Fundamentals of Information Theory

Solution 2

Problem 1 *(10 points) Suppose you're on a game show, and you've given the choice of three doors: Behind one door is a car; behind the others, goat. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice so that you have a higher chance to win the car instead of the goat?*

Solution:

Let the events C_1, C_2 and C_3 represent the car is behind door 1, 2 and 3, respectively. All these 3 events have probability 1/3.

The player picking door 1 is described by the event X_1. As the first choice of the player is independent of the position of the car, therefore the conditional probabilities are $P(C_i|X_1) = P_1(C_i) = 1/3$. For ease of notation the conditional probability given X_1 is denoted by P_1.

The host opening door 3 is described by H_3. For this event it holds:

$$\begin{aligned} P_1(H_3|C_1) &= 1/2 \\ P_1(H_3|C_2) &= 1 \\ P_1(H_3|C_3) &= 0 \end{aligned}$$

Then, if the player initially selects door 1, and the host opens door 3, the conditional probability of winning by switching is

$$\begin{aligned} P_1(C_3|H_3) &= \frac{P_1(H_3|C_2)P_1(C_2)}{P_1(H_3)} \\ &= \frac{P_1(H_3|C_2)P_1(C_2)}{P_1(H_3|C_1)P_1(C_1) + P_1(H_3|C_2)P_1(C_2) + P_1(H_3|C_3)P_1(C_3)} \\ &= \frac{P_1(H_3|C_2)}{P_1(H_3|C_1) + P_1(H_3|C_2) + P_1(H_3|C_3)} \\ &= \frac{1}{1/2 + 1 + 0} \\ &= \frac{2}{3} \end{aligned}$$

A detailed discussion can be found at `http://en.wikipedia.org/wiki/Monty_Hall_problem`.

Problem 2 *(15 points) One is given 24 coins. It is known that precisely one coin is fake, which weights differently compared with genuine coins. It is not clear whether this fake coin is heavier or lighter, though. Now we use a balance to identify the fake coin, but we do not have the weights for this balance.*

(a) What is the minimum number of the weighting operations in order to identify this fake coin.

(b) Explain briefly about your weighting procedure to identify this fake coin.

Solution:

(a) In these 24 coins, precisely one coin is fake, which weights differently compared with genuine coins. However, it is not whether this fake coin is heavier or lighter; hence, uncertainty arises. We use a balance to weight two coins. On each weighting operation, we are able to obtain certain information, eliminating some uncertainty. After several weighting operations, all the uncertainty can be removed. Then, the fake can be identified to be heavier or lighter than the genuine coins. Therefore, we define event A as one of the 24 coins is fake. Event A appears with the probability as

$$P(A) = \frac{1}{24}.$$

We define event B as the fake coin is heavier or lighter than the genuine coin. Event B has the probability as

$$P(B) = \frac{1}{2}.$$

The uncertainty (self-information) of event A is

$$I(A) = -\log_2 P(A) = \log_2 24.$$

The uncertainty (self-information) of event B is

$$I(B) = -\log_2 P(B) = \log_2 2 = 1 \text{ bit}.$$

To identify the fake coin and to make clear whether the fake coin is heavier or lighter than the genuine coin is to eliminate the uncertainty of both events. Because event A and event B are independent, the total required information is

$$I_1 = I(A) + I(B) = \log_2 24 + 1 = \log_2 48 = 5.585 \text{ bits}.$$

There are three outcomes with one weighting operation: heavier, lighter or equal, with the same probability. Hence, the probability of each outcome is $I(C) = \frac{1}{3}$. Therefore, the amount of information obtained by one weighting operation, that is, the amount of the eliminated information, is

$$I_2 = I(C) = -\log_2 P(C) = \log_2 3 = 1.585 \text{ bits}.$$

$$\frac{I_1}{I_2} = I(C) = \frac{\log_2 48}{\log_2 3} = 3.53.$$

Hence, the minimal required weighting operations is 4.

(b) There are a number of weight procedures to identify the fake coin in four times. Please try it on your own.

Problem 3 *(10 points)*

Let $p(x, y)$ be given in Table 1, i.e., $p(X = 0, Y = 1) = 0$.
Find.

(a) $H(X)$, $H(Y)$.

(b) $H(X|Y)$, $H(Y|X)$.

(c) $H(X, Y)$.

Table 1: $p(x, y)$

X\Y	0	1
0	1/3	0
1	1/3	1/3

(d) $H(Y) - H(Y|X)$.

(e) $I(X;Y)$.

(f) Draw a Venn diagram for the quantities in (a) through (e).

