The Competition on the Mathematics of Information 2024

The Chinese University of Hong Kong

24 February, 2024

Name: ______________________________

Contestant Number: ______________________________

School Name: ______________________________

Question	Points
1	70
2	70
3	70
4	70
Total	280

Instructions:

1. This examination booklet contains 5 pages, including this page. Please write your solutions on the answer sheets given to you. You should write the solution to each problem **only** on its designated answer sheet. If more answer sheets are required, please ask the invigilators. You can use this booklet as scratch paper. Please ask if you require more scratch paper.
2. You have **three (3) hours** to complete the examination. As a courtesy to your classmates, we ask that you not leave during the last fifteen minutes.
3. Calculators are **not** allowed in the exam.
4. Please clip your answer papers with the paper clip provided as well as the cover page and return to us at the end of the examination.
5. Please sign the below Honor Code statement.

In recognition of and in the spirit of the Honor Code, I certify that I will neither give nor receive unpermitted aid on this examination.

Signature: ______________________________

Question 1: Quantization

In image compression, a common technique is quantization, where the value of a pixel is converted into another similar value that has a smaller set of possibilities. For simplicity, consider a grayscale image, where each pixel has a value that is an integer in the range $0, 1, \ldots, 99, 100$. To compress the value x_0 in the range $0, 1, \ldots, 100$ by a quantizer with step size 11 and initial level 0, we compress x_0 into the value x_1 among $0, 11, 22, \ldots, 99$ that is closest to x. For example, if $x_0 = 50$, it is compressed into $x_1 = 55$, which is the number divisible by 11 within the range $0, 1, \ldots, 100$ that is closest to x_0. This way, we can reduce the 101 different possible values of x_0 into only 10 different possible values of x_1, which requires a smaller space to store in a computer.

More generally, to compress x_0 by a quantizer with step size a (which must be a positive odd integer) and initial level b (which must be an integer in the range $0, 1, \ldots, a-1$), we list all numbers in the form $an+b$ (where n is an integer) that are within the range $0, 1, \ldots, 100$, and then choose the number in the list which is the closest to x_0, and output that number as x_1.

a) What is the smallest positive odd integer a so that there are only 2 different possible values of x_1? You may choose the initial level b. Justify your assertion.

b) Compressing an image multiple times may degrade the quality of the image. Suppose now we apply a quantizer on x_0 (in the range $0, 1, \ldots, 100$) to get x_1, and then apply another quantizer on x_1 to get x_2, and so on. After we quantize k times, we get x_k. Your goal is to design a sequence of quantizers that is as bad as possible. The worst image compression software would compress any input image into an image that contains only a single colour. Therefore, you want the number of different possible values of x_k to be only 1.
For example, consider the sequence of quantizers (where $k = 3$):

1. Step size 39, initial level 30,
2. Step size 35, initial level 18,
3. Step size 51, initial level 40.

We can check that regardless of the value of x_0 (in the range $0, 1, \ldots, 100$), we must have $x_3 = 40$, and hence the goal is satisfied.

i) Consider the case where each of these quantizers must have the same step size a (you can choose a), and each of these quantizers may have a different initial level that you can choose. Fix $k = 2$ (you can quantize twice). What is the smallest positive odd integer a such that you can design a sequence of quantizers with only one possible value of x_k? Justify your assertion.

ii) Suppose now each quantizer must have a fixed step size $a = 11$, and you may only choose the initial levels of the quantizers (you may choose a different initial level each time). We allow k to be any positive integer. What is the smallest k such that you can design a sequence of quantizers with only one possible value of x_k? Justify your assertion.

iii) Suppose now each quantizer must have a fixed initial level $b = 0$. You first choose a positive odd integer c (the "maximum step size"). Then you can design a sequence of quantizers, where each quantizer has an initial level 0, and a step size at most c (you may choose a different step size each time as long as it is less than or equal to c). We allow k to be any positive integer. What is the smallest value of c such that you can design a sequence of quantizers with only one possible value of x_k? Justify your assertion.

Question 2: Unbiased Coin

a) Imagine you possess a biased coin with a 1/3 probability of landing on heads and a 2/3 probability of landing on tails. You wish to participate in a game that requires an unbiased fair coin, but the only coin you have is the aforementioned biased one. Suppose you can throw the biased coin only two times. You would like to *simulate* a fair coin, but this may not be possible. What is the "fairest" coin you can simulate? In other words, after throwing the biased coin twice, you need to "declare" H or T, and you would like the probability of H and T to be as close as possible to each other. What strategy can you employ to achieve this?

b) Repeat the previous part if you are allowed to throw the biased coin only three times.

c) Now, you have the freedom to toss the coin as many times as you desire. Can you devise a strategy that enables you to simulate a fair (unbiased) coin using the biased coin? Here's a hint: throw the biased coin twice. If both throws yield the same result (either two heads or two tails), disregard those tosses and repeat the procedure.

d) Imagine a scenario where you have two coins: one is fair and unbiased, while the other is biased. You don't have any prior knowledge about which coin is which, except that the biased coin has a 1/3 probability of landing on heads, while the unbiased coin has a 1/2 probability. Your goal is to determine which coin is the unbiased one. To make your guess, you randomly select one of the coins, toss it, and observe the result. Based on this outcome, what would be your most reasonable guess of which coin is the unbiased one?

e) If you have the option to toss both coins (each once) in order to solve the previous question, what is the best guess (based on the outcomes) to maximize the probability of correctly identifying the unbiased coin?

