Exercise Problems: Information Theory and Coding

Exercise 5

Suppose that X is a random variable whose entropy $H(X)$ is 8 bits. Suppose that $Y(X)$ is a deterministic function that takes on a different value for each value of X.

(i) What then is $H(Y)$, the entropy of Y?

(ii) What is $H(Y|X)$, the conditional entropy of Y given X?

(iii) What is $H(X|Y)$, the conditional entropy of X given Y?

(iv) What is $H(X,Y)$, the joint entropy of X and Y?

(v) Suppose now that the deterministic function $Y(X)$ is not invertible; in other words, different values of X may correspond to the same value of $Y(X)$. In that case, what could you say about $H(Y)$?

(vi) In that case, what could you say about $H(X|Y)$?

Solution:

(i) The entropy of Y: $H(Y) = 8$ bits also.

(ii) The conditional entropy of Y given X: $H(Y|X) = 0$

(iii) The conditional entropy of X given Y: $H(X|Y) = 0$ also.

(iv) The joint entropy $H(X,Y) = H(X) + H(Y|X) = 8$ bits

(v) Since now different values of X may correspond to the same value of $Y(X)$, the new distribution of Y has lost entropy and so $H(Y) < 8$ bits.

(vi) Now knowledge of Y no longer determines X, and so the conditional entropy $H(X|Y)$ is no longer zero: $H(X|Y) > 0$

Exercise 6

Suppose that the following sequence of Yes/No questions was an optimal strategy for playing the "Game of 7 questions" to learn which of the letters $\{A, B, C, D, E, F, G\}$ someone had chosen, given that their *a priori* probabilities were known:

"Is it A?"	"No."
"Is it a member of the set $\{B, C\}$?"	"No."
"Is it a member of the set $\{D, E\}$?"	"No."
"Is it F?"	"No."

1. Write down a probability distribution for the 7 letters, $p(A), ..., p(G)$, for which this sequence of questions was an optimal strategy.
2. What was the uncertainty, in bits, associated with each question?
3. What is the entropy of this alphabet?
4. Now specify a variable length, uniquely decodable, prefix code for this alphabet that would minimise the average code word length.
5. What is your average coding rate R for letters of this alphabet?
6. How do you know that a more efficient code could not be developed?

Solution:

1. Under the Asymptotic Equipartition Theorem, the following *a priori* probability distribution would make the given questioning strategy an optimal one:

$p(A)$	$p(B)$	$p(C)$	$p(D)$	$p(E)$	$p(F)$	$p(G)$
1/2	1/8	1/8	1/16	1/16	1/16	1/16

2. Each Yes/No question had 1 bit entropy (uncertainty), because both possible answers were equiprobable in each case.
3. Since entropy $= -\sum_i p_i \log_2 p_i$, the entropy of this alphabet is 2.25 bits.
4. One possible variable length, uniquely decodable, prefix code is:

A	B	C	D	E	F	G
0	110	111	1000	1001	1010	1011

5. Summing over all the letters, the probability of each letter times its code word length in bits, gives us $R = (1/2)(1) + (2/8)(3) + (4/16)(4) = 2.25$ bits per letter on average.
6. Because the coding rate equals the entropy of the source alphabet, and Shannon's Source Coding Theorem tells us that this is the lower bound for the coding rate, we know that no more efficient code could be developed.

Exercise 7

Consider a binary symmetric communication channel, whose input source is the alphabet $X = \{0, 1\}$ with probabilities $\{0.5, 0.5\}$; whose output alphabet is $Y = \{0, 1\}$; and whose channel matrix is

$$\begin{pmatrix} 1-\epsilon & \epsilon \\ \epsilon & 1-\epsilon \end{pmatrix}$$

where ϵ is the probability of transmission error.

1. What is the entropy of the source, $H(X)$?
2. What is the probability distribution of the outputs, $p(Y)$, and the entropy of this output distribution, $H(Y)$?
3. What is the joint probability distribution for the source and the output, $p(X, Y)$, and what is the joint entropy, $H(X, Y)$?
4. What is the mutual information of this channel, $I(X; Y)$?
5. How many values are there for ϵ for which the mutual information of this channel is maximal? What are those values, and what then is the capacity of such a channel in bits?
6. For what value of ϵ is the capacity of this channel minimal? What is the channel capacity in that case?

Solution:

1. Entropy of the source, $H(X)$, is $\underline{\text{1 bit}}$.
2. Output probabilities are:
   $p(y = 0) = (0.5)(1 - \epsilon) + (0.5)\epsilon = 0.5$ and $p(y = 1) = (0.5)(1 - \epsilon) + (0.5)\epsilon = 0.5$.

   Entropy of this distribution is $\underline{H(Y) = 1 \text{ bit}}$, just as for the entropy $H(X)$ of the input distribution.
3. Joint probability distribution $p(X, Y)$ is

   $$\begin{pmatrix} 0.5(1-\epsilon) & 0.5\epsilon \\ 0.5\epsilon & 0.5(1-\epsilon) \end{pmatrix}$$

   and the entropy of this joint distribution is $H(X, Y) = -\sum_{x,y} p(x, y) \log_2 p(x, y)$

   $$= -(1 - \epsilon) \log(0.5(1 - \epsilon)) - \epsilon \log(0.5\epsilon) = (1 - \epsilon) - (1 - \epsilon) \log(1 - \epsilon) + \epsilon - \epsilon \log(\epsilon)$$

   $$= \underline{1 - \epsilon \log(\epsilon) - (1 - \epsilon) \log(1 - \epsilon)}$$

4. The mutual information is $I(X; Y) = H(X) + H(Y) - H(X, Y)$, which we can evaluate from the quantities above as: $\underline{1 + \epsilon \log(\epsilon) + (1 - \epsilon) \log(1 - \epsilon)}$.
5. In the $\underline{\text{two}}$ cases of $\underline{\epsilon = 0}$ and $\underline{\epsilon = 1}$ (perfect transmission, and perfectly erroneous transmission), the mutual information reaches its maximum of $\underline{\text{1 bit}}$ and this is also then the channel capacity.
6. If $\underline{\epsilon = 0.5}$, the channel capacity is minimal and equal to $\underline{0}$.

Exercise 8

Consider Shannon's third theorem, the *Channel Capacity Theorem,* for a continuous communication channel having bandwidth W Hertz, perturbed by additive white Gaussian noise of power spectral density N_0, and average transmitted power P.

(a) Is there any limit to the capacity of such a channel if you increase its signal-to-noise ratio $\dfrac{P}{N_0 W}$ without limit? If so, what is that limit?

(b) Is there any limit to the capacity of such a channel if you can increase its bandwidth W in Hertz without limit, but while not changing N_0 or P? If so, what is that limit?

Solution:

(a) The capacity of such a channel, in bits per second, is

$$C = W \log_2 \left(1 + \frac{P}{N_0 W}\right)$$

Increasing the quantity $\frac{P}{N_0 W}$ inside the logarithm without bounds causes the channel capacity to increase monotonically and without bounds.

(b) Increasing the bandwidth W alone causes a monotonic increase in capacity, but only up to an asymptotic limit. That limit can be evaluated by observing that in the limit of small α, the quantity $\ln(1+\alpha)$ approaches α. In this case, setting $\alpha = \frac{P}{N_0 W}$ and allowing W to become arbitrarily large, C approaches the limit $\frac{P}{N_0} \log_2(e)$. Thus there are vanishing returns from endless increase in bandwidth, unlike the unlimited returns enjoyed from improvement in signal-to-noise ratio.