AUSTRALIAN NATIONAL UNIVERSITY

COMP2610/COMP6261

Information Theory, Semester 2 2022

Assignment 2 Solution

Release Date: Wednesday, 5 October 2022

Assignment 2 weighting is 20% of the course mark.

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**]

Solution:

I. Let X be a random variable that describes the height of raccoon in inches. According to the Markov's inequality, we have the upper bound:

$$p(X \geq 15) \leq E[X]/15 = 10/15 = 2/3$$

II. We want to calculate:

$$p(5 < X < 15) = p(5 - E[X] < X - E[X] < 15 - E[X])$$

Given that $E[X] = 10$,

$$p(5 < X < 15) = p(5 - 10 < X - 10 < 15 - 10)$$

According to the Chebyshev's inequality, we have:

$$p(|X - 10| \geq 5) \leq V[X]/5^2$$

With E[X] = 10 and V[X] = (Standard deviation)^2 = $(2)^2$ = 4, we have

$$p(|X - 10| \geq 5) \leq V[X]/5^2 = 4/5^2 = 4/25$$

Therefore, the lower bound of it can be calculated as;

$$p(|X - 10| < 5) = 1 - p(|X - 10| \geq 5) \geq 1 - 4/25 = 21/25 = 0.84$$

Question 1(b)

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

Solution:

I. Let the number of heads in N flips to be a random variable, which has Binomial distribution.

$$p(X = k) = \binom{N}{k} p^k (1-p)^{(N-k)}$$

Where k denotes the number of heads, and p denotes the probability of the coin lands on head in one flip, which is given by $p < 1/6$. Thus we have,

$$E[X] = Np$$

$$V[X] = Np(1-p)$$

By Markov's inequality;

$$p(X \geq N/3) \leq E[X]/(N/3) = Np/(N/3) = 3p < 3/6 = 1/2.$$

Therefore, the upper bound of the probability of observing $N/3$ or more heads is 1/2.

II. By Chebyshev's inequality,

$$p(X \geq N/3) = p(|X - E[X]| \geq N/3 - E[X]) \leq V[X]/(N/3 - E[X])^2$$

But,

$$E[X] = Np < N/6,$$

and

$$V[X] = Np(1-p) < 5N/36$$

Therefore,

$$p(X \geq N/3) = p(|X - N/6| \geq N/3 - N/6) < (5N/36)/(N/3 - N/6)^2$$

$$p(X \geq N/3) = p(|X - N/6| \geq N/6) < (5N/36)/(N^2/36) = 5/N$$

Thus, the required upper bound on the probability of observing $N/3$ or more heads using Chebyshev's inequality is given by $5/N$.

