Error-Correcting Codes: The Mathematics of Communication

Nathan Kaplan

University of California, Irvine

Museum of Mathematics: Math Encounters

July 13, 2022

I. What is Coding Theory All About?

Communication over a Noisy Channel

◦ Suppose we want to send a Message.

◦ For simplicity, I will send you a single Bit, a 0 or 1.

◦ But, there is some probability, let's say 10%, that the message you receive is NOT the message I sent.

Example

If I send a 0:

- 90% chance you receive a 0,
- 10% chance you receive a 1.

◦ Communication is accurate 90% of the time.

Maybe 0 means

Fire the Missiles!

and 1 means

DON'T Fire the Missiles!

◦ 90% may not be high enough.

We can communicate more reliably by **Repeating the Message**.

Example

If I want to send 0, I will instead send 000.
If I want to send 1, I will instead send 111.

Question

If you receive 101*, what do you do?*

- If I send 000, the probability you receive 101 is

$$\frac{1}{10} \cdot \frac{9}{10} \cdot \frac{1}{10} = \frac{9}{1000}.$$

- If I send 111, the probability you receive 101 is

$$\frac{9}{10} \cdot \frac{1}{10} \cdot \frac{9}{10} = \frac{81}{1000}.$$

- Decode the message as 111.
- This strategy decodes correctly if there are 0 errors or 1 error out of 3.

How likely is this?

Question

How likely is it that there are 0 or 1 errors in the message you receive?

- The probability of 0 errors is

$$\frac{9}{10} \cdot \frac{9}{10} \cdot \frac{9}{10} = \frac{729}{1000}.$$

- The probability of exactly 1 error out of 3 is

$$\frac{1}{10} \cdot \frac{9}{10} \cdot \frac{9}{10} + \frac{9}{10} \cdot \frac{1}{10} \cdot \frac{9}{10} + \frac{9}{10} \cdot \frac{9}{10} \cdot \frac{1}{10} = \frac{243}{1000}.$$

- So the probability that the message is received correctly is 97.2%.
- There is a Cost for this increased reliability– must send 3 bits instead of 1.

Question

- *What if instead of repeating this message* 3 *times, we repeat it* 5 *times?*
- *What if we repeat it* 100 *times?*
- *What if we repeat it* n *times?*

How do we efficiently build redundancy into our set of messages so that we can identify and correct errors?

II. The Hat Guessing Game

◦ You are in a room with two friends.

◦ Each of you has a Red Hat or a Blue Hat.

◦ You can see the other two hats but you cannot see your own.

Example

You see one of XRR, XRB, XBR, XBB.

◦ Each player has the opportunity to guess the color of their hat.

⋆ You do not have to guess.

◦ If every player that does guess picks the correct color
AND at least one player guesses, then the team wins the **BIG PRIZE**.

Example

Suppose the hats are RBB.

- Player 1 guesses Red.
- Player 2 does not guess.
- Player 3 guesses Red.

The Team **LOSES**.

∘ Team can work together before hats are given out to develop a Strategy.

⋆ Once hats are given out, no further communication is allowed.

⋆ Each player simultaneously announces their decision: Red, Blue, or Pass.

Example

Each player will guess Red.

∘ Win with probability $\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$.

Example

Players 1 and 2 will Pass.

Player 3 will guess Red.

∘ Win with probability $\frac{1}{2}$.

Strategy: If I see two hats of the same color, guess the Opposite Color.
If I see two hats of different colors, Pass.

Example

If I see XRR, I will guess Blue.

Strategy: If I see two hats of the same color, guess the Opposite Color.
If I see two hats of different colors, Pass.

Example

If our hats are BRR:

- Player 1 guesses Blue.
- Players 2 and 3 Pass.

We win the **BIG PRIZE**!

If the hats are RRR:
◦ All three players guess Blue.
◦ We LOSE!

Break: Some Questions

Question

1. *How often does this Strategy win?*

   BBB, BBR, BRB, BRR, RBB, RBR, RRB, RRR

2. *Suppose there are now 4 players.*
   I claim you can win at least as often as with the strategy above.
   Can you explain why?

