Section A

Q1

5/5

i) True. Suppose $p(X=1)=p(Y=1)$. Then

$$\begin{aligned} p(X=0) &= 1-p(X=1) \\ &= 1-p(Y=1) \\ &= p(Y=0) \end{aligned}$$

1/1

ii) False. Suppose $p(X=1 \mid Y=1) < p(X=1)$. Then

$$\cancel{p(X=1 \mid Y=0) = \frac{p(X=1 \mid Y=0)}{p(Y=0)}}$$

~~Suppose $p(X=1 \mid Y=1) < p(X=1)$. Then~~

$$\cancel{p(X=1)} = p(Y=0)\,p(X=1 \mid Y=0) + p(Y=1)\,p(X=1 \mid Y=1)$$

Suppose $p(X=1 \mid Y=0) < p(X=1)$ and $p(X=1 \mid Y=1) < p(X=1)$.
Then

$$\begin{aligned} p(X=1) &= p(Y=0)\,p(X=1 \mid Y=0) + p(Y=1)\,p(X=1 \mid Y=1) \\ &< p(Y=0)\,p(X=1) + p(Y=1)\,p(X=1) \\ &= p(X=1) \end{aligned}$$

a contradiction. 1/1

iii) True. If $X \perp Y$, then 1/1

$$\begin{aligned} & p(X=1 \mid Y=1) + p(X=0 \mid Y=0) \\ &= \frac{p(X=1, Y=1)}{p(Y=1)} + \frac{p(X=0, Y=0)}{p(Y=0)} \\ &= \frac{p(X=1)\,p(Y=1)}{p(Y=1)} + \frac{p(X=0)\,p(Y=0)}{p(Y=0)} \\ &= p(X=1) + p(X=0) \\ &= 1 \end{aligned}$$

PTO

1 iv. ~~False~~ True

$$
\begin{aligned}
p(X=0, Y=1) &= p(X=0 \mid Y=1)\, p(Y=1) \\
&= (1 - p(X=1 \mid Y=1))\, p(Y=1) \\
&= p(Y=1) - p(X=1, Y=1) \\
&= p(Y=1) - p(X=1)\, p(Y=1) \\
&= (1 - p(X=1))\, p(Y=1) \\
&= p(X=0)\, p(Y=1)
\end{aligned}
$$

By symmetry, $p(X=1, Y=0) = p(X=1)\, p(Y=0)$.

then,

$$
\begin{aligned}
p(X=0, Y=0) &= p(X=0 \mid Y=0)\, p(Y=0) \\
&= (1 - p(X=1 \mid Y=0))\, p(Y=0) \\
&= p(Y=0) - p(X=1, Y=0) \\
&= p(Y=0) - p(X=1)\, p(Y=0) \\
&= (1 - p(X=1))\, p(Y=0) \\
&= p(X=0)\, p(Y=0)
\end{aligned}
$$

1/1

$\therefore\ p(X, Y) = p(X)\, p(Y)$

v. True.

$$
\begin{aligned}
\frac{p(X=1 \mid Y=1)}{p(Y=1 \mid X=1)} &= \frac{\frac{p(X=1, Y=1)}{p(Y=1)}}{\frac{p(X=1, Y=1)}{p(X=1)}} \\
&= \frac{\frac{1}{p(Y=1)}}{\frac{1}{p(X=1)}} \\
&= \frac{p(X=1)}{p(Y=1)}
\end{aligned}
$$

1/1

PTO

Q

Q2

15/15

i. ~~p(Michael fails | ¬Glenn)~~

$=$

$p(\text{Michael fails}) = p(\text{Glenn})\, p(\text{Michael fails} \mid \text{Glenn}) + p(\neg\text{Glenn})\, p(\text{Michael fails} \mid \neg\text{Glenn})$

$\therefore p(\text{Michael fails} \mid \neg\text{Glenn}) = \dfrac{p(\text{Michael fails}) - p(\text{Glenn})\, p(\text{Michael fails} \mid \text{Glenn})}{p(\neg\text{Glenn})}$ ✓

