# The Competition on the Mathematics of Information: 2023

The Chinese University of Hong Kong

March 4, 2023

Name: ______________________________

Contestant Number: ______________________________

School Name: ______________________________

| **Question** | **Points** |
|---|---|
| **1** | 70 |
| **2** | 70 |
| **3** | 70 |
| **4** | 70 |
| **Total** | 280 |

**Instructions:**

1. This examination booklet contains 9 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.

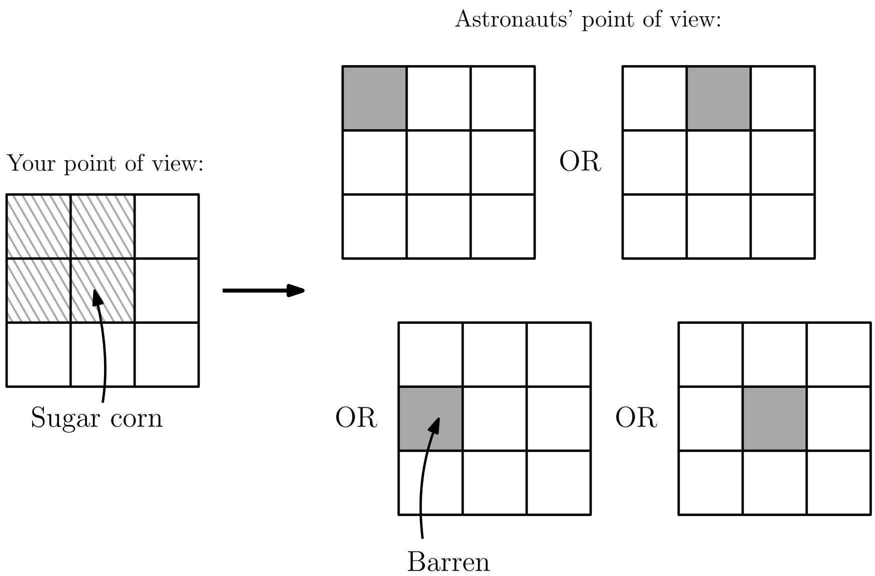
In recognition of and in the spirit of the Honor Code, I certify that I will neither give nor receive unpermitted aid on this examination.

Signature: ______________________________

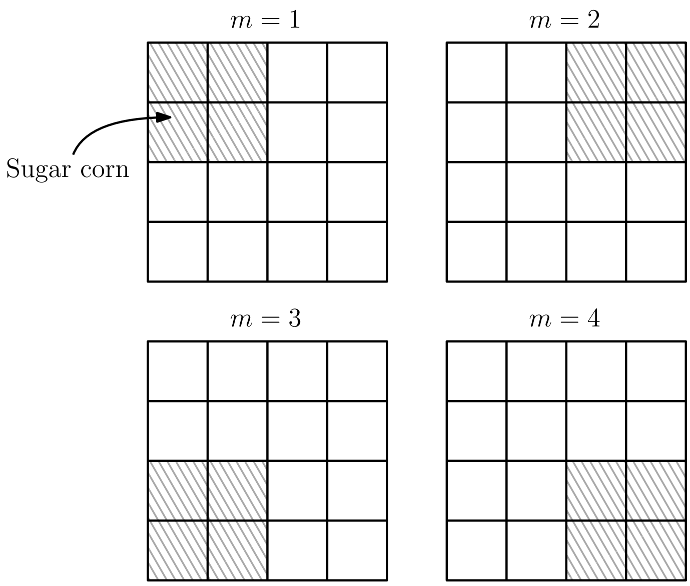
## Question 1: Primitive Space Communication

You are working at the ground control center of a space station. You must send a message $m$ which is an integer in the range $1, \ldots, k$ to the space station. Unfortunately, your satellite dishes are ruined by a lightning strike, so you are resorting to a primitive method to communicate with the astronauts.

