The Competition on the Mathematics of Information: 2023

The Chinese University of Hong Kong

March 4, 2023

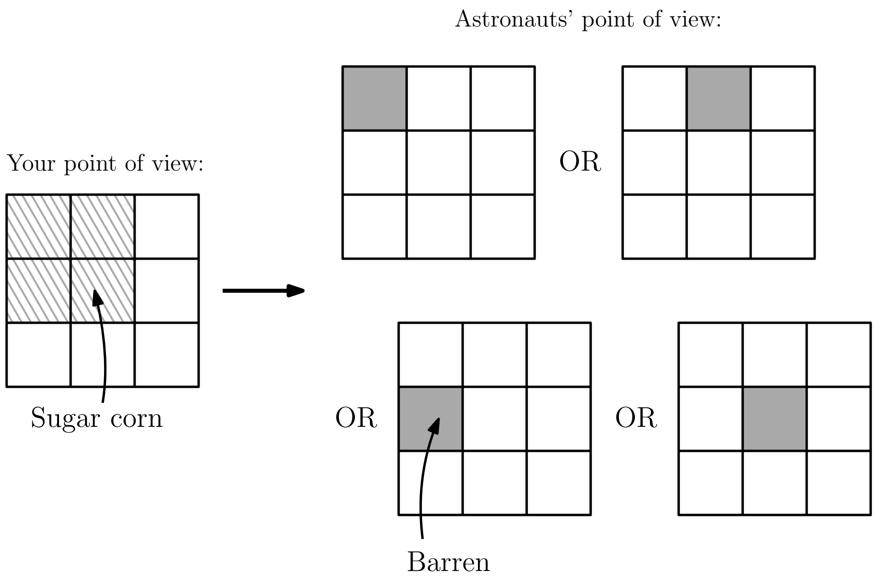
SOLUTIONS

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

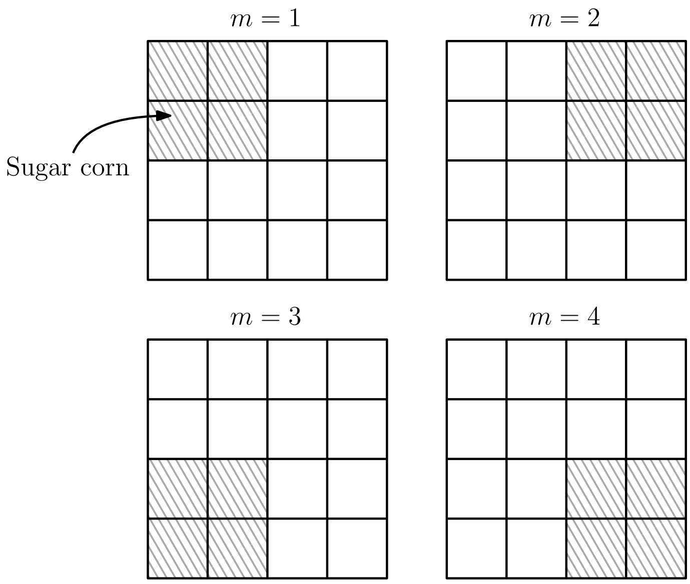
Question 1: Primitive Space Communication

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.

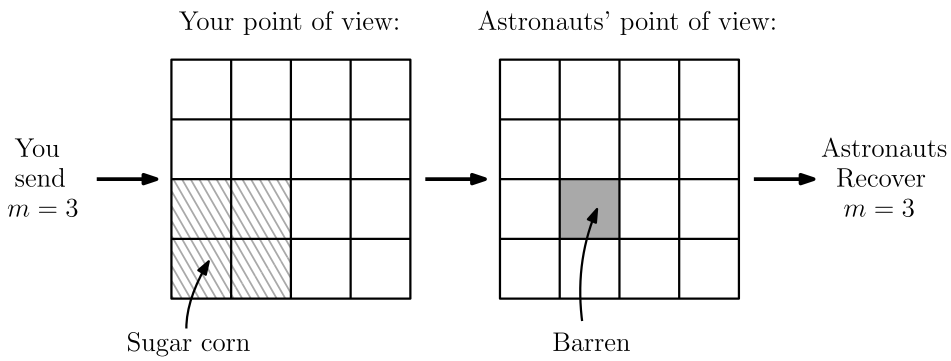
There is an $n\,\text{km} \times n\,\text{km}$ square field next to the ground control center, ripe of golden wheat. The field is divided into n^2 square patches, each measures $1\,\text{km} \times 1\,\text{km}$. You plan to spread sugar corn over a $2\,\text{km} \times 2\,\text{km}$ square covering 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 2×2 square if the message $m = 1$, the top right 2×2 square if $m = 2$, the bottom left 2×2 square if $m = 3$, and the bottom right 2×2 square if $m = 4$, as illustrated below. Both you and the astronauts know this strategy.