$= \dfrac{0.05 - 0.1 \times 0.5}{0.9}$

5/5

$= \dfrac{0}{0.9}$

$= 0$ ✓

ii. $p(\text{Glenn} \mid \text{Michael fails}) = \dfrac{p(\text{Glenn})\, p(\text{Michael fails} \mid \text{Glenn})}{p(\text{Glenn})\, p(\text{Michael fails} \mid \text{Glenn}) + p(\neg\text{Glenn})\, p(\text{Michael fails} \mid \neg\text{Glenn})}$ ✓

$= \dfrac{p(\text{Glenn})\, p(\text{Michael fails} \mid \text{Glenn})}{p(\text{Glenn})\, p(\text{Michael fails} \mid \text{Glenn}) + 0}$

$= 1$ ✓

5/5

iii. $p(\text{Sanath} \mid \text{fail}) = 0.8$

$p(\text{fail} \mid \text{Sanath}) = 0.2$

$p(\text{fail}) = 0.01$

~~p(Sanath | fail) p(fail)~~

$p(\text{Sanath}, \text{fail}) = p(\text{fail} \mid \text{Sanath})\, p(\text{Sanath})$

$= p(\text{Sanath} \mid \text{fail})\, p(\text{fail})$

PTO

2 iii.

$$\therefore\ p(\text{Sanath}) = \frac{p(\text{fail})\, p(\text{Sanath} \mid \text{fail})}{p(\text{fail} \mid \text{Sanath})}$$

$$= \frac{0.01 \times 0.8}{0.2}$$

$$= \frac{0.008}{0.2}$$

$$= \frac{8 \times 10^{-3}}{2 \times 10^{-1}}$$

5/5

$$= 4 \times 10^{1-3}$$

$$= 4 \times 10^{-2}$$

$$= 0.04$$

$\therefore \frac{4}{100}$ of Glenn's games are against Sanath

KTO

Q3

5/5

3 i. $\hat{P}_{MLE}(X=a, Y=b) = \frac{n_{ab}}{N}$ ✓

where ~~n_{ab} is the number of~~ times

~~$n_{ab} = \sum$~~

3/3

n_{ab} is the number of times that $X=a$ and $Y=b$ in the total runs. This answer is the relative frequency, and maximizes the likelihood of the given results occurring. ✓

ii. It depends on what one means by 'probability'. If 'probability' is simply Carl's subjective expectations about the future state of the world (subjective Bayesianism) ~~on~~ then Carl is correct: his probabilities for X and Y are dependent. If 'probability' is the long-run frequencies of events (frequentism) or the ~~objective~~ propensity of outcomes to occur (objective chance / propensity theory), ~~then it is possible~~ then Judith is correct: it is possible that Carl has not seen a representative sample of the data and that the probabilities that he has calculated are not the true probabilities, and that the true probabilities are statistically independent. ✓

2/2

PTO.

Section B

Q1 15/15

i. $H(X) = \frac{1}{2}\log_2 2 + \frac{1}{4}\log_2 4 + \frac{1}{8}\log_2 8 + \frac{1}{8}\log_2 8$
$= \frac{1}{2} + \frac{1}{4}\times 2 + \frac{1}{8}\times 3 + \frac{1}{8}\times 3$
$= 1 + \frac{6}{8}$
$= 1 + \frac{3}{4}$
$= \frac{7}{4}$ ✓

