Today

- COMPRESSION
  - SINGLE SHOT: SHANNON + HUFFMAN
  - UNIVERSAL: (START LEMPEL-ZIV)
  - MARKOV SOURCES + HIDDEN MARKOV MODELS

Up To Now: Asymptotic 'Compression
Compress n i.i.d. variable $X_1, \ldots, X_n$

**Single-Shot:**
- Sender + Receiver know $P_x$; design $E, D$
- Sender gets $X \sim P_x$
- Must send $E(x) \in \Sigma_01^*$ to receiver
- Receiver decodes $\hat{X} = D(E(x))$
- Want $\hat{X} = X$ & minimize $\mathbb{E} \left[ I(E(x)) \right]$
Two Solutions

1. See [Shannon] (Note this is in contrast to his solution for asymptotic compression)

Define: Let \( l_i = \sum \log \frac{1}{P_i} \quad S^2 = S_1, \ldots, S_S \)

Design (by Kraft's inequality), \( E : \mathbb{Z} \rightarrow \{0, 1\}^* \)

and \( E(i) \in \{0, 1\}^* \)

- send \( E(X) \).

Performance:

\[
E \left[ 1 \cdot E(X) \right] = \mathbb{E} \left[ \sum \log \frac{1}{P_x} \right]
\]

\[
\leq \mathbb{E} \left[ \log \frac{1}{P(x)} + 1 \right]
\]

\[
= 1 + \mathbb{E} \left[ \log \frac{1}{P(x)} \right]
\]

\[
= 1 + H(x)
\]

Optimal to within 1 bit.

(Note: follow from Shannon's asymptotic theorem that no prefix-free scheme achieves \( < H(x) \).)
Huffman

Recursive algorithm \((P_1 \ldots P_n)\)
- Sort \(P_1 \geq P_2 \geq \ldots \geq P_n\)
- Combine \(P_{n-1} \) & \(P_n\) to form new element \(Q_i \) with \(Q_i = P_i\)
  except \(Q_{n-1} = P_n + P_n\)
- \(E'\) Huffman code \((Q_1 \ldots Q_n)\);
- \(E(i) = E'(i)\) except \(E(n-1) = 0 \cdot E'(n-1)\) & \(E(n) = 1 \cdot E'(n-1)\)

Theorem: Huffman is optimal \(\forall P\).

Proof:
Ingredients 0: if \(P_1 \geq P_2 \geq \ldots \geq P_n\) then \(e_1 \leq e_2 \leq \ldots \leq e_n\)
Proof: Else swap \(e_i, e_j\) & get cost that is no more \&

2. \(e_{n-1} = e_n\): Proof look at tree. Right leaf at depth \(e_n\) must have siblings.

3. Wlog \(P_{n-1} \leq P_n\) are siblings.
"Optimal Substructure":

\[ q_1 = p_1, \ldots, q_{n-2} = p_{n-2}, q_{n-1} = p_{n-1} + p_n \]

Optimal tree for \( p_1, \ldots, p_n \) with siblings \( p_{n-1} \& p_n \)
is also optimal tree for \( q_1, \ldots, q_{n-1} \) with parent \( p_n \)
as leaf.

Universal Coding: How do gzip, zip, ... work?

- Single file to compress; No distribution?
- Produces something smaller!

Central Algorithm: Lempel-Ziv + variations
- Works well empirically
- Has a "Theorem"!

Algorithm: Input \( w \in \Sigma^* \) \( \Sigma = \text{finite alphabet} \)
- Break \( w \) into \( S_0 \circ S_1 \circ S_2 \circ S_3 \ldots \circ S_m \)
  - where \( \circ \) denotes concatenation of strings
  - \( S_0 = \text{empty string} \)
  - \( S_i = S_{j_i} \circ b_i \) where \( j_i < i \)
    - \( b_i \notin \Sigma \)
- Compression of \( w \) = Encoding of \( j_1, j_2, \ldots, j_m \)
in any reasonable prefix tree
De "compressing": \((j_1, \ldots, j_m)\)
\(b_1, \ldots, b_m\)
- Can compute for \(i = 1 \rightarrow m\)
  \(S_i = S_{j_i} \circ b_i\)
- Output \(S_1 \circ S_2 \circ \ldots \circ S_m\)

How well does this perform?

**Case 1**: \(W_1, \ldots, W_n\) i.i.d. over \(P\) distribution over \(\Sigma^*\).

**Claim**: Expected length of compression = \((1 + o(1)) \cdot H(P)\)

(Proof later)

- More impressive

**Case 2**: \(W_1, \ldots, W_n\) drawn from "Hidden Markov Model".

\[\Rightarrow\] Expected length of compression = \((1 + o(1)) \cdot H(M)\)

Entrophy of chain

Need to define:
1. "Hidden Markov Model"
2. Entropy \(H(M)\) of Hidden Markov Model.
Markov Chain

- \( Z_1, Z_2, \ldots, Z_n, \ldots \) form a (time-invariant) Markov Chain if
  \[
  \Pr[\{Z_i \mid Z_1, \ldots, Z_{i-1}\}] = \Pr[\{Z_i \mid Z_{i-1} = a_{i-1}\}]
  = \Pr[\{Z_1 \mid Z_0 = a_0\}]
  \]

- Conditioned on last "state" \( Z_{i-1} \), distribution of \( Z_i \) doesn't depend on past.

- If \( Z_i \in \Omega = \{1, \ldots, k\} \),

  - Markov Chain specified by matrix: \( M \in \mathbb{R}^{k \times k} \)
    \[
    M_{ij} = \Pr[Z_2 = i \mid Z_1 = j]
    \]

  - Pictorially:

```
  \begin{array}{ccc}
    2 & \rightarrow & 3 \\
    \downarrow & \downarrow & \downarrow \\
    3 & \rightarrow & 1 \\
  \end{array}
```

  - "et cetera."
Entropy (Rate) of Markov Chain: $M$

$$\lim_{n \to \infty} H(Z_n | Z_{n-1}, \ldots, Z_1)$$

$$= \lim_{n \to \infty} H(Z_n | Z_{n-1})$$

$$= H(Z_2 | Z_1). \quad \left[ \text{corr. \# to \# bits needed to describe } Z_n \text{ given } Z_1 \ldots Z_{n-1} \right]$$

Simplifying setting: $M$ is irreducible (path from every state to every other state) and a periodic (gcd (cycle lengths) = 1).

- Assume if so MC has a stationary distribution $\pi = \pi(M)$
  - i.e., if $Z_n \sim \pi$ then $Z_{n+1} \sim \pi$.
- Assume $Z_1 \sim \pi$.
- Theorem [Shannon]: Can compute entropy rate of Markov Chain $H(M)$ given $M$. 

Hidden Markov Models

Specified by:
1. Markov Chain \( M \) on \( \mathbb{R} \) states = \( \Gamma \)
2. Distributions \( \{P_x\}_{x \in \Gamma} \)

Generates sequence \( X_1, \ldots, X_n \) as follows:
1. Pick \( Z_1 \sim \Pi \) (stationary prob. of \( M \))
2. Generate \( Z_i | Z_{i-1} \), according to \( M \)
3. Generate \( X_i \sim P_{Z_i} \)

Example:

![Diagram of Hidden Markov Model]

- Entropy of HMM \( M \)

\[
H(M) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} H(X_i | X_{i-1}, \ldots, X_1)
\]

- Limit exists! \( \forall n \ H(X_n | X_{n-1}, \ldots, X_1) \leq H(X_{n-1} | X_{n-2}, \ldots, X_1) \)

Next Lecture: Analysis of L-2 on HMM.