AUSTRALIAN NATIONAL UNIVERSITY

COMP2610/COMP6261

Information Theory, Semester 2 2023

Assignment 1

Due Date: Monday 28 August 2023, 9:05 am

Assignment 1 weighting is 10% of the course mark.

Instructions:

Marks:

1. The mark for each question is indicated next to the question. For questions where you are asked to prove results, if you can not prove a precedent part, you can still attempt subsequent parts of the question assuming the truth of the earlier part.
2. **COMP2610 students:** Answer *Questions 2, 3, 5* and only *Section I of Question 1 and 4*. You are not expected to answer Section II of Question 1 and 4. You will be marked out of 100.
3. **COMP6261 students:** Answer *Questions 2, 3, 5* and only *Section II of Question 1 and 4*. You are not expected to answer Section I of Question 1 and 4. You will be marked out of 100.

Submission:

1. Submit your assignment together **with a cover page** as a single PDF on Wattle.
2. **Clearly mention whether you are a COMP2610 student or COMP6261 student on the cover page.**
3. A late submission attracts a penalty of **5%** per working day as per ANU Policy until a **week (5 working days)** from the due date. We will provide solutions to the assignment after a week from the due date and if you submit after that time you get zero marks **(100% penalty)**. Extensions will be considered according to the ANU Policy.
4. All assignments must be done individually. Plagiarism is a university offense and will be dealt with according to university procedures *http://academichonesty.anu.edu.au/UniPolicy.html*.

Question 1 [25 marks total]

****Section I: Only COMP 2610 students are expected to attempt this question.**

The Blue Challenge is a unique swimming competition that tests both the physical and emotional endurance of the participants and gives them an exciting opportunity to experience the oceanic wilderness.

Let's consider two universities, M (MIT) and C (Cambridge), competing in a series of 5 rounds. The contest rules are as follows:

- Each round has exactly one winner (i.e. universities cannot tie).
- In every race, both universities have equal probability ($\frac{1}{2}$) of winning.
- Once one university wins 3 rounds, no further rounds are conducted.

Let S be the random variable whose outcomes are all possible strings representing the winners of the rounds,

$$\mathcal{S} = \{\mathcal{MMM}, \mathcal{CCC}, \mathcal{MCMM}, \ldots, \mathcal{MCMCM}\}$$

Let T denote a random variable representing the total number of rounds conducted in the series, so that the possible outcomes are $\mathcal{T} = \{3, 4, 5\}$.

Calculate each of the following, showing all your working:

(a) The information in seeing the sequence $MCMM$, i.e., compute $h(S = MCMM)$ **[3 marks]**

(b) $H(S)$ **[3 marks]**

(c) $H(T)$ **[3 marks]**

(d) $H(S|T = 3)$ **[3 marks]**

(e) $H(S|T)$ **[4 marks]**

(f) $H(T|S)$. Provide an intuitive explanation of your answer. **[3 marks]**

(g) Generally, the winner of the series is simply the university that wins the last round. However, if the university who wins the first round in the series is different to the one who wins the last round, then officials are suspicious of some form of fixing. In such a series, there is a 50% chance of officials declaring the series invalid, and neither university is declared winner; the other 50% of the time, the officials decide the series is legitimate, and declare the winner as usual.

Let P be a random variable denoting the identity of the winning university, with possible outcomes $\mathcal{U} = \{\mathcal{M}, \mathcal{C}, \mathcal{N}\}$, for the winner being respectively MIT, Cambridge, or neither. Compute $H(U)$. **[6 marks]**

Solution : According to question,

$$S = \left\{ \begin{array}{l} MMM, \; CCC \rightarrow 2 \quad \text{when } T=3 \\ MCMM, \; MMCM, \; CMMM, \; CMCC, \; CCMC, \; MCCC \rightarrow 6 \quad \text{when } T=4 \\ MMCCM, \; MCCMM, \; MCMCM, \; CCMMM, \; CCMMC, \; CMMCC, \; CMCMC, \; MMCCC, \; MCCMC, \; CMMCM, \; MCMCC, \; CMCMM \rightarrow 12 \quad \text{when } T=5 \end{array} \right.$$

Total posiblities = 20

a) $P(S = MCMM) = P(W_c = M)^3 \times P(W_c = C)^1$

Here, W_R : Winner of competition

Since, $P(W_c = M) = P(W_c = C) = 1/2$

Therefore, $P(S = MCMM) = \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right) = \left(\frac{1}{2}\right)^4$

$$= \frac{1}{16}$$

$$\Rightarrow h(S = MCMM) = -\log_2(P(S = MCMM)) = -\log_2(1/16)$$

$$= -\log_2(2^{-4}) = 4 \text{ bits}$$

b) $H(S) = -\sum_{\forall s} P(s) \log_2 P(s)$ - - - - - - - (i)

When $T = 3 \Rightarrow P(s) = (1/2)^3 \rightarrow$ 2 possibility

$T = 4 \Rightarrow P(s) = (1/2)^4 \rightarrow$ 6 possibility

$T = 5 \Rightarrow P(s) = (1/2)^5 \rightarrow$ 12 possibility therefore, from eqn. (i)

$$H(S) = -\left[2\times\left(\tfrac{1}{8}\log_2 \tfrac{1}{8}\right) + 6\times\left(\tfrac{1}{16}\log_2 \tfrac{1}{16}\right) + 12\times\left(\tfrac{1}{32}\log_2 \tfrac{1}{32}\right)\right]$$

$$= \frac{3}{4} + \frac{3}{2} + \frac{15}{8} = \frac{6+12+15}{8} = \frac{33}{8} \text{ bits}$$

$$= 4.125 \text{ bits}$$

c) $H(T) = -\sum_{\forall T} P(T)\log_2 P(T)$ ------ (ii)

$$P(T=3) = \left(\begin{array}{c}\text{No. of possibilities}\\ \text{for } T=3\end{array}\right) \times \left(\begin{array}{c}\text{Probability of each}\\ \text{occurance in it}\end{array}\right)$$

$$= 2\times\left(\tfrac{1}{2}\right)^3 = \tfrac{1}{4}$$

Similarly, $P(T=4) = 6\times\left(\tfrac{1}{2}\right)^4 = \tfrac{3}{8}$

$$P(T=5) = 12\times\left(\tfrac{1}{2}\right)^5 = \tfrac{3}{8}$$

from eqn. (ii),

$$H(T) = -\left[\tfrac{1}{4}\log_2\left(\tfrac{1}{4}\right) + \tfrac{3}{8}\log_2\left(\tfrac{3}{8}\right) + \tfrac{3}{8}\log_2\left(\tfrac{3}{8}\right)\right]$$

$$= 1.5613 \text{ bits}$$

d> $H(S/T=3) = -\sum P(S/T=3)\log_2(P(S/T=3))$ --- (iii)