$H(Y) = 4 \times \frac{1}{4} \times \log_2 4$ 3/3
$= 4 \times \frac{1}{4} \times 2$
$= 2$

ii. $D_{KL}(p\|q) = \frac{1}{2}\log_2 \frac{\frac{1}{2}}{\frac{1}{4}} + \frac{1}{4}\log_2 \frac{\frac{1}{4}}{\frac{1}{4}} + \frac{1}{8}\log_2 \frac{\frac{1}{8}}{\frac{1}{4}}$
$+ \frac{1}{8}\log_2 \frac{\frac{1}{8}}{\frac{1}{4}}$
$= \frac{1}{2}\log_2 2 + \frac{1}{4}\log_2 1 + \frac{1}{8}\log_2 \frac{1}{2} + \frac{1}{8}\log_2 \frac{1}{2}$
$= \frac{1}{2} + 0 - \frac{1}{8} - \frac{1}{8}$
$= \frac{1}{4}$ ✓ 3/3

$D_{KL}(q\|p) = \frac{1}{4}\log_2 \frac{\frac{1}{4}}{\frac{1}{2}} + \frac{1}{4}\log_2 \frac{\frac{1}{4}}{\frac{1}{4}} + \frac{1}{4}\log_2 \frac{\frac{1}{4}}{\frac{1}{8}} + \frac{1}{4}\log_2 \frac{\frac{1}{4}}{\frac{1}{8}}$
$= \frac{1}{4}\log_2 \frac{1}{2} + \frac{1}{4}\times 0 + \frac{1}{4}\log_2 2 + \frac{1}{4}\log_2 2$
$= -\frac{1}{4} + \frac{1}{4} + \frac{1}{4}$
$= \frac{1}{4}$ ✓

iii. Yes. $D_{KL}(r\|p) = 0$ iff $r = p$ iff $D_{KL}(p\|r) = 0$ 2/2

iv. No. See part ii. ✓ 2/2

v. $L(C, X) = \frac{1}{2} + \frac{1}{4}\times 2 + \frac{1}{8}\times 3 + \frac{1}{8}\times 3$
$= H(X)$
$\therefore$ C is optimal for X by SCT

1b. (cont.)

$$L(C, Y) = \frac{1}{4}\times 1 + \frac{1}{4}\times 2 + \frac{1}{4}\times 3 + \frac{1}{4}\times 3$$
$$= \frac{1}{4} + \frac{1}{2} + \frac{3}{4} + \frac{3}{4}$$
$$= \frac{9}{4}$$
$$> H(Y)$$

A better code for Y would be

$$C' : Y \to \{0,1\}^2$$
$$a \mapsto 00$$
$$b \mapsto 01$$
$$c \mapsto 10$$
$$d \mapsto 11$$

Then, $L(C', Y) = \frac{1}{4}\times 2 \times 4$
$$= 2$$
$$= H(Y),$$ optimal by SCT.

5/5

PTO

Q2 (5/5)

~~a~~i. Best first question is Q_j, where $p_j \geq p_i$ for $i \neq j$ (this assigns the shortest code to the highest probability outcome) ✓ 2/2

ii. Yes. Suppose ~~there is an even distribution~~ p is the uniform distribution over X.
Then ~~$p = 1/100$~~ $H(X) = \log_2 100 > 1$
Gary could guess correctly on his first try, asking fewer than $H(X)$ questions. However, in expectation, Gary will ask $H(X)$ or more questions. ✓ 1/1

iii. No. Suppose the distribution is bimodal:

✓ 2/2

Then Gary would be better off asking Q_{50} first, which has entropy 1, than Q_{25}, which has entropy $\sim \frac{1}{4}\log_2\frac{1}{4} + \frac{3}{4}\log_2\frac{3}{4} < 1$, but would be asked first in part i

170

Q

Q3 5/5

i. $p(X=1) = p(X=1 \mid Y=0) + p(X=1 \mid Y=1)$

$= \frac{3}{8} + \frac{1}{8}$

$= \frac{1}{2}$

$p(Y=1) = p(Y=1 \mid X=0) + p(Y=1 \mid X=1)$

$= \frac{1}{8} + \frac{1}{8}$

$= \frac{1}{4}$ ✓ 1/1

ii. $p(X=1, Y=1) = \frac{1}{8}$

$= p(X=1)\, p(Y=1)$ ✓

$\therefore X \perp\!\!\!\perp Y$ (by a1, part A) ✓

$\therefore I(X;Y) = 0$, since $I(X;Y) = \sum p(x,y) \log_2 \frac{p(x,y)}{p(x)p(y)}$

