Name: .... Student ID: ....

UNSW SCHOOL OF MATHEMATICS AND STATISTICS

MATH3411 INFORMATION CODES AND CIPHERS

2017 S2 **TEST 1** **VERSION A**

- Time Allowed: **45 minutes**

For multiple choice questions, **circle the correct answer**;
each multiple choice question is worth **1 mark**.
For written answer questions, **use extra paper**.
Staple all papers together when finished.

1. There may be an error in the check digit in the ISBN number 0-76-535615-4. The correct check digit is

(a) 2 (b) 5 (c) 7 (d) X (e) None of these

2. The binary code $C = \{101010, 110011, 101101, 111111\}$ has minimum distance

(a) 1 (b) 2 (c) 3 (d) 4 (e) None of these

For the next 3 questions, let C be a binary linear code with check matrix

$$H = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 \end{pmatrix}$$

Assume that the check bits correspond to the first three columns.

3. How many codewords are there in C?

(a) 2 (b) 3 (c) 4 (d) 5 (e) 8

4. The codeword encoding the message 01 in code C has weight

(a) 1 (b) 2 (c) 3 (d) 4 (e) None of these

5. Which of the following is a generator matrix G for the code C?

(a) $\begin{pmatrix} 1&1&1&1&0 \\ 0&1&0&1&1 \end{pmatrix}$ (b) $\begin{pmatrix} 1&1&1&1&0 \\ 1&1&0&0&1 \end{pmatrix}$ (c) $\begin{pmatrix} 0&0&0&0&0 \\ 1&1&0&0&1 \end{pmatrix}$ (d) $\begin{pmatrix} 1&1&0&0&1 \\ 1&1&0&0&1 \end{pmatrix}$ (e) None of these

6. A binary code C has minimum distance $d = 8$. Suppose this is used to correct a errors and detect b errors. Which of the following pairs (a, b) **does not** give a valid strategy for decoding C?

(a) $(0, 7)$ (b) $(1, 6)$ (c) $(2, 5)$ (d) $(3, 4)$ (e) $(4, 3)$

7. Consider a compression code with codewords $\mathbf{c}_1 = 10$, $\mathbf{c}_2 = 11$, $\mathbf{c}_3 = 100$, $\mathbf{c}_4 = ?$ where $\mathbf{c}_4$ is to be chosen from the list of four possibilities below.
Which choice, if any, of $\mathbf{c}_4$ makes the resulting code uniquely decodable?

(a) $\mathbf{c}_4 = 0$ (b) $\mathbf{c}_4 = 1$ (c) $\mathbf{c}_4 = 011$ (d) $\mathbf{c}_4 = 111$ (e) None of these

8. A binary UD-code has codeword lengths (not necessarily in order) 1, 3, 3, 4, ℓ. What is the smallest value of ℓ for which such a code exists?

(a) $\ell = 1$ (b) $\ell = 2$ (c) $\ell = 3$ (d) $\ell = 4$ (e) $\ell = 5$

9. Consider the standard ternary I-code with codeword lengths 1, 3, 3, 3, 3.
The codeword $\mathbf{c}_5$ corresponding to symbol s_5 is given by

(a) 000 (b) 102 (c) 110 (d) 111 (e) None of these

10. A Markov source $S = \{s_1, s_2, s_3\}$ has transition matrix M.
The Huffman code for the equilibrium distribution is $\text{Huff}_E = [1,\ 00,\ 01]$.
(That is, $\mathbf{c}_1 = 1$, $\mathbf{c}_2 = 00$ and $\mathbf{c}_3 = 01$.)
The Huffman codes for the columns of M are given by

$$\text{Huff}_1 = [00,\ 1,\ 01] \qquad \text{Huff}_2 = [0,\ 10,\ 11] \qquad \text{Huff}_3 = [11,\ 10,\ 0]$$

The string 001101100 decodes under the Markov Huffman encoding as

(a) $s_1s_1s_1s_3s_1s_1$ (b) $s_2s_3s_3s_1s_1$ (c) $s_2s_3s_1s_3s_2s_2s_2$ (d) $s_2s_2s_1s_2s_3$ (e) None of these

11. [**5 marks**]

(a) Use the Kraft-MacMillan Theorem to show that there is no uniquely decodable **ternary** code with codeword lengths $1, 1, 2, 2, 2, 3$, respectively.

(b) Symbol s_1 of the source $S = \{s_1, s_2\}$ occurs with probability $4/5$ and symbol s_2 occurs with probability $1/5$. Find a uniquely decodable binary code of minimal average length for S^2, assuming that successive symbols occur independently, and state the average length per original source symbol of the code.

Name: .... Student ID: ....

UNSW School of Mathematics and Statistics

MATH3411 Information Codes and Ciphers

2017 S2 **TEST 1** **VERSION B**

- Time Allowed: **45 minutes**

For multiple choice questions, **circle the correct answer**; each multiple choice question is worth **1 mark**. For written answer questions, **use extra paper**. Staple all papers together when finished.

