MA3466 Tutorial Sheet 2, outline solutions[1]

16 Febuary 2010

1. (C&T 2.6) Find joint random variables X, Y and Z such that

   (a) $I(X;Y|Z) < I(X;Y)$

   (b) $I(X;Y|Z) > I(X;Y)$

   Solution: So the idea here is to show that there is no inequality between the mutual information and the conditional mutual information. To provide an example where $I(X;Y|Z) < I(X;Y)$ lets start of by making $I(X;Y)$ as big as possible. Well, we know

$$I(X;Y) = H(X) - H(X|Y) \quad (1)$$

   so for fixed X we can minimize this by making $H(X|Y)$ zero. This happens if $X = Y$, so that is a start: $I(X;X) = H(X)$. Now, we have

$$I(X;Y|Z) = H(X|Z) - H(X|Y,Z) \quad (2)$$

   but if $Y = X$, the value of x is still determined by the value of y, irrespective of what z is; hence

$$I(X;X|Z) = H(X|Z) \quad (3)$$

   and we know $H(X|Z) \leq H(X)$ so $I(X;X|Z) \leq I(X;X)$, a concrete example where $I(X;X|Z) < I(X;X)$ is provided by C& T example 2.6.1; a more extreme example is given by $Z = X$, in which case $I(X;X|X) = 0$.

   To go the other way and find an example where $I(X;Y|Z) > I(X;Y)$ lets start by making $I(X;Y)$ as small as possible; this happens when X and Y are independent: $I(X;Y) = 0$ for independant distributions. Now, knowledge if there is a third variable Z which is not independent of X and Y then the conditional distributions are not in general independent: $p(x|z)p(y|z) \neq p(x,y|z)$ in general. Let us choose $Z = X + Y$ and $\mathcal{X} = \mathcal{Y} = \{0,1\}$ with $p_X(0) = p_X(1) = 1/2$ and $p_Y(0) = p_Y(1) = 1/2$ and, of course, $p(x,y) = p(x)p(y)$. Now $p_{X|Z}(0|0) = 1$, since $z = 0$ only if $x = y = 0$. However there are two ways z can be one, $x = 1$, $y = 0$ and $x = 0$, $y = 1$, so $p_{X|Z}(0|1) = p_{X|Z}(1|1) = 1/2$. Finally $p_{X|Z}(1|1) = 1$. Now

$$I(X;Y|Z) = H(X|Z) - H(X|Z,Y) \quad (4)$$

   but the value of X is determined by the values of Y and Z; $x = z-y$, so $H(X|Z,Y) = 0$ and

$$I(X;Y|Z) = H(X|Z) - H(X|Z,Y) = H(X|Z) = \sum p(z)H(X|Z=z) \quad (5)$$

   Now, from the conditional distributions above $H(X|0) = H(X|2) = 0$ and $H(X|1) = 1$. Finally $p_Z(1) = 1/2$ so $I(X;Y|Z) = 1/2$.

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

2. (C&T 2.7) Suppose that one has n coins, among which there may or may not be one counterfeit coin. If there is a conterfeit coin it will weight either less or more than the other coins. The coins are weighed using a balance.

   (a) Find an upper bound on the number of coins n so that k weighings will find the couinterfeit coin, if any, and correctly declare it to be heavier or lighter.

   (b) What is the coin-weighing strategy for $k = 3$ weighings and 12 coins.

   Solution: So given that one of n coins is counterfeit; there are $2n$ possible configurations, numbering the coins one to n, each possibility is either of the form the ith coin is heavier, or the ith coin is lighter. Thus, assuming all possibilities are equally likely, the random variable X giving the identity and type of the bent coin has entropy $H(X) = \log 2n$. What about weighing, Y, well each weighing involves taking two groups of coins and balancing them and this has three possible outcomes: left heavier, right heavier or balanced. Obivously, depending on what we have already worked out about the coins from previous weighings, these possibilities have different outcomes, for example, at the start, given that one coin is counterfeit, weighing $n/2$ coins against $n/2$ coins can't give balanced and the entropy for this measurement will be one bit. However, we know that $H(Y) < \log 3$; the most uncertain measurement is the one where all possibilities are equally likely.

   Now, imagine drawing up a weighing statedgy, you are going to do k weighings Y_1, Y_2 to Y_k. The outcome of a weighing is determined by the value of x, the idenity and type of the bent coin, so $H(Y_1, Y_2, \ldots, Y_k|X) = 0$. We have

$$H(X)+H(Y_1,Y_2,\ldots,Y_k|X) = H(X,Y_2,\ldots,Y_k) = H(Y_1,Y_2,\ldots,Y_k)+H(X|Y_1,Y_2,\ldots,Y_k) \quad (6)$$

   If we have a statedgy that locates and types the counterfeit, there should be no uncertainty in X given the Y_i so $H(X|Y_1, Y_2, \ldots, Y_k) = 0$. So, if we are able to find and type the counterfeit

$$H(X) = H(Y_1, Y_2, \ldots, Y_k) \quad (7)$$

   but, from the independence theorem and the bound above

$$H(X) = H(Y_1, Y_2, \ldots, Y_k) \leq H(Y_i) \leq k \log 3 \quad (8)$$

   and hence

$$n \leq 3^k/2 \quad (9)$$

   Hence, if it is possible to identify and type the coin in k weighings, we know we have less than $3^k/2$ coins. This bound may not be sharp, for particular values of k it may not be possible to choose a stategy so each Y has $H(Y) = \log 3$ or so that the entropy of the joint distribution is equal to the sum of the entropies of the marginal distribution. However, we do have a bound.

