INSTRUCTOR VERSION WITH ANSWERS & FEEDBACK SHOWN

June 6, 2017

Mathematical Theory of Communication

Center for Science of Information, A National Science Foundation Science & Technology Center `http://soihub.org`

Questions on *Mathematical Theory of Communication*

Section 1: Coding

1. In the corresponding video, the messages "0", "10", "11" were assigned for sunny, rainy, and moderate, respectively. Why was the message "1" not assigned?

   Solution: The messages do not have length and the message "1" would conflict with the messages "10" and "11", causing the encryption to not be unqiuely decodable.

2. Suppose a message has four possibilities and we encode them as "00", "01", "10", "11". Why might this not be the most efficient scheme? Recall that we have not taken into account the probability of each message.

   Solution: If message encoded by "00" appears all the time, and the other three never appear, it would just be better to encode that message with a single character to improve overall efficiency.

Section 2: Entropy with Variance

3. Suppose message M consists of L words, each of which could be n different possibilities. What is the maximum entropy message M could contain?

   Solution: Each character has maximum entropy $\log_2(n)$, so the message has maximum entropy $L\log_2(n)$.

4. Suppose there are n outcomes to a random process. If this random process occurs once, and the outcome is given, what are possible conditions so that the entropy of the process is NOT $\log_2(n)$?

   Solution: If the probability of all outcomes are uniform, then the entropy for each outcome will be $\frac{1}{n}\log_2(n)$, and the entropy of entire process is then $\log_2(n)$. Thus, one possible condition for the entropy of the process to not be $\log_2(n)$ is that the probability of all outcomes are not equivalent, so that some outcomes are more likely than others.

Section 3: Huffman Coding

5. Suppose we had a fixed length encoding scheme for 7 outcomes. How many bits would be needed? What is the average length of the encoding scheme?

   Solution: Because 2 bits can only decipher $2^2 = 4$ messages, we need 3 bits to encode and decode 7 messages, as $2^3 = 8 \geq 7$. Moreover, the encoding scheme is fixed lengths, so each message, and therefore the average length, is 3 bits.

6. Now suppose we are given the following distribution for 7 outcomes. Determine a Huffman encoding scheme for the outcomes:

Outcome	Probability
A	$\frac{1}{8}$
B	$\frac{1}{4}$
C	$\frac{1}{16}$
D	$\frac{1}{16}$
E	$\frac{1}{8}$
F	$\frac{1}{8}$
G	$\frac{1}{4}$

   Solution: One possible encoding appears below. There are many possibilities, but the average length of the schemes are the same.

Outcome	Message
A	111
B	01
C	1101
D	1100
E	101
F	100
G	00

7. What is the average length of the Huffman encoding scheme? How does it compare to the fixed length encoding scheme?

   Solution: The average length is $\frac{42}{16} = \frac{21}{8}$, which is less than 3, so the Huffman encoding scheme performs better than the fixed length encoding scheme.

8. What is the entropy of the distribution? How does it compare to the average length of the Huffman encoding scheme? Is it possible for the Huffman encoding scheme to do better?

   Solution: The entropy of the distribution is

$$\frac{1}{8}\log_2\left(\frac{1}{8}\right) + \frac{1}{4}\log_2\left(\frac{1}{4}\right) + \frac{1}{16}\log_2\left(\frac{1}{16}\right) + \cdots + \frac{1}{4}\log_2\left(\frac{1}{4}\right) = \frac{21}{8},$$

   which is the same as the average length of the Huffman encoding scheme. Equality occurs because the probability of all outcomes are powers of 2. It is not possible for Huffman scheme to do better, because any encoding scheme must have average length at least as great as the entropy.

Section 4: Shannon's First Theorem

9. Suppose we are given a distribution in which 0 shows up with probability $p = 0.8$ and 1 shows up with probability $p = 0.2$. Determine a Huffman encoding scheme for the outcomes and compute the difference between the average length of the encoding scheme and the entropy of the distribution.

Solution: If we just assign the messages 0 and 1 to the corresponding outcomes, then the average length of the encoding scheme is 1. The entropy of the distribution is

$$-0.2\log_2(0.2) - 0.8\log_2(0.8) \approx 0.7219,$$

and the difference is approximately 0.2781.

10. Now, group outcomes in series of two, so that there are now 4 possible outcomes. Determine a Huffman encoding scheme for the outcomes and compute the difference between the average length of the encoding scheme and the entropy of the distribution.

Solution:

Outcome	Probability	Message
00	0.64	0
01	0.16	10
10	0.16	110
11	0.04	111

The average length is now 1.5 while the entropy is now 1.4439, so the difference is approximately 0.0561.

11. How does the difference in the first scheme compare to the difference in the second scheme? This is the idea behind Shannon's First Theorem!

Solution: The difference in the second scheme is smaller than the difference in the first scheme even though no new information was gained!

Section 5: Kraft's Inequality

12. Determine a Huffman encoding scheme for the following distribution on 4 outcomes:

Outcome	Probability
A	$\frac{1}{12}$
B	$\frac{1}{3}$
C	$\frac{1}{12}$
D	$\frac{1}{2}$

Solution: One possible scheme is:

Outcome	Message
A	111
B	10
C	110
D	0

Regardless, the encoded message for D must have length 1, the encoded message for B must have length 2, and the encoded messages for A and C must have length 3 each, for optimality.