III. Let $p = 1/9$. Therefore,

$$E[X] = Np = N/9,$$

and

$$V[X] = Np(1-p) = 8N/81$$

By Markov's inequality

$$p(X \geq N/3) \leq E[X]/(N/3) = Np/(N/3) = 3p = 1/3.$$

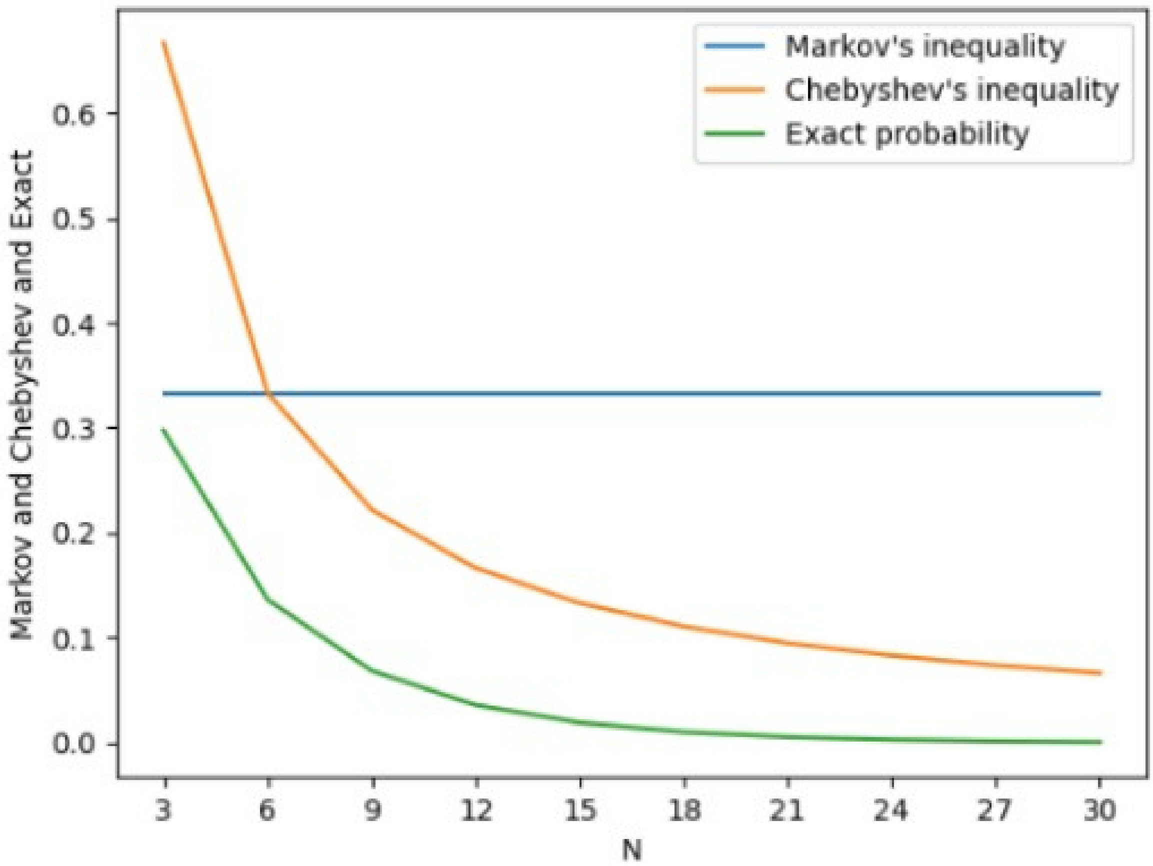
And by Chebyshev's inequality,

$$p(X \geq N/3) = p(|X - N/9| \geq N/3 - N/9) = (8N/81)/(N/3 - N/9)^2 = 2/N$$

Also, the exact probability of observing $N/3$ or more heads will be,

$$p(X \geq N/3) = \sum_{k=N/3}^{N} \binom{N}{k} 1/9^k (1 - 1/9)^{(N-k)}$$

$$=> p(X \geq N/3) = \sum_{k=N/3}^{N} \binom{N}{k} (1/9)^k (8/9)^{(N-k)}$$



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:

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

I. 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**]

II. 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**]

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

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

Solution:

I. Yes, it is possible.
According to chain rule of mutual information,

$$I(X;Y,Z) = I(X;Y) + I(X;Z/Y) = I(X;Z) + I(X;Y/z)$$

As $X \to Y \to Z$ forms a Markov Chain, hence $I(X;Z/Y) = 0$.

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

Here equality sign will hold $(i.e, I(X;Y) = I(X;Z))$ when $I(X;Y/Z)$. That is, when X and Y are also conditionally independent given Z. A simple example of this case is when data Y and data Z contains the same information of state X (i.e, Y = Z).

II. Yes, it is possible.
According to Chain rule of Mutual information,

$$I(X;Y,Z) = I(X;Y) + I(X;Z/Y) = I(X;Z) + I(X;Y/Z)$$

$$=> I(X;Y) - I(X;Z) = I(X;Y/Z) - I(X;Z/Y) \geq 0$$

Thus, $I(X;Y) \geq I(X;Z)$ provided $I(X;Y/Z) \geq I(X;Z/Y)$. An example could be when data Y gives more information about the state X than data Z does.

III.

- According to Chain rule of Mutual information,

$$I(X;Y,Z) = I(X;Y) + I(X;Z/Y) = I(X;Z) + I(X;Y/Z)$$

  As $X \rightarrow Y \rightarrow Z$ forms a Markov Chain, hence $I(X;Z/Y) = 0$.

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

  Clearly $I(X;Z) \leq I(X;Y)$, because $I(X;Y/Z) \geq 0$, as Information never hurts (i.e, Information can't be negative).

- Similarally, $I(X;Y/Z) \leq I(X;Y)$, because $I(X;Z) \geq 0$, as Information never hurts (i.e, Information can't be negative).

