Fundamentals of Information Theory

Homework 1

Name: Zhou Shouchen
Student ID: 2021533042

Due 23:59 (CST), Oct. 6, 2024

Problem 1

2.1 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 expressions may be useful:

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

(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.

Solution

(a) Let $p = \frac{1}{2}, q = 1 - p = \frac{1}{2}$, then $X \sim FS(p)$. i.e. $P(X = k) = q^{k-1}p = \frac{1}{2^k}, k = 1, 2, \dots$
Then

$$\begin{aligned}
H(X) &= -\sum_{k=1}^{\infty} P(X=k) \log P(X=k) \\
&= -\sum_{k=1}^{\infty} \frac{1}{2^k} \log \frac{1}{2^k} \\
&= -\sum_{k=1}^{\infty} \frac{1}{2^k}(-k) \\
&= \sum_{k=1}^{\infty} k \left(\frac{1}{2}\right)^k \\
&= \sum_{k=0}^{\infty} k \left(\frac{1}{2}\right)^k \\
&= \frac{\frac{1}{2}}{\left(1-\frac{1}{2}\right)^2} \\
&= 2 \ \text{ bits}
\end{aligned}$$

(b) We can ask the questions iteratively. Let $S = \Phi$ initially.
Then for the i-th iteration, Let $S \leftarrow S \cup \{i\}$, $i = 1, 2, \dots$.
Then let T be the number of questions we need to ask to determine X.
Then $P(T = t) = \frac{1}{2^t}, t = 1, 2, \dots$
So

$$\mathbb{E}(T) = \sum_{t=1}^{\infty} t \cdot \frac{1}{2^t} = 2$$

And we can find that with this method, $\mathbb{E}(T) = H(X)$.
And there might(or might not) exist other optimal ways to ask the questions. If it exists, with the more efficient method, the expected number of questions required to determine X : T' has

$$\mathbb{E}(T') \leq H(X)$$

Problem 2

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

Solution

We have known that if $\hat{\boldsymbol{p}}$ is deterministic, then

$$H(\boldsymbol{p})_{\min}=H(\hat{\boldsymbol{p}})=0$$

Since

$$H(\boldsymbol{p})=\sum_{i=1}^{n} p_i \log \frac{1}{p_i} \geq 0$$

If and only if a single $p_i=1$, and others $p_i=0$.
So all these vectors are the deterministic vectors. i.e.

$$\hat{\boldsymbol{p}}_1=(1,0,\ldots,0), \hat{\boldsymbol{p}}_2=(0,1,0,\ldots,0), \ldots, \hat{\boldsymbol{p}}_n=(0,\ldots,0,1)$$

Problem 3

2.5 Zero conditional entropy. Show that if $H(Y \mid X) = 0$, 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 $p(x, y) > 0$].

Solution

Suppose that there exist a x_1 with $p(x_1) > 0$, there are more than 1 possible values of y with $p(x_1, y) > 0$. Suppose that they are

$$p(x_1, y_1), \cdots, p(x_1, y_k), k \geq 2$$

Since there are more than 1 possible values of y with $p(x, y) > 0$, we have

$$p(y_i|x_1) = \frac{p(x_1, y_i)}{p(x_1)} \in (0, 1), i = 1, \cdots, k$$

i.e. $\log \frac{1}{p(x_1, y_i)} > 0$.

So

$$\begin{aligned} H(Y|X) &= \sum_{x,y} p(x, y) \log \frac{1}{p(y|x)} \\ &\geq \sum_{i=1}^{k} p(x_1, y_i) \log \frac{1}{p(y_i|x_1)} \\ &> 0 \end{aligned}$$

Which is a contradiction with $H(Y|X) = 0$.

So for all x with $p(x) > 0$, there is only one possible value of y with $p(x, y) > 0$.

Problem 4

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

(b) $H(X \mid Y), H(Y \mid X)$.

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

(d) $H(Y) - H(Y \mid X)$.

(e) $I(X;Y)$.

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

Solution

From the table, we can get that

$$P(X=0) = \frac{2}{3}, P(X=1) = \frac{1}{3}$$
$$P(Y=0) = \frac{2}{3}, P(Y=1) = \frac{1}{3}$$
$$P(X=0|Y=0) = 1, P(X=1|Y=0) = 0$$
$$P(X=0|Y=1) = \frac{1}{2}, P(X=1|Y=1) = \frac{1}{2}$$
$$P(Y=0|X=0) = \frac{1}{2}, P(Y=1|X=0) = \frac{1}{2}$$
$$P(Y=0|X=1) = 0, P(Y=1|X=1) = 1$$

(a)

$$H(X) = H\left(\frac{2}{3}, \frac{1}{3}\right) = \log 3 - \frac{2}{3} = 0.918 \text{ bits}$$
$$H(Y) = H\left(\frac{1}{3}, \frac{2}{3}\right) = \log 3 - \frac{2}{3} = 0.918 \text{ bits}$$

(b)

$$\begin{aligned} H(X|Y) &= P(Y=0)H(X|Y=0) + P(Y=1)H(X|Y=1) \\ &= \frac{1}{3}H(1,0) + \frac{2}{3}H\left(\frac{1}{2}, \frac{1}{2}\right) \\ &= \frac{1}{3}0 + \frac{2}{3}\log 2 \\ &= \frac{2}{3} \\ &= 0.667 \text{ bits} \end{aligned}$$

