Conditional entropy

If we are given two random variables X and Y, with their joint probability distribution $P_{(X,Y)}(x,y)$ (usually just denoted by $p(x,y)$), then we define the conditional entropy of X given Y as:

$$\boxed{H(X|Y) = \sum_{y} p(Y=y) \cdot H(X|Y=y)}$$

$$= \sum_{y} p(Y=y) \cdot \left(\sum_{x} p(X=x|Y=y) \cdot \log_2 \frac{1}{P(X=x|Y=y)} \right)$$

$$= \sum_{x,y} P(X=x, Y=y) \cdot \log_2 \frac{1}{P(X=x|Y=y)}$$

P1 Let X and Y be two random variables. Prove the chain rule for entropy which states that:

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

Sol: From the definiton of conditional probability we know that: $p(X=x, Y=y) = p(Y=y) \cdot p(X=x|Y=y)$,

and taking $\log_2$ of the equation we get:

$$\log_2 P(X=x, Y=y) = \log_2 P(Y=y) + \log_2 P(X=x|Y=y).$$

Now,

$$H(X,Y) = \sum_{x,y} P(X=x, Y=y) \cdot \log_2 \frac{1}{P(X=x, Y=y)} =$$

$$= \sum_{x,y} P(X=x, Y=y) \cdot \log_2 \frac{1}{P(Y=y)} + \underbrace{\sum_{x,y} P(X=x, Y=y) \cdot \log_2 \frac{1}{P(X=x|Y=y)}}_{=H(X|Y)}$$

$$= \sum_{y} \log_2 \frac{1}{P(Y=y)} \cdot \underbrace{\left(\sum_{x} P(X=x, Y=y) \right)}_{=P(Y=y)} + H(X|Y)$$

$$= \sum_{y} P(Y=y) \cdot \log_2 \frac{1}{P(Y=y)} + H(X|Y) = H(Y) + H(X|Y).$$

P2. The input source to a noisy communication channel is a random variable X over the four symbols a, b, c, d. The output from the channel is a r.v. Y over the same sym. The joint distribution of these two random variables is:

	$x=a$	$x=b$	$x=c$	$x=d$
$y=a$	$1/8$	$1/16$	$1/16$	$1/4$
$y=b$	$1/16$	$1/8$	$1/16$	0
$y=c$	$1/32$	$1/32$	$1/16$	0
$y=d$	$1/32$	$1/32$	$1/16$	0

a) Write the distribution of X and compute $H(X)$
b) Do the same for Y.
c) Find $H(X,Y)$!
d) Find $H(X|Y)$ and $H(Y|X)$.

Sol: Questions a)-c) are P4 in Tutorials 3.

Now, knowing a)-c) we can solve d) using the chain rule for entropy:

$$H(X|Y) = H(X,Y) - H(Y) = 3.375 - 1.75 = 1.625$$

and

$$H(Y|X) = H(X,Y) - H(X) = 3.375 - 2 = 1.375,$$

P3 Let X and Y be two random variables. Prove that

$$H(X|Y) \leq H(X).$$

Sol: We will prove the inequality above using the 'Gibbs' inequality, i.e. P7 in Tutorials 3, to estimate $H(X) - H(X|Y)$. Let $\mathcal{A}_X = \{x_1, \dots, x_n\}$ be all the values that X takes and $\mathcal{A}_Y = \{y_1, \dots, y_m\}$ be all the values that Y takes.

$$H(X) - H(X|Y) = \sum_{x \in \mathcal{A}_X} P(X=x) \log_2 \frac{1}{P(X=x)} - \sum_{\substack{y \in \mathcal{A}_Y \\ x \in \mathcal{A}_X}} P(X=x, Y=y) \cdot \log_2 \frac{1}{P(X=x|Y=y)}$$

Notice that $P(X=x) = \sum_{y \in \mathcal{A}_Y} P(X=x, Y=y)$, so we have

$$H(X)-H(X|Y)=\sum_{\substack{x\in\mathcal{A}_X\\ y\in\mathcal{A}_Y}} P(X=x,Y=y)\cdot\log_2\frac{P(X=x|Y=y)}{P(X=x)} = \Big| \text{using def of conditional probability}$$

$$=\sum_{\substack{x\in\mathcal{A}_X\\ y\in\mathcal{A}_Y}} P(X=x,Y=y)\cdot\log_2\frac{P(X=x,Y=y)}{P(X=x)\cdot P(Y=y)}$$

Now, we can think of both $\overbrace{P(X=x,Y=y)}^{\text{denote it } p(x,y)}$ and $\underbrace{P(X=x)\cdot P(Y=y)}_{\text{denote it } q(x,y)}$ as if they are probability functions defined on $\mathcal{A}_X\times\mathcal{A}_Y$, so:

$$H(X)-H(X|Y)=\sum_{x,y\in\mathcal{A}_X\times\mathcal{A}_Y} p(x,y)\cdot\log_2\frac{p(x,y)}{q(x,y)} = D_{KL}(p(x,y)\|q(x,y))$$

and by the Gibbs's inequality $D_{KL}(p(x,y)\|q(x,y))\geq 0$, so

$H(X)-H(X|Y)\geq 0$, i.e. $H(X|Y)\leq H(X)$.

