Systems and Information Theory - Tutorial SF/PHE-2

Information & Source Coding

Q1 Calculate the information content of the following:

a) A symbol with probability 0.99
b) A symbol with probability 0.01
c) A word of 6 independent symbols, half of which have probability 0.99, the other half having probability 0.01.
d) Any one of the digits 0-9 which occur with equal probability.
e) The symbols in a binary source where the probability of a '1' is twice the probability of a '0'.
f) The symbols from a 5-symbol source of maximum entropy.

(0.014 bits; 6.64 bits; 19.96 bits; 3.32 bits; 1.58 bits and 0.585 bits; 2.32 bits)

Q2 A game is to be designed for a mobile phone having a screen measuring 128 pixels wide by 64 pixels high. How much information is required to store the position of the dragon, assuming that it can appear anywhere on the screen with equal probability?

Show that it does not matter whether its position is encoded in x-y coordinates or by labelling the pixels in sequence starting at the top left hand corner.

(13 bits)

Q3 A source generates messages using six different symbols with probabilities of 1/2, 1/4, 1/8, 1/16, 1/32 and 1/32. Find the entropy of the source, and compare it with a source of six equally-likely symbols.

(1.94 bits/symbol; 2.58 bits/symbol)

Q4 A source generates binary symbols with probabilities of 0.25 and 0.75. Calculate the information rate

a) if each symbol is 1ms long

b) if the less likely symbol takes 2ms and the more likely 1ms

(815 bits/s; 652 bit/s)

Q5 A message is composed of the two symbols A and B, and a typical sequence of 20 symbols is given by

A A B B B A A A A B B A A A B B B A A A

(To give 20 pairs, assume that the next symbol is an A, since A is more likely than B)

Show (by counting the relative frequencies of successive symbol pairs i,j) that the following relationships apply:

a) $\sum_i p(i) = 1$ **b)** $\sum_i \sum_j p(i,j) = 1$

c) $\sum_j p(j\,|\,i) = 1$ **d)** $p(i) = \sum_j p(i,j)$

e) $p(i,j) = p(i).p(j\,|\,i)$ **f)** $p(j\,|\,i) = \dfrac{p(j)\,p(i\,|\,j)}{p(i)}$

nb: p(i,j) is the *joint probability* of the pair i,j and p(j|i) is the *conditional probability* that j follows i.

Calculate the entropy of the source, and its redundancy, assuming that dependency (or correlation) only extends over two successive symbols.

Q6 a) Calculate the entropy and source efficiency for the following message source.

Symbol	A	B	C	D	E	F
Probability	0.32	0.1	0.06	0.1	0.24	0.18

b) Find an instantaneous, compact binary code to represent the symbols using the Huffman method.

Calculate the average word length and hence the efficiency of the code.

c) Repeat **(b)** for a Fano code.

Is there any difference in performance between your codes?

(2.37 bits/symbol, 0.92; 2.42 bits/symbol, 0.98)

Q7 Is the binary coding shown below compact, instantaneous and uniquely decodable?

If necessary redesign the code to make it so.

Symbol	Prob.	Code
A	1/4	00
B	1/4	01
C	1/4	10
D	1/8	101
E	1/16	0110
F	1/16	1111

Q8 A binary source has symbol probabilities of 0.75 and 0.25.

Calculate the efficiency of a simple binary code, and show how the efficiency can be improved by using second- and third-extension coding, ie coding successive pairs and triples of symbols.

Assume that successive symbols are independant of each other.

(81.1%; 96%; 98%)

Time to complete: 3hours

MathsAid: Conditional Probability

1 Consider a bag containing 5 Red tokens and 5 Blue tokens: the probability of drawing a Red token is exactly 5/10 or 0.5. If a Red token **is** drawn there are now only 4 Red tokens in the bag *and if a second token is drawn the probabilities will be 4/9 Red and 5/9 Blue*.

2 Thus P{Red on first draw } = 0.5

and P{Red on second draw, given that a Red was drawn first} = 0.444...

also P{Blue on second draw, given that a Red was drawn first} = 0.555...

Or more compactly $P(R|R) = 0.444...$ and $P(B|R) = 0.555...$

In general terms,

> the ***Conditional Probability*** $P(B|A)$ is the ***Probability of B, GIVEN that A has occurred***

Note carefully that B is conditional on A.

3 If the Red token is replaced before the second draw, then we are just repeating the circumstances of the first draw and $P(R) = 0.5$ on second draw, which is now *independent* of the first event. Independent events are unconnected or uncorrelated.

Thus,

> **$P(B|A) = P(B)$ if the events are independent**

4 The joint probability

P{Red on first draw, Blue on second draw} = $P(R,B)$

$= P(R).P(B|R)$ using the product law of probabilities.

Hence, if the events are independent, $P(R,B) = P(R).P(B)$ as before.

5 The general result can be written as:

$$\mathbf{P(B \mid A) = \frac{P(A,B)}{P(A)}}$$

6 Clearly, the joint event

P{Red token first, Blue token second} = P{Red first | Blue second}.P{Blue second}
and if we now distinguish carefully between first and second events so that

$P(R_1,B_2) = P(R_1|B_2).P(B_2)$ whereas before we had

$P(R_1,B_2) = P(R_1).P(B_2|R_1)$ then, in general terms

$$\mathbf{P(B_2 \mid A_1) = \frac{P(B_2).P(A_1 \mid B_2)}{P(A_1)}}$$

which is a simplified form of ***Bayes Theorem***.