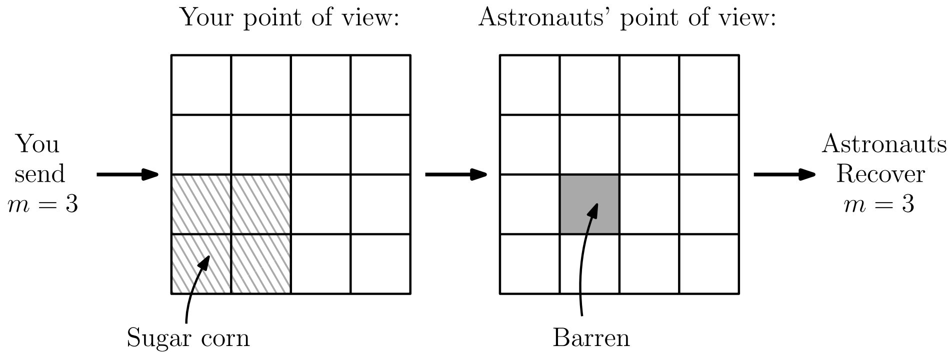
There is an $n\,\text{km} \times n\,\text{km}$ square field next to the ground control center, ripe of golden wheat. The field is divided into $n^2$ square patches, each measures $1\,\text{km} \times 1\,\text{km}$. You plan to spread sugar corn over a $2\,\text{km} \times 2\,\text{km}$ square covering 4 patches (you are not allowed to choose a square that only intersects a patch partially; the square you choose must completely contain 4 patches). The sugar corn will attract sparrows, which will flock to one of the 4 patches chosen at random, and eat all the sugar corn and wheat there, turning that patch completely barren (i.e., without any plants). The astronauts can then observe which one of the $n^2$ patches is barren, and try to recover the message you are attempting to convey. Assume you and the astronauts are allowed to agree on a communication protocol before the astronauts departs to space. Note that the astronauts do not observe which 4 patches you have chosen. They only observe the location of the barren patch.




For example, for $n = 4, k = 4$, a simple strategy is to select the top left $2 \times 2$ square if the message $m = 1$, the top right $2 \times 2$ square if $m = 2$, the bottom left $2 \times 2$ square if $m = 3$, and the bottom right $2 \times 2$ square if $m = 4$, as illustrated below. Both you and the astronauts know this strategy.



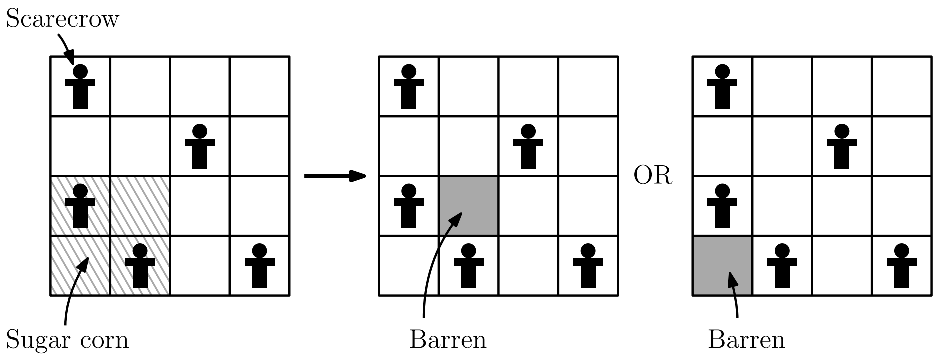
After you select the $2 \times 2$ square to spread the sugar corn, the astronauts observe the location of the barren patch (which is a random patch among the 4 selected patches), and deduce the message $m$ based on whether the barren patch lies in the top left (which implies $m = 1$), top right ($m = 2$), bottom left ($m = 3$) or bottom right ($m = 4$) $2 \times 2$ square. The figure below shows the case $m = 3$, where you spread sugar corn over the bottom left $2 \times 2$ square, which attracts sparrows to one of those 4 patches in the $2 \times 2$ square, and make that patch barren. The astronauts can then know that $m = 3$ since the barren patch is in the bottom left $2 \times 2$ square.



