Roadmap

- Types of Channels
- Jointly Typical Sequences
- Channel Coding Theorem
- Zero-Error Codes
- Fano's Inequality
- Converse to the Channel Coding Theorem
- Hamming Codes
Basic Idea of the Channel Coding Theorem

- Achievability of channel capacity
- Shannon used a number of new ideas to prove that information can be sent reliably over a channel at all rates up to the channel capacity:
  - Allowing an arbitrarily small but nonzero probability of error
  - Using the channel many times in succession, so that the law of large numbers comes into effect
  - Calculating the average of the probability of error over a random choice of codebooks
- There are many versions of the proof of the channel coding theorem.
- We will be looking at one of the simplest proofs, which uses properties of typical sequences.
  - This proof is based on – but not present in – Shannon’s original work, and was developed much later.
In the proof, we will *decode* by joint typicality.

We look for a codeword that is jointly typical with the received sequence.

If we find a unique codeword satisfying this property, we declare that word to be the transmitted codeword.

With high probability, the transmitted codeword and the received sequence are jointly typical, since they are probabilistically related.

The probability that any other codeword looks jointly typical with the received sequence is $2^{-n I(X;Y)}$.

Hence, if we have fewer then $2^{n I(X;Y)}$ codewords, then with high probability there will be no other codewords that can be confused with the transmitted codeword, and the probability of error is small.
Theorem

For a discrete memoryless channel, all rates below the capacity $C$ are achievable. That is, for every rate $R < C$, there exists a sequence of $(2^{nR}, n)$ codes with the maximum probability of error $\lambda^{(n)} \rightarrow 0$ as $n \rightarrow \infty$.

Conversely, any sequence of $(2^{nR}, n)$ codes with $\lambda^{(n)} \rightarrow 0$ must have $R \leq C$.

Proof.

The proof outline consists of showing

- Achievability
  - Setup
  - Error analysis
  - Construction of a code with arbitrarily low maximum error probability
Setup: code generation

- Fix the source distribution \( p(x) \)
- Generate a \( (2^{nR}, n) \) code at random according to the distribution \( p(x) \). That is, generate \( 2^{nR} \) codewords independently according to
  \[
  p(x^n) = \prod_{i=1}^{n} p(x_i)
  \]
- The \( 2^{nR} \) codewords can be represented as a matrix
  \[
  C = \begin{pmatrix}
  x_1(1) & x_2(1) & \cdots & x_n(1) \\
  \vdots & \vdots & \ddots & \vdots \\
  x_1(2^{nR}) & x_2(2^{nR}) & \cdots & x_n(2^{nR})
  \end{pmatrix}
  \]
- Each entry in \( C \) is generated i.i.d. according to \( p(x) \). Therefore, the probability that a particular code \( C \) is generated is
  \[
  Pr(C) = \prod_{w=1}^{2^{nR}} \prod_{i=1}^{n} p(x_i(w))
  \]
A random code $C$ is generated according to $p(x)$. It is then revealed to both sender and receiver. Both sender and receiver also know the channel transition matrix $p(y|x)$. A message $W$ is chosen according to a uniform distribution, i.e.,

$$Pr(W = w) = 2^{-nR} \quad \forall w \in \{1, 2, \ldots, 2^{nR}\}.$$ 

The $w$th codeword $X^n(w)$ (corresponding to the $w$th row of $C$) is sent across the channel. The receiver receives a sequence $Y^n$ according to the distribution

$$p(y^n|x^n(w)) = \prod_{i=1}^{n} p(y_i|x_i(w)).$$

Now, the receiver guesses which message was sent. This is done using *jointly typical decoding*.
In jointly typical decoding, the receive declares that the index $\hat{W}$ was sent if the two following conditions are met:

- $(X^n(\hat{W}), Y^n)$ is jointly typical, and
- There is no index $W' \neq \hat{W}$ such that $(X^n(W'), Y^n)$ is jointly typical.

If such a $\hat{W}$ is not found, error is declared.

In such cases, we may implement the system to make the receive $r$ output a dummy index like $-1$.

There is a **decoding error** if $\hat{W} \neq W$.

Let $\mathcal{E}$ denote the event that $\hat{W} \neq W$. 
Error Analysis Outline

- Instead of calculating the probability of error for a single code, we calculate the average over all codes generated at random according to the distribution

\[ Pr(C) = \prod_{w=1}^{2^n R} \prod_{i=1}^{n} p(x_i(w)) \]

- In our setup, the average probability of error does not depend on the index being sent over the channel.

- Also, note that there are two different sources of error when we use jointly typical decoding:
  1. the output $Y^n$ is not jointly typical with the transmitted codeword, or
  2. there is some other codeword that is jointly typical with $Y^n$.

- In the last lecture, we showed that the probability that the transmitted codeword and the received sequence are jointly typical goes to 1 (the ‘Joint AEP theorem’).
Error Analysis Outline

- The probability that the transmitted codeword and the received sequence are jointly typical goes to 1 (the Joint AEP theorem from last lecture).
- For any other codeword, the probability that it is jointly typical with $Y^n$ is approximately $2^{-nI(X;Y)}$.
- Therefore, we can use approximately $2^{nI}$ codewords and still have a low probability of error.
  - This argument will be extended to construct a code with a low maximum probability of error.

(details of error analysis done in class)