$= \sum p(x)p(y) \log_2 1$

$= 0$

This intuitively tells us that learning X tells us nothing we did not already know about Y, and vice versa. ✓ 2/2

iii. No. The ✓ Data Processing Inequality is only for Markov chains

~~$X \to Y \to Z$~~

~~where $p(X,Y,Z) = p(X)\,p(Y \mid X)\,p(Z \mid Y)$~~

2/2

$X \to Z \to Y$

where $p(X,Y,Z) = p(X)\,p(Z \mid X)\,p(Y \mid Z)$ ✓

PTO

Section C

$\frac{10}{10}$ Q1

i. $A_{X^3} = \{aaa, aab, aba, abb, baa, bab, bba, bbb\}$ ✓ 2

ii. $P(abb) = p_a p_b p_b$
$= \frac{1}{3} \times \frac{2}{3} \times \frac{2}{3}$
$= \frac{4}{27}$ ✓ 2

iii. $T_{N\beta} = \{ \underline{x} \in A_{X^3} \mid | -\frac{1}{3} \log_2 P(\underline{x}) - H(X) | < 0.2 \}$
for each $\underline{x}$,

$$-0.2 < -\frac{1}{3} \log_2 P(\underline{x}) - H(X) < 0.2$$
$$-0.2 < \frac{1}{3} \log_2 P(\underline{x}) + 0.92 < 0.2$$
~~0.72 <~~ 2
$$-1.12 < \frac{1}{3} \log_2 P(\underline{x}) < -0.72$$
$$-3.36 < \log_2 P(\underline{x}) < -2.16$$
therefore ~~only~~ abb $\in T_{N\beta}$. and aaa, bbb $\notin T_{N\beta}$ ✓

iv. $S_\delta = X^3 \setminus \{aaa\}$, since $p(aaa) = \frac{1}{27}$, ~~$p(others) > (\frac{1}{25} - \frac{1}{27})$~~ ✓
2 $p(others) > (\frac{1}{25} - \frac{1}{27})$

v. $H_\delta(X^3) = \log_2 |X^3 \setminus \{aaa\}|$ ✓
$= \log_2 7$ 2

PTO

$\frac{8}{8}$

Q2

i. $H(X) = \frac{1}{2}\log_2 2 + \frac{1}{4}\log_2 4 + \frac{1}{8}\log_2 8 + \frac{1}{16}\log_2 16 + \frac{1}{16}\log_2 16$

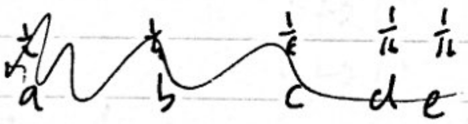
$= \frac{1}{2} + \frac{1}{4}\times 2 + \frac{1}{8}\times 3 + \frac{1}{16}\times 4 + \frac{1}{16}\times 4$

$= 1 + \frac{3}{8} + \frac{1}{4} + \frac{1}{4}$

$= 1 + \frac{3}{8} + \frac{4}{8}$

$= \frac{8+7}{8}$

$= \frac{15}{8}$ ✓

ii. 

A Huffman code for X is

$a \mapsto 0$

$b \mapsto 10$

$c \mapsto 110$ ✓

$d \mapsto 1110$

$e \mapsto 1111$

$L(C,X) = \frac{1}{2}\times 1 + \frac{1}{4}\times 2 + \frac{1}{8}\times 3 + \frac{1}{16}\times 4 + \frac{1}{16}\times 4$

$= \frac{15}{8}$ ✓

$= H(X)$

iii. No. Huffman codes are optimal, and you cannot get lower than the entropy lower bound. ✓

iv. $[\frac{1}{2}, \frac{3}{4}) = [0.5, 0.75)_{dec}$ ✓

$= [0.1, 0.11)_{bin}$

p70

Q

3/3

Q3

Yes. ~~Suppose C is a prefix-free code.~~

Let $C : \{a, b, c\} \to \{0,1\}^+$

$$a \mapsto 0$$
$$b \mapsto 10$$
$$c \mapsto 11$$

C is prefix-free and uniquely decodeable.