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]**

Solution:

1. From chain rule for mutual information,

   $$I(X; (Y, Z), T) = I(X; (Y, Z)) + I(X; T|(Y, Z))$$

   $$= I(X; T) + I(X; (Y, Z)|T)$$

   Due to Markov, X and T are independent given (Y,Z), indicating that $I(X; T|(Y, Z)) = 0$. On the other hand, $I(X; (Y, Z)|T) \geq 0$, as mutual information is always non-negative. Therefore,

   $$I(X; Y, Z) \geq I(X; T)$$

2. The equality holds when $I(X; (Y, Z)|T) = 0$, which means X and (Y,Z) are independent given T. This implies that $X \to T \to (Y, Z)$ is also a Markov chain.

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**]

Solution:

1. Let $X \in \{0, 1\}$ be a deterministic random variable with probabilities $p(x = 0) = 1$ and $p(x = 1) = 0$. Then, trivially for every $a > 0$,

$$P(X \geq a) = \frac{E[X]}{a} = 0$$

2. Let $Y \in \{-1, +1\}$ be a deterministic random variable with probabilities $p(y = -1) = \frac{1}{2}$ and $p(y = +1) = \frac{1}{2}$. Then, $E[Y] = 0$ and for any $a > 0$,

$$P(Y \geq a) \geq \frac{\mathbb{E}[Y]}{a} = 0$$

3. Yes, this is possible. We need,

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

As $V[Z] = E[(Z - E[Z])^2] \geq 0$, this quantity is non-negative. So, if we pick a random variable with the ratio not equal to 0, and for which $E[Z] = 0$, we are done. For example, let Z is uniform over $\{$-9, 3, 6$\}$. It can be easily verified that for $a < 7$, the Markov bound will be tighter.

Question 3: AEP [25 marks total]

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

Let X be an ensemble with outcomes $\mathcal{A}_X = \{\text{h, t}\}$ 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 $\mathcal{S}_\delta$ be the smallest set of N-outcome sequences with $P(\mathbf{x} \in \mathcal{S}_\delta) \geq 1-\delta$ 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**]

Solution:

a) Given $\mathcal{A}_X = \{\text{h, t}\}$ with $p_h = 0.8$ and $p_t = 0.2$.Therefore,

$$H(x) = \sum_x -p(x)\log_2 p(x) = 0.8\log_2\frac{1}{0.8} + 0.2\log_2\frac{1}{0.2} = 0.8*(0.3219) + 0.2*(2.3219) = 0.7219$$

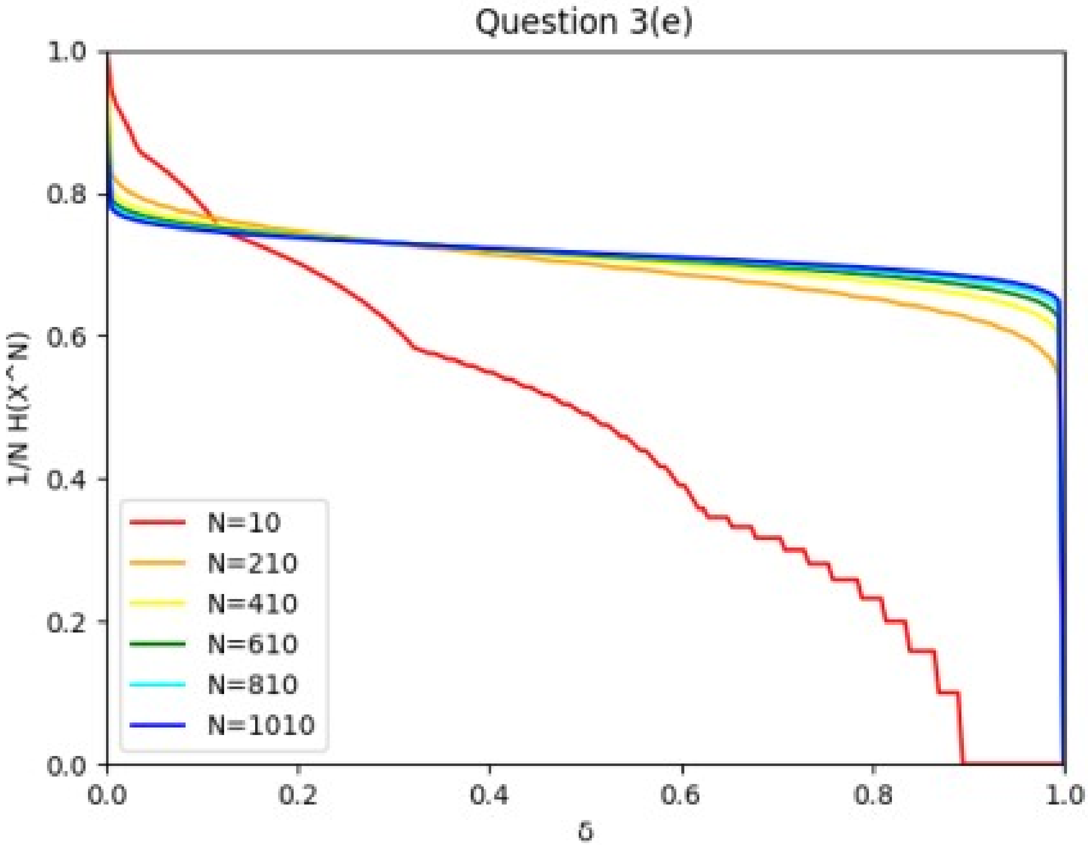
b) The size of A_{X^N} of the extended ensemble X^N will be 2^N. This is because there are two possible outcomes $\{\text{H, T}\}$. But the coin is flipped N-times, therefore the outcome set-size will be 2^N.

