Information Theory
Lecture 7
Definition and Computation of Channel Capacity

A. Basic Definitions (Section 7.5) and the Channel Capacity Theorem (Thm 7.71 pg 200)
B. Properties of capacity (Section 7.3 and Computation of Channel Capacity (Section 7.1)
C. The use of symmetry (Section 7.2) and in particular cyclic symmetry (not in the book)

Part A:
Basic Definitions (Section 7.5)

The Channel Capacity Theorem
(Theorm 7.71 pg 200)

Channel Model

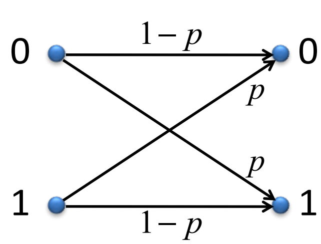
- A discrete time, discrete alphabet, memoryless channel is a set of conditional distributions on the output Y given the input X:

$$\{\mathcal{X}, p(y|x), \mathcal{Y}\}$$

$$X \longrightarrow \boxed{p(y|x)} \longrightarrow Y$$

Example: Binary Symmetric Channel (BSC)

$$\{\mathcal{X}, p(y|x), \mathcal{Y}\} = \{\{0,1\}, p(y|x), \{0,1\}\}$$



$$p(y=0 \mid x=0) = 1-p$$

$$p(y=1 \mid x=0) = p$$

$$p(y=0 \mid x=1) = p$$

$$p(y=1 \mid x=1) = 1-p$$

Block Code

- An (M, n) block code for $(\mathcal{X}, p(y|x), \mathcal{Y})$ uses *n* symbols from $\mathcal{X}$ to transmit one of M messages W_i, $i \in \{1, \ldots, M\}$.

$$W \longrightarrow \boxed{\text{encoder}} \xrightarrow{X^n(W)} \boxed{p(y|x)} \xrightarrow{Y^n} \boxed{\text{decoder}} \longrightarrow \hat{W}(Y^n)$$

- Set of possible $X^n(W)'s$ is called the codebook.
- Rate of (M,n) code : $R = \dfrac{\log M}{n}$
- An (M,n) code is a $(2^{nR}, n)$ code.

Achievable rate

- A rate R is achievable for $(\mathcal{X}, p(y|x), \mathcal{Y})$ if there is a sequence of $(2^{nR}, n)$ codes such that

$$\max_i p(\hat{W} \neq W_i \mid W = W_i) \to 0, \quad \text{as } n \to \infty$$

Channel Capacity Definition

- For channel ($\mathcal{X}$, $p(y|x)$, $\mathcal{Y}$)

$$C = \max_{p(x)} I(X;Y)$$

is the channel capacity.

Channel Capacity Theorem

- 1) all rates $R < C$ are achievable.
- 2) No rate $R > C$ is achievable.
  - We'll prove this theorem next lecture.
  - Today we'll get comfortable with the properties of C and learn how to compute it.

Part 7B:

Properties of Capacity (Section 7.3)

Computation of Channel Capacity (Section 7.1)

Properties of C (Section 7.3)

- 1) $C \geq 0$ ...since $I(X;Y) \geq 0.$
- 2 & 3) $C \leq \log|\mathcal{X}|, \quad C \leq \log|\mathcal{Y}|$

$$\boxed{\begin{aligned} C &= \max I(X;Y) \\ &= \max H(X) - H(X|Y) \\ &\leq \max H(X) \\ &= \log|\mathcal{X}| \end{aligned}}$$

- 4) $I(X;Y)$ is a continuous function of $p(x)$
- 5) Recall $I(X;Y)$ is a concave function of $p(x)$ for fixed $p(y|x)$. Theorem 2.7.4 on page 33.

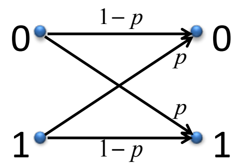
How to find C

- Finding C is a convex optimization problem:
  - EE 236B provides general techniques, we'll give a specific algorithm later in the course called "Blahut-Arimoto".

- For now we compute C using a few "tricks" that work for lots of simple channels.

- Trick #1: Find an upper bound, and then achieve it.

Capacity upper bound for BSC



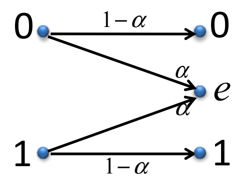
$$
\begin{aligned}
I(X;Y) &= H(Y) - H(Y \mid X) \\
&= H(Y) - \sum_x p(x) \underbrace{H(Y \mid X = x)}_{H(p)} \\
&= H(Y) - H(p) \\
&\le 1 - H(p) \qquad \text{...achieved when } H(Y) = 1
\end{aligned}
$$

Achieving upper bound for C in BSC

- $H(Y) = 1$ when y=0, y=1 are equally likely.
- To achieve this, set x=0, x=1 to be equally likely.
- $C = 1 - H(p)$ with maximizing $p(x) = \begin{cases} 1/2 & x = 0 \\ 1/2 & x = 1 \end{cases}$

Binary Erasure Channel

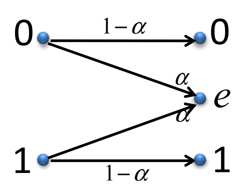
E



$$I(X;Y) = H(Y) - H(Y \mid X)$$
$$= H(Y) - H(a)$$

- At this time it looks a lot like BSC, but $H(Y) < \log 3$ in general.

