MATH3411 INFORMATION, CODES & CIPHERS

Test 2 2018 S2 SOLUTIONS

Version A

Multiple choice: **b**, **c**, **a**, **e**, **c**, **c**, **d**, **b**, **b**, **a**

1. **(b)**: This is the only codeword which has 1221 in the information positions.

2. **(c)**: If the sent codeword is $\mathbf{x}$, then $\mathbf{y} = \mathbf{x} + a\mathbf{e}_i$ for some $a \in \mathbb{Z}_3$ and $i \in \{1, \dots, 7\}$.
Since $S(\mathbf{y}) = H\mathbf{y}^T = 022^T = 2H\mathbf{e}_5^T$, twice the 5th column of H, and that $S(\mathbf{y}) = aH\mathbf{e}_i^T$, we see that $a = 2$ and $i = 5$, so $\mathbf{x} = \mathbf{y} - 2\mathbf{e}_5 = 0021010$, so $\mathbf{m} = 1010$.

3. **(a)**: 1001 and 0121 are the only codewords and, of these, only 1001 can encode 10.

4. **(e)**: By trial and error, we see that none of the four words for $\mathbf{c}_4$ are suitable:

   (a) $\mathbf{c}_4\mathbf{c}_4 = \mathbf{c}_2$ (b) $\mathbf{c}_4 = \mathbf{c}_1\mathbf{c}_2$ (c) $\mathbf{c}_4 = \mathbf{c}_1\mathbf{c}_1\mathbf{c}_1$ (d) $\mathbf{c}_4\mathbf{c}_2\mathbf{c}_4 = \mathbf{c}_1\mathbf{c}_3\mathbf{c}_3\mathbf{c}_1$

5. **(c)**: The Kraft-McMillan number $K = \sum \frac{1}{2^{\ell_i}}$ must be at most 1 for UD codes.
Testing values of $\ell = 1, 2, 3, \dots$ gives us that $\ell = 4$ is the minimum length that satisfies this.
You can also draw a decision tree.

6. **(c)**: Encode the message $ba\bullet$:

	subinterval start	**width**
begin	0	1
b	0.4	0.3
a	0.4	$0.4 \times 0.3 = 0.12$
$\bullet$	$0.4 + 0.7 \times 0.12 = 0.484$	$0.3 \times 0.12 = 0.036$

   so the message encodes as a number in the interval $[0.484, 0.52)$.

7. **(d)**: 1. b 2. a 3. aa 4. aab 5. aaa

8. **(b)**: Use one dummy symbol (and Knuth's Theorem to save time).

9. **(b)**: $\displaystyle\lim_{n\to\infty} \frac{L_3^{(n)}}{n} = H_3(S) = -0.4\log_3 0.4 - 0.3\log_3 0.3 - 0.3\log_3 0.3 - 0.3\log_3 0.3 \approx 1.16.$

10. **(a)**:

p_i	$\frac{1}{p_i}$	ℓ_i	code
0.4	2.5	2	00
0.3	3.3	2	01
0.2	5	3	100
0.1	10	4	1010

   So, the message $\mathbf{m} = s_1s_4s_2$ is encoded as 00101001.

11. (a) Let us now calculate the Huffman codes $\mathrm{Huff_E}$, $\mathrm{Huff}_{(1)}$, $\mathrm{Huff}_{(2)}$, $\mathrm{Huff}_{(3)}$:

Source	p_i	Huff_E	Source	p_i	$\mathrm{Huff}_{(1)}$	Source	p_i	$\mathrm{Huff}_{(2)}$	Source	p_i	$\mathrm{Huff}_{(3)}$
s_1	$\frac{6}{17}$	00	s_1	0.7	0	s_1	0.2	10	s_1	0.1	01
s_2	$\frac{7}{17}$	1	s_2	0.2	10	s_2	0.6	0	s_2	0.4	00
s_3	$\frac{4}{17}$	01	s_3	0.1	11	s_3	0.2	11	s_3	0.5	1

(b) The average lengths of these codes

$$L_E = \frac{27}{17} \approx 1.59 \qquad L_{(1)} = 1.3 \qquad L_{(2)} = 1.4 \qquad L_{(3)} = 1.5$$

The Markov Huffman code has average length

$$L_M = \frac{6}{17}L_{(1)} + \frac{7}{17}L_{(2)} + \frac{4}{17}L_{(3)} = \frac{6}{17}1.3 + \frac{7}{17}1.4 + \frac{4}{17}1.5 \approx 1.39$$

c) We encode $s_1 s_3 s_2 s_1$:

symbol	code to use	encoded symbol
s_1	Huff_E	00
s_3	$\mathrm{Huff}_{(1)}$	11
s_2	$\mathrm{Huff}_{(3)}$	00
s_1	$\mathrm{Huff}_{(2)}$	10

so this is encoded as 00110010 .