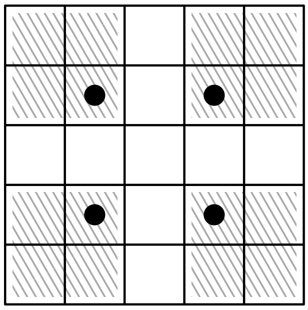


After you select the 2×2 square to spread the sugar corn, the astronauts observe the location of the barren patch (which is a random patch among the 4 selected patches), and deduce the message m based on whether the barren patch lies in the top left (which implies $m = 1$), top right ($m = 2$), bottom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



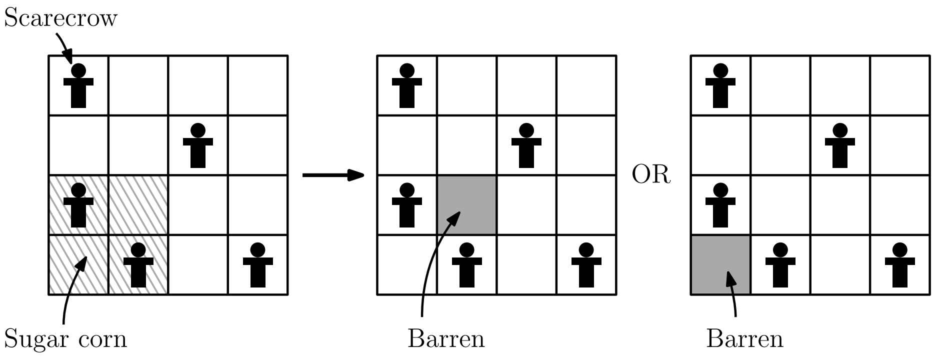
a) **(10 points)** Suppose $n = 5$. Find the largest k such that there is a communication strategy that guarantees that the astronauts can recover the message. Justify your answer.

$k = 4$. It is clear that you can fit four 2×2 squares in a 5×5 square without overlapping. To prove that this is optimal, consider the four patches at positions $(2, 2)$, $(2, 4)$, $(4, 2)$, $(4, 4)$ (where the x and y-coordinates range from 1 to 5; they are the patches with circles in the below figure). One cannot fit a 2×2 square that does not contain any of these 4 locations. Hence we can select at most four 2×2 squares without overlapping.

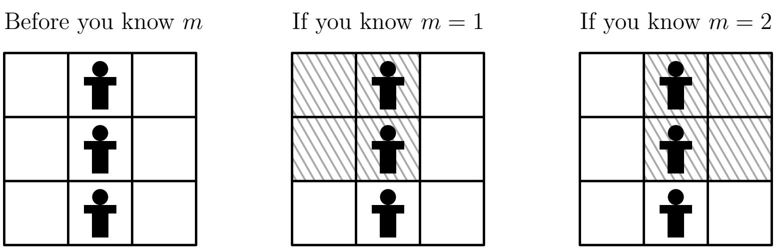


b) Before the astronauts departs to space, you decide to place at most s scarecrows in the field to deter the sparrows. The sparrows will always choose a patch without scarecrow, i.e., if some of the 4 patches with sugar corn have no scarecrows, the sparrows will choose one of those scarecrow-less patches 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 2×2 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 2×2 square, which contains 2 scarecrows, then the barren patch will be one of the remaining 2 patches in the 2×2 square.

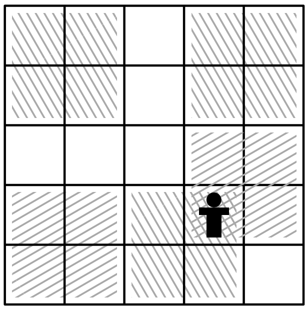


An example of a communication strategy for $n = 3$, $s = 3$, $k = 2$ is given below. Before the communication begins, you place the 3 scarecrows on the centre column. You then observe the message m to be sent. If $m = 1$, you select the top left 2×2 square. If $m = 2$, you select the top right 2×2 square. 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.

$k = 5$.



