Today:

- Channel Coding
- Definitions
- Binary Symmetric Channel
- General Channels

Next few lectures: Error Correction (with "random errors")

General Channel of Communication (Memoryless)

\[ X \rightarrow \square \rightarrow Y \]

- Given by \( P_{Y|X} \) given by \( P_x \times P_y \) matrix
- \( P_{Y|X}(\beta|\alpha) = P[Y = \beta | X = \alpha] \)

Non i.i.d. uses of channel
How much "information" per use?
Simple Example

$\text{BSC}(p)$ [Binary Symmetric Channel]

\[
x \rightarrow y = x \text{ w.p. } 1-p = 1-x \text{ w.p. } p
\]

"Capacity": rate at which information can be pushed through.

Formally: encoding + decoding functions $E_n, D_n$ achieve rate $R$ if $E_n, D_n$ with error $\epsilon$:

1. $E_n: \{0,1\}^R \rightarrow \mathbb{Z}_x^n$
2. $D_n: \mathbb{Z}_y^n \rightarrow \{0,1\}^R$
3. $Pr[\text{Decoding failure}] = Pr_{m \sim \text{Unit}(\{0,1\}^R)} \left[ D(y) \neq m \right] = \epsilon$

Capacity of channel $P_{y|x}$

\[
R = \sup \left\{ \lim_{n \to \infty} \frac{1}{n} \sum_{m \in \{0,1\}^R} E_{E_n, D_n}(0, m) \right\}
\]
Connections to Information Theory

1. Capacity \( (P_{x|y|x}) = \max_{P_x} \sum I(x; y) \)

   [operational view of Information, theorem to be proved later]

2. Information Theory gives "best" algorithm + codes!!

Today 0

\[ \text{Special Case: BSC}(p) \]

Capacity = \( 1 - \# h(p) \) \[ P_x = \text{Bern}(\frac{1}{2}) \].

\[ = H(\text{Bern}(\frac{1}{2})) - H(Y|X) \]

\[ = 1 - H(\text{Bern}(p)) \]

\[ = 1 - h(p). \]

Proof of 1 for BSC(p).

0. \( R \geq 1 - h(p) - \epsilon \) \( k = (R \cdot n) \)

Pick \( E_n: \Sigma_{0,1}^r \rightarrow \Sigma_{0,1}^n \) at random

\( D_n = \max \text{ likelihood decoding} \).
Lemma:

\[ \Pr_{E_n, m, Y \mid E_n(m)} \left[ D_n(Y) + m \right] \leq \varepsilon \]

\[ \Rightarrow \exists E_n \left[ \Pr_{m, Y \mid E_n(m)} \left[ \cdot \right] \leq \varepsilon \right] \]

Proof: Error events

1. \[ \Pr_{Y \mid X} \left[ \Delta(Y, E_n(m)) \geq (p + \varepsilon) \right] \leq \exp(-\varepsilon^2 n) \]

2. \[ \Pr \left[ \forall \exists m' \neq m; E'_n, \Delta(E'_n(m'), Y) \leq (p + \varepsilon) n \right] \]

\[ \leq 2^k \left( \frac{n}{(p + \varepsilon)n} \right) \cdot \frac{1}{2^n} \]

\[ \approx 2^k \cdot 2^{-n} \cdot 2^{-n} \]

\[ = 2 \exp(-\varepsilon n). \]

If \( E_1 \) or \( E_2 \) don't happen, then cheating right.
General Channel

Fix $P_x$.

- Pick $E_n : \{0,1\}^n \to \Omega^2_x$ by picking $E_n(m)_i$ i.i.d. $\sim P_x$ ind. for all $(m,i)$.

- Decoding ($Y$)

  If there exists unique $X \in \mathbb{S}^n_x$ such that for $X = E(m)$,

  \[ \Pr[X] = \frac{1}{2^{H(R)n}} \]

  \[ \Pr[xy] \approx \frac{1}{2^{H(P_{xy})n}} \]

  $P_{xy}$ is typical.

  Output $m$.

  Else error.
Analysis

(i) Two types of errors

(ii) \( X \) not typical, \( Y \) not typical, \( (X,Y) \) not jointly typical

(iii) \( (E(m'), Y) \) jointly typical, for some \( m' \) \( \neq m \).

\[ \Pr[{\text{(E)}}] \to 0 \quad \text{by AEP} \]

2. \( E_2 \) ? Key + useful lemma.

Lemma: Let \( P, Q \) be distributions over \( \Sigma^* \).

\[ \Pr[\Xi \text{ typical for } Q^n] \leq 2^{-D(Q||P) \cdot n} \]

In our case

\( (E(m'), Y) \) drawn from \( P_x^n \times P_y^n \)

\[ \Pr[(E(m'), Y) \text{ typical for } P_{xy}^n] \leq 2^{-D(P_{xy} || P_x \times P_y) \cdot n} = 2^{-I(Y; X) \cdot n} \]
\[ \Rightarrow \text{Can take union bound over } 2^m \text{ } \forall m \text{'s.} \]

\[ \Rightarrow \text{Rate } \geq I(x; y)! \]

Can optimize over \( P_x \)
to get

\[ \text{Capacity } \geq \sup_{P_x} \left\{ I(x; y) \right\} \]

Converse Coding Theorem:

\[ V_1: \text{if Pr[decoding failure]} \rightarrow 0 \text{ then } R \leq \text{Capacity} \]

\[ V_2: \text{for BSC}(p): \text{if Rate } = \sup_{P_x} \left\{ I(x; y) \right\} + \epsilon \text{ then Pr[decoding failure]} \geq 1 - \exp(-n) \]

[\( V_2 \text{ much stronger quantitatively; but this being true only for BSC}(p) \).]
Proof of V1: (using Fano's Inequality)

have \( m \rightarrow X^n \rightarrow Y^n \rightarrow \hat{m} \) - a Markov chain.

1. \( I(X^n; Y^n) \leq n \cdot \sup_{p_X} \{ I(X; y) \} \)

2. \( H(m) = nR \geq H(m | \hat{m}) + I(m ; \hat{m}) \)
   \[ \leq H(m | \hat{m}) + I(X^n; Y^n) \quad [DP] \]
   \[ \leq H(m | \hat{m}) + nC \]

need to bound \( H(m | \hat{m}) \)

Fano:
\[ H(m | \hat{m}) \leq h(Pr[m \neq \hat{m}]) + Pr[m \neq \hat{m}] \cdot nR \]
\[ \leq 1 + o(nR) \]

\[ \Rightarrow nR (1 - o(1)) \leq nC \]
\[ R (1 - o(1)) \leq C \]
\[ \Rightarrow R \leq C \quad \text{in the limit} \]

V2: Exercise / Ref.