MA3466 Tutorial Sheet 3, outline solutions[1]

22 March 2010

1. (C&T 2.2) Entropy of functions. Let X be a random variable taking on a finite number of values. What is the general inequality relating $H(X)$ and $H(Y)$ if

   (a) $Y = 2^X$

   (b) $Y = \cos X$

   *Solution:*The key point here is that if $Y = g(X)$ then X determines Y but the converse may or may not be true. This means that $H(Y|X) = 0$ since Y is not random is the outcome of X is known. However $H(X|Y)$ may not be zero if the function is not invertible.

   In the first example it is true because the function is one-to-one. First, because $Y = 2^X$

$$p(Y = 2^x|X = x) = 1 \tag{1}$$

   and so $H(X,Y) = H(X) + H(Y|X) = H(X)$; $H(Y|X) = 0$ because

$$H(Y|X) = \sum p(x)H(Y|X = x) \tag{2}$$

   and all the terms in the sum in

$$H(Y|X = x) = \sum p(y|x)\log p(y|x) \tag{3}$$

   are zero, either because $y = 2^x$ so that $p(y|x) = 1$ and its log is zero, or $y \neq 2^x$ and $x \log x$ goes to zero as x goes to zero. The converse is also true, $X = \log X$ and so

$$H(X,Y) = H(Y) + H(X|Y) = H(Y) \tag{4}$$

   and so $H(X) = H(Y)$. The situation is different for $Y = \cos X$, this is not, in general, an invertable function. Hence, it is still true that $H(Y|X) = 0$ because X still determines Y; however $H(X|Y)$ may not be zero, there may be y such that the set $cos^{-1}y = \{x \in \mathcal{X}|cosx = y\}$ may have more than one element and so knowing $Y = y$ tells you $x \in \cos^{-1} y$, but it doesn't tell you what X is. Hence

$$H(X) = H(X,Y) = H(Y) + H(X|Y) \geq H(Y) \tag{5}$$

   where we know $H(X|Y) \geq 0$ because entropy is always positive.

[1]Conor Houghton, houghton@maths.tcd.ie, see also http://www.maths.tcd.ie/~houghton/MA3466

To illustrate this further, lets consider two examples, first, if $\mathcal{X} = \{0, \pi\}$ then the function is invertible, $\mathcal{Y} = \{1, -1\}$ and if $Y = 1$, $X = 0$, if $Y = -1$, $X = \pi$. Here the inequality will be sharp. On the other hand, say $\mathcal{X} = \{0, \pi, 2\pi\}$ and

$$p_X(0) = p_X(\pi) = p_X(2\pi) = 1/3 \tag{6}$$

now $\mathcal{Y} = \{1, -1\}$ with $p_Y(1) = 2/3$ and $p_Y(-1) = 1/3$. Hence

$$\begin{aligned} H(X) &= \log 3 \\ H(Y) &= \log 3 - \frac{2}{3} \end{aligned} \tag{7}$$

and $H(Y) < H(X)$; the point being that $H(X|Y) \neq 0$, if $Y = 1$, X could be zero or 2π with equal probability so $H(X|Y = 1) = 1$ and

$$H(X|Y) = \frac{2}{3}H(X|Y = 1) + \frac{1}{3}H(X|Y = -1) = \frac{2}{3} \tag{8}$$

2. (C&T 2.4) Entropy of functions of a random variable. 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 by justifying the following steps

$$\begin{aligned} H(X, g(X)) &= H(X) + H(g(X)|X) \\ &= H(X), \\ H(X, g(X)) &= H(g(X)) + H(X|g(X)) \geq H(g(X)) \end{aligned} \tag{9}$$

and hence $H(g(X)) \leq H(X)$.

*Solution:*The key point here is that if $Y = g(X)$ then X determines Y but the converse may or may not be true. In the first example it is true because the function is one-to-one. First, because $Y = 2^X$

$$p(Y = 2^x|X = x) = 1 \tag{10}$$

and so $H(X, Y) = H(X) + H(Y|X) = H(X)$; $H(Y|X) = 0$ because

$$H(Y|X) = \sum p(x)H(Y|X = x) \tag{11}$$

and all the terms in the sum in

$$H(Y|X = x) = \sum p(y|x)\log p(y|x) \tag{12}$$

are zero, either because $y = 2^x$ so that $p(y|x) = 1$ and its log is zero, or $y \neq 2^x$ and $x \log x$ goes to zero as x goes to zero. The converse is also true, $X = \log X$ and so

$$H(X, Y) = H(Y) + H(X|Y) = H(Y) \tag{13}$$

and so $H(X) = H(Y)$. The situation is different for $Y = \cos X$, this is not, in general, an invertable function. Hence, it is still true that $H(Y|X) = 0$ because X still determines Y; however $H(X|Y)$ may not be zero, there may be y such that the set $cos^{-1}y = \{x \in \mathcal{X} | cosx = y\}$ may have more than one element and so knowing $Y = y$ tells you $x \in \cos^{-1} y$, but it doesn't tell you what X is. Hence

$$H(X) = H(X, Y) = H(Y) + H(X|Y) \geq H(Y) \tag{14}$$

where we know $H(X|Y) \geq 0$ because entropy is always positive.