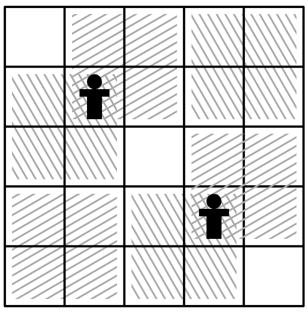
Let $k_{\max}(n, s)$ be the maximum of k given n and s. We argue that

$$k_{\max}(n, s+1) \leq k_{\max}(n, s) + 1$$

for $s \leq 2$. To prove this, assume the contrary that $k_{\max}(n, s+1) \geq k_{\max}(n, s)+2$. Consider a configuration achieving the optimum of $k = k_{\max}(n, s+1)$, where there are $s+1$ scarecrows, and k selected squares of size 2×2 that do not overlap at patches without scarecrows. Consider a scarecrow that has at most one other neighboring scarecrow (each patch has at most 4 neighbors on the top, bottom, left and right; since there are at most 3 scarecrows, there exists one with at most one neighboring scarecrow). If this scarecrow is in at most two 2×2 selected squares, then we can remove the scarecrow together with one of the selected squares, resulting in a configuration with s scarecrows and at least $k - 1$ selected squares, contradicting $k_{\max}(n, s + 1) \geq k_{\max}(n, s) + 2$. If this scarecrow is in at least three 2×2 selected squares, then these three squares will intersect at two more patches (which also requires scarecrows), resulting in this scarecrow having at least two neighboring scarecrows, which is also a contradiction. Hence $k_{\max}(n, s+1) \leq k_{\max}(n, s)+1$ for $s \leq 2$. In particular, $k_{\max}(5, 1) \leq k_{\max}(5, 0) + 1 = 5$. We know that $k = 5$ can be achieved. So $k_{\max}(5, 1) = 5$.

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.

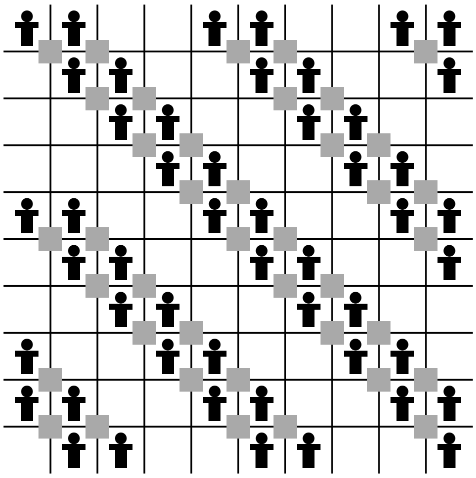
$k = 6$.



As argued in the previous part, $k_{\max}(5, 2) \leq k_{\max}(5, 1) + 1 = 6$. We know that $k = 6$ can be achieved. So $k_{\max}(5, 2) = 6$.

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.

k/n^2 can be arbitrarily close to $1/2$ for large n. We repeat the pattern below (the grey squares are the centres of the chosen 2×2 squares, which occupy half of the possible places to put 2×2 squares):



This is good enough for full score. If you are interested in why this strategy is optimal, keep reading.

We now prove that $k/n^2 \geq 1/2$ is impossible. Consider any communication strategy. Since there are an unlimited number of scarecrows, we can assume every chosen 2×2 square contains exactly one scarecrow-less patch. For a patch, its *multiplicity* is defined as the number of chosen 2×2 squares that contain that patch. Let a_i be the number of patches with scarecrows and with multiplicity i, where $i = 0, 1, 2, 3, 4$. Since each chosen 2×2 square contains one scarecrow-less patch, we have

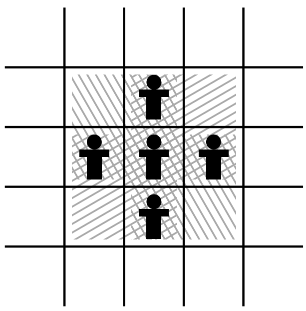
$$k + a_0 + \cdots + a_4 \leq n^2.$$

Also, since each chosen 2×2 square contains exactly three patches with scarecrows, we have

$$3k = a_1 + 2a_2 + 3a_3 + 4a_4,$$

since a patch with scarecrow and multiplicity 2 is counted in two chosen 2×2 squares.