Next, let $C' : \{a, b, c\} \to \{0,1\}^+$

$$a \mapsto 0$$
$$b \mapsto 01$$
$$c \mapsto 11$$

C' is not ~~uniquely decodeable~~ prefix-free: $C'(a)$ is a prefix of $C'(b)$. However, C' is uniquely decodeable: take a message, reverse the order of the bits, decode according to C, then reverse the order of the letters in the decoded message. ✓

PTO

$\frac{4}{4}$ Q4

i. $H(X) = -\sum_i p(X=i) \log_2 p(X=i)$ ✓

$H(X|Y)$ ~~= ...~~
$= H(X,Y) - H(Y)$ ✓

where $H(X,Y)$ is the entropy of the random variable with alphabet $A_X \times A_Y$

Given a universal prefix turing machine, U,

$K(x) = \min_p \{ l(p) \mid U(p) = x \}$ ✓

$K(x|y) = \min_p \{ l(p) \mid U(y' p) = x \}$ ✓

ii. Upper bound:

$H(X) \le \log |X|$

$K(x) <$ ~~log~~ $l(x) + 2 \log l(x) + O(1)$ ✓

Extra information:

$H(X|Y) \le H(X)$

$K(x|y) < K(x) + O(1)$ ✓

Subadditivity:

$H(X,Y) \le H(X) + H(Y)$

$K(xy) < K(x) + K(y) + O(1)$ ✓

Symmetry:

$H(X) - H(X|Y) = H(Y) - H(Y|X)$

$K(x) - K(x|y) = K(y) - K(y|x) + O(1)$ ✓

Information non-increase:

$$H(f(X)) < H(X)$$
$$K(f(x)) < K(x) + K(f) + O(1)$$ ✓

4 iii. H requires one to have a probability distribution over X.

K is ~~only~~ not computable.

H tells you about fundamental limits in ~~communication~~ compression and noisy channel coding

K tells you about how similar strings are

$$d(x,y) = \cancel{\max\{K(x|y)} \quad \frac{\min\{K(x|y), K(y|x)\}}{\min\{K(y), K(x)\}}$$

and ~~how to~~ plays a crucial role in Solomonoff induction, telling you how likely theories are a priori:

$$M(x) = \sum_{p : U(p) = x} 2^{-\ell(p)}$$ ✓

PTO

Section D

10/10

Q1

i. $\begin{bmatrix} 0.5 & 0.25 \\ 0.5 & 0.25 \\ 0 & 0.25 \\ 0 & 0.25 \end{bmatrix}$ ✓

ii. No: It cannot be partitioned into blocks like so:

✓

such that within each block, each row is a permutation of every other row and each column is a permutation of each other column.

iii

$$f_Y = Q f_X$$

$$= \begin{bmatrix} 0.5 & 0.25 \\ 0.5 & 0.25 \\ 0 & 0.25 \\ 0 & 0.25 \end{bmatrix} \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$$

$$= \begin{bmatrix} 0.375 \\ 0.375 \\ 0.125 \\ 0.125 \end{bmatrix} = \begin{bmatrix} \frac{3}{8} \\ \frac{3}{8} \\ \frac{1}{8} \\ \frac{1}{8} \end{bmatrix}$$ ✓

PTO

$\cancel{I(X;Y) = H(X) - H(X|Y)}$

$\therefore\ I(X;Y) = H(Y) - H(Y|X)$

$H(Y) = -\frac{3}{8}\log_2\frac{3}{8} \times 2 - \frac{1}{8}\log_2\frac{1}{8} \times 2$

