THE PRISONERS CIRCLE HAT PROBLEM AND ITS OPTIMAL SOLUTION

LEO GITIN AND PAUL STAHR

1. Introduction

Among popular mathematics Prisoners and Hats puzzles are known to many people. A common variant goes as follows.

There are $n$ prisoners which are arrested for conspiracy but the jail is full. The (puzzle-affine) jailer comes up with a way of solving this problem: He puts a hat, either black or white, on each prisoner’s head and depending on whether an individual prisoner guesses the color of his hat correctly, he releases him - otherwise he gets executed. Because he wants to get over this quickly, he arranges all $n$ prisoners in a circle and has them guess the hat color simultaneously. Communication is not allowed, of course, otherwise they all get executed - nevertheless everybody sees all hats but his own. However rumors spread and the prisoners got to know about all this and could agree on a strategy in order to maximize the certain number of conspirators able to reconvene in freedom. What would such a strategy look like?

Don’t read further if you never saw this puzzle and think about it yourself.

The expected value of freed prisoners does not depend on the strategy, assuming that the hat colors are chosen randomly. Fix a prisoner and observe all possible distributions of hats. There will always be two distributions where he sees the same but one time his hat color will be white and the other time black. Thus he’ll win in half the cases. Summing over all prisoners we get an expectation of $\frac{n}{2}$ released prisoners. Thus, a strategy could only achieve at most $\left\lfloor \frac{n}{2} \right\rfloor$ certainly freed prisoners.

A strategy could look like this: Assigning the value 0 to black and value 1 to white we know for sure that for a given distribution of hats the sum of values will be either odd or even. Force $\left\lfloor \frac{n}{2} \right\rfloor$ conspirators to remember “odd” and $\left\lceil \frac{n}{2} \right\rceil$ to remember “even” – they form two groups. The conspirators who remembered “odd” then guess black as their hat color if the sum of values of all others is odd and guesses white otherwise. The conspirators remembering “white” do it the other way around. This way we assure that every member of one group will be right and no one of the other. We have $\left\lfloor \frac{n}{2} \right\rfloor$ certainly saved prisoners.

Now, consider an altered story:

The conspirators heard the rumors but are overwhelmed by the complexity of the problem. They (secretly) invite a mathematician to help them figuring out a strategy. Because of the secrecy and given limitations in the temporal imprisonment the mathematician can only write down the strategy on a piece of paper which will be read separately by the conspirators. He can not distinguish

2010 Mathematics Subject Classification. 05C35.
between the individual prisoners and has to treat them as equal. How many prisoners can be certainly saved and how?

We interpret the story in the following way: Since the prisoners are indistinguishable the mathematician can only give a deterministic strategy depending on the hat colors of the other \( n - 1 \) prisoners standing in the circle, that is, a function \( S : \{0,1\}^{n-1} \rightarrow \{0,1\} \).

We prove, that \( \lfloor \frac{n-1}{2} \rfloor \) prisoners can be saved for sure and that this is best possible. The key idea is to construct a certain graph \( G \) and view every possible strategy \( S \) as an orientation of the graph \( G \). The task to optimize the certain amount of saved prisoners reduces to finding an optimal orientation of \( G \), which will be formulated in terms of Graph Theory.

### 2. The bijection between strategies and graph orientations

We construct our desired graph \( G = (V(G), E(G)) \) as follows: Each vertex is an equivalence class of hat distributions \( \{0,1\}^{n} \) which are identical up to rotation. We call vertices symmetric if they contain at least two distributions (in Figure 1 three vertices are symmetric: the left, lower middle and right one). Every prisoner sees a sequence \( s \in \{0,1\}^{n-1} \) of hat colors and does not know its own hat color. For each sequence we insert one edge \( e_s \) in our graph, connecting the two possible hat distributions. This results in two vertices being connected by at least one edge if they only differ in one hat color. If they differ in more than two (for all possible rotations) they will not be connected. Particularly, parallel edges are possible (see Figure 1) and \( G \) has \( 2^{n-1} \) edges. Let \( \gamma(v) := \sum_{i \in [n]} v_i \) be the number of white hats, in each vertex. For all Edges \( \{u,v\} \in E(G) \) we have \( |\gamma(u) - \gamma(v)| = 1 \) because for all other distances one has to change more then one hat color. We visualize the graph \( G \) by ordering the vertices by \( \gamma \).