c) Raw-bit content : $H_o(X^4) = \log_2|A_{X^4}| = \log_2 16 = 4\ bits.$

d) Since, $X^N = \{X_1, X_2, ..., X_N\}$ is a string of N independent identically distributed (i.i.d) Random Variables from a single ensemble X and entropy is additive for independent variables. So,

$$H(X^N) = N * H(X) = 0.7219N\ bits.$$

e) From the plot below, we can observe when N increases, the curve flattens to a constant value approaching the entropy of a single coin flip that is $H(x) = 0.7219$.



This is because for significantly large N, most of the sequences will be considered typical, and hence will occupy large portion of the probability mass resulting into roughly equal probabilities.

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 $\delta-$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]**

Solution:

a. Since, X is an ensemble with alphabet $A_x = a, b$ with probabilities $(\frac{2}{5}, \frac{3}{5})$. Therefore,

$$H(X) = \sum_x p(x)\log_2(\frac{1}{p(x)}) = \frac{2}{5}\log_2(\frac{5}{2}) + \frac{3}{5}\log_2(\frac{5}{3})$$

$$= 0.52877 + 0.44219 = 0.970966\ bits.$$

b. As X^N denotes an extended ensemble. So, the alphabet of the extended ensemble X^3 will be,

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

c. The examples of three sequences in the typical set (for N = 3, $T_{N\beta} = 0.2$).

$$p(aaa) = (\frac{2}{5})^3 = \frac{8}{125} = 0.064$$

$$p(aab) = p(aba) = p(baa) = (\frac{2}{5})(\frac{2}{5})(\frac{3}{5}) = \frac{12}{125} = 0.096$$

$$p(abb) = p(bab) = p(bba) = (\frac{3}{5})(\frac{3}{5})(\frac{2}{5}) = \frac{18}{125} = 0.144$$

$$p(bbb) = \frac{3}{5}^3 = \frac{27}{125} = 0.216$$

Since,

$$T_{N\beta} := \{X : |-\frac{1}{N}\log_2 p(X) - H(X)| < \beta\}$$

We have,

$$H(X) - 0.2 < (-\frac{1}{3})\ \log_2 P(x) < H(x) + 0.2$$

$$0.77096 < (-\frac{1}{3})\ \log_2 P(x) < 1.17096$$

$$2.31269 < -\log_2 P(x) < 3.51288$$

$$2^{-3.51288} < P(x) < 2^{-2.31269}$$

$$0.08760 < P(x) < 0.20128$$

Clearly, P(abb) = P(bab) = P(bba) = 0.144 is in the specified range. Hence, the required three sequence will be {abb, bab, bba}. Also, p(aab) = p(aba) = p(baa) = 0.096 is in the specified range. So, {aab, aba, baa} can be another example of three sequences in the typical set.