For $T=3$, there are 2 possibilities (MMM, CCC)
each with $P(S=s) = \left(\frac{1}{2}\right)^3 = \frac{1}{8}$

therefore, $P(S/T=3) = \frac{1}{2} \quad \forall s$

from eqn. (iii),

$$H(S/T=3) = -\sum_{T=3} P(S/T=3)\log_2(P(S/T=3))$$

$$= -\left[2\times\left(\tfrac{1}{2}\log_2\tfrac{1}{2}\right)\right] = \log_2(2) = 1 \text{ bit}$$

e> $$H(S/T) = \sum_{\forall T} P(T=t)\times H(S/T=t) \quad \text{--- (iv)}$$

$$= P(T=3)H(S/T=3) + P(T=4)\times H(S/T=4)$$

$$+ P(T=5)\times H(S/T=5)$$

since, $P(T=3) = 2\times\left(\frac{1}{2}\right)^3$ (i.e, for $T=3$, only 2 possiblities MMM, CCC)

$$P(T=4) = 6\times\left(\tfrac{1}{2}\right)^4 = \tfrac{3}{8}$$

$$P(T=5) = 12\times\left(\tfrac{1}{2}\right)^5 = \tfrac{3}{8}$$

Also, from part (d),

$$H(S|T=3) = 2\times\left(\tfrac{1}{2}\log_2 2\right) = 1 \text{ bit}$$
$$H(S|T=4) = 6\times\left(\tfrac{1}{6}\log_2 6\right) = \log_2 6 \text{ bits}$$
$$H(S|T=5) = 12\times\left(\tfrac{1}{12}\log_2 12\right) = \log_2 12 \text{ bits}$$

therefore, from eqn. (iv)

$$H(S/T) = \frac{1}{4} \times 1 + \frac{3}{8}\log_2 6 + \frac{3}{8}\log_2 12$$

$$= 2.5637 \text{ bits}$$

f) $H(T/S)$

If the outcome S is provided, by counting the number of races in S, T (total number of rounds) can be directly determined. Hence, no additional information is contained in T, if S is given.

Therefore, $H(T/S) = 0$

g)

S = { MMM , CCC → 2 when T=3

MCMM, MMCM, CMMM
CMCC, CCMC, MCCC → 6 when T=4

MMCCM, MCCMM, MCMCM, CCMMM
CCMMC, CMMCC, CMCMC, MMCCC
MCCMC, CMMCM, MCMCC, CMCMM → 12 when T=5

Total posibilities = 20

Total possible suspicious events [i.e, the university who wins the first round in the series is different to the one who wins the last round) highlighted in pink = 8

For these 8 events, 50% of U = N

therefore, $P(U=N) = \sum P(S=s)\, P(U=N/S=s)$

$$= 2 \times \left(\tfrac{1}{2}\right)^4 \times \left(\tfrac{1}{2}\right) + 6 \times \left(\tfrac{1}{2}\right)^5 \times \left(\tfrac{1}{2}\right)$$

$$= \tfrac{5}{32}$$

For $P(U=M) = \sum P(S=s)\, P(U=N/S=s)$

But $P(U=M) = P(U=C)$

Also, $\sum_{\forall i} P(U=u_i) = 1$

$$\Rightarrow P(U=M) = P(U=C) = \frac{1 - P(U=N)}{2}$$

$$= \frac{1 - 5/32}{2} = \frac{27}{64}$$

therefore, $H(U) = -\sum_{\forall i} P(U_i) \log_2 (P(U_i))$

$$= -\left[2 \times \left(\tfrac{27}{64}\right) \log_2\left(\tfrac{27}{64}\right) + \left(\tfrac{5}{32}\right) \log_2\left(\tfrac{5}{32}\right) \right]$$

$= 1.4690$ bits

****Section II: Only COMP 6261 students are expected to attempt this question.**

In ”The Trans-Galactic Olympiad”, alien civilizations from different galaxies compete in a series of 5 rounds, testing both their physical and emotional resilience while giving them a thrilling opportunity to experience the interstellar wilderness.

Let's consider three civilizations, A (Andromeda), M (Milky Way), and S (Sombrero), participating in the tournament. The tournament rules are as follows:

- Each round has exactly one winner (i.e. civilizations cannot tie).
- In every round, all civilizations have an equal probability ($\frac{1}{3}$) of winning.
- Once a civilization wins 3 rounds, no further rounds are conducted.

Let S, be the random variable whose outcomes are all possible strings representing the winners of the rounds including when there is no final winner for 5 rounds.

Assumption: Now assume that by the end of 5 rounds, one civilization must emerge as the winner. Here, we are not accounting for outcomes where all three civilizations (A, M, S) failed to clinch 3 wins out of 5 rounds.

Let P be the random variable whose outcomes are all possible strings representing the winners of the rounds as per the assumption,

$$\mathcal{P} = \{\mathcal{AAA}, \mathcal{MMM}, \mathcal{SSS}, \ldots, \mathcal{AMSAA}\}$$

Note: Here, P is the subset of sample space S such that, $S = \{P, N \text{ (no winner)}\}$.

Let Q denote a random variable representing the total number of rounds conducted in the series, so that the possible outcomes are $Q = \{3, 4, 5\}$.

Calculate each of the following, showing all your working [Hint: You may need to use any programming tool (of your choice) to calculate the sample space $\mathcal{P}$]:

(a) The information in seeing the sequence $AMSSS$, i.e., compute $h(P = AMSSS)$ **[2 marks]**

(b) $H(P)$ **[3 marks]**

(c) $H(Q)$ **[3 marks]**

(d) $H(P|Q = 4)$ **[3 marks]**

(e) $H(P|Q)$ **[3 marks]**

(f) $H(Q|P)$. Provide an intuitive explanation of your answer. **[3 marks]**

(g) Generally, the winner of the series is simply the civilization that wins the last round. However, if the civilization who wins the first round in the series is different to the one who wins the last round, then the officials are suspicious of some form of competition manipulation. In such a series, there is a 33% chance of officials declaring the series invalid, and no civilization is declared the winner; the other 66% of the time, the officials decide the series is legitimate, and declare the winner as usual.

Additionally, if a civilization loses 4 consecutive rounds at any point in the series, the officials begin a fair-play investigation. If the investigation finds evidence of unfair practices, there is a 50% chance that the series is declared void, and will be restarted.

Let G be a random variable denoting the identity of the winning civilization, with possible outcomes $\mathcal{G} = \{\mathcal{A}, \mathcal{M}, \mathcal{S}, \mathcal{N}\}$, for the winner being respectively Andromeda, Milky Way, Sombrero, or none. Compute $H(G)$. **[5 marks]**

(h) If the series is voided due to fair-play violations, the officials may decide to restart the series. Compute the conditional entropy of the series being restarted, given that a fair-play investigation was initiated, $H(R|F)$. **[3 marks]**

Solution: According to question,

$P = \{ AAA, MMM, SSS$ → 3 possiblities, when $Q=3$

AAMA, AMAA, MAAA, AASA, ASAA, SAAA
MMAM, MAMM, AMMM, MMSM, MSMM, SMMM
SSAS, SASS, ASSS, SSMS, SMSS, MSSS
} 18 possiblities when $Q=4$

MSAAA, SMAAA, AMSAA, ASMAA, AAMSA,
AASMA, MAASA, SAAMA, ASAMA, AMASA,
ASMMM, SAMMM, MASMM, MSAMM, MMASM,
MMSAM, AMMSM, SMMAM, MSMAM, MAMSM,
AMSSS, MASSS, SAMSS, SMASS, SSAMS,
SSMAS, ASSMS, MSSAS, SASMS, SMSAS,
} 72 $(6\times4\times3)$ possiblities, when $Q=5$

MMAAA, AMMAA, AAMMA, MAMAA, MAAMA, AMAMA
SSAAA, ASSAA, AASSA SASAA, SAASA, ASASA
- - - - - - - and so on

Also, $N(\text{no winner}) = 3\times\left({}^5C_2 \times {}^3C_2 \times {}^1C_1\right)$

$= 90$ (Accounts for all combination of 2, 2, 1)

As, $S = \{P, N\}$ ⇒ Total number of elements in $S = 93 + 90 = 183$

Sanity check: For sample space 'S'

$P_r(S) = 3\times\left(\tfrac{1}{3}\right)^3 + 18\times\left(\tfrac{1}{3}\right)^4 + 162\times\left(\tfrac{1}{3}\right)^5 = 1$