3. *Can you come up with a strategy like this one when there are 5 players? What about 7 players?*

4. *What does any of this have to do with the first part of the talk?*

The Hat Game with 3 Players

Strategy: If I see two hats of the same color, guess the Opposite Color.
If I see two hats of different colors, Pass.

Winners: BBR, BRB, BRR, RBB, RBR, RRB.
Losers: BBB, RRR.

◦ We win with probability $3/4$!

Question

It turns out that this is the best you can do.
That is, there is no strategy that wins with probability greater than $3/4$.
Can you explain why?

III. Coding Theory Basics

Definition

A binary code C of length n is a subset of the 2^n binary strings of length n.

Example

$C = \{000, 111\}$ is a binary code of length 3.

Definition

The Hamming distance between two binary strings of length n is the number of coordinates in which they are different.

Example

$$d(000, 111) = 3 \text{ and } d(101, 001) = 1.$$

Definition

The minimum distance of a code C is the minimum Hamming distance that occurs between two different elements of C.

Example

The minimum distance of $C = \{000, 111\}$ is 3.

Question

Why is minimum distance important?

Fact: A binary code C of length n and minimum distance d can identify and correct up to $\lfloor \frac{d-1}{2} \rfloor$ errors.

$\star$ $\lfloor \frac{d-1}{2} \rfloor$ is the floor of $\frac{d-1}{2}$.

$$\left\lfloor \frac{d-1}{2} \right\rfloor = \begin{cases} \frac{d-1}{2} & \text{if } d \text{ is odd,} \\ \frac{d}{2} - 1 & \text{if } d \text{ is even.} \end{cases}$$

Example

$C = \{000, 111\}$ can identify and correct one error.

We want codes C of large size and large minimum distance.

Main Problem in Combinatorial Coding Theory

We want codes C of **large size** and **large minimum distance**.

Definition

∘ $A_2(n, d)$ *is the maximum size of a binary code* C *of length* n *that has minimum distance at least* d*.*
∘ $A_2(n, d)$ *is the maximum number of binary strings of length* n *such that any two differ in at least* d *positions.*

Example

∘ $A_2(3, 3) = 2$.
∘ $A_2(n, n) = 2$.

Main Problem in Combinatorial Coding Theory

Compute values of $A_2(n, d)$.

On the Size of Optimal Three-Error-Correcting Binary Codes of Length 16

Patric R. J. Östergård

***Abstract*—Let $A(n, d)$ denote the maximum size of a binary code with length n and minimum distance d. It has been known for decades that $A(16, 7) = A(17, 8) = 36$ or 37, that is, that the size of optimal 3-error-correcting binary codes of length 16 is either 36 or 37. By a recursive classification via subcodes and a clique search in the final stage, it is shown that the size of optimal such codes is 36.**

attaining the lower bound have been constructed in [13], [14] (see also [10, pp. 57,58]) and the upper bound is from [3]. The problem of determining this particular value is also mentioned in [8, Research Problem 7.18]. The main result of this work is that the best known lower bound is the exact value: $A(17, 8) = 36$.

◦ The maximum number of binary strings of length 17 such that any two differ in at least 8 positions is 36.

◦ Not so difficult to show that the maximum number of binary strings of length 17 such that any two differ in at least 8 positions is at most 37.

◦ Not so difficult to find 36 binary strings of length 17 such that any two differ in at least 8 positions.

Main result of this Paper: Any collection of 37 binary strings of length 17 contains two strings that differ in at most 7 positions.