![Graph G for n = 4.](image)

If a symmetric vertex \( u \) is connected to vertex \( v \), then there are multiple ways of obtaining \( v \) by changing one hat color (in other words: there are prisoners which see the same sequence \( v \in \{0,1\}^{n-1} \)). Let \( e_s = (u,v) \) be an edge with \( \gamma(u) < \gamma(v) \) and \( s \) its corresponding sequence. We assign \( e_s \) (more precisely, the vertices \( u \) and \( v \) it is connecting), two weights \( w_-(e_s) \) and \( w_+(e_s) \): the number of prisoners in \( u \) and \( v \) seeing the sequence \( s \), respectively.

Denote \( w(v) = n \) to be the sum of weights, belonging to a vertex, which for a given orientation of \( G \) separates into \( w^+(v) \) for the edges leaving \( v \) and \( w^-(v) \) of edges entering \( v \). We are now ready to formulate the bijection between strategies \( S \) and edge orientations of \( G \):

An edge \( e_s = \{u,v\} \) with \( \gamma(u) < \gamma(v) \) ends on \( v \) if and only if \( S(s) = 1 \) otherwise it ends in \( u \).
So an edge which ends in \( v \) describes that the prisoner is saved in the distribution \( v \) and would be killed in \( u \). For a given orientation of \( G \) (and thus strategy) the number of released prisoners will be \( w^+(v) \) for \( v \) being the distribution of hats. Thus our task is the optimization of \( \min_{v \in V(G)} w^+(v) \).

3. Constructing an optimal strategy

For now we actually did nothing to solve our initial problem, we only understood better how to look at strategies in a different way.

Just before we really get into it, we have to know a bit more about the structure of \( G \), especially about the symmetric vertices. We postpone the proof of the following preposition to the last section, because it has nothing to do with the main line of our arguments.

Preposition 3.1. No two symmetric vertices are connected. Every vertex is connected to at most one symmetric vertex if \( n \) is odd and at most two if \( n \) is even.

We are now ready to prove the announced result.

Theorem 3.1. In the given Prisoners and Hats Problem \( \left\lfloor \frac{n-1}{2} \right\rfloor \) prisoners can be saved for sure and this is best possible.

Proof. We first prove that \( \left\lfloor \frac{n-1}{2} \right\rfloor \) is best possible.

Assume that every distribution of hats is chosen with probability \( \frac{1}{2^n} \). As mentioned in the introduction the expected number of saved prisoners is \( \frac{n}{2} \) and for odd \( n \) we immediately get \( \frac{n-1}{2} \) as an upper bound. For even \( n \) assume that there is a strategy which assures \( \frac{n}{2} \) saved prisoners. Consider the distribution where all hats are black. If \( S(0, \ldots, 0) = 1 \) then all prisoners get executed. If \( S(0, \ldots, 0) = 0 \) then all \( n \) conspirators are released, which is a contradiction because saving at least \( \frac{n}{2} \) in all other hat configurations would lead to a higher expectation than \( \frac{n}{2} \).

Combing the even and odd case we get an upper bound of \( \left\lfloor \frac{n-1}{2} \right\rfloor \).

In what follows we will find the desired orientation of \( G \) constructively.

Denote the set of symmetric vertices \( A \) and the complement set \( B := V(G) \setminus A \). According to Preposition 3.1 there are no edges in \( G[A] \) and for all edges \( e \) in \( G[B] \) we have \( w_-(e) = w_+(e) = 1 \). For an edge \( e = (u, v) \) with \( u \in A \) and \( v \in B \) with \( \gamma(u) < \gamma(v) \) the weight \( w_-(e) \) is greater than 1 and \( w_+(e) \) is equal to 1.