Also, P is the subset of 'S'. Hence, all probabilities in subset 'P' needs to be normalised. $P_r(P) = 3\times\left(\tfrac{1}{3}\right)^3 + 18\times\left(\tfrac{1}{3}\right)^4 + 72\left(\tfrac{1}{3}\right)^5$

$$\Rightarrow Pr(P) = \frac{1}{9} + \frac{2}{9} + \frac{72}{243} = 17/27$$

Normalising factor for probability of elements in subsets P

a) To find: $h(P = AMSSS)$

$$Pr(P = AMSSS) = \frac{(1/3)^5}{Pr(P)} = \frac{27}{17 \times 243}$$

$$= \frac{1}{153}$$

$$\Rightarrow h(P = AMSSS) = -\log_2\left(Pr(P = AMSSS)\right)$$

$$= -\log_2\left(1/153\right) = 7.2573 \text{ bits}$$

b) To find: $H(P)$

$$H(P) = -\sum_{\forall P} Pr(P)\log_2 Pr(P)$$

When $Q = 3 \Rightarrow Pr(P) = \frac{(1/3)^3}{(17/27)} = 1/17$

$Q = 4 \Rightarrow Pr(P) = \frac{(1/3)^4}{(17/27)} = 1/51$

$$Q = 5 \Rightarrow Pr(P) = \frac{(1/3)^5}{(17/27)} = 1/153$$

$$\Rightarrow H(P) = -\left[3 \times \left(\frac{1}{17} \log_2\left(\frac{1}{17}\right)\right) + 18 \times \left(\frac{1}{51} \log_2\left(\frac{1}{51}\right)\right) + 72 \times \left(\frac{1}{153} \log_2\left(\frac{1}{153}\right)\right)\right]$$

$$= \left(\frac{3}{17}\right) \log_2\left(\frac{1}{17}\right) + \left(\frac{6}{17}\right) \log_2\left(\frac{1}{51}\right) + \left(\frac{8}{17}\right) \log_2\left(\frac{1}{153}\right)$$

$$= 0.7213 + 2.0020 + 3.41524 = 6.13854 \text{ bits}$$

c) $$H(Q) = -\sum_{\forall q} Pr(Q) \log_2(Pr(Q)) \quad \text{------} \quad (22)$$

Approach 1: Using subset P

$$Pr(Q=3) = \left(\begin{array}{c}\text{No. of possiblities}\\ \text{for } Q=3\end{array}\right) \times \left(\begin{array}{c}\text{Probability of each}\\ \text{occurance in } P\end{array}\right)$$

$$= 3 \times \frac{(1/3)^3}{(17/27)} = 3/17$$

$$Pr(Q=4) = 18 \times \frac{(1/3)^4}{(27/27)} = \left(\frac{18}{81}\right) \times \left(\frac{27}{17}\right) = 6/17$$

$$Pr(Q=5) = 72 \times \frac{(1/3)^5}{(17/27)} = \frac{72 \times 27}{243 \times 17} = 8/17$$

therefore, $$H(Q) = -\left[\frac{3}{17} \log_2\left(\frac{3}{17}\right) + \frac{6}{17} \log_2\left(\frac{6}{17}\right) + \frac{8}{17} \log_2\left(\frac{8}{17}\right)\right] = 0.4416 + 0.53029 + 0.5117472$$

$$= 1.4836372 \text{ bits}$$

Approach 2: Using complete sample space 'S'

$$Pr(Q=3) = \begin{pmatrix} \text{No. of possiblities} \\ \text{for } Q=3 \end{pmatrix} \times \begin{pmatrix} \text{Probability of} \\ \text{each occurrence} \\ \text{in } S \end{pmatrix}$$

$$= 3 \times \left(\tfrac{1}{3}\right)^3 = \tfrac{1}{9}$$

$$Pr(Q=4) = 18 \times \left(\tfrac{1}{3}\right)^4 = \tfrac{2}{9}$$

$$Pr(Q=5) = (72 \times 90) \times \left(\tfrac{1}{3}\right)^5 = \frac{162}{243}$$

$$= \tfrac{2}{3}$$

therefore, $H(Q) = -\left[\tfrac{1}{9}\log_2\left(\tfrac{1}{9}\right) + \tfrac{2}{9}\log_2\left(\tfrac{2}{9}\right) + \tfrac{2}{3}\log_2\left(\tfrac{2}{3}\right)\right]$

$$= 0.3522 + 0.482205 + 0.389975$$

$$= 1.22438 \text{ bits}$$

As, question didn't explicitly says to calculate H(Q) w.r.t 'S' or 'P' (subset of S), both results will be valid.

d> To calculate : $H(P/Q=4)$

For $Q=4$, there are 18 possiblities each with $P_r(P=p) = \dfrac{\left(\frac{1}{3}\right)^4}{\left(\frac{17}{27}\right)} = \frac{1}{51}$

therefore, $P_r(P/Q=4) = \frac{1}{18} \ \forall_p$

Hence, $H(P/Q=4) = -\sum P_r(P=p/Q=4) \times \log_2(P_r(P=p/Q=4))$

$= -18\left[\left(\frac{1}{18}\right)\log_2\left(\frac{1}{18}\right)\right]$

$= \log_2 18$ bits $= 4.170340$

e> $H(P/Q) = \sum_{\forall q} P_r(Q=q)\, H(P/Q=q)$

Similar to part d, $H(P/Q=4) = \log_2 18$ bits

$H(P/Q=3) = -3 \times \left(\frac{1}{3}\log_2 \frac{1}{3}\right)$

$= \log_2 3$ bits

and $H(P/Q=5) = \log_2 72$ bits

Also, $P_r(Q=3) = 3 \times \dfrac{\left(\frac{1}{3}\right)^3}{\left(\frac{17}{27}\right)} = \frac{3}{17}$

$$Pr(Q=4) = 18 \times (1/3)^4 \times (27/17) = 6/17$$

$$Pr(Q=5) = 72 \times (1/3)^5 \times (27/17) = 8/17$$

therefore, $H(P/Q) = \sum_{\forall q} P(Q=q) \times H(P/Q=q)$

$= Pr(Q=3) \times H(P/Q=3) + Pr(Q=4) \times H(P/Q=4)$
$\quad + Pr(Q=5) \times H(P/Q=5)$

$= (3/17) \times \log_2 3 + (6/17) \times \log_2 18 + (8/17) \times \log_2 72$

$= 0.2796992 64 + 1.471738 + 2.9034941 = 4.654932$ bits

f) $H(Q/P)$

If the outcome P is provided, by counting the number of races in P, Q (total number of rounds) can be directly determined. Hence, no additional information is contained in Q, if P is given.

Therefore, $H(Q/P) = 0$

g)

P = { AAA, MMM, SSS → 3 possiblities, when Q=3

AAMA, AMAA, MAAA, AASA, ASAA, SAAA
MMAM, MAMM, AMMM, MMSM, MSMM, SMMM
SSAS, SASS, ASSS, SSMS, SMSS, MSSS } 18 possiblities when Q=4

MSAAA, SMAAA, AMSAA, ASMAA, AAMSA, MASAA
AASMA, MAASA, SAAMA, ASAMA, AMASA, SAMAA
ASMMM, SAMMM, MASMM, MSAMM, MMASM, AMSMM
MMSAM, AMMSM, SMMAM, MSMAM, MAMSM, SMAMM
AMSSS, MASSS, SAMSS, SMASS, SSAMS, ASMSS
SSMAS, ASSMS, MSSAS, SASMS, SMSAS, MSASS
MMAAA, AMMAA, AAMMA, MAMAA, MAAMA, AMAMA
SSAAA, ASSAA, AASSA, SASAA, SAASA, ASASA
AAMMM, MAAMM, MMAAM, AMAMM, AMMAM, MAMAM
SSMMM, MSSMM, MMSSM, SMSMM, SMMSM, MSMSM
AASSS, SAASS, SSAAS, ASASS, ASSAS, SASAS
MMSSS, SMMSS, SSMMS, MSMSS, MSSMS, SMSMS } for Q = 5, 72

Total number of elements in P = 93

Also, N = if the civilization who wins first and last is different.

