

IC - A2

I.

1. given,

Variable length codes

$$C(A) = 0 \quad \# \text{Tree 1}$$

$$C(B) = 100$$

$$C(C) = 101$$

$$C(D) = 110$$

$$C(E) = 111$$

$$\# \text{Tree 2}$$

$$C(A) = 00$$

$$C(B) = 01$$

$$C(C) = 10$$

$$C(D) = 110$$

$$C(E) = 111$$

(a) 01000111101

The codes are prefix free so,

$\rightarrow A \underline{1000111101}$

$\rightarrow A \underline{1000} 111101$

$\rightarrow A \underline{1000} 111101 \rightarrow AB \underline{011101}$

$\rightarrow ABA \underline{111101}$

$\rightarrow ABAEC$

(b)

$L(C)$ for Tree #1

wkT,

$$P(A) = 0.5 \quad P(B) = P(C) = P(D) = P(E) = 0.125$$

The shortest message averaged over many messages is the tree which has lower average length $L(C)$

$$L^{(C)}_{\#1} = \sum p(x) \cdot l(x)_{\#1} \quad x \in \{A, B, C, D, E\}$$

$$= 0.5 \times (17 + 0.125(3) + 0.125(3) + 0.125(3) \\ + 0.125(3))$$

$$= 0.5 + 4 \times 3 \times 0.125$$

$$= 0.5 + 4 \times 0.375$$

$$= 0.5 + 1.5$$

$$LCC = 2 \text{ bits}$$

$$L^{(C)}_{\#2} = \sum p(m) L^{(C)}_{\#2}$$

$$LCC_{\#2} = \sum p(m) \cdot l(m)$$

$$= 0.5(2) + 0.125 \times 2 + 0.125 \times 3 + 0.125 \times 2 \\ + 0.125 \times 3$$

$$= 1 + 0.25 + 0.375 + 0.25 + 0.375$$

$$= 2.25 \text{ bits}$$

\therefore Tree #1 has a lower average code length (2.0) bits than Tree #2 (2.25) bits.

$$(c) L_2 = 2.25 \text{ bits/symbol} \quad (\text{average})$$

↓
 length

for 100 symbol messages

$$\frac{f(l)}{200} = 100 * 2.25 = 225 \text{ bits}$$

For 100 symbol message the expected avg length is 225 bits

2. Initial uncertainty for a 3-digit number with no clues $H = \log_2(8) = 3$ bits

(a) Alice told number is odd.

the last digit is ~~odd~~ (1)

(b) Bob said it's not multiple of 3

then, not possible values are (0, 3, 6)

it should be (1 or 5 or 7)

(c) according to Charlie it has exactly two 1's then,

it only '5' because 1 has only 1 no. of 1's and 7 has 3 no. of 1's where 5 only has 2 no. of 1's

(a) by Alice

Possible numbers are

$$\{001, 011, 101, 111\}$$

$$H_A = \log_2(4) = 2$$

$$I = H - H_A$$

$$I = 3 - 2 = 1 \text{ bit}$$

(b) by Bob possible numbers are

$$\{001, 010, 100, 101, 111\}$$

$$H_B = \log_2(5) = 2.3219$$

$$I = H - H_B = 3 - 2.3219 \approx 0.678 \text{ bits}$$

(c) by charlie
possible values are {011, 101, 110}

$$H = \log_2 (3) \approx 1.585 \text{ bits}$$

$$I = H - H_C = 3 - 1.585$$

$$I \approx 1.415 \text{ bits}$$

(d) from above theory if all these clues are given Deb then,

- Deb can find exact answer, from - ②
- Hence the uncertainty for finding answer for Deb is 0.

$$\text{i.e } I = H - \log_2 (1) = 3 - 0 = 3 \text{ bits}$$

the info gained by deb is 3 bits.

~~(0, 1, 10, 0)~~

let,
if all N outcomes are equally likely

$$P_i = \frac{1}{N} \quad N = 2^3$$

wkt,

$$H(X) = - \sum_{i=1}^N \left(\frac{1}{N} \log_2 \frac{1}{N} \right) \xrightarrow{(00)} \sum_{i=1}^N \frac{1}{N} \log_2 \frac{1}{N}$$

$$H(X) = - \log_2 \left(\frac{1}{N} \right) \cdot \sum_{i=1}^N \frac{1}{N}$$

$$H(X) = - \log_2 \left(\frac{1}{N} \right) = \log_2 (N)$$

$I = H_{\text{before}} - H_{\text{after}}$ and for ~~H~~ H with

no clues $H_a = \log_2 (N) = \log_2 8 = 3$

given,

3.

