Information Capacity of the BSC Permutation Channel

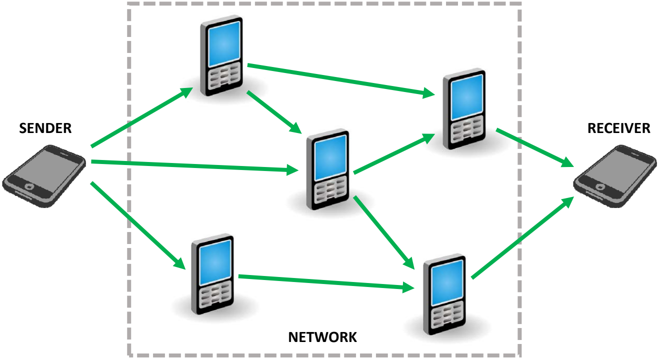
Anuran Makur

EECS Department, Massachusetts Institute of Technology

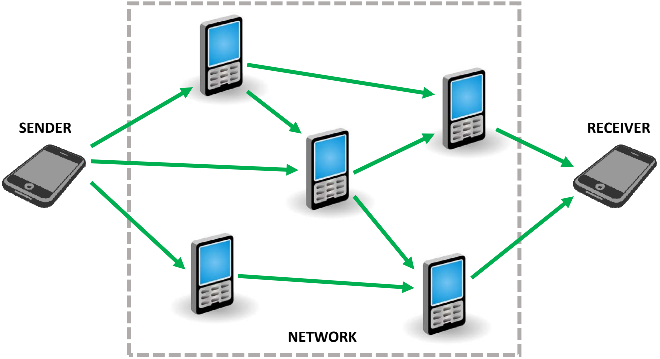
Allerton Conference 2018

Outline

Motivation: Point-to-point Communication in Networks

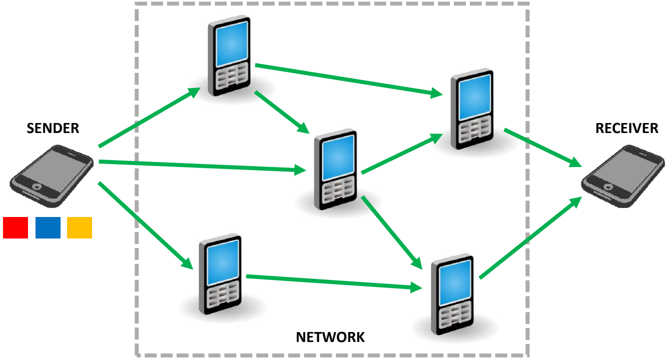


Motivation: Point-to-point Communication in Networks



Model communication network as a channel

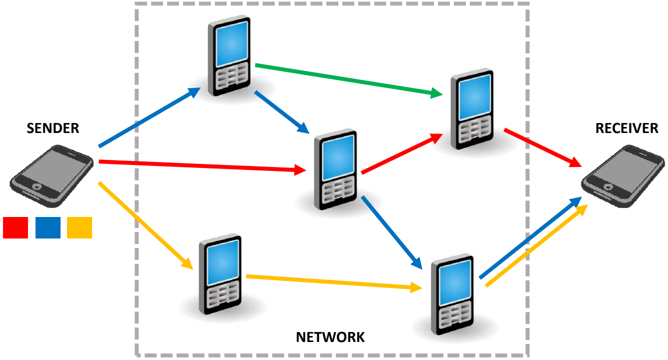
Motivation: Point-to-point Communication in Networks



Model communication network as a channel:

- Alphabet symbols = all possible L-bit packets $\Rightarrow$ 2^L input symbols

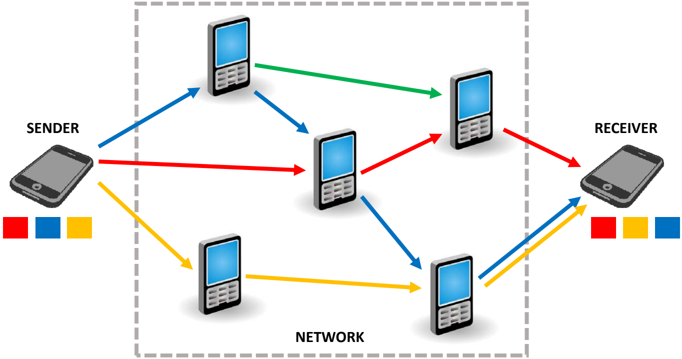
Motivation: Point-to-point Communication in Networks



Model communication network as a channel:

- Alphabet symbols = all possible L-bit packets
- multipath routed network or evolving network topology

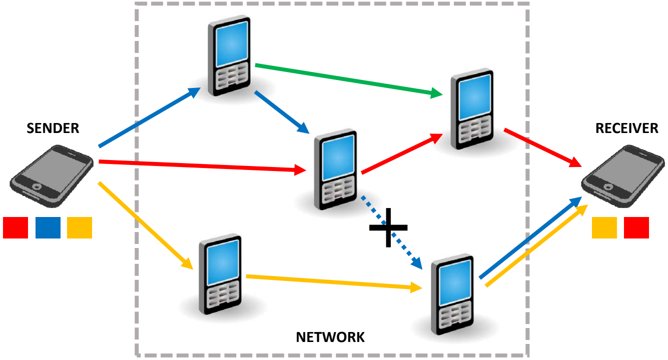
Motivation: Point-to-point Communication in Networks



Model communication network as a channel:

- Alphabet symbols = all possible L-bit packets
- multipath routed network $\Rightarrow$ packets received with transpositions

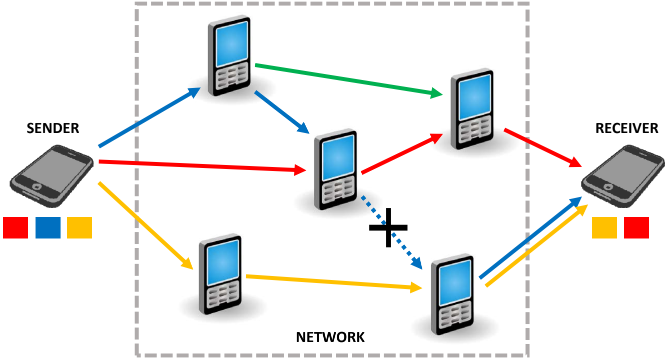
Motivation: Point-to-point Communication in Networks



Model communication network as a channel:

- Alphabet symbols = all possible L-bit packets
- multipath routed network $\Rightarrow$ packets received with transpositions
- packets are impaired (e.g. deletions, substitutions)

