AUSTRALIAN NATIONAL UNIVERSITY

COMP2610/COMP6261

Information Theory, Semester 2 2022

Assignment 2

Due Date: Monday 26 September 2022, 5:00 pm

Assignment 2 weighting is 20% 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 1-3* and *Question 4*. You are not expected to answer Question 5. You will be marked out of 100.
3. **COMP6261 students:** Answer *Questions 1-3* and *Question 5*. You are not expected to answer Question 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 in the cover page.**
3. All assignments must be done individually. Plagiarism is a university offence and will be dealt with according to university procedures *http://academichonesty.anu.edu.au/UniPolicy.html*.

Question 1: Inequalities [20 marks total]

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

Question 1(a)

Let the average height of a Raccoon is 10 inches.

1. Use Markov's inequality to derive an upper bound on the probability that a certain raccoon is at least 15 inches tall. (You may leave your answer as a fraction.) [**3 Marks**]

2. Suppose the standard deviation in raccoon's height distribution is 2 inches. Use Chebyshev's inequality to derive a lower bound on the probability that a certain raccoon is between 5 and 15 inches tall. (You may leave your answer as a fraction.) [**3 Marks**]

Question 1(b)

A coin is known to land heads with probability $(p) < 1/6$. The coin is flipped N times for some even integer N.

1. Using Markov's inequality, provide a bound on the probability of observing $N/3$ or more heads. [**3 Marks**]

2. Using Chebyshev's inequality, provide a bound on the probability of observing $N/3$ or more heads. Express your answer in terms of N. [**3 Marks**]

3. For $N \in \{3, 6, \ldots, 30\}$, in a single plot, show the bounds from part (a) and (b), as well as the exact probability of observing $N/3$ or more heads. [Note: To demonstrate, you can choose any specific value of p < 1/6. Also, you can choose any plotting tool] [**8 Marks**]

Question 2 : Markov Chain [30 marks total]

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

Question 2(a)

Random variables X, Y, Z are said to form a Markov chain in that order (denoted by $X \to Y \to Z$) if their joint probability distribution can be written as:

$$p(X, Y, Z) = p(X) \cdot p(Y|X) \cdot p(Z|Y)$$

1. Suppose (X, Y, Z) forms a Markov chain. Is it possible for $I(X; Y) = I(X; Z)$? If yes, give an example of X, Y, Z where this happens. If no, explain why not. [**3 Marks**]

2. Suppose (X, Y, Z) does not form a Markov chain. Is it possible for $I(X; Y) \geq I(X; Z)$? If yes, give an example of X, Y, Z where this happens. If no, explain why not. [**3 Marks**]

3. If $X \to Y \to Z$ then show that [**6 Marks**]

    - $I(X; Z) \leq I(X; Y)$
    - $I(X; Y|Z) \leq I(X; Y)$

Question 2(b)

Let $X \to (Y, Z) \to T$ form a Markov chain, where by Markov property we mean:

$$p(x, y, z, t) = p(x)p(y, z|x)p(t|y, z)$$

Or simply:

$$p(t|y, z, x) = p(t|y, z)$$

Do the following:

1. Prove that $I(X; Y, Z) \geq I(X; T)$. [**5 Marks**]

2. Find the condition that $I(X; Y, Z) = I(X; T)$. [**3 Marks**]

Question 2(c)

Recall that Markov's inequality states that if X is a non-negative random variable, for any $a > 0$,

$$P(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

1. Give an example of a non-negative random variable X for which Markov's statement is an *equality*, i.e.for any $a > 0$, [**3 Marks**]

$$P(X \geq a) = \frac{\mathbb{E}[X]}{a}$$

2. Given an example of a random variable Y (not necessarily non-negative) for which Markov's statement reverses, i.e. for any $a \geq 0$, [**3 Marks**]

$$P(Y \geq a) \geq \frac{\mathbb{E}[Y]}{a}$$

3. Let Z be a random variable such that $\mathbb{E}[Z] = 0$. Then, for any $a > 0$, Markov's inequality tells us that, for any $a > 0$,

$$P(|Z| \geq a) \leq \frac{\mathbb{E}[|Z|]}{a}$$

while Chebyshev's inequality tells us that

$$P(|Z| \geq a) \leq \frac{\mathbb{V}[Z]}{a^2}$$

Is it possible for the bound in Markov's inequality to be tighter than that from Chebyshev's inequality for some $a > 0$, i.e. does there exist a Z and $a > 0$ such that [**2 Marks**]

$$\frac{\mathbb{E}[|Z|]}{a} < \frac{\mathbb{V}[Z]}{a^2}?$$

If yes, provide an example of a random variable Z and a number $a > 0$ for which this is true. If no, provide a proof that this is impossible. [**2 Marks**]