I	$P(I)$	$\log_2(1/P(I))$	$P(I) \log_2(1/P(I))$
A	0.22	2.18	0.48
E	0.34	1.55	0.53
I	0.17	2.57	0.43
O	0.19	2.40	0.46
U	0.08	3.64	0.29
Total	1.00	12.34	2.19

(a) from above table

$$P(I) = 0.17$$

$$P(U) = 0.08$$

By given data

$$H_1 = 2.19$$

For probability to be $I_{\text{or}} U$

is $P(I \cup U) = 0.25$

$$P(I|I \cup U) = \frac{0.17}{0.25} = 0.68$$

$$P(U|I \cup U) = \frac{0.08}{0.25} = 0.32$$

~~$H_2 = 0.68 \log 0.68 + 0.32 \log 0.32$~~

$$H_2 = P(I|I \cup U) \log_2 P(I|I \cup U)$$

$$+ P(U|I \cup U) \log_2 (U|I \cup U)$$

$$H_2 = -0.68 \log_2 0.68 - 0.32 \log_2 0.32$$

$$H_2 = 0.378 + 0.526 \approx 0.904 \text{ bits}$$

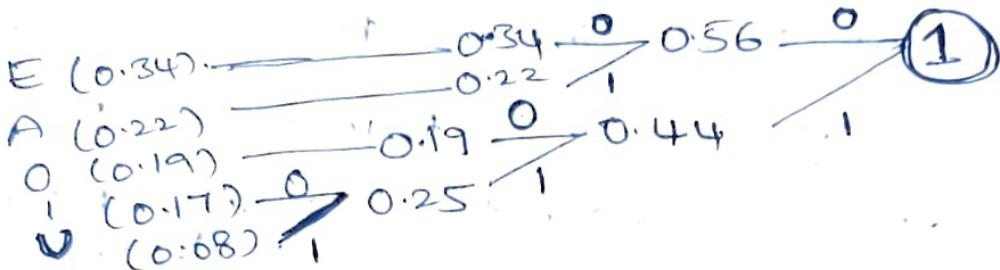
$$I = H_1 - H_2$$

$$I = 2.19 - 0.904$$

$$I = 1.286$$

So, the gain in information is 1.286 bits.

(b)



$$E = 00$$

$$A = 01$$

$$O = 10$$

$$I = 110$$

$$U = 111$$

$$L(C) = P(E)l(E) + P(A)l(A) + P(O)l(O) + P(I)l(I)$$

$$L(C) = 0.34(2) + 0.22(2) + 0.19(2) + 0.08(3) + P(U).l(U)$$

$$L(C) = 2.25 \rightarrow (1.5 + 0.75) + 0.75$$

$$L(C) = 2.25$$

(c)

Avg for 4 codeword (Vowel) is 2.25

for 100 vowels are

$$L_{100} = 2.25 \times 100 = 225 \text{ bits}$$

d) WKT the lower bound of Expectation of length $L(C)$ is entropy.

$$H(x) \leq L(C)$$

if bitdiddle transmits 1.97 bits
for 100 vowels the 1.97 per vowel
its below the level of given data
entropy.

If vowels are independent then there
definitely some certain error possibility
for data transmitted using bitdiddle
obtained code can be treated as lossy compression

Thus this ensures data distortion
during decoding i.e. decoded data can lead
to different ~~types~~ answers

so, this ensures data cannot be decoded
into unique value.

So Ben's code is non-uniquely decodable

4. (a) By given data.

$I = H - H_B$
for I to be small H_B should
be larger

$$H_B = - P(x_B) \log \frac{1}{P(x_B)}$$

For I to be change in entropy minimum
 H_B should be maximum probability department
(Information)

④ So, the least info department is EECS(VI) ✓
 ⑤ H_p must be maximum
 so if $P(c_1) \uparrow H_p \uparrow P(c_2) \propto H_p \uparrow$

Intuitively, the student department with max probability has very high uncertainty already i.e. high info even if involved (provided) it may not change much compare to others.

$$H_p = 0.38 \times \log_{10} 0.38 \Rightarrow I = H - H_p = 2.305 - 0.5304$$

course	probability	(Least possible info) $I = 1.7746$ bits	
		0.38	1
(b) VII	0.38		1
II	0.23	0.23	0.62
X	0.13	0.25	0
XVI	0.12	0	
III	0.07	0.14	
I	0.07	0	

$$I(C) = \sum_{\text{next}} P(c_i) \cdot I(c_i)$$

course code

I	0100(4)	$H_p = \sum P(c_i) \log P(c_i)$
III	0101(4)	$0.38 + 0.36 + 0.39 + 0.69$
II	0111(3)	$+ 0.28 + 0.28$
I	1(1)	$I(1) = 2.38$
X	001(3)	too handle
XII	000(3)	

(c) The average message length for 100 random group students.