Motivation: Point-to-point Communication in Networks

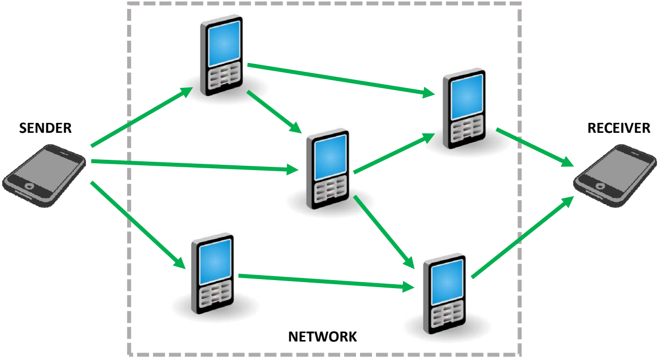


Model communication network as a channel:

- Alphabet symbols = all possible L-bit packets
- multipath routed network ⇒ packets received with transpositions
- packets are impaired ⇒ model using channel probabilities

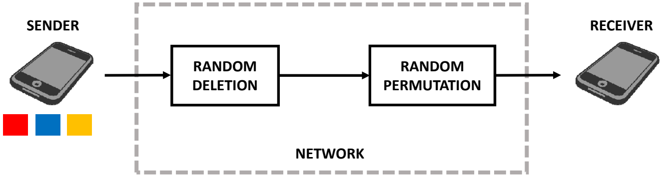
Example: Coding for Random Deletion Network

Consider a communication network where packets can be dropped:



Example: Coding for Random Deletion Network

Consider a communication network where packets can be dropped:

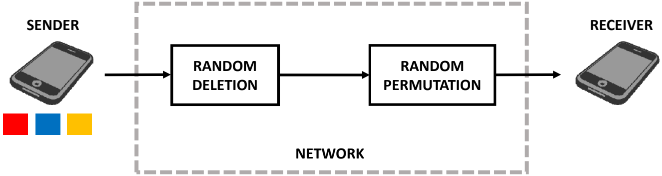


Abstraction:

- n-length codeword $=$ sequence of n packets

Example: Coding for Random Deletion Network

Consider a communication network where packets can be dropped:

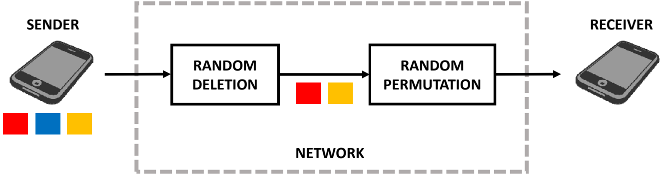


Abstraction:

- n-length codeword = sequence of n packets
- Random deletion channel: Delete each symbol/packet of codeword independently with probability $p \in (0, 1)$

Example: Coding for Random Deletion Network

Consider a communication network where packets can be dropped:

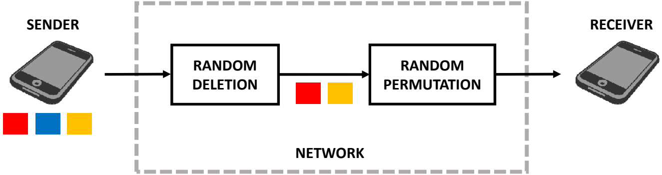


Abstraction:

- n-length codeword = sequence of n packets
- Random deletion channel: Delete each symbol/packet of codeword independently with probability $p \in (0, 1)$

Example: Coding for Random Deletion Network

Consider a communication network where packets can be dropped:

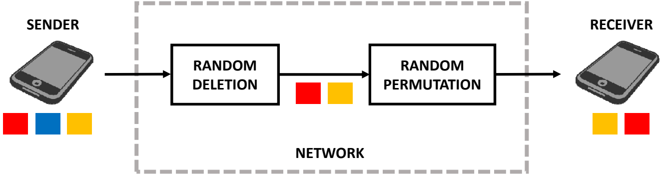


Abstraction:

- n-length codeword = sequence of n packets
- Random deletion channel: Delete each symbol/packet of codeword independently with probability $p \in (0, 1)$
- Random permutation block: Randomly permute packets of codeword

Example: Coding for Random Deletion Network

Consider a communication network where packets can be dropped:

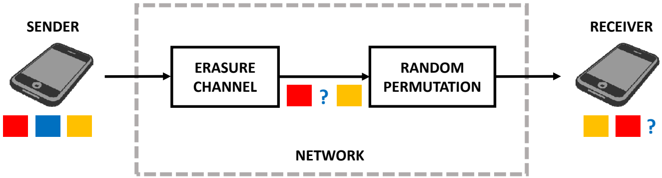


Abstraction:

- n-length codeword $=$ sequence of n packets
- Random deletion channel: Delete each symbol/packet of codeword independently with probability $p \in (0, 1)$
- Random permutation block: Randomly permute packets of codeword

Example: Coding for Random Deletion Network

Consider a communication network where packets can be dropped:

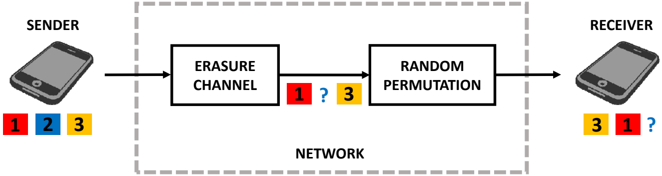


Abstraction:

- n-length codeword = sequence of n packets
- Equivalent Erasure channel: Erase each symbol/packet of codeword independently with probability $p \in (0, 1)$
- Random permutation block: Randomly permute packets of codeword

Example: Coding for Random Deletion Network

Consider a communication network where packets can be dropped:

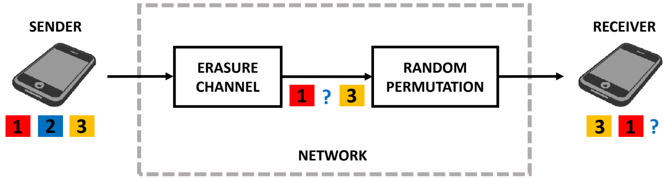


Abstraction:

- n-length codeword $=$ sequence of n packets
- Erasure channel: Erase each symbol/packet of codeword independently with probability $p \in (0, 1)$
- Random permutation block: Randomly permute packets of codeword
- Coding: Add sequence numbers (packet size $= L + \log(n)$ bits, alphabet size $= n\,2^L$)