d. When $\delta = \frac{1}{25}$, we have

$$1 - \delta = 1 - (\frac{1}{25}) = 0.96$$

Therefore,

$$S_\delta < 1 - \delta = 0.96$$

Thus, the smallest $\delta-$sufficient subset of X^3 when $\delta = 1/25$ will be,

$$S_\delta = \{aab, aba, abb, baa, bab, bba, bbb\}$$

Because,

$$= (\frac{2}{5})^2(\frac{3}{5}) + (\frac{2}{5})^2(\frac{3}{5}) + (\frac{3}{5})^2(\frac{2}{5}) + (\frac{2}{5})^2(\frac{3}{5}) + (\frac{3}{5})^2(\frac{2}{5}) + (\frac{3}{5})^2(\frac{2}{5}) + (\frac{3}{5})^3$$

$$= (\frac{12}{125}) + (\frac{12}{125}) + (\frac{18}{125}) + (\frac{12}{125}) + (\frac{18}{125}) + (\frac{18}{125}) + (\frac{27}{125}) = \frac{117}{125} = 0.936$$

Also, When $\delta = \frac{1}{10}$, we have

$$1 - \delta = 1 - (\frac{1}{10}) = 0.9$$

Therefore,

$$S_\delta < 1 - \delta = 0.9$$

Thus, the smallest $\delta-$sufficient subset of X^3 when $\delta = 1/10$ will be,

$$S_\delta = \{aab, aba, abb, baa, bab, bba\}$$

Because,

$$= (\frac{2}{5})^2(\frac{3}{5}) + (\frac{2}{5})^2(\frac{3}{5}) + (\frac{3}{5})^2(\frac{2}{5}) + (\frac{2}{5})^2(\frac{3}{5}) + (\frac{3}{5})^2(\frac{2}{5}) + (\frac{3}{5})^2(\frac{2}{5})$$

$$= (\frac{12}{125}) + (\frac{12}{125}) + (\frac{18}{125}) + (\frac{12}{125}) + (\frac{18}{125}) + (\frac{18}{125})) = \frac{90}{125} = 0.72$$

e. Since, $N = 1000$, and $\beta = 0.2$

$$T_{N\beta} := \{X : |-\frac{1}{N}\log_2 p(X) - H(X)| < \beta\}$$

We have,

$$H(X) - \beta \leq (-\frac{1}{1000})\ \log_2 P(x) \leq H(x) + \beta$$

$$H(X) - 0.2 \leq (-\frac{1}{1000})\ \log_2 P(x) \leq H(x) + 0.2$$

$$0.97096 - 0.2 \leq (-\frac{1}{1000})\ \log_2 P(x) \leq 0.97096 + 0.2$$

$$0.77096 \leq (-\frac{1}{1000})\ \log_2 P(x) \leq 1.17096$$

$$2^{-1170.96} \leq P(x) \leq 2^{-770.96}$$

Let 'a' appears k-times and 'b' appears 1001 - K times in a sequence. Hence, we have;

$$p(k) = (\frac{2}{5})^k (\frac{3}{5})^{1000-k} = \frac{2^{1000} * 3^{1000-k}}{5^{1000}}$$

Therefore,

$$2^{-1170.96} \leq \frac{2^{1000} * 3^{1000-k}}{5^{1000}} \leq 2^{-770.96}$$

$$=> 1000 - \log_2(\frac{5^{1000}}{2^{770.96} * 2^{1000}}) \leq k \leq 1000 - \log_2(\frac{5^{1000}}{2^{1170.96} * 2^{1000}})$$

$$=> 1000 - \log_2(\frac{2.5^{1000}}{2^{770.96}}) \leq k \leq 1000 - \log_2(\frac{2.5^{1000}}{2^{1170.96}})$$

$$=> 449 \leq k \leq 849$$

Thus, the required number of sequences that are typical set will be,

$$\sum_{i=449}^{849} \binom{1000}{i} = \sum_{i=0}^{849} \binom{1000}{i} - \sum_{i=0}^{449} \binom{1000}{i}$$

The total number of sequences will be 2^{1000}, and the fraction of sequences in X^N the typical set will be,

$$\frac{\sum_{i=0}^{849} \binom{1000}{i} - \sum_{i=0}^{449} \binom{1000}{i}}{2^{1000}} = 0.726$$

f. According to Asymptotic Equipartition Property (AEP), the approximate probability that it is in the N $\beta-$typical set will be 1, as

$$(\forall \beta > 0) \lim_{N \to \infty} P(| - \frac{1}{N} log_2 P(X_1, X_2, ..., X_N) - H(X)| < \beta) = 1$$