For hundred students The total average expected message length = $2.38 \times 100 = 238$ bits when encoding department for groups of randomly chosen students.

II

1. (a) Source coding: codes the messages to be transmitted, it removes redundancy from the encoding process, it assigns less length codewords to symbols with more probability of frequency within the message, these by decreasing the avg length of the message effectively decreasing the amt of bits to be transmitted while making sure that during decoding these encoded messages there is no error where 2 or more messages have the same encoded message.
Ex:- Huffman coding..

(b) Channel coding: on the other hand is performed on the encoded bit messages obtained after source coding, it adds redundancy to the encoded message such that, even if some error was generated by the channel still the message could mostly be reconstructed, the channel through which the message is sent could introduce errors into the transmitted message for which we introduce redundancy to keep track of the errors that crept in and to correct them.

Ex : Hamming Code

(b) for a $C(n, k)$ linear block code

$$u^T H = 0$$

$$\dim(V) + \dim(V^\perp) = n$$

k

$(n-k)$



$\dim(G)$

$\dim(H)$

$$H^T G = 0$$

$$c^T H = 0$$

$$(H^T c)^T = 0$$

Linear block
code

G is dual of H and vice versa

G & H are subspace of \mathbb{R}^n

Generator matrix (G):

The generator matrix is a matrix of $k \times n$ dimensions that generates n length encoded messages from k length message codewords.

Codewords of $C(n, k)$ generated from G are linear combination of rows of G .

Set of Codewords are subspace \mathbb{F}_2^n
 G & its rows are linearly independent

Parity check matrix (H):

The parity check matrix is used to detect errors in received codewords [data]. It is an $(n-k) \times n$ dim matrix that when multiplied with a valid codeword generates a null matrix.

H also can be found by Tanner graph representation

* There are limits for error detection and error corrections based no of redundant / parity bits.

2)

(5,1) REP code \rightarrow REP-5

w.r.t,

$$G = [11111] \Rightarrow [I | \underset{5 \times 1}{\text{1's}}] \Rightarrow [11111]_{1 \times 5}$$

$$C = I \cdot G$$

 I can be $[0]$ or $[1]$ Parity check matrix H

$$H = [P^T | I_{5 \times 1}]$$

$$H = \begin{bmatrix} 1 & & & \\ 1 & 1 & & \\ 1 & 1 & 1 & \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

$$H_{\text{REP-5}} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

4x5

 $H \cdot C^T$ (or) $C \cdot H^T = 0$ for above H & C ✓* G of REP-n is dual of H of SPC-n similarly H H of REP-n is dual of G of SPC-n

3) (6,5) single parity check code

 C is linear combination of any 6 digit numbers which belongs to F_2^6

$$\dim(G) = 5, F_2^5 \subset F_2^6 \quad \textcircled{O} F_2^5 \text{ is span } \{ \text{each row in } G \}$$

$$G = \left[\begin{array}{c|c} I & P \end{array} \right] = [I_5 | P]$$

$$G = \left[\begin{array}{ccccc|c} 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{array} \right]$$

wkT,

$$H = [P^T | I_{n-k}] = [P^T | I_{6-5}]$$

$$H = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}_{1 \times 6} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}_{1 \times 6}$$

as we can observe it also could be

→ It is also systematic G & H.

→ H for (6,5) single parity check matrix is
G for Rep-6 code

→ Similarly G for (6,5) single parity check
matrix is H for REP-6 code

4. given m (no. of parity bits) = 4

k = 11 (no. of message bits).

acc to hamming code

$$- k \leq 2^m - m - 1 \quad (\text{as } 2^m \geq k + m + 1)$$

$$11 \leq 2^4 - 4 - 1 \Rightarrow 16 - 5 = 11 \checkmark$$

So, we can proceed with 4 parity bits
in each code word.