1. A message is sent using a 5-character 8-bit ASCII code which encodes characters in blocks of four together with a 5th character which is used as a check character for even parity in rows and columns, similar to the 9-character 8-bit ASCII code.

The message 10101010 10010110 11001101 00111010 11000011 is received. Assuming at most two errors, which of the following bits could be incorrect?

(a) 3rd (b) 5th (c) 21th (d) 25th (e) None of these

2. Consider a binary channel with bit-error probability p, where errors in different positions are independent. Suppose that a codeword $\mathbf{x}$ is sent from a binary code with minimum distance 7 and codeword length 12. The probability that one or more errors are correctly corrected using a minimum distance decoding strategy is

$w = 12p(1-p)^{11}$ $x = 66p^2(1-p)^{10}$ $y = 110p^3(1-p)^9$ $z = 220p^3(1-p)^9$

The probability that one or more errors are correctly corrected using a minimum distance decoding strategy is

(a) w (b) $w+x$ (c) $w+x+y$ (d) $w+x+z$ (e) $x+y+z$

3. Let C be the code of all vectors $\mathbf{x} = x_1x_2x_3x_4 \in \mathbb{Z}_3^4$ satisfying the check equations

$$\begin{array}{ccccccccc} x_1 & & & + & 2x_3 & + & x_4 & \equiv & 0 \pmod 3 \\ x_1 & + & x_2 & + & & + & x_4 & \equiv & 0 \pmod 3 \end{array}$$

There are two information bits but you are not told in which positions they lie. Which of the following codewords could possibly encode the message 10?

(a) 1000 (b) 1100 (c) 1110 (d) 0111 (e) None of these

4. Let C be the ternary linear code with generator matrix

$$G = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 2 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 \end{pmatrix}$$

How many codewords are there in C?

(a) 3 (b) 8 (c) 16 (d) 27 (e) 81

5. For the code C of Question 4, assume that the first four bits are check bits. The codeword that encodes $\mathbf{m} = 001$ is then

(a) 0011101 (b) 0012101 (c) 1002001 (d) 1112001 (e) None of these

6. A binary linear code C has minimum distance $d = 4$ and length $n = 7$. The maximal possible number of information bits k for such a code is

(a) 1 (b) 2 (c) 3 (d) 4 (e) 5

7. A uniquely decodable code has codewords $\mathbf{c}_1 = 1$, $\mathbf{c}_2 = 10$, $\mathbf{c}_3 = 01$, $\mathbf{c}_4 = ?$. Which of the following codewords could $\mathbf{c}_4$ be?

(a) $\mathbf{c}_4 = 0$ (b) $\mathbf{c}_4 = 11$ (c) $\mathbf{c}_4 = 00$ (d) $\mathbf{c}_4 = 010$ (e) None of these

8. The minimum radix that would be needed to create a UD-code for the source

$$S = \{s_1, s_2, \ldots, s_7\}$$

with codeword lengths $1, 1, 2, 2, 3, 3, 3$, respectively is

(a) 2 (b) 3 (c) 4 (d) 5 (e) 6

9. Consider the standard binary I-code with codeword lengths 1, 3, 3, 3, 3. The codeword $\mathbf{c}_5$ corresponding to symbol s_5 is given by

(a) 001 (b) 010 (c) 100 (d) 110 (e) 111

10. Let $S = \{s_1, s_2\}$ be a source with probabilities $p_1 = \frac{3}{4}$, $p_2 = \frac{1}{4}$. The average length per original symbol of a radix 3 Huffman code for the second extension $S^{(2)}$ of this source is

(a) $\frac{23}{16}$ (b) $\frac{23}{32}$ (c) $\frac{27}{16}$ (d) $\frac{27}{32}$ (e) $\frac{5}{8}$

11. **[5 marks]** A Markov source $S = \{s_1, s_2, s_3\}$ has transition matrix $M = \frac{1}{10}\begin{pmatrix} 7 & 2 & 7 \\ 1 & 5 & 2 \\ 2 & 3 & 1 \end{pmatrix}$.

i) Show that $\mathbf{p} = \frac{1}{5}\begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix}$ is an equilibrium vector for M.

ii) Show that the Huffman code Huff_E for the probability distribution given by $\mathbf{p}$ is $\text{Huff}_E = [0, 10, 11]$ (that is, $\mathbf{c}_1 = 0$, $\mathbf{c}_2 = 10$ and $\mathbf{c}_3 = 11$), and find the average codeword length L_E for Huff_E.

iii) Assuming that the Huffman codes for the columns of M are given

$$\text{Huff}_1 = [0, 11, 10] \qquad \text{Huff}_2 = [01, 1, 00] \qquad \text{Huff}_3 = [0, 10, 11]$$

use Markov Huffman encoding to encode the string of source symbols $s_1 s_2 s_1 s_3 s_1$.