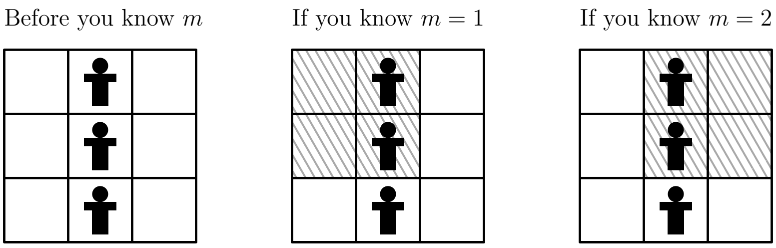

a) **(10 points)** Suppose $n = 5$. Find the largest $k$ such that there is a communication strategy that guarantees that the astronauts can recover the message. Justify your answer.

b) Before the astronauts departs to space, you decide to place at most $s$ scarecrows in the field to deter the sparrows. The sparrows will always choose a patch without scarecrow, i.e., if some of the 4 patches with sugar corn have no scarecrows, the sparrows will choose one of those scarecrow-less patches at random (if exactly one of the 4 patches has no scarecrows, the sparrows will definitely choose that one patch). You are not allowed to select a $2 \times 2$ square to spread sugar corn if all 4 patches there contain scarecrows. As a part of the communication strategy, you are allowed to choose where to place the scarecrows before the astronauts departs to space, and the astronauts know the positions of the scarecrows too. You cannot move the scarecrows after you learn about the message.

   For example, in the following figure, if you spread the sugar corn in the bottom $2 \times 2$ square, which contains 2 scarecrows, then the barren patch will be one of the remaining 2 patches in the $2 \times 2$ square.




   An example of a communication strategy for $n = 3$, $s = 3$, $k = 2$ is given below. Before the communication begins, you place the 3 scarecrows on the centre column. You then observe the message $m$ to be sent. If $m = 1$, you select the top left $2 \times 2$ square. If $m = 2$, you select the top right $2 \times 2$ square. The astronauts can recover $m$ by observing whether the barren patch is in the left column or the right column.




i) **(15 points)** Suppose $n = 5$, $s = 1$ (i.e., you can place at most 1 scarecrow). Find the largest $k$ such that there is a communication strategy that guarantees that the astronauts can recover the message. Justify your answer.

ii) **(15 points)** Suppose $n = 5$, $s = 2$. Find the largest $k$ such that there is a communication strategy that guarantees that the astronauts can recover the message. Justify your answer.

iii) **(30 points)** Now you are allowed to choose both $n$ and the communication strategy. Suppose you have an unlimited number of scarecrows. Define the *efficiency* of the communication strategy to be $r = k/n^2$ (i.e., the number of messages per km$^2$). Design a communication strategy that achieves an efficiency as large as you can. Your score depends on how large the efficiency of your communication strategy is.

## Question 2: Quality control

You work at a cake factory. You just baked $N$ boxes of cakes, numbered $1, 2, \ldots, N$. Right before you are going to ship out the cakes, you receive a shocking news that the farmer forgot to discard a bag full of rotten strawberries. Now, those bad strawberries have gone into exactly one of those $N$ boxes of cakes (all cakes in that one box are bad). You do not know which box is bad, and you want to find out before it upsets the stomach of some unfortunate soul.

a) **(15 points)** You can inspect all of the $N$ boxes one by one (sampling without replacement) until you find the bad one. If you inspect a box, you can tell with 100% accuracy whether it is bad. Assume that you use a random order to inspect the $N$ boxes. For example, if $N = 3$, the order in which the boxes are inspected can be $1, 2, 3$ (i.e., first inspect box 1, then 2, then 3), or $1, 3, 2$, or $2, 1, 3$, or $2, 3, 1$, or $3, 1, 2$, or $3, 2, 1$, and each of these 6 orders are equally likely to be chosen. Let $p_k$ be the probability that you find the bad box after exactly $k$ inspections. Compute $p_k$. Then compute the *average number of required inspections*:

$$\sum_{k=1}^{N} p_k \cdot k \;=\; p_1 + 2p_2 + \cdots + Np_N.$$

b) **(15 points)** Next, assume that you have some limited prior information about the boxes (that box smells terrible, so it is probably bad?). Using this information, you can assign a meaningful probability $q_j$ to each box $j \in \{1, \cdots, N\}$ of being bad. The bad box is no longer equally likely to be box $1, 2, \ldots, N$, but now the bad box has a probability $q_1$ to be box 1, a probability $q_2$ to be box 2, etc. You still want to inspect all boxes one by one, but instead of inspecting boxes in a completely random order, you can utilize the prior probability $q_j$. Find the best order to inspect the boxes so as to minimize the average number of inspections needed to find the bad box (assume that the probabilities satisfy $q_1 \geq q_2 \geq \cdots \geq q_N \geq 0$ and $q_1 + q_2 + ... + q_N = 1$). Remember that the average number of required inspections is equal to $\sum_{k=1}^{N} p_k \cdot k$ where $p_k$ is the probability that you find the bad box after exactly $k$ inspections.

c) Instead of doing the inspection yourself, you are now delegating the work to your coworker. In each hour, the coworker will randomly select one box to inspect, where box $i$ is selected with probability $r_i$, $0 < r_i < 1$ (where $r_1 + r_2 + ... + r_N = 1$). In the next hour, the coworker will randomly select one box again in the same manner, forgetting which boxes he/she has already inspected previously (your coworker is quite forgetful, and it is possible that he/she inspects the same box multiple times). We are interested in the average number of hours needed for the coworker to find the bad box.

We break up the calculation of this value in multiple steps:

i) **(10 points)** Every hour, the coworker uses the probabilities $r_1, r_2, ..., r_N$ to choose the box to be inspected. Let $a_m(k)$ be the probability that the box $k$ is not chosen in the first $m$ hours, and is selected for the first time in the $(m+1)$-th hour. Compute $a_m(k)$ in terms of $r_k$ and $m$.

ii) **(10 points)** The average number of hours needed to screen box $k$ is defined as

$$\begin{aligned}\mu(k) &= \sum_{m=0}^{\infty} a_m(k) \cdot (m+1) \\ &= a_0(k) + 2a_1(k) + 3a_2(k) + 4a_3(k) + \cdots .\end{aligned}$$

Prove that $\mu(k) = \dfrac{1}{r_k}$.

*Hint:* We have $\sum_{m=0}^{\infty}(m+1)x^m = \frac{1}{(1-x)^2}$ for any $x \in (-1, 1)$.

iii) **(20 points)** Assume that the probability that box $k$ is bad equals $q_k > 0$ (where $q_1 + q_2 + ... + q_N = 1$). Then, the probability that the coworker finds the bad box in the $(m+1)$-th hour is given by

$$p_m = \sum_{k=1}^{N} q_k a_m(k).$$

Therefore the average number of hours needed to find the bad box is given via the following formula

$$\mu = \sum_{m=0}^{\infty}(m+1)p_m = \sum_{m=0}^{\infty}\sum_{k=1}^{N}(m+1)q_k a_m(k) = \sum_{k=1}^{N} q_k \mu(k).$$

Using the results from the previous part, we can find a closed-form formula for $\mu$ in terms of $q_j$ and $r_j$ as follows. Since

$$\mu(k) = \frac{1}{r_k},$$

we obtain

$$\mu = \sum_{k=1}^{N} \frac{q_k}{r_k}.$$

Find the best choice of positive values $r_1, r_2, ..., r_N$ (that you need to assign) satisfying $r_1 + r_2 + \cdots + r_N = 1$, that minimizes $\mu$, and find the smallest possible $\mu$. Your answers should be in terms of $q_j$. In other words, solve the following optimization problem:

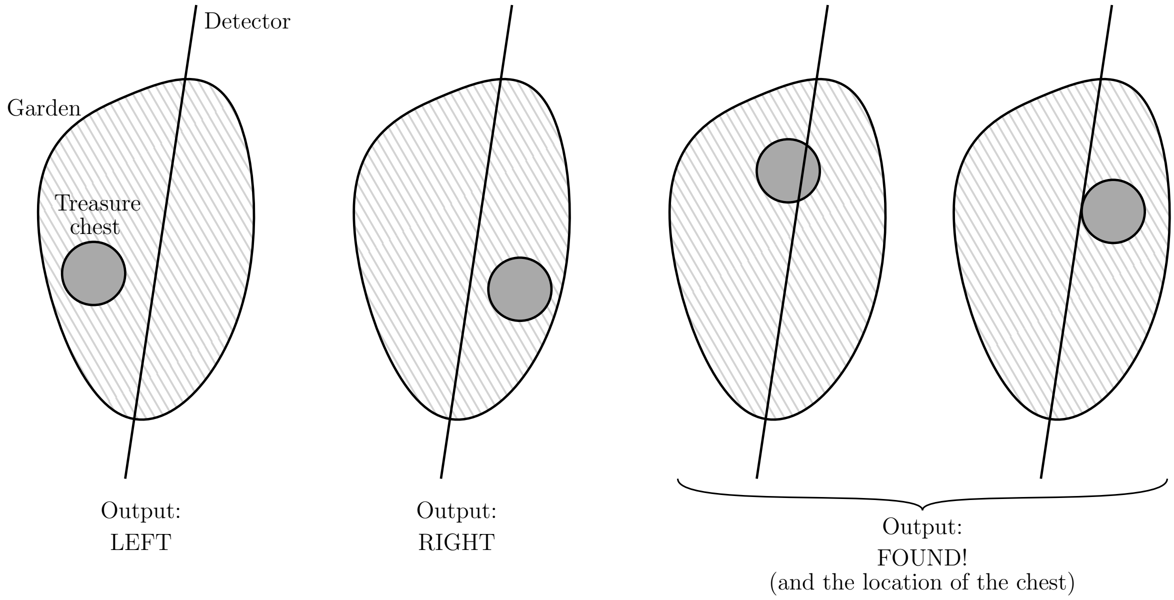
$$\min_{\substack{r_1, r_2, \cdots, r_N > 0: \\ r_1 + r_2 + \cdots + r_N = 1}} \sum_{k=1}^{N} \frac{q_k}{r_k}.$$

**Note**: A practical application of this idea is the following. For example, in an airport screening of two groups of people, if group $A$ is twice as likely as group $B$ to be terrorists, your answer should tell you what should be the optimal screening probability of groups A and B.

## Question 3: Hidden treasure

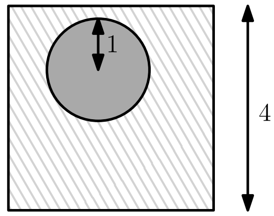
You are a treasure hunter looking for a treasure chest buried in a garden, which is a region on the 2D plane. The treasure chest has a circular shape with radius 1 metre, and it must lie completely within the garden (it is allowed to touch the boundary of the garden), though you do not know its location. You have a detector that allows you to perform the following action: You choose a line on the 2D plane, and the detector will either output the precise location of the chest if the chest intersects or touches that line, or it will output which side of the line the chest lies in if the chest does not intersect or touch that line (e.g. whether the chest is on the left or on the right of that line, or whether the chest is above or below that line if the line happens to be horizontal). Based on the output, you can then choose another line and use the detector again.

The following figure shows the four possible cases: the chest is on the left of the detection line, the chest is on the right, the chest intersects the line, and the chest touches the line. In the last two cases, the chest can be found.




You want to find the precise location of the chest after using the detector at most $k$ times. (Note that for each time, you can choose a different line depending on the previous outputs you have observed.) Find out the minimum $k$ that guarantees your success, and the detection strategy attaining the minimum $k$, for the cases in parts a, b, c, d:

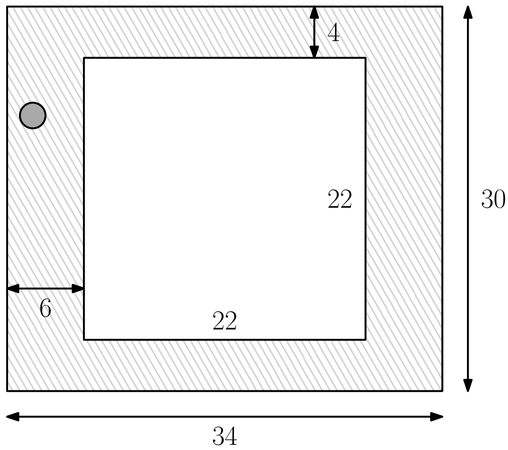
a) **(10 points)** The garden is a $4 \times 4$ square. (Hint: The answer is 1. Consider the centre of the circle. How do you ensure that you can detect the chest in one try, no matter where the chest is located in the garden?)