Example: Coding for Random Deletion Network

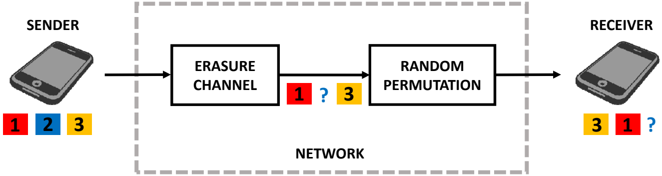
Consider a communication network where packets can be dropped:



Abstraction:

- n-length codeword = sequence of n packets
- Erasure channel: Erase each symbol/packet of codeword independently with probability $p \in (0, 1)$
- Random permutation block: Randomly permute packets of codeword
- Coding: Add sequence numbers and use standard coding techniques

Consider a communication network where packets can be dropped:

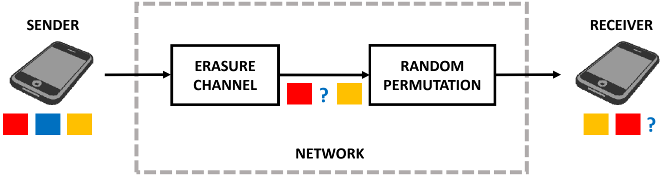


Abstraction:

- n-length codeword = sequence of n packets
- Erasure channel: Erase each symbol/packet of codeword independently with probability $p \in (0, 1)$
- Random permutation block: Randomly permute packets of codeword
- Coding: Add sequence numbers and use standard coding techniques
- More refined coding techniques *simulate* sequence numbers, e.g. [Mitzenmacher 2006], [Metzner 2009]

Example: Coding for Random Deletion Network

Consider a communication network where packets can be dropped:

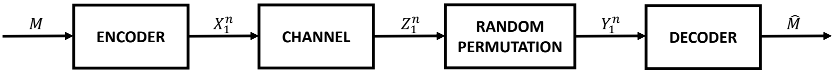


Abstraction:

- n-length codeword = sequence of n packets
- Erasure channel: Erase each symbol/packet of codeword independently with probability $p \in (0, 1)$
- Random permutation block: Randomly permute packets of codeword

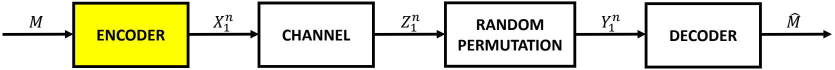
How do you code in such channels without increasing alphabet size?

The Permutation Channel Model



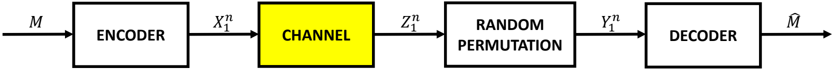
- Sender sends message $M \sim \text{Uniform}(\mathcal{M})$

The Permutation Channel Model



- Sender sends message $M \sim \text{Uniform}(\mathcal{M})$
- Possibly randomized encoder $f_n : \mathcal{M} \to \mathcal{X}^n$ produces codeword $X_1^n = (X_1, \dots, X_n) = f_n(M)$ (with block-length n)

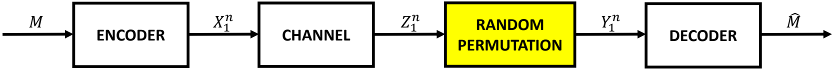
The Permutation Channel Model



- Sender sends message $M \sim \text{Uniform}(\mathcal{M})$
- Possibly randomized encoder $f_n : \mathcal{M} \to \mathcal{X}^n$ produces codeword $X_1^n = (X_1, \ldots, X_n) = f_n(M)$ (with block-length n)
- Discrete memoryless channel $P_{Z|X}$ with input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$ produces Z_1^n:

$$P_{Z_1^n|X_1^n}(z_1^n|x_1^n) = \prod_{i=1}^{n} P_{Z|X}(z_i|x_i)$$

The Permutation Channel Model

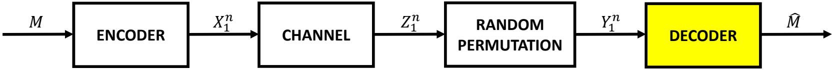


- Sender sends message $M \sim \text{Uniform}(\mathcal{M})$
- Possibly randomized encoder $f_n : \mathcal{M} \to \mathcal{X}^n$ produces codeword $X_1^n = (X_1, \dots, X_n) = f_n(M)$ (with block-length n)
- Discrete memoryless channel $P_{Z|X}$ with input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$ produces Z_1^n:

$$P_{Z_1^n|X_1^n}(z_1^n|x_1^n) = \prod_{i=1}^{n} P_{Z|X}(z_i|x_i)$$

- Random permutation generates Y_1^n from Z_1^n

The Permutation Channel Model

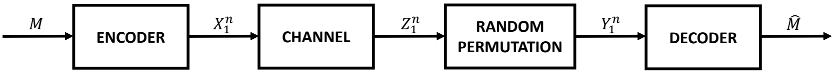


- Sender sends message $M \sim \text{Uniform}(\mathcal{M})$
- Possibly randomized encoder $f_n : \mathcal{M} \to \mathcal{X}^n$ produces codeword $X_1^n = (X_1, \ldots, X_n) = f_n(M)$ (with block-length n)
- Discrete memoryless channel $P_{Z|X}$ with input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$ produces Z_1^n:

$$P_{Z_1^n|X_1^n}(z_1^n|x_1^n) = \prod_{i=1}^{n} P_{Z|X}(z_i|x_i)$$

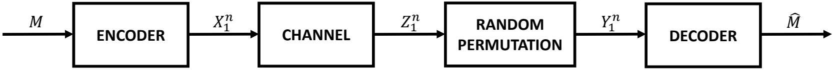
- Random permutation generates Y_1^n from Z_1^n
- Possibly randomized decoder $g_n : \mathcal{Y}^n \to \mathcal{M}$ produces estimate $\hat{M} = g_n(Y_1^n)$ at receiver

Coding for the Permutation Channel



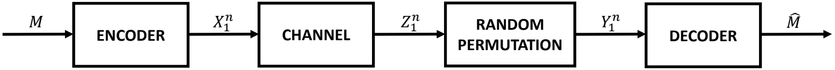
- **General Principle:**
  "Encode the information in an object that is invariant under the [permutation] transformation." [Kovačević-Vukobratović 2013]

Coding for the Permutation Channel



- **General Principle:**
  "Encode the information in an object that is invariant under the [permutation] transformation." [Kovačević-Vukobratović 2013]
- Multiset codes are studied in [Kovačević-Vukobratović 2013], [Kovačević-Vukobratović 2015], and [Kovačević-Tan 2018]

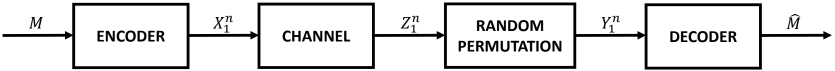
Coding for the Permutation Channel



- **General Principle:**
  "Encode the information in an object that is invariant under the [permutation] transformation." [Kovačević-Vukobratović 2013]
- Multiset codes are studied in [Kovačević-Vukobratović 2013], [Kovačević-Vukobratović 2015], and [Kovačević-Tan 2018]

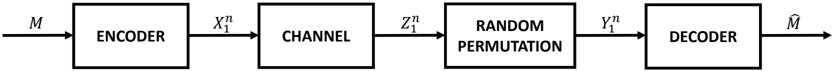
What about the information theoretic aspects of this model?

