Australian National University

ASSIGNMENT COVER SHEET

ANU College of Engineering and Computer Science
Australian National University
Canberra ACT 0200 Australia
www.anu.edu.au

+61 2 6125 5254

This coversheet must be attached to the front of your assessment

The assessment is due on 23/10/2023, 9:05 AM unless otherwise specified in the course outline.

Student ID	U7540836
Student Name	Nanthawat Anancharoenpakorn
Course Code	
Course Name	Information Theory
Assignment Item	Assignment 3
Due Date	23/10/2023
Date Submitted	22/10/2023

I declare that this work:

- ☐ upholds the principles of academic integrity, as defined in the ANU Policy: Code of Practice for Student Academic Integrity;
- ☐ is original, except where collaboration (for example group work) has been authorized in writing by the course convener in the course outline and/or Wattle site;
- ☐ is produced for the purposes of this assessment task and has not been submitted for assessment in any other context, except where authorized in writing by the course convener.
- ☐ gives appropriate acknowledgment of the ideas, scholarship, and intellectual property of others insofar as these have been used.
- ☐ in no part involves copying, cheating, collusion, fabrication, plagiarism, or recycling.

Signature Nanthawat Anancharoenpakorn

SUBJECT: NO: DATE: / /

Question 1

(a) $H(X_1) = 4 \times \frac{1}{4} \times \log\left(\frac{1}{0.25}\right) = 2$ bits

(b) $H(X_1, X_2, \ldots, X_{52}) = -\log\left(\frac{1}{52!}\right)$

$= 225.5$ bits

Question 2

(I)

code	p(x)	(huffman code)		
00	0.3	0.3	0.4	0.6
01	0.3	0.3	0.3	0.4
10	0.2	0.2	0.3	
110	0.1			
111	0.1			

- 0.2, - 0.4, - 0.6, - 1

avg. code length $= 0.3(2) + 0.3(2) + 0.2(2) + 0.1(6)$

$= 0.6 + 0.6 + 0.4 + 0.6 = 2.2$

(Shannon-Fano)

p(x)	F(x)	$\bar{F}(x)$	$\bar{F}(x)_2$	l(x)	code
0.3	0.3	0.15	0.001	3	001
0.3	0.6	0.45	0.011	3	011
0.2	0.8	0.7	0.101	3	101
0.1	0.9	0.85	0.1101	4	1101
0.1	1	0.95	0.1111	4	1111

avg. code length $= 3(0.3) \times 2 + 0.2(3) + 0.1(4) \times 2$

$= 3.2$ bits

(II) (b) $H(X) = 0.49 \log \frac{1}{0.49} + 0.26 \log \frac{1}{0.26} + 0.12 \log \frac{1}{0.12}$

$+ 0.04 \log \frac{1}{0.04} + 0.04 \log \frac{1}{0.04} + 0.03 \log \frac{1}{0.03} + 0.02 \log \frac{1}{0.02}$

$= 2.01$ bits

expected code length $= 1(0.49) + 2(0.26) + 3(0.12) + 0.04(5) \times 2 + 5 \times (0.03) + 5 \times 0.02 = 2.02$

(a)

code	x_i						
1	1	0.49	0.49	0.49	0.49	0.49	0.51
00	2	0.26	0.26	0.26	0.26	0.26	0.49
011	3	0.12	0.12	0.12	0.13	0.25	
01000	4	0.04	0.05	0.08	0.12		
010001	5	0.04	0.04	0.05			
01010	6	0.03	0.04				
01011	7	0.02					

- 1

(c)

code	x_i			
0	1	0.49	0.49	0.49
1	2	0.26	0.26	0.26
20	3	0.12	0.12	0.25
22	4	0.04	0.09	
210	5	0.04	0.04	
211	6	0.03		
212	7	0.02		

- 1

(III) (a)

		huffman	l(x)	shannon	l(x)
a	0.55	0	1	0	1
b	0.25	10	2	01	2
c	0.2	11	2	011	3

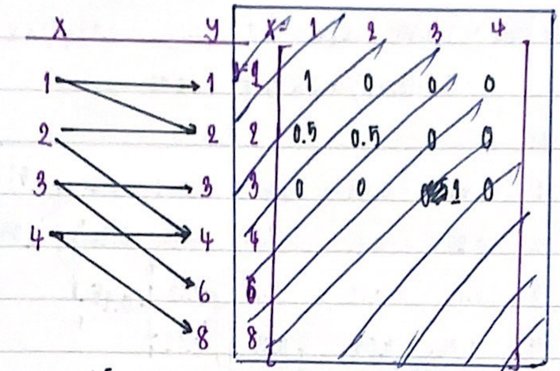
$E[l(X)] = 1.45, 1.65$

(b) $H(X) = 1.43$ the smallest integer is when $D = 3$.

