Introduction to Information Theory

1. How to measure information? How to ask the most informative questions?
   - “bit”... but: 0 vs 1
   - “entropy”

2. How to compress a data source?
   - “guess a number” game
   - lossless: FLAC, ZIP, GIF,...
   - lossy: JPG, MP3, MP4,...

3. How to reliably send information over unreliable channels?
   - LTE, Blue-ray, QR-codes,...

Mackay: How to measure information. How to ask the most informative questions? Entropy.


Origins: telecommunication + physics
- Morse (1830s)
- Bell Labs (1920s)
- Thermodynamics (1870s)
- Boltzmann, Gibbs

\[
\text{info} \sim \text{log (number of voltage levels)} \quad \circ \quad \text{log (number of possible signals)}
\]

Nyquist: abstraction.

Today: engineering + theory (efficient codes, beyond i.i.d.) + (quantum)

\[C = \text{Compression}\]

Suppose we want to compress a message in \{A, B, C, D\} = \mathcal{A}:

<table>
<thead>
<tr>
<th>Source message X \in \mathcal{A}</th>
<th>Compressor</th>
<th>Code word C \in \mathcal{C} = {0, 1}^l</th>
<th>Decompressor</th>
<th>Decompressed message X \in \mathcal{A}</th>
</tr>
</thead>
<tbody>
<tr>
<td>X = A</td>
<td></td>
<td>01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X = B</td>
<td></td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X = C</td>
<td></td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X = D</td>
<td></td>
<td>00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Want: \#X = 4, \#messages \Rightarrow \text{need } l = 2

In general: \[2^l \geq \#\mathcal{A} \Rightarrow l \geq \log_2(\#\mathcal{A})\]

Can we do better? Imagine some messages are more frequent than others...

<table>
<thead>
<tr>
<th>Code I</th>
<th>Code II</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Sunshine 44%</td>
<td>10</td>
</tr>
<tr>
<td>B Rain 55%</td>
<td>0</td>
</tr>
<tr>
<td>C Snow 0.1%</td>
<td>110</td>
</tr>
<tr>
<td>D Hurricane 0.01%</td>
<td>111</td>
</tr>
</tbody>
</table>

Can both be decoded nicely? Yes: D, B, A, C.

\[l = 2\] is not always possible.
Code I: lossless, $\text{average length} = 1.46$
Code II: lossy! $\text{error} = 0.01\%$, $\text{average length} \approx 1.45$

How to do even better? Look at blocks of messages!

Let Shannon: Optimal rate of compression is $\approx 1.06$ bits/message $= \text{entropy of source} \approx \log_2 N$.

Communicating over Noisy Channels

Examples of noisy channels & how to avoid:
* Scratch on Blu-ray disk
* Loud party
* Nail varnish crumpled
* Bad signal
* Bit flip on hard disk

Mathematical model:

\[
\begin{align*}
\text{input} & \rightarrow \text{channel} \rightarrow \text{output} \\
p(\text{output} | \text{input})
\end{align*}
\]

E.g. binary symmetric channel:

\[
\begin{align*}
0 & \xrightarrow{1-f} 0 \\
1 & \xrightarrow{1-f} 1 \\
0 & \xrightarrow{f} 1 \\
1 & \xrightarrow{f} 0
\end{align*}
\]

\[
p(1|0) = p(0|1) = f \\
p(0|0) = p(1|1) = 1-f
\]

\[f = \text{probability of bit flip}\]

Assume we know $f$!!!

How to reduce error? Introduce redundancy by encoding message!

WANT: \(S = \hat{S}\) with high probability

Repetition code $R_2$:

\[
\begin{array}{c|ccc}
S & 000 & 001 & 010 & 011 & 100 & 101 & 110 & 111 \\
\hline
x = x_1x_2x_3 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1
\end{array}
\]

\*decode: \[
\begin{align*}
\hat{S} &= \text{majority vote} \\
y &= y_1y_2y_3
\end{align*}
\]

\[y = 000, 001/010/100, 011/101/110, 111\]

\[S = 0 \]

\[\hat{S} = 0\]
* Analysis: Can deal with \( \leq 1 \) bit flip

\[
\text{Pr}_{\text{error}} = \Pr(2 \text{ or } 3 \text{ bit flips}) = 2 \cdot f^2 (1-f) + f^3 \approx 2f^2 \quad \text{if } f \text{ small}
\]

\( e.g., f = 10\% = 0.1 \), \( \Pr_{\text{error}} = 0.028 \approx 0.03 = 3\% \)

* rate = \( \frac{\# \text{ source msg bits}}{\# \text{ transmitted bits}} = \frac{1}{J} \)

Ex: Show that the decoder is optimal (if \( f \leq 50\% \)). Discuss \( f = 50\% \).

What if we repeat \( N > 3 \) times?

\[
\Pr_{\text{error}} = \Pr(C = \geq \frac{N}{2} \text{ bit flips}) \approx \sum_{k=\frac{N}{2}}^{N} \binom{N}{k} f^k (1-f)^{N-k} \approx 2^{N} f^N (1-f)^{N/2}
\]

\( e.g., f = 10\% \): \( \Pr_{\text{error}} \approx 0.6^N \)

It seems like rate \( \to 0 \) if \( \Pr_{\text{error}} \to 0 \)

**But Shannon can do better?** (See below)

How can we find more & better codes?

**Black codes:**

Encode more than one symbol at a time!

\( (7,4) - \text{Hamming code}: \)

\[
x_1 = s_1 \ldots \quad x_4 = s_4
\]

\( x_5, \ldots, x_7 \) chosen such that sum in each circle even

(“parity bits”)
Any two codewords differ by 3 or more bits!

How to decode?

1. Compute parities in all three circles: $z_i = y_1 \oplus y_2 \oplus y_3 \oplus y_5 \pmod{2}$
2. If at least one $z_i \neq 0$:
   - Flip unique bit that is only in circles with $z_i \neq 0$

<table>
<thead>
<tr>
<th>$z$ = $z_1 \  z_2 \  z_3$</th>
<th>000</th>
<th>001</th>
<th>010</th>
<th>011</th>
<th>100</th>
<th>101</th>
<th>110</th>
<th>111</th>
</tr>
</thead>
<tbody>
<tr>
<td>flipped bit</td>
<td>/</td>
<td>Y_4</td>
<td>Y_6</td>
<td>Y_5</td>
<td>Y_4</td>
<td>Y_6</td>
<td>Y_2</td>
<td>Y_3</td>
</tr>
</tbody>
</table>

$$\Rightarrow P_{\text{block error}} \leq \Pr(\geq 2 \text{ bit flips}) \sim \binom{7}{2} f^2 (1-f)^5 = 21 f^2$$

$$P_{\text{bit error}} = \frac{4}{7} \sum_{k=1}^{7} \Pr(S_k \neq S_k) \sim 9 f$$

**Shannon:** For $f = 10\%$, can reliably send at optimal rate $0.53$ P

(but...)

**Thursday:** Probability theory recap + entropy (towards compression)