$\star$ There are $\binom{131072}{37}$ such subsets. This is a **BIG number**.

	d=4	d=6	d=8	d=10	d=12	d=14	d=16
6	4	2	1	1	1	1	1
7	8	2	1	1	1	1	1
8	16	2	2	1	1	1	1
9	20	4	2	1	1	1	1
10	40	6	2	2	1	1	1
11	72	12	2	2	1	1	1
12	144	24	4	2	2	1	1
13	256	32	4	2	2	1	1
14	512	64	8	2	2	2	1
15	1024	128	16	4	2	2	1
16	2048	256	32	4	2	2	2
17	2816-3276	258-340	36	6	2	2	2
18	5632-6552	512-673	64	10	4	2	2
19	10496-13104	1024-1237	128	20	4	2	2
20	20480-26168	2048-2279	256	40	6	2	2
21	40960-43688	2560-4096	512	42-47	8	4	2
22	81920-87333	4096-6941	1024	64-84	12	4	2
23	163840-172361	8192-13674	2048	80-150	24	4	2
24	327680-344308	16384-24106	4096	136-268	48	6	4
25	2^{19}-599184	17920-47538	4096-5421	192-466	52-55	8	4
26	2^{20}-1198368	32768-84260	4104-9275	384-836	64-96	14	4
27	2^{21}-2396736	65536-157285	8192-17099	512-1585	128-169	28	6
28	2^{22}-4792950	131072-291269	16384-32151	1024-2817	178-288	56	8

Figure: Brouwer's tables of upper and lower bounds for $A_2(n, d)$

Question

What is $A_2(17, 6)$?

IV. The Hat Game with 7 Hats

◦ 7 people are in a room. Each one is given a Red Hat or a Blue Hat. Each person can see the hats of the 6 others, but not their own hat.

◦ Each player has the opportunity to guess the color of their hat.
⋆ You do not have to guess.

◦ If every player that does guess picks the correct color AND at least one player guesses, then the team wins the BIG PRIZE.

◦ Team can work together before hats are given out to develop a Strategy.
⋆ Once hats are given out, no further communication is allowed.
⋆ Each player simultaneously announces their decision: Red, Blue, or Pass.

The [7,4] binary Hamming code

Choose an assignment: Red is 0 and Blue is 1.

Consider the 16 codewords

0	0	0	0	0	0	0
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
1	1	0	0	0	1	1
1	0	1	0	1	0	1
1	0	0	1	0	0	1
0	1	0	1	0	1	0
0	1	1	0	1	1	0
0	0	1	1	1	0	0
1	1	1	0	0	0	0
1	1	0	1	1	0	0
1	0	1	1	0	1	0
0	1	1	1	0	0	1
1	1	1	1	1	1	1

Strategy: Look at the other 6 hats.

- If there is a choice for your hat color so that the 7 hats give one of these 16 words, choose the **OTHER COLOR**.
- If neither choice for your hat color gives one of these 16 words, **PASS**.

0	0	0	0	0	0	0
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
1	1	0	0	0	1	1
1	0	1	0	1	0	1
1	0	0	1	0	0	1
0	1	0	1	0	1	0
0	1	1	0	1	1	0
0	0	1	1	1	0	0
1	1	1	0	0	0	0
1	1	0	1	1	0	0
1	0	1	1	0	1	0
0	1	1	1	0	0	1
1	1	1	1	1	1	1

Question

If the hats are

$$BRBBBRR$$

what does each player do?

◦ Player 1 sees $X011100$. Since 0011100 is one of the 16 codewords, Player 1 chooses BLUE, the color corresponding to 1.

◦ Player 2 sees $1X11100$. Since 1111100 and 1011100 are not among the 16 codewords, Player 2 chooses PASS.

◦ Every other Player chooses PASS. The team WINS!

Strategy: Look at the other 6 hats.

- If there is a choice for your hat color so that the 7 hats give one of these 16 words, choose the **OTHER COLOR**.
- If neither choice for your hat color gives one of these 16 words, **PASS**.

Question

How often does this strategy win?

◦ When the hats form one of the 16 words of the $[7, 4]$ binary Hamming code, **everyone guesses** and **everyone is** WRONG!

◦ When we lose, we lose BIG.

Question

What happens for the other $128 - 16 = 112$ *possible hat configurations?*

Strategy: Look at the other 6 hats.

- If there is a choice for your hat color so that the 7 hats give one of these 16 words, choose the **OTHER COLOR**.
- If neither choice for your hat color gives one of these 16 words, **PASS**.