Consider any patch with scarecrow and multiplicity 4, i.e., it is contained in four chosen 2×2 squares. There is only one possible configuration given below:



Note that the centre patch with multiplicity 4 has four neighbor patches with multiplicity 2, and those four patches cannot have another neighbor with multiplicity 4. Therefore,

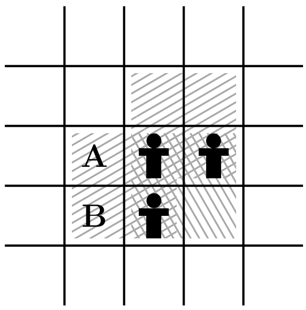
$$4a_4 \leq a_2.$$

Hence,

$$\begin{aligned} k &= \frac{1}{3}a_1 + \frac{2}{3}a_2 + a_3 + \frac{4}{3}a_4 \\ &\le \frac{1}{3}a_1 + \frac{2}{3}a_2 + a_3 + \frac{2}{3}a_4 + \frac{1}{6}a_2 \\ &\le a_0 + a_1 + a_2 + a_3 + a_4. \end{aligned} \tag{1}$$

This, combined with $k + a_0 + \cdots + a_4 \le n^2$, gives $k/n^2 \le 1/2$.

We then show that $k/n^2 = 1/2$ is impossible. Assume the contrary and consider a strategy with $k/n^2 = 1/2$. We must have $k = a_0 + \cdots + a_4 = n^2/2$, and hence the equality in (1) holds, meaning that $a_0 = a_1 = a_2 = a_4 = 0$. All patches with scarecrows have multiplicity 3. Consider the leftmost scarecrow (choose the topmost one in case of a tie). Since it has multiplicity 3, it must look like the scarecrow at the centre below:



Since every chosen 2×2 square has three scarecrows, one of the two patches A and B must have a scarecrow, contradicting the assumption that the scarecrow at the centre is the leftmost scarecrow. Hence, $k/n^2 = 1/2$ is impossible, and we always have $k/n^2 < 1/2$.

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*:

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

The probability that you find the bad box after exactly k inspections is equal to $1/N$ because of symmetry (the bad box can be at any of the locations in the random order). Therefore,

$$\sum_{k=1}^{N} \frac{k}{N} = \frac{N+1}{2}.$$

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.

Let π_i denote the index of the box inspected at the i-th time. Here π is a permutation of $\{1, 2, .., N\}$. Then the average number of inspections for this ordering is given by

$$\sum_{k=1}^{N} i q_{\pi_i}.$$

Suppose $\pi_k > \pi_{k+1}$ for some k. Then, note that by swapping the locations of inspections π_k and π_{k+1}, the expected waiting time reduces by

$$k q_{\pi_{k+1}} + (k+1) q_{\pi_k} - (k q_{\pi_k} + (k+1) q_{\pi_{k+1}}) = q_{\pi_k} - q_{\pi_{k+1}} \geq 0.$$

Therefore, in one of the best strategies, we can have $\pi_k \leq \pi_{k+1}$ for all k, or that $\pi = (1, 2, .., N)$, i.e., $\pi_k = k$.

More generally, the best strategy is to order the q_j values from largest to smallest, and begin by inspecting the boxes according to their probability, with the box with the highest probability inspected first, and the box with the lowest probability inspected last.

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.

The probability of not choosing box k in each hour is $1 - r_k$. Therefore, the probability of not choosing box k for m hours, and then choosing box k in the next hour, is

$$a_m(k) = (1 - r_k)^m r_k.$$

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

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

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 $x \in (-1, 1)$.

$$\mu(k) = \sum_{m=0}^{\infty} a_m(k) \cdot (m+1) = \sum_{m=0}^{\infty} (1-r_k)^m r_k \cdot (m+1) = \frac{1}{r_k}.$$

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 μ 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 μ, and find the smallest possible μ. 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.

We wish to minimize

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

subject to $r_k > 0$ and $\sum_k r_k = 1$. Using the Cauchy-Schwartz inequality we have

$$\left(\sum_{k=1}^{N} \frac{q_k}{r_k}\right)\left(\sum_{k=1}^{N} r_k\right) \geq \left(\sum_{k=1}^{N} \sqrt{q_k}\right)^2$$

Equality occurs when $q_k/r_k = cr_k$ for some constant c. Thus, the expression is minimized at

$$r_k = \frac{\sqrt{q_k}}{\sum_{i=1}^{N} \sqrt{q_i}}$$

which yields the following value for the minimum μ:

$$\mu = \left(\sum_{i=1}^{N} \sqrt{q_i} \right)^2$$

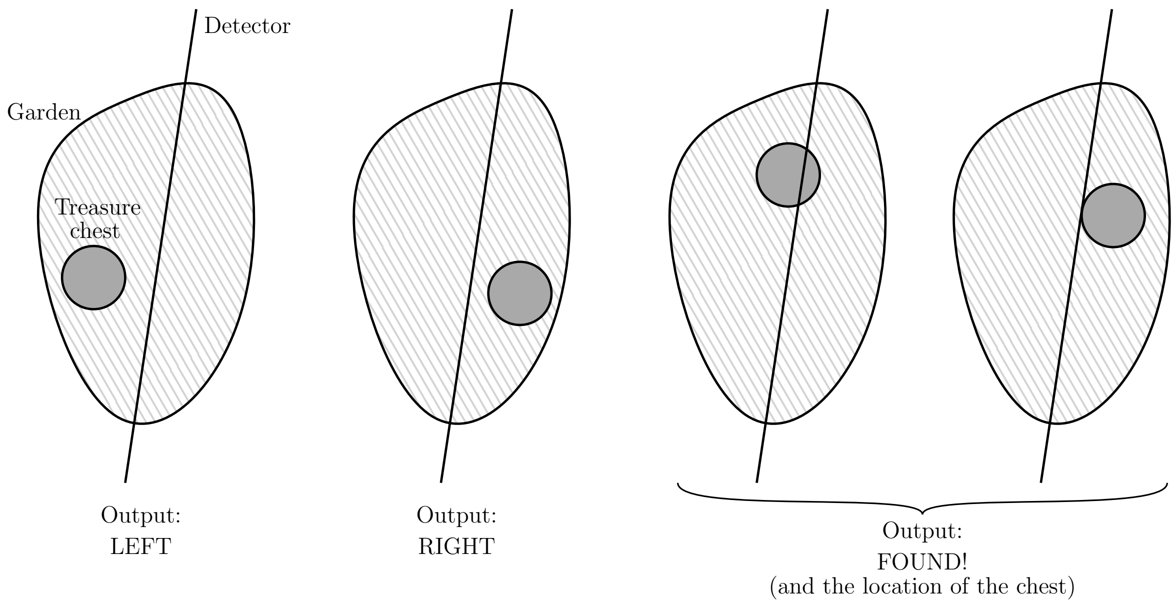
Thus, the square root of a person's prior probability should be used to pick them for screening.

Remark: A practical application of this idea is the following. For example, in an airport screening, if group A is four times more likely than group B to be terrorists, we should screen individuals in group A with $\sqrt{4} = 2$ times the frequency of individuals in group B. In particular, the screening probability should not be proportional to the risk posed by the group, but rather be proportional to the square root of the risk.

Question 3: Hidden treasure

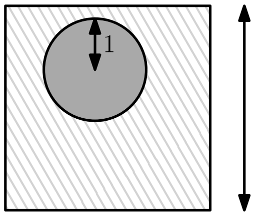
You are a treasure hunter looking for a treasure chest buried in a garden, which is a region on the 2D plane. The treasure chest has a circular shape with radius 1 metre, and it must lie completely within the garden (it is allowed to touch the boundary of the garden), though you do not know its location. 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.

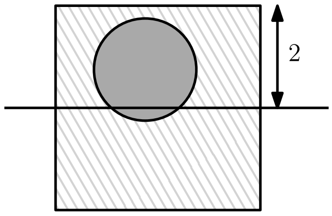


You want to find the precise location of the chest after using the detector at most k times. (Note that for each time, you can choose a different line depending on the previous outputs you have observed.) Find out the minimum k that guarantees your success, and the detection strategy attaining the minimum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

a) **(10 points)** The garden is a 4×4 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?)



Choose the horizontal line passing through the centre of the square. It is clear any any circle in the square intersects or touches that line.

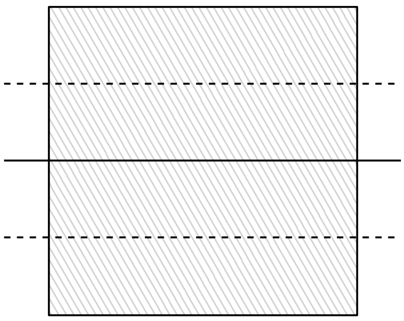


b) **(10 points)** The garden is a 8×8 square. (Hint: The answer is 2. Find a detection strategy which requires at most two uses of the detector, regardless of the position of the chest in the garden.)

Consider the y-coordinate of the centre of the circle $y_0 \in [1, 7]$. First choose the horizontal line $y = 4$. If $3 \leq y_0 \leq 5$, the chest is found. If $1 \leq y_0 < 3$ (the detector says that the chest is below the line), then choose the horizontal line $y = 2$, which is guaranteed to intersect or touch the circle. If $5 < y_0 \leq 7$ (the detector says that the chest is above the line), then choose the horizontal line $y = 6$, which is guaranteed to intersect or touch the circle.