Version B

Multiple choice: **a**, **d**, **d**, **d**, **b**, **b**, **e**, **a**, **a**, **c**

1. **(a)**: This is the only codeword which has 1221 in the information positions.
2. **(d)**: If the sent codeword is $\mathbf{x}$, then $\mathbf{y} = \mathbf{x} + a\mathbf{e}_i$ for some $a \in \mathbb{Z}_3$ and $i \in \{1, \dots, 7\}$.
   Since $S(\mathbf{y}) = H\mathbf{y}^T = 002^T = H\mathbf{e}_3^T$, the 3rd column of H, and $S(\mathbf{y}) = aH\mathbf{e}_i^T$,
   we see that $a = 1$ and $i = 3$, so $\mathbf{x} = \mathbf{y} - \mathbf{e}_3 = 1101111$, so $\mathbf{m} = 1101$.
3. **(d)**: 1201 is the only codeword here that can encode 10.
4. **(d)**: This choice of $\mathbf{c}_4$ gives an I-code and thus a UD-code.
5. **(b)**: The Kraft-McMillan number $K = \sum \frac{1}{2^{\ell_i}}$ must be at most 1 for UD codes.
   Testing values of $\ell = 1, 2, 3, \dots$ gives us that $\ell = 3$ is the minimum length that satisfies this.
   You can also draw a decision tree.
6. **(b)**: Encode the message $ab\bullet$:

	subinterval start	**width**
begin	0	1
a	0	0.4
b	$0 + 0.4 * 0.4 = 0.16$	$0.4 \times 0.4 = 0.16$
$\bullet$	$0.16 + 0.8 \times 0.16 = 0.288$	$0.2 \times 0.16 = 0.032$

   so the message encodes as a number in the interval $[0.288, 0.32)$.
7. **(e)**: 1. b 2. ba 3. baa 4. $baab$ 5. $baaa$

8. **(a)**: Use one dummy symbol (and Knuth's Theorem to save time).
9. **(a)**: $\lim_{n\to\infty} \frac{L_4^{(n)}}{n} = H_4(S) = -0.4\log_4 0.4 - 0.2\log_4 0.2 - 0.2\log_4 0.2 - 0.1\log_4 0.1 - 0.1\log_4 0.1 \approx 1.06.$

10. **(c)**:

p_i	$\frac{1}{p_i}$	ℓ_i	code
0.4	2.5	2	00
0.2	5	3	010
0.2	5	3	011
0.1	10	4	1000
0.1	10	4	1001

   So, the message $\mathbf{m} = s_1 s_4 s_2$ is encoded as 001000010.

11. (a) Let us now calculate the Huffman codes $\mathrm{Huff}_{\mathrm{E}}$, $\mathrm{Huff}_{(1)}$, $\mathrm{Huff}_{(2)}$, $\mathrm{Huff}_{(3)}$:

Source	p_i	Huff_E	Source	p_i	$\mathrm{Huff}_{(1)}$	Source	p_i	$\mathrm{Huff}_{(2)}$	Source	p_i	$\mathrm{Huff}_{(3)}$
s_1	$\frac{10}{27}$	00	s_1	$\frac{1}{4}$	10	s_1	$\frac{1}{2}$	1	s_1	$\frac{1}{4}$	00
s_2	$\frac{13}{27}$	1	s_2	$\frac{2}{3}$	0	s_2	$\frac{1}{3}$	00	s_2	$\frac{1}{2}$	1
s_3	$\frac{4}{27}$	01	s_3	$\frac{1}{12}$	11	s_3	$\frac{1}{6}$	01	s_3	$\frac{1}{4}$	01

(b) The average lengths of these codes are

$$L_{(1)} = \frac{4}{3} \qquad L_{(2)} = 1.5 \qquad L_{(3)} = 1.5$$

The Markov Huffman code has average length

$$L_M = \frac{10}{27}L_{(1)} + \frac{13}{27}L_{(2)} + \frac{4}{27}L_{(3)} = \frac{10}{27}\frac{4}{3} + \frac{13}{27}1.5 + \frac{4}{27}1.5 = \frac{233}{162} \approx 1.44$$

c) We encode $s_1 s_3 s_2 s_1$:

symbol	code to use	encoded symbol
s_1	Huff_E	00
s_3	$\mathrm{Huff}_{(1)}$	11
s_2	$\mathrm{Huff}_{(3)}$	1
s_1	$\mathrm{Huff}_{(2)}$	1

so this is encoded as 001111.