To illustrate this further, lets consider two examples, first, if $\mathcal{X} = \{0, \pi\}$ then the function is invertible, $\mathcal{Y} = \{1, -1\}$ and if $Y = 1$, $X = 0$, if $Y = -1$, $X = \pi$. Here the inequality will be sharp. On the other hand, say $\mathcal{X} = \{0, \pi, 2\pi\}$ and

$$p_X(0) = p_X(\pi) = p_X(2\pi) = 1/3 \tag{15}$$

now $\mathcal{Y} = \{1, -1\}$ with $p_Y(1) = 2/3$ and $p_Y(-1) = 1/3$. Hence

$$\begin{aligned} H(X) &= \log 3 \\ H(Y) &= \log 3 - \frac{2}{3} \end{aligned} \tag{16}$$

and $H(Y) < H(X)$; the point being that $H(X|Y) \neq 0$, if $Y = 1$, X could be zero or 2π with equal probability so $H(X|Y = 1) = 1$ and

$$H(X|Y) = \frac{2}{3} H(X|Y = 1) + \frac{1}{3} H(X|Y = -1) = \frac{2}{3} \tag{17}$$

3. (C&T 2.8) Drawing with and without replacement. An urn contains r red, w white and b black balls. Which has higher entropy, drawing $k \geq 2$ balls from the urn with replacement or without replacement?

*Solution:*So the answer to this question relies on the fact that the probability distribution for the n drawing is the same irrespive of whether there is replacement or not. Lets use X to denote drawing from an urn with r red balls, w white balls and b black balls, so, with $n = b + r + w$

$$\begin{aligned} p_X(c_r) &= \frac{r}{n} \\ p_X(c_w) &= \frac{w}{n} \\ p_X(c_b) &= \frac{b}{n} \end{aligned} \tag{18}$$

whre c_r is red and so on. Now, if X_i is the ith drawing with replacement, then clearly the X_i are independent and $p_{X_i}(x) = p_X(x)$ for $x \in \mathcal{X} = \{c_r, c_b, c_w\}$.

Now, let Y_i be the ith drawing with replacement: although the Y_i are not independent $p_{Y_i}(x) = p_X(x)$ for $x \in \mathcal{X}$. To see this, note $Y_1 = X$ and assume it is true for Y_i and consider Y_{i+1}:

$$p_{Y_{i+1}}(c_r) = p_{(Y_{i+1}, Y_i)}(c_r, c_r) + p_{(Y_{i+1}, Y_i)}(c_r, c_w) + p_{(Y_{i+1}, Y_i)}(c_r, c_b)$$

$$
\begin{aligned}
&= p_{Y_{i+1}|Y_i}(c_r|c_r)p_{Y_i}(c_r) \\
&\quad + p_{Y_{i+1}|Y_i}(c_r|c_w)p_{Y_i}(c_w) + p_{Y_{i+1}|Y_i}(c_r|c_b)p_{Y_i}(c_b) \\
&= \frac{r-1}{n-1}\frac{r}{n} + \frac{r}{n-1}\frac{w}{n} + \frac{r}{n-1}\frac{b}{n} = \frac{r}{n} = p_X(c_r) \qquad (19)
\end{aligned}
$$

This means, using the chain rule and the conditioning theorem

$$
\begin{aligned}
H(Y_1, Y_2, \ldots, Y_n) &= H(Y_1) + H(Y_2|Y_1) + H(Y_3|Y_2, Y_1) + \ldots + H(Y_n|Y_{n-1}, \ldots, Y_1) \\
&\leq \sum H(Y_i) = nH(X) = H(X_1, X_2, \ldots, X_n) \qquad (20)
\end{aligned}
$$

with equality if and only if the Y_i were independent which they aren't, hence

$$H(Y_1, Y_2, \ldots, Y_n) < H(X_1, X_2, \ldots, X_n) \qquad (21)$$

4. (C&T 2.14) Enropy 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 and Y are independent then $H(Y) \leq H(Z)$ and $H(X) \leq H(Z)$. Thus the addition of independent random variables add uncertainy.

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

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

   *Solution:*So, given X, Y determines Z and visa versa, so $H(Z|X) = H(Y|X)$. Now, we know that

$$H(Y|X) = H(Z|X) \leq H(Z) \qquad (22)$$

   but, if X and Y are independent, $H(Y|X) = H(Y)$, so $H(Y) \leq H(Z)$; $H(X) \leq H(Z)$ follow by a similar arguement. Thus, if we want $H(X) > H(Z)$, we need X and Y dependent. In fact, we want X and Y to be dependent in such a way that adding them gives something less uncertain; as an example, let $Y = -X$ so $Z = 0$ always and so, $H(Z) = 0$ and is less than $H(X) = H(Y)$ for any non-trivial choice of X. Finally,

$$H(X + Y) = H(X, Y) \qquad (23)$$

   if the addition is invertible, that is, if there are unique X and Y for any $X + Y$; this would happen, for example, if $\mathcal{X} = \{1, 2\}$ and $\mathcal{Y} = \{1, 3\}$ since the possible values of the sum are 2, 3, 4 and 5 and each correponds to a different choice of X and Y; however, if $\mathcal{X} = \{1, 2\}$ and $\mathcal{Y} = \{1, 2\}$ then $X = 1$, $Y = 2$ and $X = 2$, $Y = 1$ both give $X + Y = 3$. Now

$$H(X, Y) = H(X) + H(Y|X) \qquad (24)$$

   and $H(Y|X) + H(Y)$ if X and Y are independent.