Problem Set 2

Suggested Solutions

1. The joint distribution of two binary random variables X and Y is $p(X = 0, Y = 0) = p(X = 0, Y = 1) = p(X = 1, Y = 1) = 1/3$, $p(X = 1, Y = 0) = 0$. Compute
   a) $H(X, Y)$;
   b) $H(X)$, $H(Y)$;
   c) $H(Y \mid X)$, $H(Y \mid X = 0)$, $H(Y \mid X = 1)$;
   d) $H(X \mid Y)$, $H(X \mid Y = 1)$, $H(X \mid Y = 1)$;
   e) $I(X; Y)$.

Sol. a) $H(X, Y) = \log_2 3$

b) $p(X = 0) = p(X = 0, Y = 0) + P(X = 0, Y = 1) = 2/3$, $p(X = 1) = 1/3$;

$p(Y = 0) = p(X = 0, Y = 0) + P(X = 1, Y = 0) = 1/3$, $p(Y = 1) = 2/3$;

so $H(X) = H(Y) = 2/3*\log_2(3/2) + 1/3*\log_2 3 = \log_2 3 - 2/3$

c) $H(Y \mid X) = H(X, Y) - H(X) = 2/3$,

the distribution of Y given X = 0 is $p_{Y|X=0} = (1/2, 1/2)$, so $H(Y \mid X = 0) = 1$,

the distribution of Y given X = 1 is $p_{Y|X=1} = (0, 1)$, so $H(Y \mid X = 1) = 0$.

d) $H(X \mid Y) = H(X, Y) - H(Y) = 2/3$,

the distribution of X given Y = 0 is $p_{X|Y=0} = (1, 0)$, so $H(X \mid Y = 0) = 0$,

the distribution of X given Y = 1 is $p_{X|Y=1} = (1/2, 1/2)$, so $H(X \mid Y = 1) = 1$.

e) $I(X; Y) = H(X) - H(X \mid Y) = \log_2 3 - 4/3$

2. Consider the same random variables X and Y as in Problem 1. Define a new binary random variable Z as $Z = X \oplus Y$. Compute
   a) $H(Z)$;
   b) $H(Z \mid X, Y)$, $H(X, Y, Z)$ (Hint: Since $Z = X \oplus Y$, Z is deterministic given X and Y);
   c) $H(Y \mid X, Z)$, $H(X \mid Y, Z)$, $H(Y, Z)$, $H(X, Z)$ (Hint: $Y = Z \oplus X$, $X = Y \oplus Z$, and use the results you get in a) and b));
   d) $H(Y \mid Z)$, $H(X \mid Z)$, $H(X, Y \mid Z)$ (Hint: instead of calculating by definition,

make use of the results you get in a)-c));

e) $I(X; Z)$, $I(X; Y \mid Z)$ (Hint: instead of calculating by definition, make use of the results you get in a)-d)).

Sol. a) $p(Z = 0) = p(X = 0, Y = 0) + p(X = 1, Y = 1) = 2/3$, $p(Z = 1) = 1/3$,

$H(Z) = \log_2 3 - 2/3$

b) $H(Z \mid X, Y) = 0$, $H(X, Y, Z) = H(X, Y) + H(Z \mid X, Y) = \log_2 3$

c) As Y is deterministic on X and Z, and X is deterministic on Y and Z,

$$H(Y \mid X, Z) = H(X \mid Y, Z) = 0$$

$H(X, Z) = H(X, Y, Z) - H(Y \mid X, Z) = H(X, Y, Z) = \log_2 3$

$H(Y, Z) = H(X, Y, Z) - H(X \mid Y, Z) = H(X, Y, Z) = \log_2 3$

d) $H(Y \mid Z) = H(Y, Z) - H(Z) = 2/3$, $H(X \mid Z) = H(X, Z) - H(Z) = 2/3$

$H(X, Y \mid Z) = H(X, Y, Z) - H(Z) = 2/3$.

e) $I(X; Z) = H(X) - H(X \mid Z) = \log_2 3 - 4/3$,

$I(X; Y \mid Z) = H(X \mid Z) + H(Y \mid Z) - H(X, Y \mid Z) = 2/3$

(alternatively, $I(X; Y \mid Z) = H(X \mid Z) - H(X \mid Y, Z) = 2/3$)

3. Assume that $X \rightarrow Y \rightarrow Z$ forms a Markov Chain. What is

a) $I(X; Z \mid Y)$?

b) $H(X \mid Y, Z) - H(X \mid Y)$?