For N, number of suspecious events = (6+36) = 42, with 2/3 chances the officials declaring series invalid.

therefore, $Pr(G=N) = \sum Pr(P=p) \times Pr(W=N, | P=p)$

$$= \frac{6 \times (2/3)^4 \times (1/3)}{(17/27)} + \frac{36 \times (1/3)^5 \times (2/3)}{(17/27)} = \frac{2}{51} + \frac{4}{51} = 6/51$$

$$= 2/17$$

Since, $Pr(G=A) = Pr(G=M) = Pr(G=S)$

Also, $\sum_{\forall i} Pr(G=G_i) = 1$

therefore, $Pr(G=A) = Pr(G=M) = Pr(G=S) = \dfrac{1 - Pr(G=N)}{3}$

$$= \frac{1 - 6/51}{3} = 15/51 = 5/17$$

$$H(G) = -\sum_{\forall i} P(G_i) \log_2 (P(G_i))$$

$$= -\left[3 \times \left(5/17\right) \log_2 \left(5/17\right) + \left(2/17\right) \log_2 \left(2/17\right)\right]$$

$$= 1.557824 + 0.3632309228$$

$$= 1.921054 \text{ bits}$$

b) Since, 'R' is the Random Variable representing series being restarted or not. Also, if the investigation finds evidence of unfair practices, there is a 50% chance that the series is declared void and will be restarted.

$$\Rightarrow Pr(R = Restart/F) = Pr(R = not\text{-}Restart/F) = 1/2$$

$$\Rightarrow H(R/F) = -\left[Pr(R = Restart/F) \log_2 (Pr(R = Restart/F)) + Pr(R = not\text{-}Restart/F) \log_2 (Pr(R = not\text{-}Restart/F))\right]$$

$$= -1/2 \log_2 (1/2) - 1/2 \log_2 (1/2) = \log_2 2 = 1 \text{ bits}$$

Question 2 [30 marks total]

****All students are expected to attempt this question.**

A standard deck of cards contains 4 *suits* — $\heartsuit, \diamondsuit, \clubsuit, \spadesuit$ ("hearts", "diamonds", "clubs", "spades") — each with 13 *values* — A, 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K (The A, J, Q, K are called "Ace", "Jack", "Queen", "King"). Each card has a *colour*: hearts and diamonds are coloured red; clubs and spades are black. Cards with values J, Q, K are called *face cards*.

Each of the 52 cards in a deck is identified by its value V and suit S and denoted VS. For example, 2$\heartsuit$, J$\clubsuit$, and 7$\spadesuit$ are the "two of hearts", "Jack of clubs", and "7 of spades", respectively. The variable c will be used to denote a card's colour. Let $f = 1$ if a card is a face card and $f = 0$ otherwise.

1. A card is drawn at random from a thoroughly shuffled deck. Calculate:
   (a) The information $h(c = \text{red}; f = 1)$ in observing a red face card **[2 marks]**
   (b) The conditional information $h(V = K|f = 1)$ in observing a King given a face card was drawn. **[3 marks]**
   (c) The entropies $H(S)$ and $H(V, S)$. **[3 marks]**
   (d) The mutual information $I(V; S)$ between V and S. **[3 marks]**
   (e) The mutual information $I(V; C)$ between the value and colour of a card using the last result and the data processing inequality **[4 marks]**

2. Now consider that Twenty cards are removed from a standard deck: All 13 $\spadesuit$s; the A$\heartsuit$, 2$\heartsuit$, 3$\heartsuit$, 4$\heartsuit$, and 5$\heartsuit$; and the A$\diamondsuit$ and K$\diamondsuit$. A card is drawn at random from the reduced deck.
   (a) Calculate the entropies $H(S)$ and $H(V, S)$. **[5 marks]**
   (b) Calculate $I(V; S)$. Explain why it is different from the $I(V; S)$ when a card is drawn at random from the standard deck of 52 cards (i.e. prior to the removal). **[5marks]**
   (c) Calculate $I(V; S|C)$. Give an intuitive explanation as to why knowing the color of a card changes the mutual information between its value and suit. **[5 marks]**

Solution: According to question,

Suits = 4 ('hearts', 'diamonds', 'clubs', 'spades')

Values = 13 (A, 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K)

Colour = 2 (Red, Black)

face/not face = 2 $\begin{cases} f=1, \text{ if face card, otherwise} \\ f=0, \text{ if not face card} \end{cases}$

face cards = {J, Q, K}

1. (a) To find: $h(C=red, f=1)$

$$Pr(C=red, f=1) = \frac{\text{No. of } (C=red, f=1)}{\text{Total no. of cards in deck}}$$

$$= \frac{2 \times 3}{52} = \frac{3}{26}$$

therefore, $h(C=red, f=1) = -\log_2(Pr(C=red, f=1))$

$$= -\log_2\left(\frac{3}{26}\right) = 3.115477217 \text{ bits}$$

b) To find: $h(V=K \mid f=1)$

$$Pr(V=K \mid f=1) = \frac{Pr(V=K, f=1)}{Pr(f=1)}$$

Since, $Pr(V=K, f=1) = \frac{4}{52} = \frac{1}{13}$

$$Pr(f=1) = \frac{12}{52} = \frac{3}{13}$$

$$\Rightarrow Pr(V=K \mid f=1) = \frac{1/13}{3/13} = \frac{1}{3}$$

therefore, $h(V=K \mid f=1) = -\log_2(P_r[V=K \mid f=1])$

$$= -\log_2\left(\tfrac{1}{3}\right) = \log_2(3)$$

$$= 1.584962 \text{ bits}$$

c) To compute: $H(S)$ and $H(V,S)$

Since, there are 4 suits { 'heart', 'diamond', 'club', and 'spade' } each with a probability $\frac{13}{52} = \frac{1}{4}$.

therefore, $H(S) = -\sum P_r(S) \log_2(P_r(S))$

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

$$= \log_2 4 = 2 \text{ bits}$$

Also, there are 52 cards in the deck each with $P_r(V,S) = \frac{1}{52}$

therefore, $H(V,S) = -\sum_V \sum_S P_r(V,S) \log_2 P_r(V,S)$

$$= -52 \times \left(\frac{1}{52} \log_2\left(\tfrac{1}{52}\right)\right) = \log_2 52$$

$$= 5.70 \text{ bits}$$

d) To find : $I(V;S)$

Since, 'V' and 'S' are independent events.

Hence, $P(V,S) = P(V) \times P(S)$ ---------- (i)

Also, $I(V;S) = \sum_{\forall V} \sum_{\forall S} P(V,S) \log_2 \left(\frac{P(V,S)}{P(V) \times P(S)} \right)$

$$= \sum_{\forall V} \sum_{\forall S} P(V,S) \log_2 \left(\frac{P(V) \times P(S)}{P(V) \times P(S)} \right) \text{ (from (i))}$$

$$= \sum_{\forall V} \sum_{\forall S} P(V,S) \log_2 1$$

$= 0$ bits, because $\log_2 1 = 0$

e) To find : $I(V;C)$

Since, $I(V;C) = \sum_{\forall V} \sum_{\forall C} P_r(V,C) \log_2 \left(\frac{P_r(V,C)}{P_r(V) \times P_r(C)} \right)$

As, 'V' and 'C' are independent events.