Solution:

(a)

$$\begin{aligned}
H(X) &= -\sum_{x\in\mathcal{X}} p(x)\log\left[p(x)\right] \\
&= H\left(\frac{1}{3}, \frac{2}{3}\right) \\
&= -\left[\frac{1}{3}\log(\frac{1}{3}) + \frac{2}{3}\log(\frac{2}{3})\right] \\
&= 0.918 \text{ bits}
\end{aligned}$$

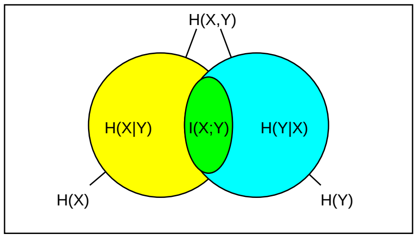
$$\begin{aligned}
H(Y) &= -\sum_{y\in\mathcal{Y}} p(y)\log\left[p(y)\right] \\
&= H\left(\frac{2}{3}, \frac{1}{3}\right) \\
&= -\left[\frac{2}{3}\log(\frac{2}{3}) + \frac{1}{3}\log(\frac{1}{3})\right] \\
&= 0.918 \text{ bits}
\end{aligned}$$

(b)

$$\begin{aligned}
H(X|Y) &= \sum_{y\in\mathcal{Y}} p(y)H(X|Y=y) \\
&= \sum_{y\in\mathcal{Y}} p(y)\sum_{x\in\mathcal{X}} p(x|y)\log\left[p(x|y)\right] \\
&= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log\left[p(x|y)\right] \\
&= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log\left[\frac{p(x,y)}{p(y)}\right] \\
&= -\left[\ \tfrac{1}{3}\log\tfrac{\frac{1}{3}}{\frac{2}{3}} + 0\log\tfrac{0}{\frac{1}{4}} + \tfrac{1}{3}\log\tfrac{\frac{1}{3}}{\frac{2}{3}} + \tfrac{1}{3}\log\tfrac{\frac{1}{3}}{\frac{1}{3}}\ \right] \\
&= 0.667 \text{ bit}
\end{aligned}$$

Due to the symmetry between X and Y, $H(Y|X) = H(X|Y) = 0.667$ bit.

(c) $H(X,Y) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log\left[p(x,y)\right] = 3 * 1/3 * \log_2 3 = 1.585$ bits.



(d) $H(Y) - H(Y|X) = 0.251$ bit.

(e) $I(X;Y) = H(Y) - H(Y|X) = 0.251$ bit.

(f)

Problem 4 *(10 points) Let the random variable X have three possible outcomes $\{a, b, c\}$. Consider two distributions on this random variable:*

Symbol	$p(x)$	$q(x)$
a	*1/2*	*1/3*
b	*1/4*	*1/3*
c	*1/4*	*1/3*

Calculate $H(p(x))$, $H(q(x))$, $D(p(x)||q(x))$, and $D(q(x)||p(x))$. Verify that in this case, $D(p(x)||q(x)) \neq D(q(x)||p(x))$.

Solution:

$$H(p(x)) = -\sum_x p(x) \log p(x) = \frac{1}{2}\log 2 + \frac{1}{4}\log 4 + \frac{1}{4}\log 4 = 1.5 \text{ bits}$$

$$H(q(x)) = -\sum_x q(x) \log q(x) = \frac{1}{3}\log 3 + \frac{1}{3}\log 3 + \frac{1}{3}\log 3 = 1.58 \text{ bits}$$

$$D(p(x)||q(x)) = \sum_x p(x) \log \frac{p(x)}{q(x)} = \frac{1}{2}\log\frac{3}{2} + \frac{1}{4}\log\frac{3}{4} + \frac{1}{4}\log\frac{3}{4} = 0.085 \text{ bit}$$

$$D(q(x)||p(x)) = \sum_x q(x) \log \frac{q(x)}{p(x)} = \frac{1}{3}\log\frac{2}{3} + \frac{1}{3}\log\frac{4}{3} + \frac{1}{3}\log\frac{4}{3} = 0.082 \text{ bit}$$

We observe that $D(p(x)||q(x)) \neq D(q(x)||p(x))$.

Problem 5 *(10 points)*

Consider a non-uniform random variable X with $M = 8$ possible outcomes and probabilities (1/2, 1/4, 1/8, 1/16, 1/64, 1/64, 1/64, 1/64). We use one symbol to represent one outcome, so we have 8 types of symbols (A, B, C, D, E, F, G, H). Then we transmit 1024 symbols from point s to point d. In this 1024-symbol sequence, it turns out that the numbers of these 8 type symbols are $(516, 255, 126, 68, 17, 18, 10, 14)$, respectively

(a) Compute the entropy $H(X)$ in bits.

(b) Consider a coding scheme c_1, $A \to 000$, $B \to 001$, $C \to 010$, $D \to 011$, $E \to 100$, $F \to 101$, $G \to 110$, $H \to 111$, to transmit these 1024 symbols. Compute the total bits using scheme c_1.

(c) Similarly, consider another coding scheme c_2, $A \to 0$, $B \to 10$, $C \to 110$, $D \to 1110$, $E \to 11110$, $F \to 111110$, $G \to 1111110$, $H \to 11111110$, to transmit these 1024 symbols. Compute the total bits using scheme c_2. On average, how many bits are used to transmit one symbol in this transmitted sequence?