Double A

SUBJECT : NO : DATE : / /

Question 3 (I) (a)



where $\xrightarrow{0.5}$ each

$$
\begin{array}{c|cccc}
 & x=1 & 2 & 3 & 4 \\
\hline
y=1 & 0.5 & 0 & 0 & 0 \\
2 & 0.5 & 0.5 & 0 & 0 \\
3 & 0 & 0 & 0.5 & 0 \\
4 & 0 & 0.5 & 0 & 0.5 \\
6 & 0 & 0 & 0.5 & 0 \\
8 & 0 & 0 & 0 & 0.5
\end{array}
\quad p(y|x)
$$

(b) since y is deterministic function of X and it is memoryless channel, $H(X|Y)=0$ then given prob of X are p_1, p_2, p_3 and $1-p_1-p_2-p_3$ where p_i represent prob of X in order.

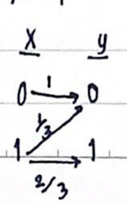
$I(X;Y) = H(X) - H(X|Y) = H(X)$ Then

$I(X;Y) = -p_1\log(p_1) - p_2\log(p_2) - p_3\log(p_3)$
$\quad - (1-p_1-p_2-p_3)\log(1-p_1-p_2-p_3)$ #

(c) In order to achieve max $I(X,Y)$ with zero error, we can select elements from X s.t. output or y is not overlap. For example, $\{1,3,4\}$ as input If we get 1 or 2 as output, we know that X must be 1. If we get 6 as output, we know that X is 3 for sure. Since there is no error, the max $I(X,Y)$ is max $H(X)$ which is 1.58. This is the maximum since we pick maximum number of input st. no overlap.

(II)

transition matrix $p(y|x)$ =
$$
\begin{array}{c|cc}
 & x=0 & 1 \\
\hline
y=0 & 1 & \frac{1}{3} \\
1 & 0 & \frac{2}{3}
\end{array}
$$



where $f = \frac{1}{3}$

$p = \dfrac{1(1-f)}{1+2^{H_2(f)/(1-f)}} = 0.416$ where $f = \frac{1}{3}$ than

we get capacity by using $H_2((1-f)p) - pH_2(f)$ then

$C = H_2(0.27) - 0.416 \times H_2(\frac{1}{3}) = 0.469$ #

(III) $I(X_1, X_2, \dots, X_n; Y_1, Y_2, \dots Y_n) = H(Y_1, Y_2, \dots Y_n) - H(Y_1, Y_2 \dots Y_n | X_1, X_2 \dots X_n)$ since $X \perp Z$ and $Y = X \oplus Z$ we get

$I(X_1, X_2, \dots X_n; Y_1, Y_2 \dots Y_n) = H(Y_1, Y_2, \dots Y_n) - H(Z_1, Z_2, \dots, Z_n)$

since $H(Y_i|X_i) \leq Z_i$, $H(Y_1, Y_2, \dots Y_n) - H(Z_1, Z_2, \dots Z_n) \geq n(H(Y_i|X_i) - H(Z_i)) \geq nH(p)$ where $C = 1 - H(p, 1-p)$

Question 4 (I)

since there are 3 random variables, and $p(x^n, y^n, z^n)$ are independent P lower bound $= 2^{-n(H(X)+H(Y)+H(Z)+7e)}$

P upper bound $= 2^{-n(H(X,Y,Z)-7e)}$

(II)

	occurance	symbol	prob
(a)	7	a	0.4
	5	b	0.33
	3	c	0.27

$H(X^n) = \frac{1}{15} \times 7\log(0.4) + 5\log(0.33) + 3\log(.27)$
$= 1.527$ bits

(b) $H(X) = -[0.4\log(0.4) + 0.33\log(0.33) + 0.27\log(0.27)]$
$= 1.566$

$|\tilde{H}(x) - H(x)| \leq \epsilon = |1.527 - 1.566| = 0.038 < 0.05$, then the sequence x is e-typical

(c) $|\bar{H}(X,Y) - H(X,Y)| \leq e$

$H(X,Y) = -\sum_{x,y} p(x,y)\log_2 p(x,y) = 3.058$

$$
\begin{array}{c|ccc}
 & d & e & f \\
\hline
a & 4 & 1 & 2 \\
b & 1 & 2 & 2 \\
c & 2 & 0 & 1
\end{array}
\quad \bar{H}(X,Y) = 2.955
$$

Hence,

$|2.955 - 3.058| = 0.1 > 0.05$, the sequence is not e-typical.

Double A