→ data bits $k = 11$

so, we can write parity check matrix H
const all non-zero binary (11) we can derive parity
column vectors length $m = 4$ egn from

$$H = [I - 1^5]$$



By labelling
D's inside boxes

$$P_1 = D_1 \oplus D_2 \oplus D_4 \oplus D_5 \oplus D_7 \oplus D_9 \oplus D_{11}$$

$$P_2 = D_1 \oplus D_3 \oplus D_4 \oplus D_{6,2} \oplus D_7 \oplus D_{10} \oplus D_4,$$

$$P_3 = D_2 \oplus D_3 \oplus D_4 \oplus D_8 \oplus D_9 \oplus D_{10} \oplus D_4$$

$$P_4 = D_5 \oplus D_6 \oplus D_7 \oplus D_8 \oplus D_9 \oplus D_{10} \oplus D_4$$

$$H_1 = \left[\begin{array}{ccccccccc|c} 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \end{array} \right] \quad \begin{matrix} 1000 \\ 0100 \\ 0010 \\ 0001 \end{matrix}$$

H(15,4)

(or) [Both ways can be done]

of finding H

4x15

$$H_2 = \left[\begin{array}{ccccccccc|c} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 \end{array} \right] \quad \begin{matrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{matrix}$$

4x15

$$H = \begin{bmatrix} I_4 & P \end{bmatrix}_{4 \times 15}$$

wkT,

$$G_2 = \begin{bmatrix} P^T & I_{n-k} \end{bmatrix}$$

$$G_1 = \begin{bmatrix} P^T & I_{11} \end{bmatrix}_{11 \times 15}$$

after writing, I_{11} on Right of 11×15

matrix P can be written correspondingly by previously obtained equations or else Just write P^T of H_2 and I_{11} .

$$G_1 = \begin{bmatrix} P_{H_2}^T & I_{11} \end{bmatrix}_{11 \times 15}, \text{ (or)} \quad \begin{bmatrix} I_{11} & P^T \end{bmatrix}_{11 \times 15}$$

$$G(154) \stackrel{?}{=} G_1 = \left[\begin{array}{c|c} 1000000000000 & 1110 \\ 0100000000000 & 1101 \\ 0010000000000 & 1100 \\ 0001000000000 & 1011 \\ 0000100000000 & 1010 \\ 0000010000000 & 1001 \\ 0000001000000 & 1000 \\ 0000000100000 & 0111 \\ 0000000010000 & 0110 \\ 0000000001000 & 0101 \\ 0000000000100 & 0100 \end{array} \right]_{11 \times 15}$$

⑤ Given $G_1 = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}$

(a) rate = $\frac{\text{no. of info bits}}{\text{total bits}}$

as given data G_1 is 5×6 matrix

→ by this redundancy PS 1 of Blocklength (SPC-6.) is 6.

→ C is $(6, 5)$ & looks like a single parity

→ No. of Information bits (k) = 5 (rows of G_1)
 rate = $\frac{5}{6} = 0.83$ check code

(b) $[w \ x \ y \ z \ u] \cdot \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}_{5 \times 6} = [w \ x \ y \ z \ p] = [C]_{5 \times 6}$

Set of all code words are $[w \ x \ y \ z \ p]$

where p is F_2 sum(ROR) of w, x, y, z

$P = w \oplus x \oplus y \oplus z$ (or) $P = w + x + y + z + u$

$+ \in F_2$

$$C = [m_1, m_2, m_3, m_4, m_5, P]$$

$$P = \sum_{i=1}^5 m_i \pmod{2} \rightarrow \mathbb{F}_2 \text{, } (+, 0, 1)$$

$$\text{Code Word for } (10010) = [1, 0, 0, 1, 0, (1+0+0+1+0)]$$

$$= [1, 0, 0, 1, 0, 0]$$

$$= [100100]$$

w.k.t, $G = [I_5 | P]$ since we noticed

~~row~~ 1 acts as a parity in parity ~~column~~ ~~row~~

for $C(6,5)$ The parity check equation

~~row~~ let $[c_1, c_2, c_3, c_4, c_5, c_6]$ be code word

w.k.t, for above parity check eqn $c * [11111] \Rightarrow P_e \pmod{2}$

$$H = [P^T | I_{6-5}] = [11111 | 1]$$

$$H = [11111] \quad [c * [11111]^T \Rightarrow P_e] \quad \begin{array}{l} \text{where, should} \\ P_e = 0 \\ \text{for no error} \\ \text{transmission} \end{array}$$

For any valid codeword $H \cdot C^T = 0 \pmod{2}$

$$C \cdot H^T = 0$$

b) Given

$$G = [I_3 | P]$$

$$P = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

$$H = [P^T | I_{n-k}]$$

by the we can say

$$\text{than } n = 7$$

$$k = I_3 = 3$$

$$P_n = 4$$

$$H = [P^T | I_{7 \times 3}] = [P^T | I_4]_{4 \times 7}$$

$$H = \left[\begin{array}{ccc|ccccc} 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 \end{array} \right]_{4 \times 7}$$

$$H = \left[\begin{array}{ccc|cccc} 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 \end{array} \right]$$

$$C \cdot H^T = 0 \quad [P^T P = 0]$$

$$\boxed{H^T} \times 4 \checkmark$$

G & H are systematic