$\Rightarrow P_r(V,C) = P_r(V) \times P_r(C)$

$$\Rightarrow \log_2 \left(\frac{P_r(V,C)}{P_r(V) \times P_r(C)} \right) = \log_2 \left(\frac{P_r(V) \times P_r(C)}{P_r(V) \times P_r(C)} \right)$$

$$= \log_2 (1) = 0$$

Then, similar to part (d), $I(V;C)$ will also be 0.

2. a) After the removal,

$\heartsuit$	$\diamondsuit$	$\clubsuit$	$\spadesuit$
8	11	13	0

Total = 32

To calculate : $H(S)$ and $H(V,S)$

$$H(S) = -\sum P_r(S)\log_2(P_r(S))$$

$$= -\left[\frac{8}{32}\log_2\left(8/32\right) + \frac{11}{32}\log_2\left(11/32\right) + \frac{13}{32}\log_2\left(13/32\right)\right]$$

$$= 0.5 + 0.5295 + 0.52794636$$

$$= 1.55744636 \text{ bits}$$

$$H(V,S) = -\sum\sum P_r(V,S)\log_2(P_r(V,S))$$

$$= -32\left(1/32\,\log_2\left(1/32\right)\right)$$

$$= \log_2(32)$$

$$= \mathbf{5} \text{ bits}$$

b) To find : $I(V;S) = H(S) + H(V) - H(V,S)$

From part (a), $H(S) = 1.557446$ bits

and, $H(V,S) = 5$ bits

for $H(V)$:

$$P_r(V) = \begin{cases} 2/32 & ; \text{ when } V = \{2,3,4,5,K\} \\ 1/32 & ; \text{ when } V = \{A\} \\ 3/32 & ; \text{ elsewhere} \end{cases}$$

therefore, $H(V) = -\sum_{\forall v} P_r(v) \log_2(P_r(v))$

$$= -\left(\frac{2}{32}\log_2\left(\frac{2}{32}\right)\times 5 + \frac{3}{32}\log_2\left(\frac{3}{32}\right)\times 7 + \frac{1}{32}\log_2\left(\frac{1}{32}\right)\right) = 1.25 + 2.2411 + 0.15625 = 3.6473$$

Thus, $I(V;S) = H(S) + H(V) - H(V,S)$

$$= (1.55744636 + 3.64735 - 5) \text{ bits}$$

$$= 0.204796 \text{ bits}$$

When a card is drawn at random from a standard deck of 52 cards (i.e., prior to the removal of 20 cards), then 'V' and 'S' will be independent, resulting in $P_r(V,S) = P_r(V) \times P_r(S)$ and $I(V;S) = 0$.

However, when 20 cards are removed from the deck, then 'V' and 'S' will no longer remain indepen-dent resulting in $I(V;S) \neq 0$.

c) To calculate : $I(V;S|C)$

$$H(V/C) = \sum_{\forall c} P_r(C=c)\, H(V/C=c)$$

<u>For Red:</u>

$$P_r(C=r) = \frac{19}{32}$$

$$P_r(V/C=r) = \begin{cases} \frac{1}{19}; & V = 2,3,4,5,K \\ 0; & V = A \\ \frac{2}{19}; & \text{elsewhere} \end{cases}$$

<u>For Black:</u>

$$P_r(C=b) = \frac{13}{32}$$

$$P_r(V/C=b) = \frac{1}{13}$$

$$H(V/C) = -\frac{19}{32}\left[\frac{1}{19}\log_2\left(\frac{1}{19}\right)\times 5 + \frac{2}{19}\log_2\left(\frac{2}{19}\right)\times 7\right] -$$

$$\frac{13}{32}\left[\frac{1}{13}\log_2\left(\frac{1}{13}\right)\times 13\right]$$

$$= \frac{19}{32}\left[1.11787566 + 2.393209747\right] +$$

$$\frac{13}{32}\left[3.700439718\right]$$

$$= 2.08470696 + 1.503303635$$

$$= 3.588010595 \text{ bits}$$

Also, $H(V/S,C) = H(V/S)$, because if we know 'S', then 'C' is already known

Since, $Pr(S = \heartsuit) = 8/32$

$Pr(S = \diamondsuit) = 11/32$

$Pr(S = \clubsuit) = 13/32$

$Pr(S = \spadesuit) = 0/32$

And, $Pr(V/S = \heartsuit) = 1/8$

$Pr(V/S = \diamondsuit) = 1/11$

$Pr(V/S = \clubsuit) = 1/13$

$Pr(V/S = \spadesuit) = 0$

therefore, $H(V/S) = \frac{8}{32}\left[8 \times \left(\frac{1}{8}\log_2 8\right)\right] + \frac{11}{32}\left[11 \times \left(\frac{1}{11}\log_2 11\right)\right] + \frac{13}{32}\left[13 \times \left(\frac{1}{13}\log_2 13\right)\right]$

$$= \frac{3}{4} + \frac{11}{32}\log_2 11 + \frac{13}{32}\log_2 13 = 2.253303638$$

$= 3.4424832$ bits

By using chain rule,

$$I(V; S/C) = H(V/C) - H(V/S, C)$$

$$= H(V/C) - H(V/S) = 3.588010 - 3.44248$$

$$= 0.145527395$$

Question 3 [20 marks total]

****All students are expected to attempt this question.**

(a) Consider a fair coin flip. What is the mutual information between the top side and the bottom side of the coin **[3 marks]**

(b) A 6-sided fair die is rolled. What is the mutual information between the top side and the front face (the side most facing you)? **[5 marks]**

(c) Show that the entropy of the probability distribution, $(p_1, \cdots, p_i, \cdots, p_j, \cdots, p_m)$, is less than the entropy of the distribution $(p_1, \cdots, \frac{p_i+p_j}{2}, \cdots, \frac{p_i+p_j}{2}, \cdots, p_m)$. **[6 marks]**

(d) Let X be a random variable with m different outcomes $\mathcal{X} = \{1, 2, \cdots, m\}$ and corresponding nonzero probabilities $\{q_1, q_2, \cdots, q_m\}$ arranged in decreasing order. Now define a new random variable Y with m+1 different outcomes $\{\frac{q_1}{3}, \frac{q_2}{3}, \cdots, \frac{q_{m-1}}{3}, \frac{q_m}{3}, \frac{2}{3}\}$. Find the entropy of $H(Y)$ in terms of $H(X)$. **[6 marks]**

Solution:

a) let 'T' be the random variable representing 'Top' and 'B' be the random variable representing 'Bottom'.

To find: $I(T;B) = H(B) - H(B/T)$
$= H(T) - H(T/B)$

Since, bottom side of coin can either take 'Head' or 'Tail'. Therefore,

$$H(B) = - \sum_{\forall b_i = \{Head, Tail\}} P_r(b_i) \log_2(P_r(b_i))$$

$$= 2 \times \left(\tfrac{1}{2} \log_2 2\right) = \log_2(2) = 1 \text{ bits}$$

As, T and B are dependent events.

$$\therefore P_r(B=b_i / T=t_i) = \begin{cases} 0, & \text{if } b_i = t_i \\ 1, & \text{elsewhere} \end{cases}$$

$$\Rightarrow H(B/T) = 0$$

and $I(T;B) = H(B) - H(B/T)$

$$= \log_2(2) - 0 = 1 \text{ bits}$$

b) let 'T' be the random variable representing 'Top' and 'F' be the random variable representing the 'Front Face'.

Since, T has uniform distribution on $\{1, 2, 3 \ldots, 6\}$

therefore, $H(T) = 6 \times \left(\frac{1}{6}\right) \log_2 6 = \log_2 6$

Also, note that having observed a side of the cube facing us F, there are four possiblities for the top T, which are equally probable. Thus,

$$I(T;F) = H(T) - H(T/F)$$
$$= \log_2(6) - \log_2(4)$$
$$= 2.585 - 2$$
$$= 0.586 \text{ bits}$$

c) This problem depends on the Convexity of the log function.

