CIE6020 Assignment 1

Due: 23:55, 25 Jan 2019

1. If the base of the logarithm is b, we denote the entropy as $H_b(X)$. Show that $H_b(X) = (\log_b a)H_a(X)$.

Solution:

Proof. Using $\log_b p = \log_b a \log_a p$, we can write

$$\begin{aligned} H_b(X) &= -\sum_x p(x)\log_b p(x) \\ &= -\sum_x p(x)\log_b a \log_a p(x) \\ &= -\log_b a \sum_x p(x)\log_a p(x) \\ &= -\log_b a H_a(X). \end{aligned}$$

□

2. *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:

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

   (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) The probability of $\{X = n\}$ is

$$P(X = n) = \left(\frac{1}{2}\right)^n.$$

Then the entropy in bits is

$$
\begin{aligned}
H(X) &= -\sum_{n=1}^{\infty}\left(\frac{1}{2}\right)^n \log\left(\frac{1}{2}\right)^n \\
&= -\sum_{n=1}^{\infty}\frac{1}{2^n}(-n) \\
&= \sum_{n=1}^{\infty}\frac{n}{2^n} \\
&= \frac{\frac{1}{2}}{(1-\frac{1}{2})^2} \\
&= 2.
\end{aligned}
$$

(b) The ith question is "Is X i?". The probability $P(X=i) = (\frac{1}{2})^i$ Then the expected number of questions required to determine X is

$$
\begin{aligned}
E(X) &= \sum_{i=1}^{\infty} iP(X=i) \\
&= \sum_{i=1}^{\infty}\frac{i}{2^i} \\
&= \frac{\frac{1}{2}}{(1-\frac{1}{2})^2} \\
&= 2.
\end{aligned}
$$

Thus $E(X) = H(X)$.

3. *Entropy of functions.* Let X be a random variable taking on a finite number of values. What is the (general) inequality relationship of $H(X)$ and $H(Y)$ if

   (a) $Y = 2^X$?

   (b) $Y = \cos(X)$?

Solution:

(a) We have $H(Y) - H(Y|X) = H(X) - H(X|Y)$. As Y is a function of X, $H(Y|X) = 0$. Therefore, $H(Y) \leq H(X)$.

We can also write $X = \log Y$, which implies $H(X) \leq H(Y)$. Therefore, $H(X) = H(Y)$.

(b) For $Y = cos X$, if Y and X is one-to-one mapping, as a result $H(Y) = H(X)$

In general, Y and X is not one-to-one mapping, we only have $H(Y) \leq H(X)$. For example, when X takes value $\{-1, +1\}$ with equal probability (i.e., $H(X) = 1$), we have $Y = \cos(1)$, a constant (i.e., $H(Y) = 0$).

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

Solution: If there exists $p_i = 1$, then $H(\mathbf{p}) = 0$ achieves the minimum.

5. Let X be a discrete random variable. Show that the entropy of a function of X is less than or equal to the entropy of X, i.e., $H(g(X)) \leq H(X)$. (Hint: apply chain rule on $H(X, g(X))$.)

Solution: Using chain rule,

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

Obviously,

$$\begin{aligned} H(g(X)|X) &= 0 \\ H(X|g(X)) &\geq 0. \end{aligned}$$

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

6. Let $p(x, y)$ be given by

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

Find by definition: (a) $H(X)$, $H(Y)$. (b) $H(X|Y)$, $H(Y|X)$. (c) $H(X, Y)$. (d) $I(X; Y)$. Check that $H(X) + H(Y|X) = H(Y) + H(X|Y)$, and $H(X) - H(X|Y) = H(Y) - H(Y|X)$. Draw a Venn diagram (information diagram) for the quantities in parts (a) through (d).