Sol. a) As X and Z are independent given Z (i.e., $X \perp Z | Y$), $I(X; Z \mid Y) = 0$;

b) Since $X \rightarrow Y \rightarrow Z$ forms a Markov Chain, $Z \rightarrow Y \rightarrow X$ forms a Markov Chain too. We have $p(X = x_i \mid Y = y_j, Z = z_k) = p(X = x_i \mid Y = y_j)$ for all possible i, j, k. Thus $H(X \mid Y, Z) = H(X \mid Y)$ and $H(X \mid Y, Z) - H(X \mid Y) = 0$.

4. Give an example to show that $I(X; Y) = 0$ and $I(X; Y \mid Z) = 0$ does not imply each other.

(Hint: Use the random variables which are pairwise independent but not mutually independent to show $I(X; Y) = 0$ does not imply $I(X; Y \mid Z) \neq 0$. Conversely, design three binary random variables X, Y, Z as follows: X, Y are the respective input and output of a Binary Symmetric Channel with $p(X = 0) = 1/2$, $p(Y = 1 \mid X$

= 0) = 1/4, Z = Y. Show $I(X; Y \mid Z) = 0$ but $I(X; Y) \neq 0$ for these X, Y, Z)

Sol. Let $X = \{0, 1\}$, $Y = \{0, 1\}$, $p(X = 0) = p(X = 1) = p(Y = 0) = p(Y = 1) = 1/2$, $X \perp Y$, $Z = X \oplus Y$. In this case, $X \perp Z$. Then, $I(X; Y) = 0$, $I(X; Y \mid Z) = H(X \mid Z) - H(X \mid Y, Z) = H(X) - 0 = 1$.

Next consider binary random variables X, Y, Z with $p(X = 0) = p(X = 1) = 1/2$, $p(Y = 1 \mid X = 0) = p(Y = 0 \mid Y = 1) = 1/4$, $p(Y = 0 \mid X = 0) = p(Y = 1 \mid Y = 1) = 3/4$. Then, $p(X = 0, Y = 0) = p(Y = 0 \mid X = 0)p(X = 0) = 3/8$, and similarly $p(X = 0, Y = 1) = p(X = 1, Y = 0) = 1/8$, $p(X = 1, Y = 1) = 1/8$. Moreover, $p(Y = 0) = p(Y = 0 \mid X = 0)p(X = 0) + p(Y = 0 \mid X = 1)p(X = 1) = 1/2$. Consequently,

$H(X) = H(Y) = 1$, $H(X, Y) = 2\times1/8\times\log_2 8 + 2\times3/8\times\log_2(8/3) = 3 - 3/4\times\log_2 3 = 1.8133$, $I(X; Y) = H(X) + H(Y) - H(X, Y) = 0.1867 \neq 0$.

On the other hand, as Z = Y, we have $H(Y \mid Z) = 0$, $H(Y \mid X, Z) = 0$. Thus, $I(X; Y \mid Z) = H(Y \mid Z) - H(Y \mid X, Z) = 0 - 0 = 0$.

5. Let X be a function of Y. Prove that $H(X) \leq H(Y)$. (Hint: as X is a function of Y, X is deterministic given Y).

Sol. As X is a function of Y, $H(X \mid Y) = 0$. Thus,

$$H(X) \leq H(X, Y) = H(X \mid Y) + H(Y) = H(Y)$$

6. Based on the basic inequalities, prove the following inequalities on random variables X, Y, Z, and state the condition where equality holds:

   a) $H(X, Y \mid Z) \geq H(Y \mid X, Z)$

   b) $I(X, Y; Z) \geq I(X; Z)$

   c) $H(X, Y, Z) - H(X, Y) \leq H(X, Z) - H(X)$

   d) $I(X; Z \mid Y) \geq I(Z; Y \mid X) - I(Z; Y) + I(X; Z)$

   (Hint: the basic inequalities we know about information measures are: $H(X) \geq 0$, $H(X, Y) \geq 0$, $H(Y \mid X) \geq 0$, $I(X; Y) \geq 0$, $I(X; Y \mid Z) \geq 0$, $H(Y) \geq H(Y \mid X)$.)