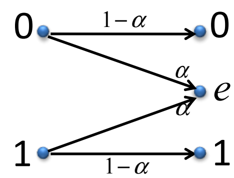
BEC: Introducing E



- Define a new random variable E as a function of Y

$$E = \begin{cases} 1 & \text{when } y = e \\ 0 & \text{otherwise} \end{cases} .$$

BEC: A tighter bound on $H(Y)$



$$
\begin{aligned}
H(Y) &= H(Y,E) \qquad \ldots \text{in general, } H\big(Y\big)=H\big(Y,f\big(Y\big)\big) \\
&= H(E)+H(Y\mid E) \\
&= H(\alpha)+(1-\alpha)H(X) \\
&\le H(\alpha)+(1-\alpha)
\end{aligned}
$$

achievable with $p(x)=\begin{cases} 1/2 & x=0 \\ 1/2 & x=1 \end{cases}$

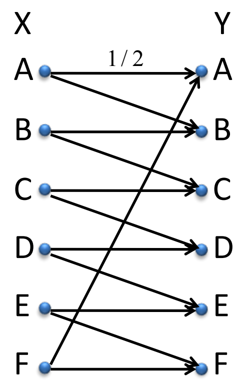
Binary Erasure Channel (cont.)

- $$
\begin{aligned}
I(X;Y) &= H(Y)-H(\alpha) \\
&\le \big[H\big(\alpha\big)+1-\alpha\big]-H\big(\alpha\big) \\
&\le 1-\alpha
\end{aligned}
$$

- $C=1-\alpha$ with maximizing $p\big(x\big)$ distribution

  $p(x)=\begin{cases} 1/2 & x=0 \\ 1/2 & x=1 \end{cases}$

Noisy typewriter



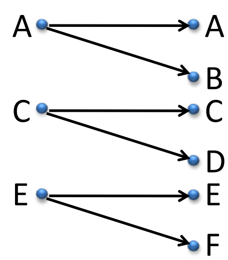
- Example:
  - Send A and Y is equally likely to be A or B

Capacity of noisy typewriter

- $$\begin{aligned} I(X;Y) &= H(Y) - H(Y \mid X) \\ &= H(Y) - 1 \\ &\leq \log|\mathcal{Y}| - 1 \end{aligned}$$
- Achieved with equality by a uniform distribution on $\mathcal{X}$.

Another good input distribution

- Another distribution allows zero-error transmission with a finite blocklength.



Zero-error Capacity

- "Zero-error capacity" always ≤ C is the rate at which

$$\max_i p(\hat{W} \neq W_i \mid W = W_i) = 0 \quad \text{for a finite } n$$

Part C:

The Use of Symmetry (Section 7.2)

Cyclic Symmetry (not in the book)

Capacity-achieving distribution

- In all of our simple examples, the capacity achieving distribution is a uniform over the input.
- Cyclic Symmetry and Weak Symmetry are two ways to know a uniform distribution will achieve capacity that can be applied to our simple examples and some more complicated ones.

Cyclic Symmetry

- Define a channel to have cyclic symmetry if the mutual information is invariant to cyclic shifts in the input distribution.

- Suppose

$$p(x) = \begin{cases} p_1 & x = a \\ p_2 & x = b \\ p_3 & x = c \end{cases} \xrightarrow{\text{a cyclic shift of p(x)}} p^{(1)}(x) = \begin{cases} p_3 & x = a \\ p_1 & x = b \\ p_2 & x = c \end{cases}$$

Cyclic Symmetry

- The BSC, BEC and noisy typewriter all have cyclic symmetry.

Theorem

- If a channel has cyclic symmetry, the uniform distribution achieves capacity.
- Proof of theorem follows on the next three slides.

Average of cyclic shifts is the uniform.

- Select any input distribution $p(x)$ and note that the average of all its cyclic shifts is the uniform $u(x)$:

$$u(x) = \frac{1}{|\mathcal{X}|}\sum_{i=1}^{|\mathcal{X}|} p^{(i)}(x) \quad .$$

Recall Jensen's inequality

– Jensen (with inequality switched because $I(X;Y)$ is <u>concave</u> in $p(x)$)

$$E[f(x)] \le f(E[x]) \quad \text{for } f(\bullet) \text{ a concave function.}$$

Apply Jenson to show optimality of $u(x)$.

$$I_{p(x)}(X;Y) = I_{p^{(i)}(x)}(X;Y)$$

$$I_{p(x)}(X;Y) = \sum_{i=1}^{|\mathcal{X}|} \frac{1}{|\mathcal{X}|} I_{p^{(i)}(x)}(X;Y) \quad \text{by cyclic symmetry}$$

$$\le I_{u(x)} \quad \text{by Jensen}$$

– Hence $u(x)$ maximizes I.

Weak symmetry

- $$p(y\,|\,x) = \begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.5 & 0.3 & 0.2 \\ 0.2 & 0.5 & 0.3 \end{bmatrix}$$

$$\begin{array}{cccc} & y_1 & y_2 & y_3 \\ x_1 & 0.3 & 0.2 & 0.5 \\ x_2 & 0.5 & 0.3 & 0.2 \\ x_3 & 0.2 & 0.5 & 0.3 \end{array}$$

- Entry in x^{th} row and y^{th} column is $p(y|x)$.
- $p\left(y_2 \,|\, x_3\right) = 0.5$
- A channel is weakly symmetric if all the rows are permutations of each other and the column sums are equal.

Weak symmetry

- For a weakly symmetric channel,

$$C = \log|\mathcal{Y}| - H(r)$$

pmf of any row

which is achievable by a uniform distribution.

- $$\begin{aligned} I(X;Y) &= H(Y) - H(Y\,|\,X) \\ &= H(Y) - H(r) \\ &\le \log|\mathcal{Y}| - H(Y\,|\,X) \end{aligned}$$