Refer to the figure below for an illustration. The strategy is given as:

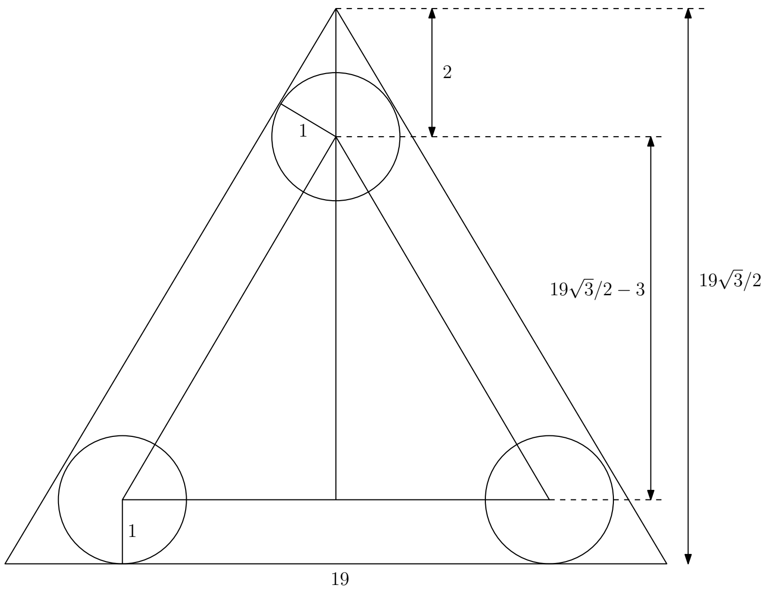
- First choose Line 0.
  - If above, choose Line 1.
  - If below, choose Line 2.



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

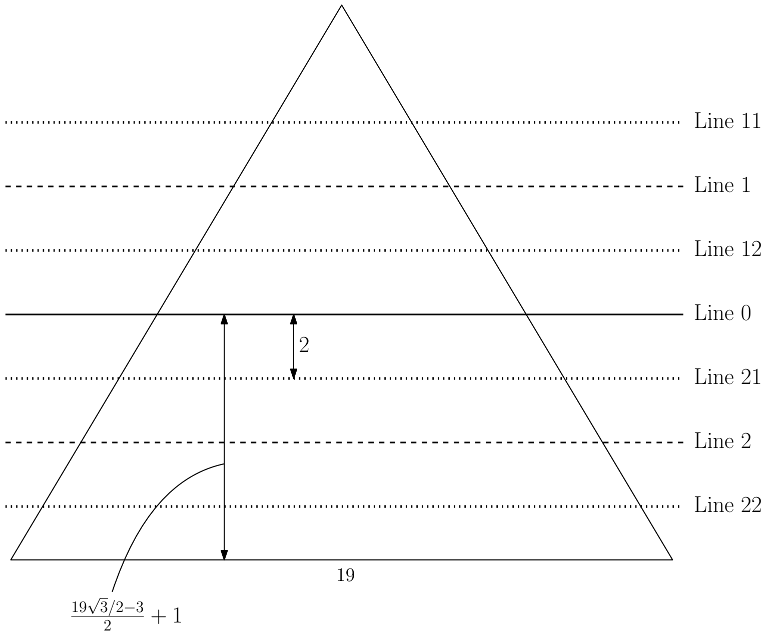
$k = 3$. Consider the y-coordinate of the centre of the circle y_0, which lies in an interval of length a. When only one use of the detector is allowed, we can find the chest if $a \leq 2$ as demonstrated in part (a). When 2 uses are allowed, we can find the chest if $a \leq 2 \times 2 + 2 = 6$ (as demonstrated in part (b)) since we can first choose the horizontal line in the middle of the interval, which can detect the chest if the centre lies in an interval of length 2. Otherwise we resort to the strategy when only one use is allowed, applied on the upper or lower remaining interval. By the same logic, when 3 uses are allowed, we can find the chest if $a \leq 6 \times 2 + 2 = 14$.

Place the triangle so that one of its sides is horizontal. Consider the y-coordinate of the centre of the circle y_0. As calculated in the figure below, y_0 lies in an interval with length $19\sqrt{3}/2 - 3 \approx 13.454 \leq 14$. Hence $k = 3$ can be achieved.