P4 Show that if we have two random variables X and Y such that $H(Y|X)=0$, then Y is a function of X. (i.e. for every value x s.t. $P(X=x)>0$ there exists just one y s.t. $P(X=x,Y=y)>0$.)

Sol: From the definition of conditional entropy:

$H(Y|X)=\sum_x p(X=x)\cdot H(Y|X=x)$, and since $p(X=x)\geq 0$

and $H(Y|X=x)\geq 0$ for all x, if $p(X=x)>0$, then it has to be that $H(Y \mid X=x) = 0$, and so:

$$\sum_y P(Y=y \mid X=x) \cdot \log_2 \frac{1}{P(Y=y \mid X=x)} = 0 .$$

Assume that there are two y_1 and y_2 such that $P(X=x, Y=y_1) > 0$ and $P(X=x, Y=y_2) > 0$, then:

$P(X=x) \cdot P(Y=y_1 \mid X=x) > 0$ and $P(X=x) \cdot P(Y=y_2 \mid X=x) > 0$, so $P(Y=y_1 \mid X=x) > 0$ and $P(Y=y_2 \mid X=x) > 0$, and it also has to be the case that $P(Y=y_1 \mid X=x) < 1$ and $P(Y=y_2 \mid X=x) < 1$, hence:

$$P(Y=y_1 \mid X=x) \cdot \log_2 \frac{1}{P(Y=y_1 \mid X=x)} > 0 \text{ and } P(Y=y_2 \mid X=x) \cdot \log_2 \frac{1}{P(Y=y_2 \mid X=x)} > 0$$

But since $P(Y=y \mid X=x) \cdot \log_2 \frac{1}{P(Y=y \mid X=x)} \geq 0$ for all x and y, we have:

$$\sum_y P(Y=y \mid X=x) \cdot \log \frac{1}{P(Y=y \mid X=x)} \geq (\Delta)$$

$$(\Delta) \geq P(Y=y_1 \mid X=x) \cdot \log_2 \frac{1}{P(Y=y_1 \mid X=x)} + P(Y=y_2 \mid X=x) \cdot \log_2 \frac{1}{P(Y=y_2 \mid X=x)} > 0,$$

but we know that

$$\sum_y P(Y=y \mid X=x) \cdot \log_2 \frac{1}{P(Y=y \mid X=x)} = 0,$$

and that is a contradiction. Hence if $P(X=x) > 0$, there is just an y s.t. $P(X=x \mid Y=y) > 0$.

The mutual information between two random variables X and Y is defined as:

$$I(X;Y) = H(X) - H(X|Y)$$

$$= \sum_{x,y} P(X=x, Y=y) \cdot \log_2 \frac{P(X=x, Y=y)}{P(X=x) \cdot P(Y=y)}$$

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

$$= H(X) + H(Y) - H(X,Y)$$

P5 Let the joint probability mass function $p(x,y)$ for two random variables X and Y be given by:

	$Y=0$	$Y=1$
$X=0$	$1/3$	$1/3$
$X=1$	0	$1/3$

Find: a) $H(X)$ and $H(Y)$;
b) $H(X|Y)$ and $H(Y|X)$;
c) $H(X,Y)$;
d) $I(X;Y)$.

Sol: a) First we need to find the probability mass functions of X and of Y. We do that as follows:

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

Similarly for Y we have:

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

So $H(X) = \frac{1}{3}\log_2 3 + \frac{2}{3}\log_2\frac{3}{2} = \log_2 3 - \frac{2}{3} = 0{,}918$
and
$$H(Y) = \frac{2}{3}\log_2\frac{3}{2} + \frac{1}{3}\log 3 = 0{,}918.$$

b) To find $H(X|Y)$ we need the probability mass functions $P(X|Y=0)$ and $P(X|Y=1)$, and we find them as follows:

$$P(X=0|Y=0) = \frac{P(X=0, Y=0)}{P(Y=0)} = \frac{1/3}{1/3} = 1, \text{ and}$$
$$P(X=1|Y=0) = \frac{P(X=1, Y=0)}{P(Y=0)} = \frac{0}{1/3} = 0; \text{ so}$$