13. Confirm Kraft's Inequality for the previous problem.

Solution: We have

$$2^{-3} + 2^{-2} + 2^{-3} + 2^{-1} = 1,$$

so Kraft's Inequality holds, and the encryption is uniquely decodable.

14. Prove the following lengths in bits for the encoding of each of the following 9 possible outcomes as instantaneous codes is not uniquely decodable given a binary alphabet.

Outcome	Length in Bits of Encoding
A	3
B	4
C	2
D	5
E	6
F	3
G	3
H	2
I	5

Solution: In a binary alphabet, we use the base 2 in Kraft's Inequality:

$$2^{-3} + 2^{-4} + 2^{-2} + 2^{-5} + 2^{-6} + 2^{-3} + 2^{-3} + 2^{-2} + 2^{-5} = \frac{65}{64} \geq 1.$$

Hence, the code is not uniquely decodable.

Section 6: Channels

15. What problems might a lossy channel present? Although it may not be possible to guarantee correctness, what are some ways we could increase accuracy of the encoding and decoding process?

Solution: A lossy channel can flip bits in a transmission, resulting in the incorrect message being received. Accuracy can be increased with more redundancy in the message, for example, if the message were repeated twice.

16. Let X and Y be random variables representing a roll of a die such that Y is the result of the die and $X = 0$ if $Y = 6$ and $X = 1$ otherwise. What is the entropy of Y given X, $H(Y|X)$? What is the entropy of X given Y, $H(X|Y)$?

Solution: The entropy of Y given X is

$$-\frac{1}{6}\log_2(1) - \frac{5}{6}\log_2\frac{1}{5} \approx 1.935.$$

On the other hand, if Y is known, then X must also be known, so $H(X|Y) = 0$.

Section 7: Conditional and Joint Entropy

17. A cruel and unusual teacher distributes grades according to the flip of a coin. Suppose X is a random variable which represents the outcome of the coin. That is, $X = 1$ if the coin is heads and $X = 0$ if the coin is tails. Now, suppose Y is a random variable representing the grade given by the teacher. If the coin is heads, $Y = 100$ with probability $p = 0.5$, $Y = 80$ with probability $p = 0.25$ and $Y = 60$ with probability $p = 0.25$. On the other hand, if the coin is tails, $Y = 100$ with probability 0, $Y = 80$ with probability $p = \frac{1}{3}$ and $Y = 60$ with probability $p = \frac{2}{3}$.

    (a) Determine the conditional entropy of Y given X, $H(Y|X)$.

Solution: The conditional entropy of Y is

$$\frac{1}{2}\left(-0.5\log_2 0.5 - 0.25\log_2 0.25 - 0.25\log_2 0.25\right) +$$

$$\frac{1}{2}\left(-\frac{1}{3}\log_2\frac{1}{3} - \frac{2}{3}\log_2\frac{2}{3}\right) \approx 1.209148$$

(b) Determine the conditional entropy of X given Y, $H(X|Y)$.

Solution: The probaility of getting $Y = 100$ is

$$\frac{1}{2}\cdot\frac{1}{2} = \frac{1}{4}.$$

The probability of getting $Y = 80$ is

$$\frac{1}{2}\left(\frac{1}{4}+\frac{1}{3}\right) = \frac{7}{24}.$$

The probability of $X = 1$ given $Y = 80$ is

$$\frac{\frac{1}{4}}{\frac{1}{4}+\frac{1}{3}} = \frac{3}{7}.$$

The probability of getting $Y = 60$ is

$$\frac{1}{2}\left(\frac{1}{4}+\frac{2}{3}\right) = \frac{11}{24}.$$

The probability of $X = 1$ given $Y = 60$ is

$$\frac{\frac{1}{4}}{\frac{1}{4}+\frac{2}{3}} = \frac{3}{11}.$$

The conditional entropy of X is

$$\frac{1}{4}\log_2 1 +$$

$$\frac{7}{24}\left(-\frac{3}{7}\log_2\left(\frac{3}{7}\right) - \frac{4}{7}\log_2\left(\frac{4}{7}\right)\right) +$$

$$\frac{11}{24}\left(-\frac{3}{11}\log_2\left(\frac{3}{11}\right) - \frac{8}{11}\log_2\left(\frac{8}{11}\right)\right) \approx 0.674810$$

(c) Determine the joint entropy $H(X, Y)$ and confirm that

$$H(X, Y) = H(Y|X) + H(X) = H(X|Y) + H(Y).$$

Solution: The entropy of X is 1 bit, and the entropy of Y is

$$-\frac{1}{4}\log_2\frac{1}{4} - \frac{7}{24}\log_2\frac{7}{24} - \frac{11}{24}\log_2\frac{11}{24} \approx 1.5343372$$

so $H(X) + H(Y|X) \approx 1 + 1.209148 = 2.209148$ and

$$H(Y) + H(X|Y) \approx 1.534337 + 0.674810 = 2.209147$$

and the two values are indeed the same.