Let $P_1 = (p_1, p_2, \ldots \ldots p_i, \ldots \ldots p_j, \ldots \ldots p_m)$

$$P_2 = \left(p_1, \ldots \ldots \frac{p_i + p_j}{2}, \ldots \ldots, \frac{p_j + p_i}{2}, \ldots p_m\right)$$

then, by the log sum inequality

$$H(P_2) - H(P_1) = -2\left(\frac{p_i + p_j}{2}\right)\log_2\left(\frac{p_i + p_j}{2}\right) + p_i \log_2 p_i + p_j \log_2 p_j$$

$$= -(p_i + p_j)\log_2\left(\frac{p_i + p_j}{2}\right) + p_i \log_2 p_i + p_j \log_2 p_j$$

$$\geq 0$$

Thus, $H(P_2) \geq H(P_1)$, or equivalently $H(P_1) \leq H(P_2)$

Therefore, we prove that the entropy of the probability distribution $(P_1, \_\_\_, P_i \_\_\_, P_j \_\_\_, P_m)$, is less than the entropy of the distribution

$$\left(P_1, \_\_\_, \frac{P_i + P_j}{2}, \_\_\_, \frac{P_i + P_j}{2}, \_\_, P_m\right)$$

d) According to question,

X is a random variable with 'm' different outcomes i.e. $X = \{1, 2, \ldots m\}$. Also, the non-zero probabilities corresponding to X in decreasing order are $\{q_1, q_2, \ldots q_m\}$

let Y be a random variable with $m+1$ different outcomes $\left\{\frac{q_1}{3}, \frac{q_2}{3}, \ldots, \frac{q_{m-1}}{3}, \frac{q_m}{3}, \frac{2}{3}\right\}$.

To find : $H(Y)$ in terms of $H(X)$.

$$H(X) = -\sum_{i=1}^{m} q_i \log_2(q_i) \quad \text{------} \quad (i)$$

$$Y = \left\{\frac{q_1}{3}, \frac{q_2}{3}, \ldots, \frac{q_{m-1}}{3}, \frac{q_m}{3}, \frac{2}{3}\right\}$$

$$H(Y) = -\left[\sum_{i=1}^{m} \left(\frac{q_i}{3}\right) \log_2\left(\frac{q_i}{3}\right) + \left(\frac{2}{3}\right) \log_2\left(\frac{2}{3}\right)\right]$$

$$= -\frac{1}{3}\left[\sum_{i=1}^{m} q_i \log_2\left(\frac{q_i}{3}\right) + (2/3) \log_2(2/3)\right]$$

$$= -\frac{1}{3}\left[\sum_{i=1}^{m} q_i (\log_2 q_i - \log_2 3) + 2(\log_2 2 - \log_2 3)\right]$$

$$= -1/3 \sum_{i=1}^{m} q_i (\log_2 q_i - \log_2 3) - 1/3(2 - 2\log_2 3)$$

$$= \underbrace{-\frac{1}{3}\sum_{i=1}^{m} q_i \log_2 q_i}_{H(X)} + \frac{1}{3}\sum_{i=1}^{m} q_i \log_2 3 - \frac{2}{3} + \left(\frac{2}{3}\right)\log_2 3$$

$$= \frac{H(X)}{3} + \frac{1}{3}\left[\sum_{i=1}^{m} q_i \log_2 3 + 2\log_2 3\right] - \frac{2}{3}$$

$$= \frac{H(X)}{3} + \frac{\log_2 3}{3}\left[\underbrace{\sum_{i=1}^{m} q_i}_{=1} + 2\right] - \frac{2}{3}$$

OR $= 1$ [Sum of all probabilities]

$$= \frac{1}{3}H(X) + \frac{\log_2 3}{3}[1+2] - \frac{2}{3}$$

$$= \frac{1}{3}H(X) + \log_2 3 - \frac{2}{3}$$

Question 4 [15 marks total]

****Section I: Only COMP 2610 students are expected to attempt this question.**

Let Z be a random variable with possible outcomes $0, 2$, and $p(Z = 2) = 1/2$. Let W be a random variable with possible outcomes $W = \{d, e, f\}$. Define

$$\mathbf{r} = (p(W = d|Z = 2), p(W = e|Z = 2), p(W = f|Z = 2))$$

$$\mathbf{s} = (p(W = d|Z = 0), p(W = e|Z = 0), p(W = f|Z = 0)).$$

(a) Suppose that

$$\mathbf{r} = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$$

$$\mathbf{s} = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$$

.

Compute $I(W; Z)$. **[3 marks]**

(b) Using the definition of mutual information, show that for any choice of $\mathbf{r}$, $\mathbf{s}$,

$$I(W; Z) = \frac{D_{\mathrm{KL}}(\mathbf{r}||\mathbf{n} + \mathbf{s}||\mathbf{n})}{2}$$

where $\mathbf{n} = \frac{\mathbf{r}+\mathbf{s}}{2}$. **[5 marks]**

(c) Explain the relationship between mutual information, entropy, and conditional entropy. You can use mathematical derivations to explain. **[2 marks]**

(d) Why is mutual information considered a measure of the dependence between two random variables? **[2 marks]**

(e) How does the concept of Kullback-Leibler divergence relate to mutual information? **[3marks]**

Solution:

According to question,

$$Z \in \{0, 2\} \text{ and } P_r(z=2) = 1/2$$

$$\Rightarrow P_r(z=0) = 1 - P_r(z=2) = 1/2$$

Also, $W \in \{d, e, f\}$, and

$$r = (P(W=d|z=2), P(W=e|z=2), P(W=f|z=2))$$

$$s = (P(W=d|z=0), P(W=e|z=0), P(W=f|z=0))$$

Given, $r = (1/4, 1/2, 1/4)$ and $s = (1/4, 1/2, 1/4)$

a) since, $D_{KL}(r||m) = -\sum_{W} r \log_2 (r/m)$

Here, $m = \frac{r+s}{2}$

$$\Rightarrow m = \{1/4, 1/2, 1/4\}$$

therefore, $D_{KL}(r||m) = \left[\frac{1}{4}\log_2\left(\frac{1}{4}\times 4\right)\right.$

$$\left. + \frac{1}{2}\log_2\left(\frac{1}{2}\times 2\right) + \frac{1}{4}\log_2\left(\frac{1}{4}\times \frac{4}{1}\right)\right]$$

$$= \left[\frac{1}{4}\log_2(1) + \frac{1}{2}\log_2(1) + \frac{1}{4}\log_2(1)\right]$$

$$= 0$$

Similarally, $D_{KL}(S||N) = \sum_{W} s \log_2 (s/n)$

$$\Rightarrow D_{KL}(S||N) = \left[\frac{1}{4}\log_2\left(\frac{1}{4}\times 4\right) + \frac{1}{2}\log_2\left(\frac{1}{2}\times 2\right) + \frac{1}{4}\log_2\left(\frac{1}{4}\times\frac{4}{1}\right)\right]$$

$$= \left[\frac{1}{4}\log_2(1) + \frac{1}{2}\log_2(1) + \frac{1}{4}\log_2(1)\right]$$

$$= 0$$

Therefore, $I(W;Z) = \dfrac{D_{KL}(r||n + s||n)}{2}$

$$= \frac{0}{2} = 0$$

b) To show: For any choice of r, s,

$$I(W;Z) = \frac{D_{KL}(r||n + s||n)}{2}$$

Where, $n = \dfrac{r+s}{2}$

Since, $I(W;Z) = D_{KL}\left(P_r(W,Z) \,||\, P_r(W) \times P_r(Z)\right)$

$$= -\sum_{\forall W,Z} P_r(W,Z) \log_2\left(\frac{P_r(W,Z)}{P_r(W)\times P_r(Z)}\right)$$