Question 3: AEP [25 marks total]

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

Let X be an ensemble with outcomes $\mathcal{A}_X = \{\text{h}, \text{t}\}$ with $p_h = 0.8$ and $p_t = 0.2$. Consider X^N - e.g., N i.i.d flips of a bent coin.

a) Calculate $H(X)$. [**3 Marks**]

b) What is the size of the alphabet $\mathcal{A}_{X^N}$ of the extended ensemble X^N? [**3 Marks**]

c) What is the Raw bit content $H_0(X^4)$? [**4 Marks**]

d) Express Entropy $H(X^N)$ as a function of N. [**5 Marks**]

e) Let S_δ be the smallest set of N-outcome sequences with $P(\mathbf{x} \in S_\delta) \geq 1 - \delta$ where $0 \leq \delta \leq 1$. Use any program language of your choice to plot $\frac{1}{N} H_\delta(X^N)$ ('Normalised Essential Bit Content') vs δ for various values of N (include some small values of N such as 10 as well as large values greater than 1000. Describe your observations and comment on any insights. [**10 Marks**]

Question 4: AEP [25 marks total]

****Only COMP2610 students are expected to attempt this question.**

Let X be an ensemble with alphabet $\mathcal{A}_X = \{a, b\}$ and probabilities $(\frac{2}{5}, \frac{3}{5})$

a) Calculate $H(X)$. **[3 Marks]**

b) Recall that X^N denotes an extended ensemble. What is the alphabet of the extended ensemble X^3? **[2 Marks]**

c) Give an example of three sequences in the typical set (for $N = 3$, $T_{N\beta} = 0.2$). **[5 Marks]**

d) What is the smallest δ−sufficient subset of X^3 when $\delta = 1/25$ and when $\delta = 1/10$? **[5 Marks]**

e) Suppose $N = 1000$, what fraction of the sequences in X^N are in the typical set (at $\beta = 0.2$) ? **[7 Marks]**

f) If $N = 1000$, and a sequence in X^N is drawn at random, what is the (approximate) probability that it is in the N, β-typical set? **[3 Marks]**

Question 5: AEP [25 marks total]

****Only COMP6261 students are expected to attempt this question.**

Suppose a music collection consists of 4 albums: the album *Alina* has 7 tracks; the album *Beyonce* has 12; the album *Cecilia* has 15; and the album *Derek* has 14.

1. How many bits would be required to uniformly code:
   (a) all the albums? Give an example uniform code for the albums. [**3 Marks**]
   (b) only the tracks in the album Alina. Give an example of a uniform code for the tracks assuming they are named "Track 1", "Track 2", etc. [**3 Marks**]
   (c) all the tracks in the music collection? [**2 Marks**]
2. What is the *raw bit content* required to distinguish all the tracks in the collection? [**2 Marks**]
3. Suppose every track in the music collection has an equal probability of being selected. Let A denote the album title of a randomly selected track from the collection.
   (a) Write down the ensemble for A – that is, its alphabet and probabilities. [**2 Marks**]
   (b) What is the raw bit content of A^4? [**2 Marks**]
   (c) What is the smallest value of δ such that the smallest δ-sufficient subset of A^4 contains fewer than 256 elements? [**2 Marks**]
   (d) What is the largest value of δ such that the essential bit content $H_\delta(A^4)$ is strictly greater than zero? [**2 Marks**]
4. Suppose the album titles ensemble A is as in part (c).
   (a) Compute an approximate value for the entropy $H(A)$ to two decimal places (you may use a computer or calculator to obtain the approximation but write out the expression you are approximating). [**2 Marks**]
   (b) Approximately how many elements are in the typical set $T_{N\beta}$ for A when $N = 100$ and $\beta = 0.1$? [**3 Marks**]
   (c) Is it possible to design a uniform code to send large blocks of album titles with a 95% reliability using at most 1.5 bits per title? Explain why or why not. [**2 Marks**]