

THE HAT PROBLEM

Brody Dylan Johnson

SAINT LOUIS UNIVERSITY

The Hat Problem:

Suppose n people attend a party and each person is adorned with either a red or black hat. Each individual is not allowed to see their own hat, but can see the hats of the other party-goers. There is equal likelihood of each hat being red or black. A game is played in which each guest is asked to identify the color of their hat. Guests may “pass” but at least one of the guests must attempt to identify their hat’s color. The guests play as a team and win whenever at least one person correctly identifies their hat’s color with no other guest making an incorrect guess. The guests may strategize about how to play the game before the party, but are forbidden to communicate in any way while playing. What strategy should the players adopt to maximize their chances of winning when $n = 3$? What about $n = 7$ or $n = 2^m - 1$?

Important aspects of the game:

- hat colors are independent events and $P(\text{red}) = P(\text{black}) = \frac{1}{2}$;
- an acceptable strategy must always result in at least one player making a guess;
- the players WIN when at least one correct guess is made and no incorrect guesses are made;
- each player can see the others' hats, but not their own;
- no communication is allowed between players at the party, including whether or not other guests have elected to PASS or GUESS.

Three guest strategy ($n = 3$):

- Strategy: If a player sees two hats of the same color (s)he guesses the opposite color. (Valid strategy because there must be two hats of the same color.)
- Let 0 = red and 1 = black. Each possible outcome of hat colors can be expressed as a 3-digit binary number. The outcomes are:

$$\begin{array}{cccc} 010 & 100 & 101 & 111 \\ 001 & 110 & 011 & 000 \\ \underbrace{\hspace{10em}} & & \underbrace{\hspace{2em}} & \\ \text{WIN} & & \text{LOSE} & \end{array}$$

Thus, the probability of winning with this strategy is $\frac{3}{4}$.

Three guest strategy ($n = 3$):

- This strategy is optimal.

Proof. Each guess has a 50-50 chance of being correct. This means that when we consider possible outcomes there must be an equal number of correct and incorrect guesses. Notice that in each WIN there is a single correct guess and in each LOSS there are *three* incorrect guesses. Winning one more game requires one more correct guess (a total of 7), but now we have only 1 possible game to lose and 7 incorrect guesses to make. There are only 3 players so this is impossible. \square

- Can we adapt this strategy to $n = 7$?

Seven guest strategy ($n = 7$):

- With seven hats we are assured of the fact that at least four hats will have the same color. (At least three guests will see four hats of the same color, too.)
- The obvious generalization of the $n = 3$ strategy would be: if a player sees four hats of the same color (s)he guesses the opposite color.
- When does the team win?

Seven guest strategy ($n = 7$):

- Assume that the first four hats are all red. Each of the other three guests will guess that their hat is black, so the only time they win is when all three of those hats *are* black. In other words, they win when exactly four hats have the same color.
- There are $\binom{7}{4} = 35$ outcomes where four hats are red and an equal number where four hats are black. There are $2^7 = 128$ possible outcomes so the probability of winning with this strategy is $70/128 \approx 0.547$.
- Is this the best we can hope for? (not if we use algebraic coding theory!)

Algebraic coding theory:

- Consider the field $\mathbb{F}_2 = \{0, 1\}$ with the addition/multiplication tables:

$+$	0	1	\times	0	1
0	0	1	0	0	0
1	1	0	1	0	1

- \mathbb{F}_2^n is an n -dimensional vector space over \mathbb{F}_2 . We may write vectors in \mathbb{F}_2^n as n -digit binary numbers, but must be careful to add digit-wise. (If $n = 3$ then $101 + 001 = 100$ not 110 .)

Algebraic coding theory:

- A *linear code*, \mathcal{C} , is a subspace of \mathbb{F}_p^n , where p is a prime ($p = 2$ for us) and n is a positive integer. (A subspace is collection of vectors that is closed under linear combinations.)
- The *Hamming weight* of a vector $\vec{c} \in \mathbb{F}_2^n$, $w(\vec{c})$, is the number of non-zero digits in the vector.
- The *Hamming distance* between $\vec{c}_1, \vec{c}_2 \in \mathbb{F}_2^n$ is

$$d(\vec{c}_1, \vec{c}_2) = w(\vec{c}_1 - \vec{c}_2).$$

This distance is the number of bits that must be *flipped* to switch one vector into the other.

Algebraic coding theory:

- Triangle Inequality: $d(\vec{c}_1, \vec{c}_3) \leq d(\vec{c}_1, \vec{c}_2) + d(\vec{c}_2, \vec{c}_3)$.

Proof. Each bit in which \vec{c}_3 differs from \vec{c}_1 falls into one of two groups: (a) $\vec{c}_3(k) = \vec{c}_2(k)$ which implies $\vec{c}_2(k) \neq \vec{c}_1(k)$ and (b) $\vec{c}_3(k) \neq \vec{c}_2(k)$. \square

- Example: Let $n = 3$ and choose \mathcal{C} to be the subspace $\{000, 111\}$. (2-dimensional) It takes three flips to change one codeword into the other. Any other word (3-digit binary number) is one flip away from either 000 or 111.
- The $n = 3$ strategy for the game corresponds to assigning each player a bit-number and telling them that if they see something that COULD be a codeword to make a guess that prevents the word from being a codeword.

