GI12/4C59 – Homework #2 (Due 12am, October 21, 2004)

Aim: To get familiarity with the basic concepts of Information Theory (entropy, mutual information, etc..). Presentation, clarity, and synthesis of exposition will be taken into account in the assessment of these exercises.

This document is available at `http://www.cs.ucl.ac.uk/staff/M.Pontil/courses/IT-homework2.pdf`

1. **[50 pts]** Let X and Y be two random variables with values in the sets $\mathcal{X} = \{0,1,2\}$ and $\mathcal{Y} = \{0,1\}$ respectively. Define the probability distribution p on $\mathcal{X} \times \mathcal{Y}$ by the table

$$p = \begin{bmatrix} & X=0 & X=1 & X=2 \\ Y=0 & \frac{1}{4} & \frac{1}{12} & \frac{1}{6} \\ Y=1 & \frac{1}{12} & \frac{1}{4} & \frac{1}{6} \end{bmatrix}$$

   (a) Compute the joint entropy of X and Y, $H(X,Y)$.

   (b) Find the marginal distribution of X and the conditional distribution of Y given X. Then, use these quantities to compute the entropy of X, $H(X)$ and the relative entropy of Y given X, $H(Y|X)$.

   (c) Verify the entropy results above by using the chain rule which relates $H(X,Y)$ to $H(X)$ and $H(Y|X)$.

   (d) Compute the mutual information between X and Y.

   (e) Can you swap two entries in the table of $p(x,y)$ defined above so that the mutual information increases? Explain your observation.

2. **[20 pts]** *Average entropy.* Let $H(p) = -p\log_2 p - (1-p)\log_2(1-p)$ be the binary entropy function.

   (a) Evaluate $H(1/4)$ using the fact that $\log_2 3 \approx 1.584$. *Hint:* Consider an experiment with four equally likely outcomes, one of which is more interesting than the others.

   (b) Calculate the average entropy $H(p)$ when the probability p is chosen uniformly in the range $0 \le p \le 1$.

3. **[30 pts]**

   The entropy rate of a dog looking for a bone. A dog walks on the integers, possibly reversing direction at each step with probability $p = \frac{1}{10}$. Let $X_0 = 0$. The first step is equally likely to be positive or negative. A typical walk might look like this:

$$(X_0, X_1, \ldots) = (0, -1, -2, -3, -4, -3, -2, -1, 0, 1, \ldots)$$

   (a) Find $H(X_1, X_2, \ldots, X_n)$. (Hint: note that for $i > 1$ the next position depends on the two previous positions...).

   (b) The entropy rate of this process is defined as

$$\lim_{n\to\infty} \frac{H(X_0, X_1, \ldots, X_n)}{n+1}$$

   Find the entropy rate of the browsing dog.

   (c) What is the expected number of steps the dog takes before reversing direction?

GI12/4C59 – Solutions to Homework #2

1. (a) $H(X,Y) = \frac{5}{3} + \frac{1}{2}\log 3$.

   (b) $p(x) = \frac{1}{3}$, $x = 0, 1, 2$. $H(X) = \log 3$.

$$p(y|x) = \begin{bmatrix} \frac{3}{4} & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{4} & \frac{3}{4} & \frac{1}{2}. \end{bmatrix}$$

   $H(Y|X) = \frac{5}{3} - \frac{1}{2}\log 3$.

   (c) The above result verifies the formula $H(X,Y) = H(X) + H(Y|X)$.

   (d) $I(X;Y) = \frac{1}{2}\log 3 - \frac{2}{3}$.

   (e) If we swap any two entries in the table of $p(x,y)$, $H(X,Y)$ remains the same. On the other hand, since in the initial configuration table X and Y have uniform marginal distributions, $H(X)$ and $H(Y)$ can only decrease or remain the same after we swap two elements in the table. Thus, using the formula

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

   we see that I can never increase.

2. *Average Entropy.*

   (a) We can generate two bits of information by picking one of four equally likely alternatives. This selection can be made in two steps. First we decide whether the first outcome occurs. Since this has probability 1/4, the information generated is $H(1/4)$. If not the first outcome, then we select one of the three remaining outcomes; with probability 3/4, this produces $\log_2 3$ bits of information. Thus

$$H(1/4) + (3/4)\log_2 3 = 2$$

   and so $H(1/4) = 2 - (3/4)\log_2 3 = 2 - (.75)(1.585) = 0.811$ bits.

   (b) If p is chosen uniformly in the range $0 \le p \le 1$, then the average entropy (in nats) is

$$-\int_0^1 p\ln p + (1-p)\ln(1-p)dp = -2\int_0^1 x\ln x\, dx = -2\left(\frac{x^2}{2}\ln x - \frac{x^2}{4}\right)\Big|_0^1 = \tfrac{1}{2}\,.$$

   Therefore the average entropy is $\frac{1}{2}\log_2 e = 1/(2\ln 2) = .721$ bits.

3. *The entropy rate of a dog looking for a bone.*

   (a) By the chain rule,

$$\begin{aligned} H(X_0, X_1, \ldots, X_n) &= \sum_{i=0}^{n} H(X_i|X^{i-1}) \\ &= H(X_0) + H(X_1|X_0) + \sum_{i=2}^{n} H(X_i|X_{i-1}, X_{i-2}), \end{aligned}$$

   since, for $i > 1$, the next position depends only on the previous two (i.e., the dog's walk is 2nd order Markov, if the dog's position is the state). Since $X_0 = 0$ deterministically, $H(X_0) = 0$ and since the first step is equally likely to be positive or negative, $H(X_1|X_0) = 1$. Furthermore for $i > 1$,

$$H(X_i|X_{i-1}, X_{i-2}) = H(.1, .9).$$

   Therefore,

$$H(X_0, X_1, \ldots, X_n) = 1 + (n-1)H(.1, .9).$$

(b) From a),

$$\begin{aligned} \frac{H(X_0, X_1, \ldots X_n)}{n+1} &= \frac{1 + (n-1)H(.1, .9)}{n+1} \\ &\to H(.1, .9). \end{aligned}$$

(c) The dog must take at least one step to establish the direction of travel from which it ultimately reverses. Letting S be the number of steps taken between reversals, we have

$$\begin{aligned} E(S) &= \sum_{s=1}^{\infty} s(.9)^{s-1}(.1) \\ &= 10. \end{aligned}$$