First we get rid of all circles in \( G[B] \) by orienting them in a cycle. For the remaining forest of not orientated edges we find an orientation iteratively, by selecting two leafs of a tree and orientating the path in between in an arbitrary direction until all edges of \( G[B] \) are oriented.

In the end we orientate all edges which have one vertex in \( A \) and one in \( B \) to \( A \).

We will now show that this orientation already satisfies \( w^+(v) \geq \left\lfloor \frac{n-1}{2} \right\rfloor \) for all \( v \in V(G) \).

First, we observe that \( w^+(v) = n \) for all \( v \in A \). For \( v \in B \) all weights are 1 and thus \( \deg_G(v) = n \). Assume \( n \) is odd. Having Proposition 3.1 in mind we know that \( \deg_T(v) \) will be odd if \( v \) only has edges in \( G[B] \) and even if there is an edge \( \{v, u\} \) with \( u \in A \). In both cases we have \( w^+(v) \geq w^-(v) - 1 \) and thus \( w^+(v) \geq \frac{n-1}{2} \) since \( w^+(v) + w^-(v) = n \).
If \( n \) is even we get \( w^+(v) \geq w^-(v) - 2 \) by Proposition 3.1 and then \( w^+(v) \geq \frac{n}{2} - 1 \). Finally, \( w^+(v) \geq \left\lfloor \frac{n-1}{2} \right\rfloor \) for all \( v \in G \). \( \Box \)

4. Discrepancy between symmetric vertices

It is left to show Preposition 3.1. This means we have to think about how different two symmetric vertices have to look.

Define the discrepancy between to hat-distributions \( v \) and \( w \) to be the minimal number of hats in \( v \) whose color has to be inverted to obtain \( w \). Or equivalently: Fix the position of the hat-distribution \( v \), superimpose it with \( w \) and count the number of mismatches. The minimal value of mismatches over all possible rotations of \( w \) is the discrepancy.

If a vertex \( u \) is connected to two symmetric vertices \( v \) and \( w \), then the \( v \) and \( w \) have discrepancy 2. Hence we can reformulate Preposition 3.1.

**Preposition 4.1.** Let \( v \) and \( w \) be two distinct symmetric vertices. The discrepancy between \( v \) and \( w \) is greater than 1 and if \( n \) is odd greater than 2. If \( n \) is even and some vertex \( u \) is connected to \( v \) and \( w \) there are no further symmetric vertices connected to \( u \).

**Proof.** \( v \) and \( w \) are symmetric if there are subsequences \( s_1 \) and \( s_2 \) of 0's and 1's of minimal length \( l \) and \( m \) respectively, which are repeated cyclically in the distribution of hats. Particularly, \( l|n \) and \( m|n \) with \( l,m < n \). First, let \( n \) be odd.

Let \( \text{lcm} : \mathbb{N}^2 \rightarrow \mathbb{N} \) be the least common multiple. If \( \text{lcm}(l,m) < n \) we automatically have \( \text{lcm}(l,m) \leq \frac{n}{3} \) because \( l,m,n \) are odd. Interpret \( v \) and \( w \) being symmetric in sequences \( s'_1 \) and \( s'_2 \) of length \( \text{lcm}(l,m) \). \( s'_1 \) and \( s'_2 \) have to differ in at least one place, otherwise \( v \) and \( w \) are identical. But then the discrepancy is at least \( \frac{n}{\text{lcm}(l,m)} \geq 3 \).

Now, consider \( \text{lcm}(l,m) = n \) and write \( l = d \cdot l' \), \( m = d \cdot m' \), \( n = d \cdot l' \cdot m' \) with coprime \( l' \) and \( m' \). Assume \( d = 1 \) and hence \( n = l \cdot m \). If we superimpose \( v \) and \( w \) as mentioned above, the number of mismatches does not depend upon the rotation of \( w \): Every element of \( s_1 \) overlaps exactly once with every element of \( s_2 \). The number of mismatches is therefore at least \( \min(l,m) \geq 3 \), except if both \( s_1 \) and \( s_2 \) only consist either out of 0's or 1's, in which case \( v \) and \( w \) are identical (i.e. monochromatic).