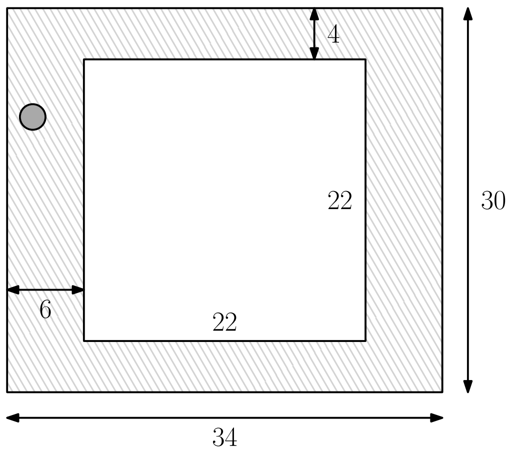
Refer to the figure below for an illustration of the strategy. The strategy is given as:

- First choose Line 0.
  - If above, choose Line 1.
    - If above, choose Line 11.
    - If below, choose Line 12.
  - If below, choose Line 2.
    - If above, choose Line 21.
    - If below, choose Line 22.



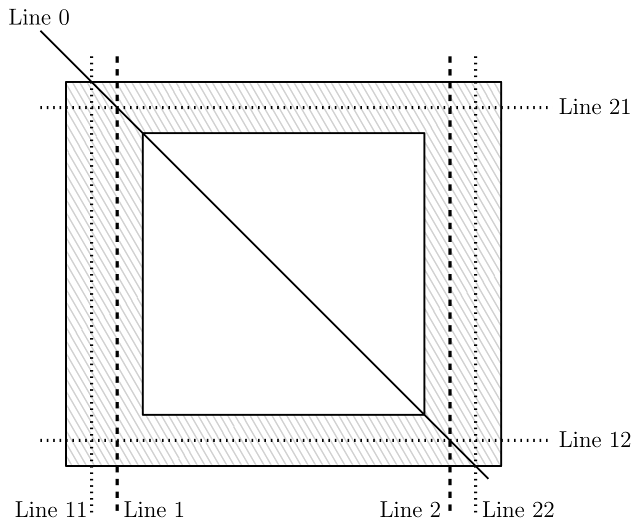
Remark: Choosing vertical lines instead of horizontal lines would result in a suboptimal strategy that requires 4 uses of the detector.

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



$k = 3$. Refer to the figure below.

- First choose Line 0.
  - If left, choose Line 1.
    - If left, choose Line 11.
    - If right, choose Line 12.
  - If right, choose Line 2.
    - If left, choose Line 21.
    - If right, choose Line 22.



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.)

Yes. First choose an arbitrary line that divides the garden into two halves, each with area $25/2$. If the chest is not found yet, consider the half which contains the chest, and again choose a line that divide it into two halves, each with area $25/4$. If the chest is not found yet, consider the half which contains the chest, and again choose a line that divide it into two halves, each with area $25/8$. This is guaranteed to find the chest since $25/8 = 3.125 < \pi$, so the circle cannot be contained in a half with area $25/8$.

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:

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

Putting $x = y$, we obtain that

$$f_1(2x) = f_1(x)^2 f_2(x)^2. \tag{2}$$

Putting $y = -x$, we obtain that (using even nature of the function)

$$f_2(2x) = f_1(x)^2 f_2(x)^2.$$

Therefore $f_1(x) = f_2(x)$, and let us call the function $f(x)$. By (2), we have $f(x) \geq 0$ and

$$f(2x) = f(x)^4. \tag{3}$$

Now if $f(x) = 0$ for some x, then $f\left(\frac{x}{2}\right) = 0$ from (3) and proceeding by induction we obtain that $f\left(\frac{x}{2^n}\right) = 0$ for all $n \geq 1$. However $\frac{x}{2^n} \to 0$ but $f(0) = 1$. This contradicts the continuity of the function. Therefore we must have $f(x) > 0$ for all $x \in \mathbb{R}$.

Let $c = f(1) > 0$. Then $f(2) = c^4$ (from (2)) by setting $x = y = 1$.

Let $x = ny$, for $n \in \mathbb{N}, n \geq 1$. Then, by the functional equation,

$$f((n+1)x)f((n-1)x) = f(nx)^2 f(x)^2. \tag{4}$$

We now use induction to prove the claim that for $x \in \mathbb{R}$, then $f(nx) = f(x)^{n^2}$ for $n \in \mathbb{N}$, $n \geq 1$. This is clearly true for $n = 1$ and $n = 2$ (see (3)). If the claim is true for $1, 2, \ldots, n$ for some $n \geq 2$, then by (4),