Information Capacity of the Permutation Channel



- Average probability of error $P_{\text{error}}^n \triangleq \mathbb{P}(M \neq \hat{M})$

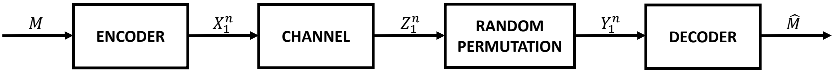
Information Capacity of the Permutation Channel



- Average probability of error $P_{\text{error}}^n \triangleq \mathbb{P}(M \neq \hat{M})$
- "Rate" of encoder-decoder pair (f_n, g_n):

$$R \triangleq \frac{\log(|\mathcal{M}|)}{\log(n)}$$

Information Capacity of the Permutation Channel

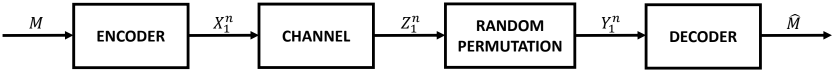


- Average probability of error $P_{\text{error}}^n \triangleq \mathbb{P}(M \neq \hat{M})$
- "Rate" of encoder-decoder pair (f_n, g_n):

$$R \triangleq \frac{\log(|\mathcal{M}|)}{\log(n)}$$

- $|\mathcal{M}| = n^R$

Information Capacity of the Permutation Channel

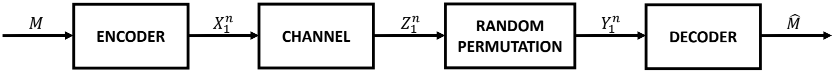


- Average probability of error $P^n_{\text{error}} \triangleq \mathbb{P}(M \neq \hat{M})$
- "Rate" of encoder-decoder pair (f_n, g_n):

$$R \triangleq \frac{\log(|\mathcal{M}|)}{\log(n)}$$

- $|\mathcal{M}| = n^R$ because number of empirical distributions of Y_1^n is $\mathit{poly}(n)$

Information Capacity of the Permutation Channel

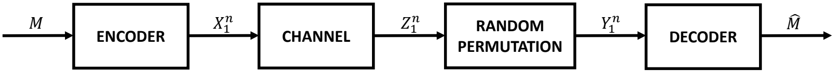


- Average probability of error $P_{\text{error}}^n \triangleq \mathbb{P}(M \neq \hat{M})$
- "Rate" of encoder-decoder pair (f_n, g_n):

$$R \triangleq \frac{\log(|\mathcal{M}|)}{\log(n)}$$

- $|\mathcal{M}| = n^R$
- Rate $R \geq 0$ is achievable $\Leftrightarrow \exists \{(f_n, g_n)\}_{n \in \mathbb{N}}$ such that $\lim_{n \to \infty} P_{\text{error}}^n = 0$

Information Capacity of the Permutation Channel



- Average probability of error $P^n_{\text{error}} \triangleq \mathbb{P}(M \neq \hat{M})$
- "Rate" of encoder-decoder pair (f_n, g_n):

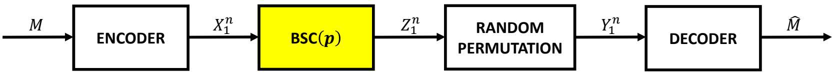
$$R \triangleq \frac{\log(|\mathcal{M}|)}{\log(n)}$$

- $|\mathcal{M}| = n^R$
- Rate $R \geq 0$ is achievable $\Leftrightarrow \exists \{(f_n, g_n)\}_{n \in \mathbb{N}}$ such that $\lim_{n \to \infty} P^n_{\text{error}} = 0$

Definition (Permutation Channel Capacity)

$$C_{\text{perm}}(P_{Z|X}) \triangleq \sup\{R \geq 0 : R \text{ is achievable}\}$$

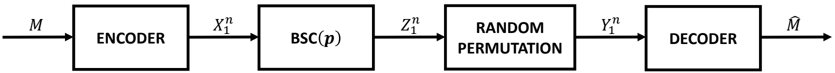
Capacity of the BSC Permutation Channel



- Channel is binary symmetric channel, denoted BSC(p):

$$\forall z, x \in \{0,1\}, \quad P_{Z|X}(z|x) = \begin{cases} 1-p, & \text{for } z = x \\ p, & \text{for } z \neq x \end{cases}$$

Capacity of the BSC Permutation Channel

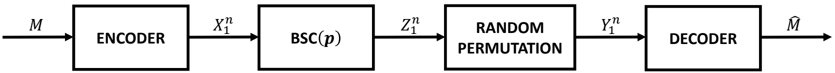


- Channel is binary symmetric channel, denoted BSC(p):

$$\forall z, x \in \{0,1\}, \quad P_{Z|X}(z|x) = \begin{cases} 1-p, & \text{for } z = x \\ p, & \text{for } z \neq x \end{cases}$$

- Alphabets are $\mathcal{X} = \mathcal{Y} = \{0,1\}$

Capacity of the BSC Permutation Channel

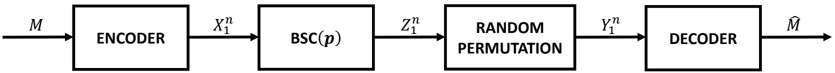


- Channel is binary symmetric channel, denoted BSC(p):

$$\forall z, x \in \{0,1\}, \;\; P_{Z|X}(z|x) = \begin{cases} 1-p, & \text{for } z = x \\ p, & \text{for } z \neq x \end{cases}$$

- Alphabets are $\mathcal{X} = \mathcal{Y} = \{0,1\}$
- Assume crossover probability $p \in (0,1)$ and $p \neq \frac{1}{2}$

Capacity of the BSC Permutation Channel



- Channel is binary symmetric channel, denoted BSC(p):

$$\forall z, x \in \{0,1\}, \;\; P_{Z|X}(z|x) = \begin{cases} 1-p, & \text{for } z = x \\ p, & \text{for } z \neq x \end{cases}$$

- Alphabets are $\mathcal{X} = \mathcal{Y} = \{0,1\}$
- Assume crossover probability $p \in (0,1)$ and $p \neq \frac{1}{2}$

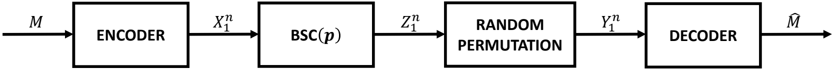
Main Question

What is the permutation channel capacity of the BSC?

