Information Theory Solutions to Homework #1

(Cover and Thomas, Chaps 2 and 4) (H.-M. Hang; 2009/10)

(1)

2. **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: Let $y = g(x)$. Then

$$p(y) = \sum_{x:\, y=g(x)} p(x).$$

Consider any set of x's that map onto a single y. For this set

$$\sum_{x:\, y=g(x)} p(x) \log p(x) \le \sum_{x:\, y=g(x)} p(x) \log p(y) = p(y) \log p(y),$$

since log is a monotone increasing function and $p(x) \le \sum_{x:\, y=g(x)} p(x) = p(y)$. Extending this argument to the entire range of X (and Y), we obtain

$$\begin{aligned} H(X) &= -\sum_x p(x) \log p(x) \\ &= -\sum_y \sum_{x:\, y=g(x)} p(x) \log p(x) \\ &\ge -\sum_y p(y) \log p(y) \\ &= H(Y), \end{aligned}$$

with equality iff g is one-to-one with probability one.

   (a) $Y = 2^X$ is one-to-one and hence the entropy, which is just a function of the probabilities (and not the values of a random variable) does not change, i.e., $H(X) = H(Y)$.

   (b) $Y = \cos(X)$ is not necessarily one-to-one. Hence all that we can say is that $H(X) \ge H(Y)$, with equality if cosine is one-to-one on the range of X.

(2)

12. **Example of joint entropy.** Let $p(x,y)$ be given by

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

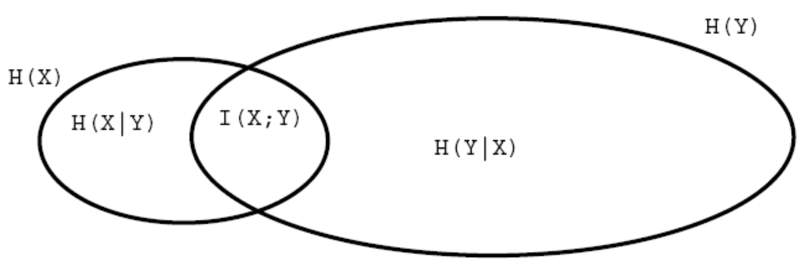
Find

(a) $H(X), H(Y)$.
(b) $H(X \mid Y), H(Y \mid X)$.
(c) $H(X,Y)$.
(d) $H(Y) - H(Y \mid X)$.
(e) $I(X;Y)$.
(f) Draw a Venn diagram for the quantities in (a) through (e).

Solution: *Example of joint entropy*

(a) $H(X) = \frac{2}{3}\log\frac{3}{2} + \frac{1}{3}\log 3 = 0.918$ bits $= H(Y)$.
(b) $H(X|Y) = \frac{1}{3}H(X|Y=0) + \frac{2}{3}H(X|Y=1) = 0.667$ bits $= H(Y|X)$.
(c) $H(X,Y) = 3 \times \frac{1}{3}\log 3 = 1.585$ bits.
(d) $H(Y) - H(Y|X) = 0.251$ bits.
(e) $I(X;Y) = H(Y) - H(Y|X) = 0.251$ bits.
(f) See Figure 1.

Figure 2.1: Venn diagram to illustrate the relationships of entropy and relative entropy



(3)

14. **Entropy of a sum.** Let X and Y be random variables that take on values $x_1, x_2, \ldots, x_r$ and $y_1, y_2, \ldots, y_s$, 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: *Entropy of a sum.*

(a) $Z = X + Y$. Hence $p(Z = z|X = x) = p(Y = z - x|X = x)$.

$$
\begin{aligned}
H(Z|X) &= \sum p(x) H(Z|X = x) \\
&= -\sum_x p(x) \sum_z p(Z = z|X = x) \log p(Z = z|X = x) \\
&= \sum_x p(x) \sum_y p(Y = z - x|X = x) \log p(Y = z - x|X = x) \\
&= \sum p(x) H(Y|X = x) \\
&= H(Y|X).
\end{aligned}
$$

If X and Y are independent, then $H(Y|X) = H(Y)$. Since $I(X;Z) \geq 0$, we have $H(Z) \geq H(Z|X) = H(Y|X) = H(Y)$. Similarly we can show that $H(Z) \geq H(X)$.

(b) Consider the following joint distribution for X and Y Let

$$
X = -Y = \begin{cases} 1 & \text{with probability } 1/2 \\ 0 & \text{with probability } 1/2 \end{cases}
$$

Then $H(X) = H(Y) = 1$, but $Z = 0$ with prob. 1 and hence $H(Z) = 0$.

(c) We have

$$
H(Z) \leq H(X, Y) \leq H(X) + H(Y)
$$

because Z is a function of (X, Y) and $H(X, Y) = H(X) + H(Y|X) \leq H(X) + H(Y)$. We have equality iff (X, Y) is a function of Z and $H(Y) = H(Y|X)$, i.e., X and Y are independent.