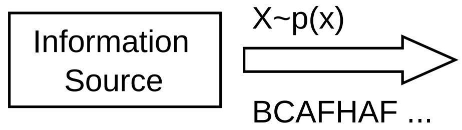


Figure 1: Source output sequence

Solution:

(a) We have 8 types of symbols (A, B, C, D, E, F, G, H) with the symbol probabilities (1/2, 1/4, 1/8, 1/16, 1/64, 1/64, 1/64, 1/64).

$$\begin{aligned} H(X) &= \sum_{x \in \mathcal{X}} p(x) \log [p(x)] \\ &= -\left(\frac{1}{2} * \log_2 \frac{1}{2} + \frac{1}{4} * \log_2 \frac{1}{4} + \frac{1}{8} * \log \frac{1}{8} + \frac{1}{16} * \log_2 \frac{1}{16} + 4 * \frac{1}{64} * \log_2 \frac{1}{64} \right) \\ &= 2 \text{ bits} \end{aligned}$$

(b) In scheme c_1: $A \to 000$, $B \to 001$, $C \to 010$, $D \to 011$, $E \to 100$, $F \to 101$, $G \to 110$, $H \to 111$, the total bits $= 3bits/symbol * 1024symbol = 3072$ bits.

(c) In scheme c_2: $A \to 0$, $B \to 10$, $C \to 110$, $D \to 1110$, $E \to 11110$, $F \to 111110$, $G \to 1111110$, $H \to 11111110$, the total bits $= 1*516+2*255+3*126+4*68+5*17+6*18+7*10+8*14 = 2051$ bits.

On average,

$$\frac{\text{total bits}}{\text{total symbols}} = \frac{2051 \text{ bits}}{1024 \text{ symbol}} = 2 \text{ bits/symbol}.$$

Problem 6 *(10 points) Let X_1 and X_2 be identically distributed but not necessarily independent. Let*

$$\rho = 1 - \frac{H(X_2|X_1)}{H(X_1)}.$$

(a) Show that $\rho = \frac{I(X_1;X_2)}{H(X_1)}$.

(b) Show that $0 \le \rho \le 1$.

(c) When is $\rho = 0$?

(d) When is $\rho = 1$?

Solution:

(a) Since X_1, X_2 are identically distributed, $H(X_1) = H(X_2)$.

$$\rho = 1 - \frac{H(X_2|X_1)}{H(X_1)} = \frac{H(X_1) - H(X_2|X_1)}{H(X_1)} = \frac{H(X_2) - H(X_2|X_1)}{H(X_1)} = \frac{I(X_1;X_2)}{H(X_1)}.$$

(b) $H(X_1) \neq 0, H(X_2|X_1) \geqslant 0$ and $H(X_2|X_1) \leqslant H(X_2)$. Thus, $0 \leqslant \rho \leqslant 1$. If $H(X_1) = 0$, then $H(X_2) = H(X_2|X_1) = 0$, the limit of $\rho = 0$. Thus, $0 \leqslant \rho \leqslant 1$.

(c) Assume $\rho = 0$ and $H(X_1) \neq 0$, $H(X_2|X_1) = H(X_2)$, $I(X_1;X_2) = 0$, that is, $p(x_{2i}|x_{1j}) = p(x_{1j}), \forall x_{1j} \in X_1, x_{2i} \in X_2$. Thus, X_1 and X_2 are independent.

(d) Assume $\rho = 1$ and $H(X_1) \neq 0$, $H(X_2|X_1) = 0$. Then, $I(X_1;X_2) = H(X_1) = H(X_2)$. Thus, $X_1 = X_2$, that is, X_1 and X_2 are one-to-one mapping.

Problem 7 *(10 points)*

Let X, Y and Z be joint random variables. Prove the following inequalities and find conditions for equality.

(a) $H(X,Y|Z) \geq H(X|Z)$.

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

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

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

Solution: Inequalities.

(a) Using the chain rule for conditional entropy,

$$H(X,Y|Z) = H(X|Z) + H(Y|X,Z) \geq H(X|Z)$$

with equality $H(Y|X,Z) = 0$, that is, when Y is a function of X and Z.

(b) Using the chain rule for mutual information,

$$I(X,Y;Z) = I(X;Z) + I(Y;Z|X) \geq I(X;Z),$$

with equality $I(Y;Z|X) = 0$, that is, when Y and Z are conditionally independent given X.

(c) Using first the chain rule for entropy and then the definition of conditional mutual information,

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

with equality $I(Y;Z|X) = 0$, that is, when Y and Z are conditionally independent given X.

(d) Using the chain rule for mutual information,

$$I(X;Z|Y) + I(Z;Y) = I(X,Y;Z) = I(Z;Y|X) + I(X;Z)$$

and therefore

$$I(X;Z|Y) = I(X,Y;Z) = I(Z;Y|X) - I(Z;Y) + I(X;Z).$$

We see that this inequality is actually an equality in all cases.