$H(X|Y=0) = 1\log\frac{1}{1} + 0\log\frac{1}{0} = 0$. Similarly

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

$H(X|Y=1) = \log_2 2 = 1$, hence:

$$H(X|Y) = P(Y=0)\cdot H(X|Y=0) + P(Y=1)\cdot H(X|Y=1) =$$
$$= \frac{1}{3}\cdot 0 + \frac{2}{3}\cdot 1 = \frac{2}{3}.$$

For $H(Y|X)$ we follow the same steps:

$$P(Y=0|X=0) = \frac{P(X=0, Y=0)}{P(X=0)} = \frac{1/3}{2/3} = \frac{1}{2}; \text{ and}$$
$$P(Y=1|X=0) = \frac{P(X=0, Y=1)}{P(X=0)} = \frac{1/3}{2/3} = \frac{1}{2}, \text{ so:}$$

$H(Y|X=0) = \frac{1}{2}\log_2 2 + \frac{1}{2}\log_2 2 = 1$; similarly

$P(Y=0|X=1) = \frac{0}{\frac{1}{3}} = 0$ and $P(Y=1|X=1) = \frac{\frac{1}{3}}{\frac{1}{3}} = 1$; hence

$H(Y|X=1) = 0\log_2\frac{1}{0} + 1\log_2\frac{1}{1} = 0$. Finally

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

c) From the definition of entropy:

$$H(X,Y) = \sum_{x,y} p(x,y)\log_2\frac{1}{p(x,y)} =$$

$$= \frac{1}{3}\log_2 3 + \frac{1}{3}\log_2 3 + 0\cdot\log_2\frac{1}{0} + \frac{1}{3}\log_2 3 = \log_2 3 = 1.585$$

d) From the definition of mutual information:

$$I(X;Y) = \sum_{x,y} p(x,y)\cdot\log\frac{p(x,y)}{p(x)\cdot p(y)} =$$

$$= \frac{1}{3}\cdot\log_2\frac{\frac{1}{3}}{\frac{2}{3}\cdot\frac{1}{3}} + \frac{1}{3}\log_2\frac{\frac{1}{3}}{\frac{2}{3}\cdot\frac{2}{3}} + 0\cdot\log_2\frac{0}{\frac{1}{3}\cdot\frac{1}{3}} + \frac{1}{3}\log_2\frac{\frac{1}{3}}{\frac{1}{3}\cdot\frac{2}{3}}$$

$$= \frac{2}{3}\log_2\frac{3}{2} + \frac{1}{3}\log_2\frac{3}{4} = \log_2 3 - \frac{2}{3} - \frac{1}{3}\cdot 2 = \log_2 3 - \frac{4}{3}$$

$$= 0.2516.$$

P6 We want to find out if a treasure lies behind doors $X = 1, 2$ or 3. We know that the probability that it is behind the door number 1 is $1/2$, behind 2 is $1/4$, and behind 3 is $1/4$. We are allowed to take a subset S of $\{1,2,3\}$ and ask if the treasure is behind the doors in S. You will receive the answer, and with that knowledge you will need to guess where the treasure is. Which S would you choose to maximize (on average) your chances of finding the treasure?

Sol: Let Y be the random variable which is 1 if the treasure is behind the doors in S, and 0 if it is not. To maximize our chances of winning, we want to maximize mutual information between X and Y. Notice that if we know where the treasure is, we know what the value of Y is going to be, i.e. if we know X, we know Y, hence $H(Y|X) = 0$. Using this and the definition of mutual information we have:

$$I(X;Y) = H(Y) - \underbrace{H(Y|X)}_{=0} = H(Y).$$

Hence to maximize $I(X;Y)$ we need to maximize $H(Y)$. Y is a binary random variable (i.e. it takes only two values), so $I(X;Y) = H(Y) \leq \log_2 2 = 1$ (tutorials 3, P6). If we set $S = \{1\}$, that is we ask if the treasure is behind the door 1, then:

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

so in this case $H(Y) = 1$, so $I(X;Y) = 1$, and we have proved that this is the maximum possible value, so this question is going to maximize our chances of finding the treasure. ($S = \{2,3\}$ would also work.)

P7 Let X_1 and X_2 be identically distributed random variables. Let:

$$\rho = 1 - \frac{H(X_2|X_1)}{H(X_1)}$$

a) Show that $\rho = \dfrac{I(X_1;X_2)}{H(X_1)}$;

b) Show that $0 \leq \rho \leq 1$;

c) When is $\rho = 0$?

d) When is $\rho = 1$?