(4)

38. **The value of a question** Let $X \sim p(x)$, $x = 1, 2, \ldots, m$. We are given a set $S \subseteq \{1, 2, \ldots, m\}$. We ask whether $X \in S$ and receive the answer

$$
Y = \begin{cases} 1, & \text{if } X \in S \\ 0, & \text{if } X \notin S. \end{cases}
$$

Suppose $\Pr\{X \in S\} = \alpha$. Find the decrease in uncertainty $H(X) - H(X|Y)$.

Apparently any set S with a given α is as good as any other.

Solution: *The value of a question.*

$$\begin{aligned} H(X) - H(X|Y) &= I(X;Y) \\ &= H(Y) - H(Y|X) \\ &= H(\alpha) - H(Y|X) \\ &= H(\alpha) \end{aligned}$$

since $H(Y|X) = 0$.

(5)

42. **Inequalities.** Which of the following inequalities are generally $\geq, =, \leq$? Label each with $\geq, =,$ or $\leq$.

(a) $H(5X)$ vs. $H(X)$

(b) $I(g(X);Y)$ vs. $I(X;Y)$

(c) $H(X_0|X_{-1})$ vs. $H(X_0|X_{-1}, X_1)$

(d) $H(X_1, X_2, \ldots, X_n)$ vs. $H(c(X_1, X_2, \ldots, X_n))$, where $c(x_1, x_2, \ldots, x_n)$ is the Huffman codeword assigned to $(x_1, x_2, \ldots, x_n)$.

(e) $H(X,Y)/(H(X)+H(Y))$ vs. 1

Solution:

(a) $H(X) = H(5X)$ since $X \to 5X$ is a one-to-one mapping.

(b) By data processing inequality, $I(g(X);Y) \leq I(X;Y)$.

(c) Because conditioning reduces entropy, $H(X_0|X_{-1}) \geq H(X_0|X_{-1}, X_1)$.

(d) $H(X_1, X_2, ..., X_n) \geq H(c(X_1, X_2, ..., X_n))$.

(e) $\dfrac{H(X,Y)}{H(X)+H(Y)} \leq 1$.

(6)

43. **Mutual information of heads and tails.**

(a) Consider a fair coin flip. What is the mutual information between the top side and the bottom side of the coin?

(b) A 6-sided fair die is rolled. What is the mutual information between the top side and the front face (the side most facing you)?

Solution:

Mutual information of heads and tails.

To prove (a) observe that

$$\begin{aligned} I(T;B) &= H(B) - H(B|T) \\ &= \log 2 = 1 \end{aligned}$$

since $B \sim Ber(1/2)$, and $B = f(T)$. Here B, T stand for Bottom and Top respectively.

To prove (b) note that having observed a side of the cube facing us F, there are four possibilities for the top T, which are equally probable. Thus,

$$\begin{aligned} I(T;F) &= H(T) - H(T|F) \\ &= \log 6 - \log 4 \\ &= \log 3 - 1 \end{aligned}$$

since T has uniform distribution on $\{1, 2, \ldots, 6\}$.

(7)

7. **Entropy rates of Markov chains.**

(a) Find the entropy rate of the two-state Markov chain with transition matrix

$$P = \begin{bmatrix} 1 - p_{01} & p_{01} \\ p_{10} & 1 - p_{10} \end{bmatrix}.$$

(b) What values of p_{01}, p_{10} maximize the rate of part (a)?

(c) Find the entropy rate of the two-state Markov chain with transition matrix

$$P = \begin{bmatrix} 1 - p & p \\ 1 & 0 \end{bmatrix}.$$

Solution: *Entropy rates of Markov chains.*

(a) The stationary distribution is easily calculated. (See EIT pp. 62–63.)

$$\mu_0 = \frac{p_{10}}{p_{01} + p_{10}}, \quad \mu_0 = \frac{p_{01}}{p_{01} + p_{10}}.$$

Therefore the entropy rate is

$$H(X_2|X_1) = \mu_0 H(p_{01}) + \mu_1 H(p_{10}) = \frac{p_{10} H(p_{01}) + p_{01} H(p_{10})}{p_{01} + p_{10}}.$$

(b) The entropy rate is at most 1 bit because the process has only two states. This rate can be achieved if (and only if) $p_{01} = p_{10} = 1/2$, in which case the process is actually i.i.d. with $\Pr(X_i = 0) = \Pr(X_i = 1) = 1/2$.

(c) As a special case of the general two-state Markov chain, the entropy rate is

$$H(X_2|X_1) = \mu_0 H(p) + \mu_1 H(1) = \frac{H(p)}{p + 1}.$$

--- The End ---