Outline

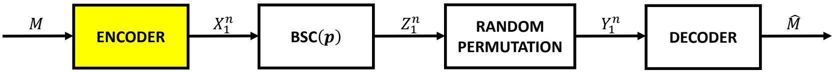
1. Introduction
2. Achievability
   - Encoder and Decoder
   - Testing between Converging Hypotheses
   - Intuition via Central Limit Theorem
   - Second Moment Method for TV Distance
3. Converse
4. Conclusion

Warm-up: Sending Two Messages



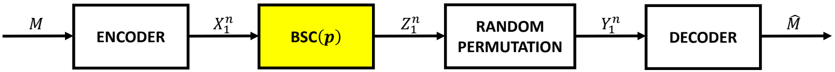
- Fix a message $m \in \{0, 1\}$

Warm-up: Sending Two Messages



- Fix a message $m \in \{0,1\}$, and encode m as $f_n(m) = X_1^n \overset{\text{i.i.d.}}{\sim} \text{Ber}(q_m)$

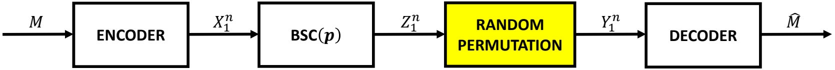
Warm-up: Sending Two Messages



- Fix a message $m \in \{0, 1\}$, and encode m as $f_n(m) = X_1^n \overset{\text{i.i.d.}}{\sim} \text{Ber}(q_m)$

- Memoryless BSC(p) outputs $Z_1^n \overset{\text{i.i.d.}}{\sim} \text{Ber}(p * q_m)$, where $p * q_m \triangleq p(1 - q_m) + q_m(1 - p)$ is the convolution of p and q_m

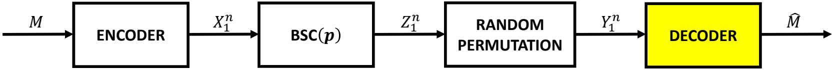
Warm-up: Sending Two Messages



- Fix a message $m \in \{0, 1\}$, and encode m as $f_n(m) = X_1^n \overset{\text{i.i.d.}}{\sim} \text{Ber}(q_m)$

- Memoryless BSC(p) outputs $Z_1^n \overset{\text{i.i.d.}}{\sim} \text{Ber}(p * q_m)$, where $p * q_m \triangleq p(1 - q_m) + q_m(1 - p)$ is the convolution of p and q_m
- Random permutation generates $Y_1^n \overset{\text{i.i.d.}}{\sim} \text{Ber}(p * q_m)$

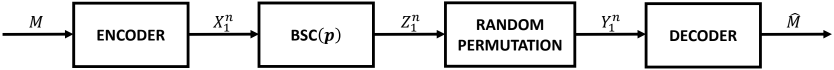
Warm-up: Sending Two Messages



- Fix a message $m \in \{0, 1\}$, and encode m as $f_n(m) = X_1^n \overset{\text{i.i.d.}}{\sim} \text{Ber}(q_m)$

- Memoryless BSC(p) outputs $Z_1^n \overset{\text{i.i.d.}}{\sim} \text{Ber}(p * q_m)$, where $p * q_m \triangleq p(1 - q_m) + q_m(1 - p)$ is the convolution of p and q_m
- Random permutation generates $Y_1^n \overset{\text{i.i.d.}}{\sim} \text{Ber}(p * q_m)$
- Maximum Likelihood (ML) decoder: $\hat{M} = \mathbb{1}\left\{\frac{1}{n}\sum_{i=1}^{n} Y_i \geq \frac{1}{2}\right\}$

Warm-up: Sending Two Messages



- Fix a message $m \in \{0, 1\}$, and encode m as $f_n(m) = X_1^n \overset{\text{i.i.d.}}{\sim} \text{Ber}(q_m)$

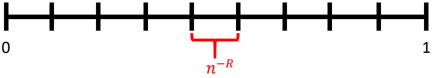
- Memoryless BSC(p) outputs $Z_1^n \overset{\text{i.i.d.}}{\sim} \text{Ber}(p * q_m)$, where $p * q_m \triangleq p(1 - q_m) + q_m(1 - p)$ is the convolution of p and q_m
- Random permutation generates $Y_1^n \overset{\text{i.i.d.}}{\sim} \text{Ber}(p * q_m)$
- Maximum Likelihood (ML) decoder: $\hat{M} = \mathbb{1}\left\{\frac{1}{n}\sum_{i=1}^{n} Y_i \geq \frac{1}{2}\right\}$
- $\frac{1}{n}\sum_{i=1}^{n} Y_i \to p * q_m$ in probability as $n \to \infty$ [WLLN]
  $\Rightarrow \lim_{n\to\infty} P_{\text{error}}^n = 0$ as $p * q_0 \neq p * q_1$

Encoder and Decoder

- Suppose $\mathcal{M} = \{1, \ldots, n^R\}$ for some $R > 0$

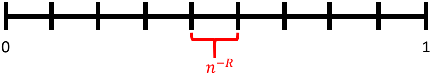
Encoder and Decoder

- Suppose $\mathcal{M} = \{1, \ldots, n^R\}$ for some $R > 0$
- **Randomized encoder:** Given $m \in \mathcal{M}$, $f_n(m) = X_1^n \overset{\text{i.i.d.}}{\sim} \text{Ber}\left(\frac{m}{n^R}\right)$



Encoder and Decoder

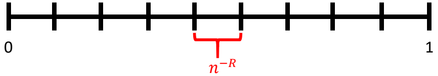
- Suppose $\mathcal{M} = \{1, \ldots, n^R\}$ for some $R > 0$
- **Randomized encoder:** Given $m \in \mathcal{M}$, $f_n(m) = X_1^n \overset{\text{i.i.d.}}{\sim} \text{Ber}\left(\frac{m}{n^R}\right)$



- Given $m \in \mathcal{M}$, $Y_1^n \overset{\text{i.i.d.}}{\sim} \text{Ber}\left(p * \frac{m}{n^R}\right)$ (as before)