Question 3: List Decoding

You have a ternary alphabet, say $\{0, 1, 2\}$. Consider

$$\{0,1,2\}^n = \{(x_1, x_2, \ldots, x_n) : x_i \in \{0,1,2\}, i = 1, 2, ..., n\}$$

be the set of sequences of length n from the ternary alphabet.

Adam has a collection of sequences $\mathcal{C}_n \subseteq \{0,1,2\}^n$ and wants to send a sequence $(x_1, x_2, \ldots, x_n)$ from $\mathcal{C}_n$ to his friend Eve. (Note that if a sequence is in $\mathcal{C}_n$, Adam can be asked to send that sequence; so $\mathcal{C}_n$ is not allowed to have unused sequences.)

However, during transmission, there is always noise, and if Adam sends i, for $i \in \{0, 1, 2\}$, at a certain position, Eve will receive one of the other two symbols at that position with probability $\frac{1}{2}$. (For example, if 1 was sent at some position, then Eve will receive either 0 or 2, each with probability $\frac{1}{2}$.) Assume that the noise at each position is independent of the noise at the other positions.

Below is an example: suppose $n = 2$, and Adam sends the sequence $(x_1, x_2) = (0, 1)$ to Eve. Then, Eve receives one of the following four sequences $(1, 0), (2, 0), (1, 2), (2, 2)$ with equal probability $1/4$. Similarly, if Adam sends the sequence $(x_1, x_2) = (0, 0)$ to Eve, she will receive one of the following four sequences $(1, 1), (2, 1), (1, 2), (2, 2)$ with an equal probability of $1/4$. Now, assume that $\mathcal{C}_2 = \{(0, 0), (0, 1)\}$ contains two possible sequences that Adam may send. Observe that if Eve receives $(2, 2)$, she will not be able to determine which sequence from $\mathcal{C}_2$ was transmitted. In this case, we say that an error occurs in the communication.

a) Eve wants to figure out exactly (with zero chance of error) which sequence was sent by Adam based on the sequence she receives. In this case, how large can the size of $\mathcal{C}_n$ be?

b) Disappointed by the above part, Eve is slightly less ambitious, and as long as she can narrow down the sequence sent by Adam to two possibilities, she is happy. **Prove** that in this case if $\mathcal{C}_n$ has the following property, then Eve can successfully complete her mission: For any three different sequences $\mathbf{a} = (a_1, \ldots, a_n), \mathbf{b} = (b_1, \ldots, b_n)$, and $\mathbf{c} = (c_1, \ldots, c_n)$ in $\mathcal{C}_n$, there is some location $i, 1 \leq i \leq n$ such that (a_i, b_i, c_i) is a permutation of $(0, 1, 2)$ (that is, all the three values are distinct). We shall call such a collection of sequences, $\mathcal{C}_n$ to be *trifferent*.

c) Prove the converse to the above statement, i.e. if Eve can successfully complete her mission, then the collection of sequences $\mathcal{C}_n$ is trifferent.

d) Let $T(n)$, $n \geq 1$. be defined as the largest value of the size of $\mathcal{C}_n$, where $\mathcal{C}_n$ is a set of sequences of length n that is trifferent. Then show the following:

- $T(1) = 3$
- $T(2) = 4$

e) Show that, for any $n \geq 2$,

$$T(n) \leq \frac{3}{2} T(n-1).$$

f) Determine the value of $T(4)$.

Remark: This question is related to perfect hashing and the best bound known today comes from the concatenation of the elements of $T(4)$. One can show that asymptotically $T(n)$ grows at least as fast as $\left(\frac{9}{5}\right)^{\frac{n}{4}}$. The key point is that once Eve allows for even an uncertainty of two symbols, the communication rate drastically increases.

Question 4: Periodic functions

A function $f(x)$ is called periodic if there is a value $T > 0$, such that $f(x+T) = f(x)$ for every x. Any positive value T such that $f(x+T) = f(x)$ for every x, is called a *period* of the function.

a) Show that if 33 and 27 are periods of $f(x)$, then 3 is also a period of $f(x)$.

b) Show that if $\sqrt{2}$ and 1 are periods of a continuous function $f(x)$, then $f(x)$ must be the constant function.

c) Construct a non-constant function (that need not be continuous) such that $\sqrt{2}$ and 1 are both its periods.

d) Prove that $f(x) = \sin(x) + \sin(\sqrt{2}\,x)$ is not periodic.