Sol. a) $H(X, Y \mid Z) = H(X \mid Z) + H(Y \mid X, Z) \geq H(Y \mid X, Z)$. When X is deterministic on Z, the equality holds.

b) I(X, Y; Z) = I(X; Z) + I(Y; Z | X) ≥ I(X; Z). When $Y \perp Z | X$, the equality holds.

c) H(X, Y, Z) – H(X, Y) = H(Z | X, Y) = H(Z | X) – I(Y; Z | X) ≤ H(Z | X) = H(X, Z) – H(X). When $Y \perp Z | X$, the equality holds.

d) The left-hand side can be expressed as H(Z | Y) – H(Z | X, Y) = H(Y, Z) – H(Y) – (H(X, Y, Z) – H(X, Y)). The right-hand side can be expressed as H(Z | X) + H(Y | X) – H(Y, Z | X) – (H(Y) + H(Z) – H(Y, Z)) + (H(X) + H(Z) – H(X, Z)) = (H(X, Z) – H(X)) + (H(X, Y) – H(X)) – (H(X, Y, Z) – H(X)) – (H(Y) + H(Z) – H(Y, Z)) + (H(X) + H(Z) – H(X, Z)) = H(X, Y) – H(X, Y, Z) – H(Y) + H(Y, Z). Thus, this inequality is actually equality.

7. A random memoryless source $X \in \{0, 1, 2\}$ with probability distribution {1/4, 1/4, 1/2}. Two experiments are designed to observe this source, with respective outcome random variables $Y_1 \in \{0, 1\}$, $Y_2 \in \{0, 1\}$. The respective conditional probability matrix of Y_1 and Y_2 given X is provided by the transition matrix

$$P_{Y_1|X} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1/2 & 1/2 \end{bmatrix}, P_{Y_2|X} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

a) Compute $I(X; Y_1)$ and $I(X; Y_2)$. Which experiments is better?

b) Compute $I(X; Y_1, Y_2)$. How much additional information on X you can obtain by doing both experiments compared with doing only experiment Y_1 or only experiment Y_2?

c) Compute $I(X; Y_1 | Y_2)$ and $I(X; Y_2 | Y_1)$. How to interpret these two information measures?

Sol. $p(Y_1 = 0) = 1/4*1 + 1/4*0 + 1/2*1/2 = 1/2$, $p(Y_1 = 1) = 1/2$, $H(Y_1) = 1$

$p(Y_2 = 0) = 1/4*1 + 1/4*1 + 1/2*0 = 1/2$, $p(Y_2 = 1) = 1/2$, $H(Y_2) = 1$

$p(Y_1 = 0, Y_2 = 0) = p(X = 0) = 1/4$, $p(Y_1 = 1, Y_2 = 0) = p(X = 1) = 1/4$

$p(Y_1 = 0, Y_2 = 1) = p(X = 2)p(Y_1 = 0, Y_2 = 1 | X = 2) = 1/4$,

$p(Y_1 = 1, Y_2 = 1) = p(X = 2)p(Y_1 = 1, Y_2 = 1 | X = 2) = 1/4$,

$H(Y_1, Y_2) = 2$, $H(X, Y_1, Y_2) = H(1/4, 1/4, 1/4, 1/4) = 2$,

$H(Y_1 \mid X)$ = 1/4*H(1) + 1/4*H(1) + 1/2*H(1/2) = 1/2,

$H(Y_2 \mid X)$ = 1/4*H(1) + 1/4*H(1) + 1/2*H(1) = 0. Then we can obtain

a) $I(X; Y_1) = H(Y_1) - H(Y_1 \mid X) = 1/2$, $I(X; Y_2) = H(Y_2) - H(Y_2 \mid X) = 1$, Experiment 2 is better than experiment 1 because $I(X; Y_2) > I(X; Y_1)$, i.e., the (weighted) information of X is revealed more by Y_2 compared with by Y_1.

b) $I(X; Y_1, Y_2) = H(X) + H(Y_1, Y_2) - H(X, Y_1, Y_2) = H(X) = 3/2$, 1 bit and 1/2 bit additional information of X can be obtained by doing both experiments compared with doing only experiment Y_1 and only experiment Y_2, respectively.

c) $I(X; Y_1 \mid Y_2) = I(X; Y_1, Y_2) - I(X; Y_2) = 1/2$, $I(X; Y_2 \mid Y_1) = I(X; Y_1, Y_2) - I(X; Y_1) = 1$. $I(X; Y_1 \mid Y_2)$ represents the additional information of X obtained by doing experiment Y_1 after experiment Y_2.