Encoder and Decoder

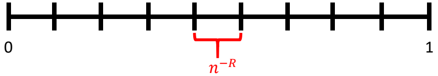
- Suppose $\mathcal{M} = \{1, \ldots, n^R\}$ for some $R > 0$
- **Randomized encoder:** Given $m \in \mathcal{M}$, $f_n(m) = X_1^n \overset{\text{i.i.d.}}{\sim} \text{Ber}\left(\frac{m}{n^R}\right)$



- Given $m \in \mathcal{M}$, $Y_1^n \overset{\text{i.i.d.}}{\sim} \text{Ber}\left(p * \frac{m}{n^R}\right)$
- **ML decoder:** For $y_1^n \in \{0,1\}^n$, $g_n(y_1^n) = \underset{m \in \mathcal{M}}{\arg\max}\, P_{Y_1^n|M}(y_1^n|m)$

Encoder and Decoder

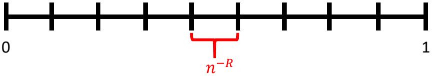
- Suppose $\mathcal{M} = \{1, \ldots, n^R\}$ for some $R > 0$
- **Randomized encoder:** Given $m \in \mathcal{M}$, $f_n(m) = X_1^n \overset{\text{i.i.d.}}{\sim} \text{Ber}\left(\frac{m}{n^R}\right)$



- Given $m \in \mathcal{M}$, $Y_1^n \overset{\text{i.i.d.}}{\sim} \text{Ber}\left(p * \frac{m}{n^R}\right)$
- **ML decoder:** For $y_1^n \in \{0,1\}^n$, $g_n(y_1^n) = \underset{m \in \mathcal{M}}{\arg\max}\, P_{Y_1^n|M}(y_1^n|m)$
- **Challenge:** Although $\frac{1}{n}\sum_{i=1}^n Y_i \to p * \frac{m}{n^R}$ in probability as $n \to \infty$, *consecutive messages become indistinguishable* i.e. $\frac{m}{n^R} - \frac{m+1}{n^R} \to 0$

Encoder and Decoder

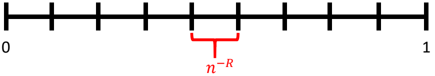
- Suppose $\mathcal{M} = \{1, \ldots, n^R\}$ for some $R > 0$
- **Randomized encoder:** Given $m \in \mathcal{M}$, $f_n(m) = X_1^n \overset{\text{i.i.d.}}{\sim} \text{Ber}\left(\frac{m}{n^R}\right)$



- Given $m \in \mathcal{M}$, $Y_1^n \overset{\text{i.i.d.}}{\sim} \text{Ber}\left(p * \frac{m}{n^R}\right)$
- **ML decoder:** For $y_1^n \in \{0,1\}^n$, $g_n(y_1^n) = \arg\max_{m \in \mathcal{M}} P_{Y_1^n|M}(y_1^n|m)$
- **Challenge:** Although $\frac{1}{n}\sum_{i=1}^n Y_i \to p * \frac{m}{n^R}$ in probability as $n \to \infty$, *consecutive messages become indistinguishable* i.e. $\frac{m}{n^R} - \frac{m+1}{n^R} \to 0$
- **Fact:** Consecutive messages distinguishable $\Rightarrow \lim_{n\to\infty} P_{\text{error}}^n = 0$

Encoder and Decoder

- Suppose $\mathcal{M} = \{1, \ldots, n^R\}$ for some $R > 0$
- **Randomized encoder:** Given $m \in \mathcal{M}$, $f_n(m) = X_1^n \overset{\text{i.i.d.}}{\sim} \text{Ber}\left(\frac{m}{n^R}\right)$



- Given $m \in \mathcal{M}$, $Y_1^n \overset{\text{i.i.d.}}{\sim} \text{Ber}\left(p * \frac{m}{n^R}\right)$
- **ML decoder:** For $y_1^n \in \{0,1\}^n$, $g_n(y_1^n) = \underset{m \in \mathcal{M}}{\arg\max}\, P_{Y_1^n|M}(y_1^n|m)$
- **Challenge:** Although $\frac{1}{n}\sum_{i=1}^n Y_i \to p * \frac{m}{n^R}$ in probability as $n \to \infty$, *consecutive messages become indistinguishable* i.e. $\frac{m}{n^R} - \frac{m+1}{n^R} \to 0$
- **Fact:** Consecutive messages distinguishable $\Rightarrow \lim_{n\to\infty} P_{\text{error}}^n = 0$

What is the largest R such that two consecutive messages can be distinguished?

Testing between Converging Hypotheses

Binary Hypothesis Testing:

- Consider hypothesis $H \sim \text{Ber}\left(\frac{1}{2}\right)$ with uniform prior

Testing between Converging Hypotheses

Binary Hypothesis Testing:

- Consider hypothesis $H \sim \text{Ber}\left(\frac{1}{2}\right)$ with uniform prior
- For any $n \in \mathbb{N}$, $q \in (0,1)$, and $R > 0$, consider likelihoods:

$$\text{Given } H = 0 : X_1^n \overset{\text{i.i.d.}}{\sim} P_{X|H=0} = \text{Ber}(q)$$

$$\text{Given } H = 1 : X_1^n \overset{\text{i.i.d.}}{\sim} P_{X|H=1} = \text{Ber}\left(q + \frac{1}{n^R}\right)$$

Testing between Converging Hypotheses

Binary Hypothesis Testing:

- Consider hypothesis $H \sim \text{Ber}\left(\frac{1}{2}\right)$ with uniform prior
- For any $n \in \mathbb{N}$, $q \in (0,1)$, and $R > 0$, consider likelihoods:

$$\text{Given } H = 0 : X_1^n \overset{\text{i.i.d.}}{\sim} P_{X|H=0} = \text{Ber}(q)$$
$$\text{Given } H = 1 : X_1^n \overset{\text{i.i.d.}}{\sim} P_{X|H=1} = \text{Ber}\left(q + \frac{1}{n^R}\right)$$

- Define the *zero-mean* sufficient statistic of X_1^n for H:

$$T_n \triangleq \frac{1}{n}\sum_{i=1}^{n} X_i - q - \frac{1}{2n^R}$$

Testing between Converging Hypotheses

Binary Hypothesis Testing:

- Consider hypothesis $H \sim \text{Ber}\left(\frac{1}{2}\right)$ with uniform prior
- For any $n \in \mathbb{N}$, $q \in (0,1)$, and $R > 0$, consider likelihoods:

$$\text{Given } H = 0 : X_1^n \overset{\text{i.i.d.}}{\sim} P_{X|H=0} = \text{Ber}(q)$$

$$\text{Given } H = 1 : X_1^n \overset{\text{i.i.d.}}{\sim} P_{X|H=1} = \text{Ber}\left(q + \frac{1}{n^R}\right)$$