This idea can be extended to \( d > 1 \). View \( v \) to consist of \( d \) symmetric hat-distributions \( v_1, \ldots, v_d \) with \( \frac{n}{d} \) hats each, obtained by taking every \( d \)th hat in \( v \). The same applies to \( w \), we divide it into \( w_1, \ldots, w_d \). Superimposing \( v \) and \( w \) with minimal mismatches means superimposing \( v_1 \) with \( w_i \), \( v_2 \) with \( w_{i+1}, \ldots, v_d \) with \( w_{i-1} \), where \( 1 \leq i \leq d \) and \( v_k = v_{k \mod d} \). We know that \( l', m' > 1 \), otherwise \( \text{lcm}(l,m) \leq \max(l,m) < n \). Using the previous case we know that \( v_1 \) and \( w_i \) are identical or already have at least 3 mismatches. Assuming that \( v \) and \( w \) have discrepancy smaller 3 we know that all pairs \( v_1 \) and \( w_i \), \( v_2 \) and \( w_{i+1} \) etc. are identical and thus are \( v \) and \( w \), contradiction.

Assume \( n \) is even. If \( \text{lcm}(l,m) < \frac{n}{2} \) we get discrepancy of at least 3, as before in the odd case. If \( \text{lcm}(l,m) = \frac{n}{2} \) we can interpret \( v \) and \( w \) being symmetric in sequences of length \( \frac{n}{2} \) (diametrically opposed hats are the same). \( v \) and \( w \) cannot have discrepancy 1 and will have discrepancy 2 only if these two sequences differ in exactly one element. Particularly \( v \) and \( w \) vary only in two diametrically opposed hats.

For \( \text{lcm}(l,m) = n \) we proceed analogously as in the odd case. Now, \( v_1 \) and \( w_i \) are identical or have \( \min(l', m') \geq 2 \) mismatches. The same applies for \( v_2, w_{i+1} \) etc. Then, \( v \) and \( w \)
cannot have discrepancy 1 and will have discrepancy 2 only if amongst the pairs \(v_1\) and \(w_i\), \(v_2\) and \(w_{i+1}\) etc. all but one are identical. Without loss of generality say \(v_1\) are \(w_i\) have 2 mismatches and the rest is pairwise identical (and monochromatic). This is only possible if the recurring sequences of \(v_1\) and \(w_i\) are ‘00’ and ‘0...01’ or ‘11’ and ‘1...10’, implying that one is monochromatic and the other only has two opposite hats with the other color. Joining \(v_1\) and \(w_i\), \(v_2\) and \(w_{i+1}\) etc. back together we see that in \(v\) and \(w\) all diametrically opposed hats are the same and thus \(\text{lcm} \ l, m \leq \frac{n}{2}\), contradiction.

We are almost done: Assume a vertex \(u\) is connected to two different symmetric vertices \(v\) and \(w\). Then \(v\) and \(w\) have discrepancy 2 which we have shown is only possible when \(n\) is even and when diametrically opposed hats in \(v\) and \(w\) are the same. \(u\) and \(v\) have discrepancy 1, \(u\) cannot be symmetric. Moreover, there are two diametrically opposed hats \(h_1\) and \(h_2\) in \(u\), one white and one black. There are only two possibilities to restore symmetry in \(u\) by changing one hat color: Inverting either \(h_1\) or \(h_2\). There cannot be any other symmetric vertices other than \(v\) and \(w\) connected to \(u\). □

References

Department of Mathematics, University of Bonn, Bonn

Email address: leo.gitin@uni-bonn.de

Department of Mathematics, University of Bonn, Bonn

Email address: paul.stahr@gmx.de