Sol: Since X_1 and X_2 are identically distributed we have $H(X_1) = H(X_2)$, so:

a)

$$\rho = 1 - \frac{H(X_2|X_1)}{H(X_1)} = \frac{H(X_1) - H(X_2|X_1)}{H(X_1)} = \frac{H(X_2) - H(X_2|X_1)}{H(X_1)} = \frac{I(X;Y)}{H(X_1)}$$

b) Since entropy is non-negative $\rho \leq 1$, and since mutually information is also non-negative, we have from a) that $\rho \geq 0$ as well, so $0 \leq \rho \leq 1$.

c) $\rho = 0$ if and only if $I(X_1;X_2) = 0$ (follows from a) and $I(X_1;X_2) = 0$ if and only if X_1 and X_2 are independent.

d) $\rho = 1$ if and only if $H(X_2|X_1) = 0$, and from
P4 $H(X_2|X_1) = 0$ if and only if X_2 is a function of X_1. Since X_1 and X_2 are identically distributed this means also that X_1 is a function of X_2, hence we conclude that X_1 and X_2 have a one-to-one relationship

---

P8 a) Consider a fair coin flip. What is the mutual information between the top and bottom sides of the coin?
b) A six-sided fair die is rolled. What is the mutual information between the top side and the front face (the side most facing you)?

Sol: a) Let X be the random variable that represents top of the coin, and Y be the bottom. The top side is completely determined by the bottom side of the coin and vice versa, and this implies that $H(X|Y) = 0$, so:

$$I(X;Y) = H(X) - H(X|Y) = H(X) = \tfrac{1}{2}\log_2 2 + \tfrac{1}{2}\log_2 2 = 1.$$

b) In this case let X be the top side, and let Y be the front side. If we know the top side, we know the bottom side as well, so we are left with four options for the front side, and all the 4 options are equally likely since the die is fair so $H(Y|X) = \log_2 4 = 2$, and so:

$$I(X;Y) = H(Y) - H(Y|X) = \log_2 6 - 2 = \log_2 3 - 1$$
$$= 0.585.$$

P9 Prove that if X and Y are independent random variables, then

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

Sol: If X and Y are independent, then

$$P(X=x, Y=y) = P(X=x) \cdot P(Y=y)$$

for all possible values x of X and y of Y, so we have:

$$I(X;Y) = \sum_{x,y} P(X=x, Y=y) \log_2 \frac{P(X=x, Y=y)}{P(X=x) \cdot P(Y=y)} =$$

$$= \sum_{x,y} P(X=x, Y=y) \cdot \log_2 \frac{\cancel{P(X=x)} \cdot \cancel{P(Y=y)}}{\cancel{P(X=x)} \cdot \cancel{P(Y=y)}} = \sum_{x,y} P(X=x, Y=y) \cdot \underbrace{\log_2 1}_{=0} = 0$$

P10 Consider the Monty Hall problem: There are three doors; behind one of them is a car and behind the others goats. You pick one of the doors, and then the host opens up another one showing you a goat, and asks you if you want to switch. Using entropy show that it is not the same whether you switch or not.

Sol: Without loss of generality suppose we chose the door number 1. Let X be the random variable taking values in $\{1, 2, 3\}$ such that it is 1 if the car is behind door 1, 2 if it is behind door 2, and 3 if it is behind door 3. Let Y be the random variable representing the number of the door the host opened. Now to understand whether it is the same if we switch or not after the host opened the door, we need to study $H(X|Y)$.

Using the chain rule for entropy $H(X,Y) = H(Y) + H(X|Y)$, and so $H(X|Y) = H(X,Y) - H(Y)$. Now, considering all the options over the whole sample space, the host is equally likely to open the doors 2 or 3, so $H(Y) = \frac{1}{2}\log_2 2 + \frac{1}{2}\log_2 2 = 1$. Our sample space is:

$\Omega = \{(1,2),(1,3),(2,2),(2,3),(3,2),(3,3)\}$ with the following probabilities (here the first coordinate represents where the car is and the second which door the host opened):

$P(1,2) = P(1,3) = \frac{1}{2}\cdot\frac{1}{3}$; $P(2,2) = 0$, $P(2,3) = \frac{1}{3}$,

$P(3,2) = \frac{1}{3}$ and $P(3,3) = 0$, and so:

$$H(X,Y) = 2\cdot\frac{1}{3}\log_2\frac{1}{\frac{1}{3}} + 2\cdot\frac{1}{6}\log_2\frac{1}{\frac{1}{6}} =$$

$$= \frac{2}{3}\log_2 3 + \frac{1}{3}\log_2 3 + \frac{1}{3}\log_2 2 = \log_2 3 + \frac{1}{3}.$$

Finally:

$H(X|Y) = H(X,Y) - H(Y) = \log_2 3 + \frac{1}{3} - 1 = \log_2 3 - \frac{2}{3}$ $= 0.919$. Now, since $H(X|Y) = 0.919$, we conclude that it is not the same whether we switch or not after the host opened the door, since if it was the same, then $H(X|Y)$ would be $\frac{1}{2}\log_2 2 + \frac{1}{2}\log_2 2 = 1$.