- Define the *zero-mean* sufficient statistic of X_1^n for H:

$$T_n \triangleq \frac{1}{n}\sum_{i=1}^{n} X_i - q - \frac{1}{2n^R}$$

- Let $\hat{H}_{\text{ML}}^n(T_n)$ denote the ML decoder for H based on T_n with minimum probability of error $P_{\text{ML}}^n \triangleq \mathbb{P}(\hat{H}_{\text{ML}}^n(T_n) \neq H)$

Testing between Converging Hypotheses

Binary Hypothesis Testing:

- Consider hypothesis $H \sim \text{Ber}\left(\frac{1}{2}\right)$ with uniform prior
- For any $n \in \mathbb{N}$, $q \in (0,1)$, and $R > 0$, consider likelihoods:

$$\text{Given } H = 0 : X_1^n \overset{\text{i.i.d.}}{\sim} P_{X|H=0} = \text{Ber}(q)$$

$$\text{Given } H = 1 : X_1^n \overset{\text{i.i.d.}}{\sim} P_{X|H=1} = \text{Ber}\left(q + \frac{1}{n^R}\right)$$

- Define the *zero-mean* sufficient statistic of X_1^n for H:

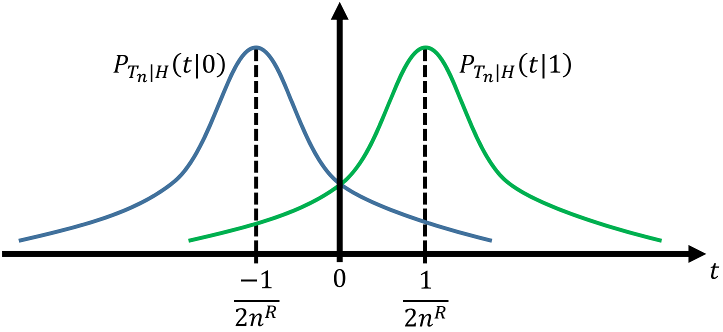
$$T_n \triangleq \frac{1}{n}\sum_{i=1}^{n} X_i - q - \frac{1}{2n^R}$$

- Let $\hat{H}^n_{\text{ML}}(T_n)$ denote the ML decoder for H based on T_n with minimum probability of error $P^n_{\text{ML}} \triangleq \mathbb{P}(\hat{H}^n_{\text{ML}}(T_n) \neq H)$
- **Want:** Largest $R > 0$ such that $\lim_{n\to\infty} P^n_{\text{ML}} = 0$?

Intuition via Central Limit Theorem

- For large n, $P_{T_n|H}(\cdot|0)$ and $P_{T_n|H}(\cdot|1)$ are Gaussian distributions [CLT]

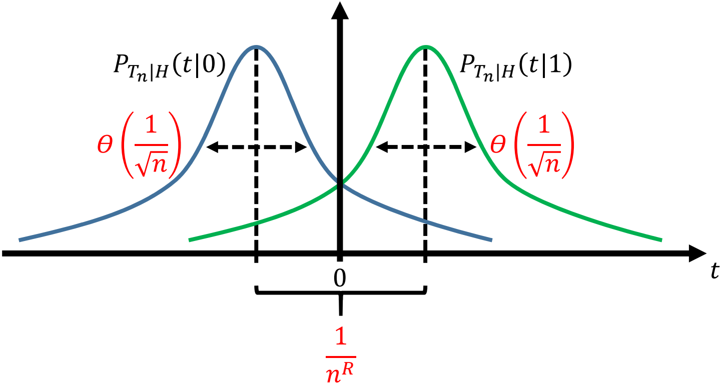
Figure:



Intuition via Central Limit Theorem

- For large n, $P_{T_n|H}(\cdot|0)$ and $P_{T_n|H}(\cdot|1)$ are Gaussian distributions [CLT]
- $|\mathbb{E}[T_n|H=0] - \mathbb{E}[T_n|H=1]| = 1/n^R$

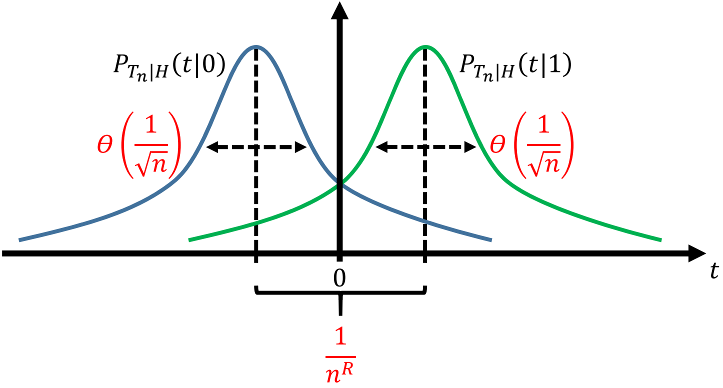
Figure:



Intuition via Central Limit Theorem

- For large n, $P_{T_n|H}(\cdot|0)$ and $P_{T_n|H}(\cdot|1)$ are Gaussian distributions [CLT]
- $|\mathbb{E}[T_n|H=0] - \mathbb{E}[T_n|H=1]| = 1/n^R$
- Standard deviations are $\Theta(1/\sqrt{n})$

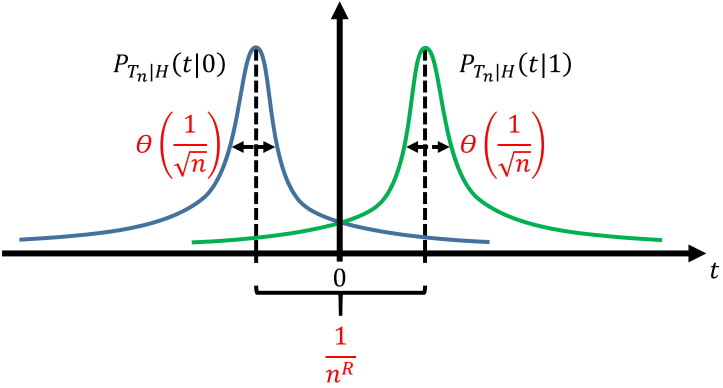
Figure:



Intuition via Central Limit Theorem

- For large n, $P_{T_n|H}(\cdot|0)$ and $P_{T_n|H}(\cdot|1)$ are Gaussian distributions [CLT]
- $|\mathbb{E}[T_n|H=0] - \mathbb{E}[T_n|H=1]| = 1/n^R$
- Standard deviations are $\Theta(1/\sqrt{n})$