(7,4) Hamming code:

- Let $n = 7$ and consider the code \mathcal{C} corresponding to the subspace generated by $\{1000110, 0100101, 0010011, 0001111\}$. (Every codeword is a linear combination of these four vectors.)
- Our message will be the four coefficients used to construct a codeword as a linear combination. Let $\vec{a} = [a_1 \ a_2 \ a_3 \ a_4]$. (4-digit binary message) We will transmit the codeword

$$\vec{c}_{\vec{a}} = \vec{a}G = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix} .$$

(7,4) Hamming code:

- There is another subspace of \mathbb{F}_2^7 , \mathcal{C}^\perp , so that the combined span of \mathcal{C} and \mathcal{C}^\perp is all of \mathbb{F}_2^7 and, moreover, every vector in \mathcal{C} is orthogonal to every vector in \mathcal{C}^\perp . (\mathcal{C}^\perp is 3-dimensional)
- In our example \mathcal{C}^\perp is generated by $\{1101100, 1011010, 0111001\}$. The orthogonality condition tells us that if $\vec{c} \in \mathcal{C}$, then

$$\begin{bmatrix} 0 & 0 & 0 \end{bmatrix} = \vec{c}H^T = \vec{c} \begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}^T.$$

- Notice that columns 3, 6, and 7 of H sum to $\vec{0}$. Also observe that no two columns of H sum to $\vec{0}$.

How close can two codewords be?

Proposition 1. *There is a nonzero codeword $\vec{c} \in \mathcal{C}$ such that $w(\vec{c}) \leq d$ if and only if there exists a set of d columns of H that are linearly dependent.*

Proof. First observe that $\vec{c}H^T = \vec{0}$ is a fancy way of writing a linear combination of columns of H that equals $\vec{0}$.

(\implies): Suppose there is a vector \vec{c} with $w(\vec{c}) \leq d$. The linear combination only involves $w(\vec{c}) \leq d$ columns and $\vec{c}H^T = \vec{0}$ is merely expressing the linear dependence of these $w(\vec{c})$ columns.

(\impliedby): Suppose there exist d columns $\vec{H}_{k_1}, \dots, \vec{H}_{k_d}$ of H and scalars a_{k_1}, \dots, a_{k_d} (at least one nonzero) so that

$$\sum_{n=1}^d a_{k_n} \vec{H}_{k_n} = \vec{0}.$$

Choose \vec{c} so that $\vec{c}(k_n) = a_{k_n}$ with $\vec{c}(k) = 0$ otherwise. Then \vec{c} belongs to the code because $\vec{c}H^T = \vec{0}$ and $w(\vec{c}) \leq d$ by construction. \square

(7, 4) Hamming code:

- In our (7, 4) Hamming code no two columns are linearly dependent, but there are triplets of columns which are dependent. Proposition 1 implies that no nonzero codeword \vec{c} has $w(\vec{c}) \leq 2$.
- Recall that $d(\vec{c}_1, \vec{c}_2) = w(\vec{c}_1 - \vec{c}_2)$, so we know that $d(\vec{c}_1, \vec{c}_2) > 2$ for all the codewords in our code. In other words, we must change at least *three* bits to change one codeword into another.
- **Conclusion:** *If a word is one flip away from a codeword, then that codeword is the unique codeword closest to the given word.*

Dissecting the (7, 4) Hamming code:

- How many codewords are there? 4-dimensional $\Rightarrow 2^4 = 16$ codewords
- How many words are one flip away from a given codeword? We have 7 digits, so there are seven words that differ from a codeword by 1. (None of these can be another codeword.)
- How many words do we have? $2^7 = 128 = 8 \times 16 = 16 + 7 \times 16$.

Conclusion: *Every word is either a codeword or one flip away from a uniquely defined codeword.*

Seven guest strategy ($n = 7$):

- Assign each guest a bit number and give each guest a list of codewords and the associated one-flip-away words. If a player sees six correct digits then (s)he guesses that his or her bit will result in a non-codeword. If a player sees that the word cannot be a codeword, (s)he simply elects not to guess.
- LOSING: In 16 cases (the codewords) everyone sees 6 correct digits and guesses incorrectly that the full word is not a codeword. (All seven players guess wrong.)
- WINNING: In the remaining 112 cases only one digit actually is “wrong” so only the player wearing this hat makes a guess. In this case the single guess is correct and the team wins.
- The probability of winning is $\frac{112}{128} = \frac{7}{8}$ with this strategy and is maximal by the same argument as before, i.e., there are 112 correct guesses and 16 incorrect guesses so there is no room for more right guesses and the correspondingly larger number of wrong guesses.

Stuff that could be thought about:

- One can define similar Hamming codes whenever $n = 2^m - 1$ and solve the hat problem using an analogous approach.
- What would happen if the probability of one hat color was greater than the other? What if more colors were used?
- Is there a way to handle the cases where $n \neq 2^m - 1$?