Hamming Codes
Some announcements

- Announcements for game due (via email) on Wednesday, March 15
- Homework 6 due on March 15
- Exam 3 on March 17
Today’s Goals

- Learn about error correcting codes and how they work.
- Construct Hamming Codes and use them to transmit simple messages.
- Examine how these codes are relevant to optimal strategies for selecting lottery tickets.
Hamming Codes

Hamming codes are a way to transmit messages that may contain errors and correct for them.

Note: this is different then transmitting messages secretly (RSA), though the two may be used together.

Hamming codes are written in bits. A bit is a digit which is either zero or one. This is also called binary or mod 2.

A string is a sequence of bits.

Attached to each string in a Hamming code is an additional parity check string (also in bits).
Hamming Codes Example

We begin with a 4-bit string. Call the string $a_1a_2a_3a_4$. Three check digits, $c_1$, $c_2$, and $c_3$, will be attached to the 4-bit string to produce a 7-bit string.

c1, c2, and c3 are chosen as follows:

\[
c_1 = \begin{cases} 
0 & \text{if } a_1 + a_2 + a_3 \text{ is even} \\
1 & \text{if } a_1 + a_2 + a_3 \text{ is odd}
\end{cases}
\]

\[
c_2 = \begin{cases} 
0 & \text{if } a_1 + a_3 + a_4 \text{ is even} \\
1 & \text{if } a_1 + a_3 + a_4 \text{ is odd}
\end{cases}
\]

\[
c_3 = \begin{cases} 
0 & \text{if } a_2 + a_3 + a_4 \text{ is even} \\
1 & \text{if } a_2 + a_3 + a_4 \text{ is odd}
\end{cases}
\]
The sums shown above are called parity check sums - so named because their purpose is to ensure that the sum of various components of the encoded message is even.

For example, $c_1$ is defined in so that $a_1 + a_2 + a_3 + c_1$ is even.
Construct the Hamming code word corresponding to the 4 bit string 0101

\[
a_1 + a_2 + a_3 = 0 + 1 + 0 \text{ is odd so } c_1 = 1
\]
\[
a_1 + a_3 + a_4 = 0 + 0 + 1 \text{ is odd so } c_2 = 1
\]
\[
a_2 + a_3 + a_4 = 1 + 0 + 1 \text{ is even so } c_3 = 0.
\]

The Hamming code word corresponding to 4-bit string 0101 is 0101110.
Every possible sequence of 7 bits is either a correct message (corresponds to a Hamming code word) or contains exactly one correctable error.
Sometimes, due to noisy transmission, code words contain errors. The Hamming Code is designed to detect and correct errors in 4-bit transmissions.

Suppose a message is received as 111010. Is this a Hamming code word? If not, what word should it have been?
In order to determine if the message received is a Hamming Code word, we simply scan the code. If it is one of the 16 code words, we know the message is received as sent.

If it is not among the 16 code words, we compare the message received with each code word and compute the Hamming distance for each.

The Hamming distance is defined as the number of times a bit in the received message differs from the bit in the code word.
Example

Compare the code words 0001011 with 1111010.

These words differ in 4 positions.

We say the **Hamming distance** is 4.
Once all the distances are computed, we locate the Hamming code which produces the shortest distance for 1111010. We also call this the nearest code word. This code will be the code used to correct the transmission error.

If there is more than one shortest distance, we do not correct the message.
The Transylvania Lottery

This game is like a usual pick 3 from the numbers 1 through 7 (no repeats) with a jackpot for hitting all three and a smaller prize for hitting two out of three (deuce).
The Transylvania Lottery

How many possible combinations are there of tickets?

$$7 \binom{3} = \frac{7!}{3!(7-3)!} = \frac{7 \cdot 6 \cdot 5}{3 \cdot 2} = 35.$$ 

Say you are going to buy seven tickets. What are your chances of hitting the Jackpot?

$$\frac{7}{35} = \frac{1}{5} = 20\%.$$ 

This is true no matter what tickets you choose.

Say you wanted to maximize the number of deuces? Does it matter what tickets you select? Yes!
This is the Fano plane. It is an example of a finite geometry (seven points). It satisfies two rules:

(1) every pair of points is contained in exactly one common line and

(2) every pair of lines contains exactly one common point.
There are exactly 7 lines in the Fano plane. They are:

124, 135, 176, 236, 257, 437, 456

These are exactly the numbers we should pick for the Translyania Lottery.
The Transylvania Lottery

Why these numbers?

124, 135, 176, 236, 257, 437, 456.

Try picking any three numbers.

There are either three deuces or no deuces.

But, if there are no deuces then we win the jackpot!
What does this have to do with Hamming codes? Here’s another codebook from the Ellenberg text:

<table>
<thead>
<tr>
<th>Original</th>
<th>Encoded</th>
</tr>
</thead>
<tbody>
<tr>
<td>000</td>
<td>0000000</td>
</tr>
<tr>
<td>010</td>
<td>0101011</td>
</tr>
<tr>
<td>101</td>
<td>1011010</td>
</tr>
<tr>
<td>100</td>
<td>1001101</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Original</th>
<th>Encoded</th>
</tr>
</thead>
<tbody>
<tr>
<td>001</td>
<td>0010111</td>
</tr>
<tr>
<td>011</td>
<td>0111100</td>
</tr>
<tr>
<td>110</td>
<td>1100110</td>
</tr>
<tr>
<td>111</td>
<td>1110001</td>
</tr>
</tbody>
</table>

This one satisfies a special property that no code with one error is close to two different codes from the codebook so we can always figure out the correct message from Hamming distance.
Hamming Codes and the Fano Plane

Say we translate each line in the Fano Plane to binary. We put a 0 in the $n$ spot if point $n$ is on the line and a 1 if it is not.

Then the line 124 is 0010111.

The line 257 is 1011010.

You can check that each line corresponds to one code from the codebook.