \tfrac{1}{4}\log\tfrac{1}{4} + \tfrac{1}{4}\log\tfrac{1}{4}\right) \\ & -\tfrac{1}{3}\left(\tfrac{1}{4}\log\tfrac{1}{4} + \tfrac{1}{2}\log\tfrac{1}{2} + \tfrac{1}{4}\log\tfrac{1}{4}\right) \\ & -\tfrac{1}{3}\left(\tfrac{1}{4}\log\tfrac{1}{4} + \tfrac{1}{4}\log\tfrac{1}{4} + \tfrac{1}{2}\log\tfrac{1}{2}\right) \\ = \quad & \tfrac{3}{2} \end{aligned}$$

$$\log|X| = \log 3$$

$$P_e \geq \frac{1.5 - 1}{\log 3} \approx 0.316$$

12. [Prob. 2.35.]

- Calculate $H(p)$ and $H(q)$

$$\begin{aligned} H(p) \quad &= \quad -\sum_{x \in X} p(x) \log p(x) \\ &= \quad -\tfrac{1}{2}\log\tfrac{1}{2} - \tfrac{1}{4}\log\tfrac{1}{4} - \tfrac{1}{4}\log\tfrac{1}{4} \\ &= \quad \tfrac{3}{2} \end{aligned}$$

$$\begin{aligned} H(q) \quad &= \quad -\sum_{x \in X} q(x) \log q(x) \\ &= \quad -\tfrac{1}{3}\log\tfrac{1}{3} - \tfrac{1}{3}\log\tfrac{1}{3} - \tfrac{1}{3}\log\tfrac{1}{3} \\ &= \quad \log 3 \end{aligned}$$

- Calculate $D\left(p \parallel q\right)$ and $D\left(q \parallel p\right)$

$$\begin{aligned} D\left(p \parallel q\right) &= \sum_{x \in X} p\left(x\right) \log \frac{p(x)}{q(x)} \\ &= \frac{1}{2} \log \frac{\frac{1}{2}}{\frac{1}{3}} + \frac{1}{4} \log \frac{\frac{1}{4}}{\frac{1}{3}} + \frac{1}{4} \log \frac{\frac{1}{4}}{\frac{1}{3}} \\ &= -\frac{3}{2} + \log 3 \\ D\left(q \parallel p\right) &= \sum_{x \in X} q\left(x\right) \log \frac{q(x)}{p(x)} \\ &= \frac{1}{3} \log \frac{\frac{1}{3}}{\frac{1}{2}} + \frac{1}{3} \log \frac{\frac{1}{3}}{\frac{1}{4}} + \frac{1}{3} \log \frac{\frac{1}{3}}{\frac{1}{4}} \\ &= \frac{5}{3} - \log 3 \end{aligned}$$

- Thus,

$$D\left(p \parallel q\right) \neq D\left(q \parallel p\right)$$