Similarly, we can get that

$$\begin{aligned}
H(Y|X) &= P(X=0)H(Y|X=0) + P(X=1)H(Y|X=1) \\
&= \frac{2}{3}H\left(\frac{1}{2},\frac{1}{2}\right) + \frac{1}{3}H\left(0,1\right) \\
&= 0.667 \text{ bits}
\end{aligned}$$

(c)

$$\begin{aligned}
H(X,Y) &= H(X) + H(Y|X) \\
&= \left(\log 3 - \frac{2}{3}\right) + \frac{2}{3} \\
&= \log 3 \\
&= 1.585 \text{ bits}
\end{aligned}$$

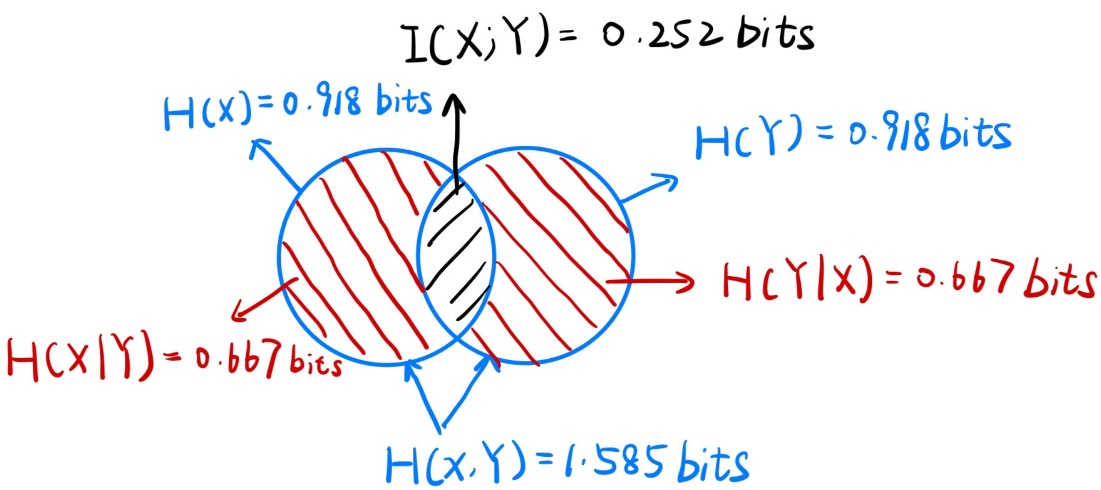
(d)

$$\begin{aligned}
H(Y) - H(Y|X) &= H(Y) - (P(X=0)H(Y|X=0) + P(X=1)H(Y|X=1)) \\
&= \left(\log 3 - \frac{2}{3}\right) - \frac{2}{3} \\
&= 0.252 \text{ bits}
\end{aligned}$$

(e)

$$I(X;Y) = H(Y) - H(Y|X) = 0.252 \text{ bits}$$

(f) The Venn diagram for the quantities are shown below.



Problem 5

2.14 Entropy of a sum. Let X and Y be random variables that take on values $x_1, x_2, \ldots, x_r$ and $y_1, y_2, \ldots, y_s$, respectively. Let $Z = X + Y$.

(a) Show that $H(Z \mid X) = H(Y \mid X)$. Argue that if X, Y are independent, then $H(Y) \leq H(Z)$ and $H(X) \leq H(Z)$. Thus, the addition of independent random variables adds uncertainty.

(b) Give an example of (necessarily dependent) random variables in which $H(X) > H(Z)$ and $H(Y) > H(Z)$.

(c) Under what conditions does $H(Z) = H(X) + H(Y)$?

Solution

(a) <1>.

$$\begin{aligned}
H(Z|X) &= \sum_{x,z} P(X=x, Z=z) \log \frac{1}{p(Z=z|X=x)} \\
&= \sum_{x,z} P(X=x, Y=z-x) \log \frac{1}{p(Y=z-x|X=x)} \\
&= \sum_{y,z} P(X=x, Y=y) \log \frac{1}{p(Y=y|X=x)} \quad \text{(Let } y = z-x\text{)} \\
&= H(Y|X)
\end{aligned}$$

<2>. Since $X \perp Y$, so $H(Y|X) = H(Y)$. And from <1>, we have known that $H(Z|X) = H(Y|X)$. So

$$H(Z) \geq H(Z|X) = H(Y|X) = H(Y)$$

Similarly, we can get $H(Z|Y) = H(X|Y), H(X|Y) = H(X)$, so

$$H(Z) \geq H(Z|Y) = H(X|Y) = H(X)$$

So above all, we have proved that

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

(b) We can let $Y = -X$, then $Z = 0$, which is deterministic. So $H(Z) = 0$.
So for any non-deterministic random variables X and Y, we have $H(X) > H(Z)$ and $H(Y) > H(Z)$.

(c) From the chain rule of entropy, we have

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

i.e.

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

When $H(X,Y|Z) = 0$ takes the equal. And since $H(X,Y) = H(X) + H(Y) - I(X;Y) \leq H(X) + H(Y)$, when $I(X;Y) = 0$ takes the equal. So

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

If and only if when $H(X,Y|Z) = 0, I(X;Y) = 0$, $H(Z) = H(X) + H(Y)$.
So above all, when $X \perp Y$ and (X,Y) is deterministic when given Z, we have $H(Z) = H(X) + H(Y)$.