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 $\mathcal{A}$ – that is, its alphabet and probabilities. [**2 Marks**]
   (b) What is the raw bit content of $\mathcal{A}^4$? [**2 Marks**]
   (c) What is the smallest value of δ such that the smallest δ-sufficient subset of $\mathcal{A}^4$ contains fewer than 256 elements? [**2 Marks**]
   (d) What is the largest value of δ such that the essential bit content $H_\delta(\mathcal{A}^4)$ is strictly greater than zero? [**2 Marks**]
4. Suppose the album titles ensemble $\mathcal{A}$ is as in part (3).
   (a) Compute an approximate value for the entropy $H(\mathcal{A})$ 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 $\mathcal{A}$ 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**]

Solution:

Lets denote the albums by their first letters and let $\mathcal{A} = \{A, B, C, D\}$.

a) Bits required to uniformly code:

   i) Two bits ($= \lceil log_2 4 \rceil$). Example uniform code: $C = \{00, 01, 10, 11\}$.

   ii) The album Alina has 7 tracks, so the number of bits required to code will be 3 (as, $\lceil log_2 7 \rceil = 3$). Example uniform code : $C = \{000, 001, 010, 100, 101, 110, 011\}$

   iii) There are 48 tracks in total. So, $\lceil log_2 48 \rceil = 6$ bits are required.

b) The raw bit content required to distinguish all the tracks in the collection will be $log_2 48 \approx 5.5849$ bits.

c) Let $\mathcal{A}$ denote ensemble of album chosen when tracks across all albums are picked uniformly at random.

i) $\mathcal{A} = \{A, B, C, D\}$ and $p = \{\frac{7}{48}, \frac{12}{48}, \frac{15}{48}, \frac{14}{48}\}$

ii) Raw bit content of $\mathcal{A}^4 = log_2|\mathcal{A}|^4 = log_2 4^4 = 8$ bits.

iii) Since, $\mathcal{A}^4$ has < 256 elements. We need a δ equal to the smallest probability for element of $\mathcal{A}^4$. As album alina (A) has smallest probability, so too will $AAAA \in \mathcal{A}^4$. Thus, choose $\delta = P(AAAA) = (\frac{7}{48})^4 \approx 0.000452$.

iv) $H_\delta(\mathcal{A}^4)$ will be zero when S_δ contains only one element. So, if δ is a set such that there are two elements in S_δ, then we are done. Since, cecilia has the highest probability. Hence, the sequence $CCCC \in \mathcal{A}^4$ will have the highest probability $P(CCCC) = (\frac{15}{48})^4 \approx 0.009536$.
The next largest probability in any sequence with three C_s and one D in it. The required probability will be $P(CCCD) = (\frac{15}{48})^3 * (\frac{14}{48}) \approx 0.008900$

d) i)

$$H(\mathcal{A}) = \sum_{\mathcal{A}} p(\mathcal{A}) \log_2(\frac{1}{p(\mathcal{A})})$$

$$= (\frac{7}{48}) * (log_2(\frac{48}{7})) + (\frac{12}{48}) * (log_2(\frac{48}{12})) + (\frac{15}{48}) * (log_2(\frac{48}{15})) + (\frac{14}{48}) * (log_2(\frac{48}{14}))$$

$$= (\frac{7}{48}) * (2.77760) + (\frac{1}{4}) * (2) + (\frac{15}{48}) * (1.6780) + (\frac{14}{48}) * (1.77760) \approx 1.94790 \text{ bits.}$$

ii) The probability of sequences of length N = 100 in $T_{N\beta}$ for $\beta = 0.1$ is no more than $2^{-100(H(\mathcal{A})-0.1)} \approx 0.235866 * 10^{-55}$, and no less than $2^{-100(H(\mathcal{A})+0.1)} \approx 2.2494 * 10^{-62}$. This means,

$$\frac{1}{0.235866 * 10^{-55}} \approx 4.2396 * 10^{55} \leq |T_{N\beta}| \leq \frac{1}{2.2494 * 10^{-62}} \approx 0.44456 * 10^{62}.$$

iii) No, because the rate is below the entropy of 1.94790 bits. So, by the source coding theorem, for large blocks a rate of 1.5 bits per title (with 95% reliability) is not possible.