∘ If the hats do not form one of the 16 words of the $[7, 4]$ binary Hamming code, then **exactly one person guesses** and **they are** right.

∘ **WIN** with probability $\frac{112}{128} = \frac{7}{8}$.

Question

What is going on here?

Hamming codes are perfect

◦ The $[7,4]$ binary Hamming code C is a Perfect 1-error-correcting code.

◦ Every binary string of length 7 is either an element of C or is Hamming distance 1 away from a unique element of C.

◦ Take each of the 16 codewords together with the 7 binary strings that you get from changing 1 of the 7 bits.

⋆ These Hamming balls exactly cover the set of all binary strings of length 7 with no overlaps.

◦ Note that $128 = 16 + 7 \cdot 16$.

What are the elements of the $[7,4]$ binary Hamming code?

0	0	0	0	0	0	0
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
1	1	0	0	0	1	1
1	0	1	0	1	0	1
1	0	0	1	0	0	1
0	1	0	1	0	1	0
0	1	1	0	1	1	0
0	0	1	1	1	0	0
1	1	1	0	0	0	0
1	1	0	1	1	0	0
1	0	1	1	0	1	0
0	1	1	1	0	0	1
1	1	1	1	1	1	1

- The first 4 bits are any binary string of length 4: $abcd$.
- Call the next three bits efg. Then
  - $e = 0$ if $a + b + d$ is even and $e = 1$ if $a + b + d$ is odd.
  - $f = 0$ if $a + c + d$ is even and $f = 1$ if $a + c + d$ is odd.
  - $g = 0$ if $b + c + d$ is even and $g = 1$ if $b + c + d$ is odd.

For example, the next to last codeword has

$$a = 0, \ b = 1, \ c = 1, \ d = 1, \ e = 0, \ f = 0, \ g = 1.$$

V. 20 Questions with a Lie

◦ Choose a number between 0 and 15.

Question

How many YES/NO questions do I need to ask to identify your number?

◦ Choose a number between 0 and 15.

Question

How many YES/NO questions do I need to ask to identify your number?

Question

1. *Is your number between* 0 *and* 7*?*
2. *Is your number between* 0 *and* 3*?*
3. *Is your number between* 0 *and* 1*?*
4. *Is your number* 1*?*

∘ Choose a number between 0 and 15.

Question

How many YES/NO questions do I need to ask to identify your number?

Write your number in **binary** $abcd$:

$$0 = 0000, \quad 1 = 0001, \quad 2 = 0010, \quad 3 = 0011,$$
$$4 = 0100, \quad 5 = 0101, \quad 6 = 0110, \quad 7 = 0111,$$
$$8 = 1000, \quad 9 = 1001, \quad 10 = 1010, \quad 11 = 1011,$$
$$12 = 1100, \quad 13 = 1101, \quad 14 = 1110, \quad 15 = 1111.$$

Question

1. *Is* $a = 0$*?*
2. *Is* $b = 0$*?*
3. *Is* $c = 0$*?*
4. *Is* $d = 0$*?*

20 Questions with a Lie

- Choose a number between 0 and 15.
- I ask a series of YES/NO questions about your number.
- You must answer honestly except you can choose one question and **LIE**.

Question

How many YES/NO questions do I need to ask to identify your number?

20 Questions with a Lie

Write your number in **binary** $abcd$:

$$0 = 0000,\quad 1 = 0001,\quad 2 = 0010,\quad 3 = 0011,$$
$$4 = 0100,\quad 5 = 0101,\quad 6 = 0110,\quad 7 = 0111,$$
$$8 = 1000,\quad 9 = 1001,\quad 10 = 1010,\quad 11 = 1011,$$
$$12 = 1100,\quad 13 = 1101,\quad 14 = 1110,\quad 15 = 1111.$$

Compute

$$e = a + b + d,\quad f = a + c + d,\quad g = b + c + d.$$

Question

1. *Is $a = 0$?*
2. *Is $b = 0$?*
3. *Is $c = 0$?*
4. *Is $d = 0$?*
5. *Is e even?*
6. *Is f even?*
7. *Is g even?*

Question

Why does this strategy work?