Solution: (a) From the given $p(x, y)$, we have $p_X(0) = \frac{1}{3}$ and $p_Y(0) = \frac{2}{3}$. Then

$$\begin{aligned} H(X) &= -\frac{1}{3}\log\frac{1}{3} - \frac{2}{3}\log\frac{2}{3} \\ &= \log 3 - \frac{2}{3}. \\ H(Y) &= \log 3 - \frac{2}{3}. \end{aligned}$$

(b) From $p(x,y)$, we can calculate the conditional distribution

$$p_{X|Y}(0|0) = \frac{1}{2}, \quad p_{X|Y}(0|1) = 0$$
$$p_{Y|X}(0|0) = 1, \quad p_{Y|X}(0|1) = \frac{1}{2}.$$

Hence,

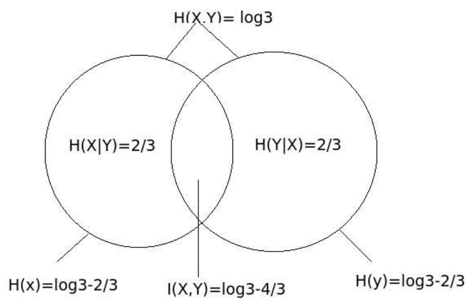
$$\begin{aligned}
H(X|Y) &= -\sum_{x,y} p(x,y) \log p(x|y) \\
&= -\frac{1}{3}\log\frac{1}{2} - 0\log 0 - \frac{1}{3}\log\frac{1}{2} - \frac{1}{3}\log 1 \\
&= \frac{2}{3}. \\
H(Y|X) &= \frac{2}{3}.
\end{aligned}$$

(c) $H(X,Y) = -3 \times \frac{1}{3}\log\frac{1}{3} = \log 3$.

(d)

$$\begin{aligned}
I(X;Y) &= \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \\
&= \frac{1}{3}\log\frac{1/3}{1/3 \times 2/3} + \frac{1}{3}\log\frac{1/3}{2/3 \times 2/3} + \frac{1}{3}\log\frac{1/3}{1/3 \times 2/3} \\
&= \log 3 - \frac{4}{3}.
\end{aligned}$$

Thus $H(X|Y) + H(Y) = \log 3 = H(Y|X) + H(X)$ and $H(X) - H(X|Y) = \log 3 - \frac{4}{3} = H(Y) - H(Y|X)$.
The Venn Diagram is shown as follows:



7. *Chain rule for conditional entropy.* Show that

$$H(X_1, X_2, ..., X_n|Y) = \sum_{i=1}^{n} H(X_i|X_1, ..., X_{i-1}, Y).$$

Solution:

$$\begin{aligned}
&H(X_1, X_2, ..., X_n|Y) \\
&= -\sum_{x_1,...,x_n,y} p(x_1, ..., x_n, y) \log p(x_1, ..., x_n|y) \\
&= -\sum_{x_1,...,x_n,y} p(x_1, ..., x_n, y) \log \prod_{i=1}^{n} p(x_i|x_1, ..., x_{i-1}, y) \\
&= -\sum_{i=1}^{n} \sum_{x_1,...,x_n,y} p(x_1, ..., x_n, y) \log p(x_i|x_1, ..., x_{i-1}, y) \\
&= \sum_{i=1}^{n} H(X_i|X_1, ..., X_{i-1}, Y).
\end{aligned}$$

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

   (a) Show that $H(Z|X) = H(Y|X)$. 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) Using the chain rule, we get

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

Since $Z = X + Y$, we have $H(Z|Y, X) = 0$,$H(Y|Z, X) = 0$. Thus, $H(Z|X) = H(Y|X)$.

If X and Y are independent, then $H(Y) = H(Y|X) = H(Z|X) \leq H(Z)$. Similarly, we can get $H(X) \leq H(Z)$.

(b) Consider $p(x, y)$ be given as follows:

$Y \backslash X$	-1	1
-1	0	$\frac{1}{2}$
1	$\frac{1}{2}$	0

In this case, $H(X) = H(Y) = 1$ and $H(Z) = 0$.

(c) Note that $H(Z) \leq H(X,Y) \leq H(X) + H(Y)$ as Z is a function of X and Y. Thus, to have $H(Z) = H(X) + H(Y)$, we require X and Y are independent.

We claim that under the condition that X and Y are independent, and $x_i + y_j \neq x_{i'} + y_{j'}$ whenever $i \neq i'$ or $j \neq j'$, $H(Z) = H(X) + H(Y)$. Under the above condition, we have that Z is a one-to-one function of (X, Y) so that $H(Z) = H(X,Y) = H(X) + H(Z)$.