8. Let $X_1, X_2 \in \{0, 1\}$ be two independent binary random variables with identical probability distribution {1/2, 1/2}. Define a new random variable $Y \in \{0, 1, 2\}$ to be the sum of X_1 and X_2. Define another random variable as follows. If Y is even, then Z = 0; if Y is odd, then Z = 1. Obviously, $(X_1, X_2) \to Y \to Z$ forms a Markov chain. Respectively calculate $I(X_1, X_2; Y)$, $I(Y; Z)$ and $I(X_1, X_2; Z)$. Compare their values in terms of the data processing theorem.

Sol. $p_Y = \{1/4, 1/2, 1/4\}$. $p_Z = \{1/2, 1/2\}$.

$I(X_1, X_2; Y) = H(Y) - H(Y \mid X_1, X_2) = H(Y) = 3/2$

$I(Y; Z) = H(Z) - H(Z \mid Y) = H(Z) = 1$

$I(X_1, X_2; Z) = H(Z) - H(Z \mid X_1, X_2) = H(Z) = 1$

This result is in line with the data processing theorem as $I(X_1, X_2; Z) \le I(Y; Z)$ and $I(X_1, X_2; Z) \le I(X_1, X_2; Y)$

9. Consider a stationary Markov source $X_1, X_2, \ldots X_j, \ldots$ (Markov source means $p(X_{j+1} \mid X_j, \ldots, X_2, X_1) = p(X_{j+1} \mid X_j)$ for all $j \ge 1$). Every random variable X_j in the stationary source has an identical probability distribution (1/3, 1/4, 1/4, 1/6). The conditional probability distribution of X_{j+1} given X_j is known via the transition

matrix

$$P_{X_{j+1}|X_j} = \begin{bmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1/3 & 1/3 & 1/3 & 0 \\ 1/3 & 1/3 & 0 & 1/3 \\ 1/2 & 0 & 1/2 & 0 \end{bmatrix}$$

Calculate $H(X_j)$, $H(X_j, X_{j+1})/2$ and entropy rate H_X of this source. Compare these values and what conclusion you can make?

Sol. $H(X_j) = H(1/3, 1/4, 1/4, 1/6) = 7/6 + 0.5\log3$. Assume the sample space of X_j is $\{0, 1, 2, 3\}$. Then,

$H(X_{j+1} \mid X_j)$

$= 1/3*H(X_{j+1} \mid X_j = 0)+1/4*H(X_{j+1} \mid X_j = 1)+1/4*H(X_{j+1} \mid X_j = 2)+1/6*H(X_{j+1} \mid X_j = 0)$

$= 1/3*H(1/4, 1/4, 1/4, 1/4) + 1/4*H(1/3, 1/3, 1/3, 0) + 1/4*H(1/3, 1/3, 0, 1/3) + 1/6*H(1/2, 0, 1/2, 0) = 1/3*\log4 + 1/4*\log3 + 1/4*\log3 + 1/6*\log2 = 5/6 + 0.5\log3$

We now have $H(X_j, X_{j+1})/2 = (H(X_j) + H(X_{j+1} \mid X_j))/2 = 1+ 0.5\log3$, and the entropy rate $H_X = \lim_{j\to\infty}H(X_{j+1} \mid X_j, X_{j-1}, \ldots, X_1) = H(X_{j+1} \mid X_j) = 5/6 + 0.5\log3$, where the second last equality holds due to the (1st order) Markovian property of the source.

Among these values, $H(X_j) > H(X_j, X_{j+1})/2 > H_X$, i.e., the more random variables considered together, the average entropy will be nonincreasing and approach to the entropy rate.