$$= -\sum_{\forall W,Z} P_r(W/Z)\, P_r(Z) \log_2\left(\frac{P_r(W/Z)\cancel{P_r(Z)}}{P_r(W)\times \cancel{P_r(Z)}}\right) \quad \text{(By chain rule)}$$

$$= - \sum_{\forall w,z} Pr(W/Z)\, Pr(Z) \log_2 \left(\frac{Pr(W/Z)}{Pr(W)} \right)$$

$$= - \left[\sum_{\forall W} Pr(W/Z=0)\, Pr(Z=0) \log_2 \left(\frac{Pr(W/Z=0)}{Pr(W)} \right) + \sum_{\forall W} Pr(W/Z=2)\, Pr(Z=2) \log_2 \left(\frac{Pr(W/Z=2)}{Pr(W)} \right) \right.$$

$$= \frac{D_{KL}(r \,||\, P(W)) + D_{KL}(s \,||\, P(W))}{2}$$

Since, $P(W) = \sum_{\forall z} P(W/Z)\, P(Z)$

$$= P(W/Z=0)\, P(Z=0) + P(W/Z=2)\, P(Z=2)$$

$$= P(W/Z=0) \times 1/2 + P(W/Z=2) \times 1/2$$

$$= \frac{P(W/Z=0) + P(W/Z=2)}{2} = n$$

therefore,

$$I(W;Z) = \frac{D_{KL}(r \,||\, n + s \,||\, n)}{2}, \text{ where } n = \frac{r+s}{2}$$

c) Entropy (H): The entropy $H(X)$ of a random variable X is a measure of the amount of uncertainity associated with it.

$$H(X) = -\sum P(X) \log_2 P(X), \text{ for all } X$$

Conditional Entropy $(H(X|Y))$: Conditional Entropy of random variable X, given that the value of another random variable Y is known.

$$H(X|Y) = -\sum P(X,Y) \log_2 P(X|Y), \text{ for all } X, Y$$

Mutual Information (I): It is the measure of mutual dependence between two random variable (let say X and Y). It quantifies the "amount of information" about one random variable (X), through the other random variable (Y)

$$I(X;Y) = \sum P(X,Y) \log_2 (P(X,Y) / P(X) P(Y)), \text{ for all } (x,y) \in (X,Y)$$

Relationship between mutual information $(I(X;Y))$ and Entropy (H):

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

d7 Since, $I(X;Y) = H(X) - H(X|Y)$
Mutual information $(I(X;Y))$ is considered the measure of dependence between two random variables (X, Y) because it quantifies the reduction in uncertai-nity about one random variable after observing other. In other words, it measures how much knowing the value of one random variable helps in predicting the value of the other.

e) The Kullback - Leibler (KL) divergence, also known as relative entropy is a measure of how one probability distribution is different from other (the reference probability distribution). The Kullback - Leibler divergence of two probability distributions P and Q is defined as,

$$D(P \| Q) = \sum P(x) \log_2 (P(x)/Q(x)), \forall x \in X$$

- Kullback - Leibler divergence (D_{KL}) is not symmetric, meaning $D(P \| Q) \neq D(Q \| P)$.
- Mutual Information ($I(X;Y)$) can be expressed as the KL divergence between the joint distribution $P(X,Y)$ and the product of the marginal distribution $P(X)P(Y)$.

i.e., $I(X;Y) = D(P(X,Y) \| P(X)P(Y))$

****Section II: Only COMP 6261 students are expected to attempt this question.**

Let Y be a random variable with possible outcomes 0, 1, and $p(Y = 1) = 1/2$. Let X be a random variable with possible outcomes $X = \{a, b, c\}$. Define

$$\mathbf{p} = (p(X = a|Y = 1), p(X = b|Y = 1), p(X = c|Y = 1))$$

$$\mathbf{q} = (p(X = a|Y = 0), p(X = b|Y = 0), p(X = c|Y = 0)).$$

(a) Suppose that

$$\mathbf{p} = (\frac{3}{5}, \frac{1}{5}, \frac{1}{5})$$

$$\mathbf{q} = (\frac{1}{5}, \frac{3}{5}, \frac{1}{5})$$

.

Compute $I(X; Y)$. **[3 marks]**

(b) Using the definition of mutual information, show that for any choice of **p**, **q**,

$$I(X; Y) = \frac{D_{\mathrm{KL}}(\mathbf{p}||\mathbf{m} + \mathbf{q}||\mathbf{m})}{2}$$

where $\mathbf{m} = \frac{\mathbf{p}+\mathbf{q}}{2}$. **[3 marks]**

(c) Let Z be a random variable with outcomes $x \in \mathtt{X}$, such that for all $x \in \mathtt{X}$ and $y \in \{0, 1\}$, $p(Z = x|Y = y) = p(X = x|Y = 1 - y)$. Using part (b) compute $I(Z; Y)$. Explain intuitively the relation of your answer to part (b). **[3 marks]**

(d) Suppose **p** and **q** are as in part (a). Using part (b), or otherwise, give an example of a random variable Z with possible outcomes $\mathtt{X}$ satisfying $I(X; Y) > I(Z; Y)$. Explain your answer in terms of the data-processing inequality. **[3 marks]**

(e) Suppose **p** and **q** are as in part (a). Give an example of a random variable Z with possible outcomes $x \in \mathtt{X}$ satisfying $I(X; Y) < I(Z; Y)$. Explain your answer in terms of the data-processing inequality. **[3marks]**

Solution:

According to question,

$$Y \in \{0, 1\} \text{ and } P_r(Y=1) = 1/2$$

$$\Rightarrow P_r(Y=0) = 1 - P_r(Y=1) = 1/2$$

Also, $X \in \{a, b, c\}$, and

$\mathbf{p} = (P(X=a \mid Y=1), P(X=b \mid Y=1), P(X=c \mid Y=1))$

$\mathbf{q} = (P(X=a \mid Y=0), P(X=b \mid Y=0), P(X=c \mid Y=0))$

Given, $\mathbf{p} = (3/5, 1/5, 1/5)$ and $\mathbf{q} = (1/5, 3/5, 1/5)$

a) since, $D_{KL}(p \| m) = \sum_x p \log_2 (p/m)$

Here, $m = \dfrac{p+q}{2}$

$$\Rightarrow m = \{2/5, 2/5, 1/5\}$$

therefore, $D_{KL}(p \| m) = \left[3/5 \log_2 \left(\dfrac{3/5}{2/5} \right) + 1/5 \log_2 \left(\dfrac{1/5}{2/5} \right) + 1/5 \log_2 \left(\dfrac{1/5}{1/5} \right) \right]$

$$= \left[3/5 \log_2 (3/2) + 1/5 \log_2 (1/2) + 1/5 \log_2 (1) \right]$$

$$= \left[3/5 \log_2 (3/2) - 1/5 + 0 \right] = 0.350977 - 0.2$$

$$= 0.150977$$

Similarally, $D_{KL}(q||m) = \sum_{X} q \log_2 (q/m)$

$$= \frac{1}{5} \log_2 \left(\frac{1/5}{2/5} \right) + \frac{3}{5} \log_2 \left(\frac{3/5}{2/5} \right) + \frac{1}{5} \log_2 \left(\frac{1/5}{1/5} \right)$$

$$= \frac{1}{5} \log_2 (1/2) + \frac{3}{5} \log_2 (3/2) + \frac{1}{5} \log_2 (1)$$

$$= -\frac{1}{5} + \frac{3}{5} \log_2 (3/2) + 0$$

$$= 0.150977$$

Hence, $I(X;Y) = \dfrac{D_{KL}(P||m + q||m)}{2}$

$$= \frac{0.150977 + 0.150977}{2}$$

$$= 0.150977$$

Note: This is one approach. However, this question can be solved in many different ways.