◦ If you answered all 7 questions honestly, your binary string $abcdefg$ would be an element of the $[7, 4]$ binary Hamming code.

◦ Since you can lie 1 time, we get a binary string with Hamming distance at most 1 away from an element of the Hamming code.

◦ This closest element is **unique**.

◦ Decode as the number corresponding to the first four bits of this closest element.

Question

1. *How does this idea adapt to larger numbers?*
2. *How does this idea adapt to more lies?*

VI. What else is there?

◦ It is usually difficult to compute $A_2(n, d)$ exactly.

◦ **Lower Bounds**: Come up with interesting examples of codes with large size and large minimum distance.

◦ **Upper Bounds**: Prove that certain codes are as large as possible.

Example

Singleton Bound: $A_2(n, d) \leq 2^{n-(d-1)}$.

◦ There is a perfect 1-error-correcting binary Hamming code of length $2^n - 1$ for each $n \geq 2$.
⋆ It has size $2^{2^n - n - 1}$.

Example

1. $C = \{000, 111\}$ $(n = 2)$.
2. $C = [7, 4]$ binary Hamming code $(n = 3)$.

Question

How do we describe the 2^{11} codewords of the binary Hamming code of length 15?

Codes over Other Alphabets

- We can consider codes whose symbols are not just bits, 0s and 1s.

Example

- Suppose 22 teams play 11 soccer matches.
- Each match has 3 possible outcomes:
  Either team could win or they could draw.
- A Bet consists of choosing an outcome for each match.
- If you choose all the matches correctly, you win 1st PRIZE.
  If you choose all the matches correctly except 1, you win 2nd PRIZE.
  If you choose all the matches correctly except 2, you win 3rd PRIZE.
  And so on...
- If you make all of the $3^{11} = 177147$ possible bets, then you are guaranteed to win 1st Prize.

Question

How many bets do you need to make to guarantee that you win at least one 2nd Prize?

The Finnish Football Pool

◦ In 1947, the Finnish football magazine *Veikkaaja* published $3^6 = 729$ bets that guarantee at least the 3rd Prize.

◦ These bets correspond to elements of the ternary Golay code of length 11.

Key Idea: The Covering Radius of this code is 2.
$\star$ That is, every string with symbols $\{0, 1, 2\}$ of length 11 has Hamming distance at most 2 from an element of this code.

Note that $3^{11} = 3^6 + \binom{11}{1} \cdot 2 \cdot 3^6 + \binom{11}{2} \cdot 2^2 \cdot 3^6$.

Question

What other interesting schemes like this can we develop?

References

1. *A Hat Trick of Hat Puzzles* by Pradeep Mutalik. *Quanta*, March 9, 2016.
Available online: `https://tinyurl.com/mutalikhatpuzzles`
2. *Hat Tricks* by Joe Buhler. *The Mathematical Intelligencer* **24**, 44–49 (2002).
Available online: `https://tinyurl.com/buhlerhattricks`
3. *Hat Problems* featuring Joe Buhler. *Numberphile* video.
Available online: `https://www.youtube.com/watch?v=laAtv310pyk`
4. *Coding theory applied to a problem of Ulam* by Ivan Niven. *Mathematics Magazine* **61** (1988), no. 5, 275–281.
Available online: `https://tinyurl.com/niven20qs`
5. *Why Mathematicians Now Care About Their Hat Color* by Sara Robinson. *New York Times*, April 10, 2001.
Available online: `https://tinyurl.com/robinsonhatcolor`
6. *The Hat Problem and Some Variations* by Wenge Guo, Subramanyam Kasala, M. Bhaskara Rao, and Brian Tucker.
In: *Advances in Distribution Theory, Order Statistics, and Inference.*
Available online: `https://tinyurl.com/guokasalaraotucker`
7. *Football Pools – A Game for Mathematicians* by Heikki Hämäläinen, Iiro Honkala, Simon Litsyn, and Patric Östergård. *American Mathematical Monthly* **102** (1995), no. 7, 579–588.
Available online: `https://tinyurl.com/footballpoolmath`