$= -\frac{3}{4}\log_2 3 + \frac{3}{4}\log_2 8 + \frac{1}{4}\log_2 8$

$= \log_2 8 - \frac{3}{4}\log_2 3$

$= 3 - 0.75 \times 1.6$

≈ 1.8

$H(Y|X) = p(X=0)\ H(Y|X=0) + p(X=1)\ H(Y|X=1)$

$= \frac{1}{2}\left(-2\times\frac{1}{2}\times\log_2\frac{1}{2}\right) + \frac{1}{2}\left(-4\times\frac{1}{4}\times\log_2\frac{1}{4}\right)$

$= \frac{1}{2} + \frac{1}{2}\times 2$

$= 1.5$

$\therefore\ I(X;Y) = 4.5 - \frac{3}{4}\log_2 3$

≈ 0.3 ✓

v. No.

$C_Q = \max_{P_X}\{I(X;Y)\}$

$\geq I(X;Y)$ for P_X uniform ✓

$= 0.3$

17/20

$\frac{8}{8}$ Q2

i. $\cancel{N = l(x^i) = 4}$

$\cancel{K = \log_2(4)}$

$N = l(x^{(1)}) = 4$

$K = \log_2(\text{no. of codewords})$

$= \log_2 4$

$= 2$

$R = \frac{K}{N} = \frac{2}{4} = \frac{1}{2}$ ✓

ii. i. $p(y = 0010 \mid s=2) = 1 \times 1 \times 0.75 \times 1$

$= \frac{3}{4}$ ✓

ii. $p(y = 1111 \mid s=2) = 0 \times 0 \times \frac{3}{4} \times 0$

$= 0$ ✓

iii. $p(y = 0000 \mid s=2) = 1 \times 1 \times \frac{1}{4} \times 1$

$= \frac{1}{4}$ ✓

iv. $p(y = 1000 \mid s=2) = 0 \times 1 \times \frac{1}{4} \times 1$

$= 0$ ✓

iii. $p(\hat{s} \neq s) = \sum_i p(s = s^i)\, p(\hat{s} \neq s \mid s = s^{(i)})$

$= 1 - p(\hat{s} = s \mid s = s^{(i)})$ ✓

$= 1 - 1 \times 1 \times \frac{3}{4} \times 1$

$= \frac{1}{4}$

PTO

3/3

Q3

Y u change question sub-numbering convention?

a) $H(Y) = 1 \log_2 1 = 0$

$H(Y|X) = 1 \log_2 1 = 0$

$\therefore I(X;Y) = 0$

b) Q' symmetric

$\therefore I(X;Y)$ maximised over uniform input distribution

$$f_Y = Q f_X$$

$$= \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{bmatrix} \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix}$$

$$= \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$$

$$H(Y) = 3 \times \frac{1}{3} \times \log_2 3$$
$$= \log_2 3$$
$$\doteq 1.6$$

$$H(Y|X) = \frac{1}{2} \times \log_2 2 + 2 \times \frac{1}{4} \times \log_2 4$$
$$\cancel{= 1 + \tfrac{1}{4} \times 2}$$
$$= \frac{1}{2} + \frac{1}{2} \times 2$$
$$= 1.5$$

$$\therefore C = I(X;Y)$$
$$= H(Y) - H(Y|X)$$
$$\doteq 0.1$$

17/20

4/4

Q4

$$C = I(X;Y)$$
$$= H(X) - H(X|Y)$$
$$\leq H(X)$$
$$\leq \log_2 |X|$$
$$= \log_2 N$$

Similarly, $C = I(X;Y)$
$$= H(Y) - H(Y|X)$$
$$\leq H(Y)$$
$$\leq \log_2 |Y|$$
$$= \log_2 M$$

$\therefore$ $C \leq \min\{\log_2 N, \log_2 M\}$ ✓

Q