   For $k = 3$ we have $n \leq 13$, in fact, there doesn't seem to be a solution for $n = 13$; there is one for $n = 12$. Lets start by numbering the coins from one to 12. The first weighing is $g_1 = \{1,2,3,4\}$ versus $g_2 = \{5,6,7,8\}$. If g_1 is heavier; then weigh $g_3 = \{1,2,5\}$ versus $g_4 = \{3,4,6\}$. Thus g_3 and g_4 each have two coins which must be heavier if they are counterfeit and the remaining two coins, 7 and 8, must be lighter. If g_3 is heavier than g_4 this can only be because either 1 or 2 is heavier, or 6 is lighter; weighing 1 or 2 settles this, if one is heavier than the other, it is the bent coin, if they balance, 5 is. If g_3 and g_4 balance then the counterfeit is either 7 or 8 and weighing them gives the answer. Finally, if g_1 and g_2 balance the counterfeit coin must be one of $\{9,10,11,12\}$; start by weighting $g_5 = \{9,10\}$ against $g_6 = \{11,1\}$: 1 is known not to be counterfiet. If g_5 and g_6 balance then the coin can only be 12 and weighing this against 1 gives the answer, otherwise, say g_5 is heavier than, either one of 9 and 10 is heavy, or 11 is light, weighing 9 against 10 sorts this out.

3. (C&T 2.10) Let X_1 and X_2 be discrete random variables drawn accorind to distributions p_1 and p_2 from their respective alphabets $\mathcal{X}_1 = \{1,2,\dots,m\}$ and $\mathcal{X}_2 = \{m+1, m+2, \dots, n\}$. Let

$$X = \begin{cases} X_1 & \text{with probability } \alpha \\ X_2 & \text{with probability } 1-\alpha \end{cases} \tag{10}$$

   (a) Find $H(X)$ in terms of $H(X_1)$ and $H(X_2)$.

   (b) Maximize over α to show that

$$2^{H(X)} \le 2^{H(X_1)} + 2^{H(X_2)} \tag{11}$$

*Solution:*We calculate the entropy directly using $p_X(x) = \alpha p_{X_1}(x)$ for $x \in \mathcal{X}_1$, and so on:

$$\begin{aligned} H(X) &= -\sum_{i=1}^{m} \alpha p_{X_1}(x=i) \log \alpha p_{X_1}(x=i) \\ &\quad - \sum_{i=m+1}^{n} (1-\alpha) p_{X_2}(x=i) \log (1-\alpha) p_{X_2}(x=i) \\ &= -\alpha \sum_{i=1}^{m} p_{X_1}(x=i) \left[\log \alpha + \log p_{X_1}(x=i)\right] \\ &\quad -(1-\alpha) \sum_{i=m+1}^{n} p_{X_2}(x=i) \left[\log (1-\alpha) + \log p_{X_2}(x=i)\right] \\ &= -\alpha \log \alpha - (1-\alpha) \log (1-\alpha) + \alpha H(X_1) + (1-\alpha) H(X_2) \end{aligned} \tag{12}$$

To maximize this over α we differentiate

$$\frac{dH(X)}{d\alpha} = \log \frac{1-\alpha}{\alpha} + H(X_1) - H(X_2) \tag{13}$$

and setting this equal to zero gives

$$\frac{1-\alpha}{\alpha} = 2^{H(X_2)-H(X_1)} \tag{14}$$

so

$$\alpha = \frac{1}{1 + 2^{H(X_2)-H(X_1)}} = \frac{2^{H(X_1)}}{2^{H(X_1)} + 2^{H(X_2)}} \tag{15}$$

and substituting back in gives a maximum for $H(X)$ so, writing $H_1 = H(X_1)$ and $H_2 = H(X_2)$ gives

$$H(X) \le \frac{2^{H_1}}{2^{H_1} + 2^{H_2}} \left(H_1 - \log \frac{2^{H_1}}{2^{H_1} + 2^{H_2}} \right) + \frac{2^{H_2}}{2^{H_1} + 2^{H_2}} \left(H_2 - \log \frac{2^{H_2}}{2^{H_1} + 2^{H_2}} \right) \tag{16}$$

and expanding out the logs gives

$$H(X) \le \log \left(2^{H_1} + 2^{H_2}\right) \tag{17}$$

or, since the log is monotonic

$$2^{H(X)} \le 2^{H(X_1)} + 2^{H(X_2)}. \tag{18}$$

4. (C&T 2.12). Let $p(x,y)$ be given by $p(0,0) = p(0,1) = p(1,1) = 1/3$ and $p(1,0) = 0$. Find $H(X)$, $H(Y)$, $H(X|Y)$, $H(Y|X)$, $H(X,Y)$, $H(Y) - H(Y|X)$ and $I(X;Y)$.

5. Prove the equals part of Jensen's inequality: if f is stricly cup-like on an interval which includes all outcomes

$$\langle f(X) \rangle = f(\langle X \rangle) \tag{19}$$

if and only if $X = \langle X \rangle$ with probability one.