b) **(10 points)** The garden is a $8 \times 8$ square. (Hint: The answer is 2. Find a detection strategy which requires at most two uses of the detector, regardless of the position of the chest in the garden.)

c) **(15 points)** The garden is an equilateral triangle with side length 19. (Hint: $\sqrt{3} \approx 1.732$, and $19\sqrt{3}/2 \approx 16.454$.)

d) **(15 points)** The garden is a $34 \times 30$ rectangle, minus a $22 \times 22$ square placed at the centre of the $34 \times 30$ rectangle, with sides that are parallel to the sides of the $34 \times 30$ rectangle. The garden is a "hollow" rectangle.




e) **(20 points)** Is it true that for every valid garden with area 25 (the garden can be of any shape as long as it can contain at least one circle of radius 1), it is possible to find the chest using the detector at most 3 times? If yes, describe your detection strategy. (Your strategy may depend on the shape of the garden.)

## Question 4: Functional Equation

a) **(30 points)** Let $f_1 : \mathbb{R} \to \mathbb{R}$ and $f_2 : \mathbb{R} \to \mathbb{R}$ be two continuous functions that map real numbers to real numbers. Further assume that both functions are even, i.e. $f_1(x) = f_1(-x)$, and $f_2(x) = f_2(-x)$ for all $x \in \mathbb{R}$. Further assume that $f_1(0) = f_2(0) = 1$. Determine all possible functions $f_1, f_2$ that satisfy the above conditions as well as satisfying:

$$f_1(x+y)f_2(x-y) = f_1(x)f_1(y)f_2(x)f_2(y) \quad \text{for all } x, y \in \mathbb{R}.$$

b) **(40 points)** Let $f_1 : \mathbb{R} \to \mathbb{C}$ and $f_2 : \mathbb{R} \to \mathbb{C}$ be two continuous functions that map real numbers to complex numbers. Further assume that both the functions satisfy $f_1(-x) = \overline{f_1(x)}$, and $f_2(-x) = \overline{f_2(x)}$ for all $x \in \mathbb{R}$. Here $\overline{z}$ denotes the complex conjugate of $z$. Further assume that $f_1(0) = f_2(0) = 1$. Determine all possible functions $f_1, f_2$ that satisfy the above conditions as well as satisfying:

$$f_1(x+y)f_2(x-y) = f_1(x)f_1(y)f_2(x)f_2(-y) \quad \text{for all } x, y \in \mathbb{R}.$$

**Note 1**: A complex number $z \in \mathbb{C}$ is in the form $z = a + bi$, where $i$ is a constant with $i^2 = -1$. The conjugate of $z$ is given by $\overline{z} = a - bi$. A useful property is that any complex number $z \in \mathbb{C}$ can be expressed in the form $z = re^{i\theta}$, where $r = |z| = \sqrt{a^2 + b^2} \in [0, \infty)$ is the modulus of $z$, $\theta \in (-\pi, \pi]$ is called the argument of $z$, and $e^{i\theta} = \cos\theta + i \sin\theta$. However, one could also consider extensions of the range of theta if needed, as $e^{i\theta} = e^{i(\theta + 2\pi)}$. If $f(x) \neq 0$ for all $x \in \mathbb{R}$ and is a continuous complex valued function, then it can be expressed in the form $f(x) = r(x)e^{i\theta(x)}$ where we can assume that $r(x)$ is continuous and $\theta(x)$ is continuous.

**Note 2**: Sometimes functional equations can be used to characterize probability distributions. The ideas in this question have their origins in Fourier transforms (or characteristic functions). Fourier analysis, or harmonic analysis, is an area that is studied extensively by mathematicians, and also forms the basis of wireless communication systems.