$$f((n+1)x)f((n-1)x) = f(nx)^2 f(x)^2,$$

$$f((n+1)x)f(x)^{(n-1)^2} = f(x)^{2n^2+2},$$

$$f((n+1)x) = f(x)^{(n+1)^2},$$

and hence the claim is also true for $n+1$. Hence $f(nx) = f(x)^{n^2}$ for $n \in \mathbb{N}$, $n \geq 1$. Setting $ny = x$ we obtain that

$$f(x) = f(ny) = f(y)^{n^2} = f\left(\frac{x}{n}\right)^{n^2},$$

or in other words

$$f\left(\frac{x}{n}\right) = f(x)^{\frac{1}{n^2}}.$$

Note that for $m, n \in \mathbb{N}$, and $n, m \geq 1$, as $f(nx) = f(x)^{n^2}$ and $f\left(\frac{x}{n}\right) = f(x)^{\frac{1}{n^2}}$, we have

$$f\left(\frac{m}{n}\right) = \left(f\left(\frac{1}{n}\right)\right)^{m^2} = f(1)^{\frac{m^2}{n^2}} = c^{\frac{m^2}{n^2}}.$$

Therefore, $f(x) = c^{x^2}$ for every rational x (since every rational $x > 0$ can be written as $x = n/q$ for some $n, q \in \mathbb{N}$. The case $x < 0$ is handled by the assumption that the function is even). From continuity, it follows that $f(x) = c^{x^2}$ for all $x \in \mathbb{R}$ (since we can approximate any real number arbitrarily well by rational numbers), where $c > 0$ is some constant.

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:

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

Let $g_1(x) = |f_1(x)|$ and $g_2(x) = |f_2(x)|$. Observe that g_1, g_2 satisfy the conditions in part (a). Therefore $g(x) = g_1(x) = g_2(x) = c^{x^2} = e^{rx^2}$, for some $c \neq 0$, where $r = \ln(c)$.

We can define continuous functions $\theta_1(x), \theta_2(x)$ such that $f_1(x) = g(x)e^{i\theta_1(x)}$ and $f_2(x) = g(x)e^{i\theta_2(x)}$. From the conjugate symmetry, we ensure that $\theta_1(-x) = -\theta_1(x), \theta_2(-x) = -\theta_2(x)$. Therefore $\theta_1(0) = \theta_2(0) = 0$. Notice that the given functional equation reduces to

$$\theta_1(x+y) + \theta_2(x-y) = \theta_1(x) + \theta_1(y) + \theta_2(x) - \theta_2(y)$$

Putting $x = y$ and $x = -y$ we obtain (using the given properties of the functions) that

$$\theta_1(2x) = 2\theta_1(x), \ \ \theta_2(x) = 2\theta_2(x).$$

We will use induction to prove that $\theta_1(nx) = n\theta_1(x)$ and $\theta_2(nx) = n\theta_2(x)$ for all $n \geq 1$. This claim clearly holds for $n = 1$, and we have shown that this claim also holds for $n = 2$. Assume that this claim holds for n and $n+1$, i.e., $\theta_1(nx) = n\theta_1(x)$ and $\theta_1((n+1)x) = (n+1)\theta_1(x)$, and $\theta_2(nx) = n\theta_2(x)$ and $\theta_2((n+1)x) = (n+1)\theta_2(x)$ for some n (which is true for $n = 1$). We will show that this claim also holds for $n+2$. The functional equation also yields, substituting the pair $(x,y) = ((n+1)x, x)$,

$$\theta_1((n+2)x) + \theta_2(nx) = \theta_1((n+1)x) + \theta_1(x) + \theta_2(nx) - \theta_2(x),$$

which by the induction hypothesis yields that $\theta_1((n+2)x) = (n+2)\theta_1(x)$. Similarly, substituting $(x,y) = (n+1)x, -x$, we obtain $\theta_2((n+2)x) = (n+2)\theta_2(x)$. Hence, by induction, $\theta_1(nx) = n\theta_1(x)$ and $\theta_2(nx) = n\theta_2(x)$ for all $n \geq 1$.

We now have

$$\theta_1\left(\frac{1}{n}\right) = \frac{1}{n}\theta_1(1), \quad \theta_1\left(\frac{p}{q}\right) = \frac{p}{q}\theta_1(1), \quad n,p,q \in \mathbb{N}$$

By continuity, we have $\theta_1(x) = x\theta_1(1)$. Similarly $\theta_2 = x\theta_2(1)$.

Therefore $f_1(x) = e^{rx^2+ic_1x}$ and $f_2(x) = e^{rx^2+ic_2x}$ are the only solutions (where we let $c_1 = \theta_1(1)$, $c_2 = \theta_2(1)$).

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.