b) To show: For any choice of P, q

$$I(X;Y) = \frac{D_{KL}(P||m + q||m)}{2}$$

Where, $m = \dfrac{P+q}{2}$

Since, $I(X;Y) = D_{KL}\left(P(X,Y) \,||\, P(X)\,P(Y)\right)$

$$= -\sum_{\forall xy} P(X,Y) \log_2 \left(\frac{P(X,Y)}{P(X)\,P(Y)}\right)$$

$$= -\sum_{\forall xy} P(X/Y)\, P(Y) \log_2 \left(\frac{P(X/Y)}{P(X)}\right)$$

$$= -\left[\sum_{\forall x} P(X/Y=0)\, P(Y=0) \log_2 \left(\frac{P(X/Y=0)}{P(X)}\right) + \sum_{\forall x} P(X/Y=1)\, P(Y=1) \log_2 \left(\frac{P(X/Y=1)}{P(X)}\right)\right]$$

$$= \frac{D_{KL}(p \,||\, P(X)) + D_{KL}(q \,||\, P(X))}{2}$$

Since, $P(X) = \sum_{\forall Y} P(X/Y)\, P(Y)$

$$= P(X/Y=0)\, P(Y=0) + P(X/Y=1)\, P(Y=1)$$

$$= P(X/Y=0) \times 1/2 + P(X/Y=1) \times 1/2$$

$$= \frac{P(X/Y=0) + P(X/Y=1)}{2}$$

$$\Rightarrow P(X) = \frac{p+q}{2} = m$$

therefore,

$$I(X;Y) = \frac{D_{KL}(P \| m) + D_{KL}(q \| m)}{2}$$

c) $p(Z=x / Y=y) = P(X=x / Y=1-y)$

This states that, if

$$\mathbb{P}' = \{ P(Z=a / Y=1), \text{------} \}$$

$$q' = \{ P(Z=a / Y=0), \text{-----} \}$$

$$\mathbb{P}' = q \;\&\; q' = p \text{ and } m' = \frac{p' + q'}{2} = m \quad \text{---} \text{R}$$

Using results from part b,

$$I(Z;Y) = \frac{D_{KL}(q' \| m') + D_{KL}(p' \| m')}{2}$$

Using eqn. R, we get

$$I(Z;Y) = I(X;Y)$$

Proved

Q4.

d.

We can construct a Markov chain $Y \to X \to Z$, which would imply that $I(X;Y) \geq I(Z;Y)$. To enforce strict equality, we must ensure that Z somehow discards information from X. For example, let Z be uniformly from $\{\mathsf{a}, \mathsf{b}\}$ if $X = \mathsf{a}$, uniformly from $\{\mathsf{a}, \mathsf{c}\}$ if $X = \mathsf{b}$, and uniformly from $\{\mathsf{b}, \mathsf{c}\}$ if $X = \mathsf{c}$. Then, we have that

$$p(Z = \mathsf{a}|Y = y) = \frac{p(X = \mathsf{a}|Y = y) + p(X = \mathsf{b}|Y = y)}{2}$$

$$p(Z = \mathsf{b}|Y = y) = \frac{p(X = \mathsf{a}|Y = y) + p(X = \mathsf{c}|Y = y)}{2}$$

$$p(Z = \mathsf{c}|Y = y) = \frac{p(X = \mathsf{b}|Y = y) + p(X = \mathsf{c}|Y = y)}{2},$$

so that

$$\mathbf{p}' = (2/5, 2/5, 1/5)$$

$$\mathbf{q}' = (2/5, 2/5, 1/5)$$

It can now be checked, using the formula from part (ii), that

$$I(Z;Y) \qquad < I(X;Y).$$

e.

We can construct a Markov chain $Y \to Z \to X$, where now X is a noisy version of Z. For example, let Z have conditional probability vectors

$$\mathbf{p}' = (1/2, 1/2, 0)$$

$$\mathbf{q}' = (1/2, 0, 1/2)$$

and say that X is uniform from $\{\mathsf{a}, \mathsf{b}\}$ if $Z = \mathsf{a}$, uniform from $\{\mathsf{a}, \mathsf{c}\}$ if $Z = \mathsf{b}$, and uniform from $\{\mathsf{b}, \mathsf{c}\}$ if $Z = \mathsf{c}$. It may be verified that this results in the same $\mathbf{p}, \mathbf{q}$ for X as in part (i); it may further be seen that $I(Z;Y) = +\infty > I(X;Y)$, with the infinite mutual information owing to the KL divergence blowing up due to the presence of zero entries in $\mathbf{p}', \mathbf{q}'$.

Question 5 [10 marks total]

****All students are expected to attempt this question.**

Suppose X is a real valued random variable with $\mu = E(X) = 0$

(a) Show that for any $t > 0$,

$$E(e^{tX}) \leq e^{g(t(b-a))},$$

where $g(u) = -\gamma u + \log(1 - \gamma + \gamma e^u)$, with $\gamma = -a/(b-a)$. **[6marks]**

(Hint: write X as a convex combination of a and b, where the convex weighting parameter depends upon X. Exploit the convexity of the function $x \to e^{tx}$ and the fact that inequalities are preserved upon taking expectations of both sides, since expectations are integrals.)

(b) By using Taylor's theorem, show that for all $u > 0$,

$$g(u) \leq \frac{u^2}{8}.$$

Furthermore show that

$$E(e^{tX}) \leq e^{\frac{t^2(b-a)^2}{8}}.$$

[4marks]

Q5.

a.

Consider $X = [a, b]$. So we can write $X = \alpha b + (1 - \alpha)a$, where $\alpha = \frac{X-a}{b-a}$. Note that $\alpha < 1$. It can be proved that e^{tX} is convex in X. Therefore,

$$e^{tX} \leq \alpha e^{ta} + (1 - \alpha e^{tb}) = \frac{X - a}{b - a} e^{ta} + \frac{b - X}{b - a} e^{tb}.$$

When taking the expectation, we obtain

$$E(e^{tX}) \leq \frac{-a}{b - a} e^{ta} + \frac{-X}{b - a} e^{tb},$$

since $E(X) = 0$. It can also be proven that

$$e^{g(t(b-a))} = \frac{-a}{b - a} e^{ta} + \frac{-X}{b - a} e^{tb},$$

when $g(u) = -\gamma u + \log(1 - \gamma + \gamma e^u)$ and $\gamma = -a/(b - a)$. Therefore,

$$E(e^{tX}) \leq e^{g(t(b-a))}.$$

b.

Using Taylor's Theorum,

$$g(u) = g(o) + \frac{g^1(0)}{1!} + + \frac{g^2(0)}{2!} + \cdots .$$

It can be proven that $g(0) = g^1(0) = 0$. Also, we obtain $g^2(u) = \frac{\alpha(1-\alpha)e^u}{1-\alpha+\alpha e^u}$. At $u = 0$, $g^2(0) = \alpha(1 - \alpha)$. It can be proven that $\alpha(1 - \alpha) < 1/4$. Combining all these results, we obtain

$$g(u) < 0 + 0 + \frac{u^2/4}{2} = \frac{u^2}{8}.$$

Thereafter, substituting this results in (a), we obtain

$$E(e^{tX}) \leq e^{g(t(b-a))} = e^{\frac{t^2(b-a)^2}{8}}.$$