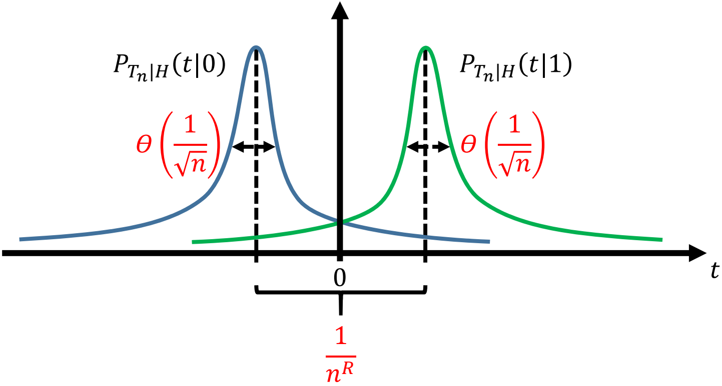
Case $R < \frac{1}{2}$**:**



Intuition via Central Limit Theorem

- For large n, $P_{T_n|H}(\cdot|0)$ and $P_{T_n|H}(\cdot|1)$ are Gaussian distributions [CLT]
- $|\mathbb{E}[T_n|H=0] - \mathbb{E}[T_n|H=1]| = 1/n^R$
- Standard deviations are $\Theta(1/\sqrt{n})$

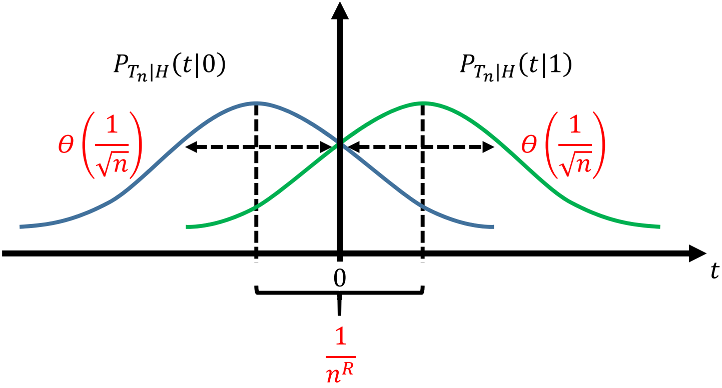
Case $R < \frac{1}{2}$**:** Decoding is possible ☺



Intuition via Central Limit Theorem

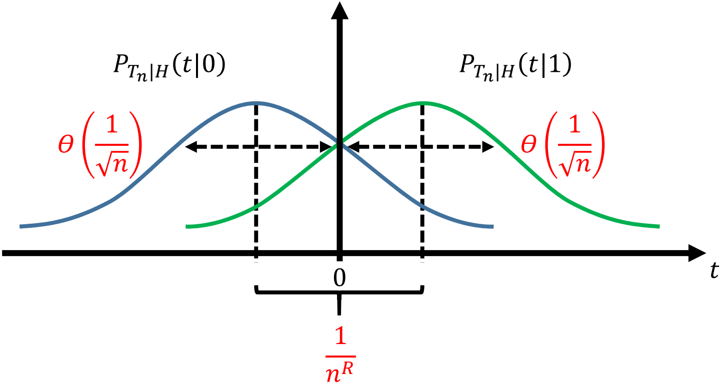
- For large n, $P_{T_n|H}(\cdot|0)$ and $P_{T_n|H}(\cdot|1)$ are Gaussian distributions [CLT]
- $|\mathbb{E}[T_n|H=0] - \mathbb{E}[T_n|H=1]| = 1/n^R$
- Standard deviations are $\Theta(1/\sqrt{n})$

Case $R > \frac{1}{2}$:



- For large n, $P_{T_n|H}(\cdot|0)$ and $P_{T_n|H}(\cdot|1)$ are Gaussian distributions [CLT]
- $|\mathbb{E}[T_n|H=0] - \mathbb{E}[T_n|H=1]| = 1/n^R$
- Standard deviations are $\Theta(1/\sqrt{n})$

Case $R > \frac{1}{2}$**:** Decoding is impossible ☹



Second Moment Method for TV Distance

Lemma (2nd Moment Method [Evans-Kenyon-Peres-Schulman 2000])

$$\left\| P_{T_n|H=1} - P_{T_n|H=0} \right\|_{\mathsf{TV}} \geq \frac{\left(\mathbb{E}[T_n|H=1] - \mathbb{E}[T_n|H=0]\right)^2}{4\,\mathbb{VAR}(T_n)}$$

where $\|P - Q\|_{\mathsf{TV}} = \frac{1}{2}\|P - Q\|_{\ell^1}$ is the *total variation (TV) distance* between the distributions P and Q.

Second Moment Method for TV Distance

Lemma (2nd Moment Method [Evans-Kenyon-Peres-Schulman 2000])

$$\left\|P_{T_n|H=1} - P_{T_n|H=0}\right\|_{\mathsf{TV}} \geq \frac{(\mathbb{E}[T_n|H=1] - \mathbb{E}[T_n|H=0])^2}{4\,\mathbb{VAR}(T_n)}$$

where $\|P - Q\|_{\mathsf{TV}} = \frac{1}{2}\|P - Q\|_{\ell^1}$ is the *total variation (TV) distance* between the distributions P and Q.

Proof: Let $T_n^+ \sim P_{T_n|H=1}$ and $T_n^- \sim P_{T_n|H=0}$

$$\left(\mathbb{E}\left[T_n^+\right] - \mathbb{E}\left[T_n^-\right]\right)^2 = \left(\sum_t t\left(P_{T_n|H}(t|1) - P_{T_n|H}(t|0)\right)\right)^2$$

Second Moment Method for TV Distance

Lemma (2nd Moment Method [Evans-Kenyon-Peres-Schulman 2000])

$$\left\| P_{T_n|H=1} - P_{T_n|H=0} \right\|_{\mathsf{TV}} \geq \frac{\left(\mathbb{E}[T_n|H=1] - \mathbb{E}[T_n|H=0]\right)^2}{4\,\mathbb{VAR}(T_n)}$$

where $\|P - Q\|_{\mathsf{TV}} = \frac{1}{2}\|P - Q\|_{\ell^1}$ is the *total variation (TV) distance* between the distributions P and Q.

Proof: Let $T_n^+ \sim P_{T_n|H=1}$ and $T_n^- \sim P_{T_n|H=0}$

$$\left(\mathbb{E}\left[T_n^+\right] - \mathbb{E}\left[T_n^-\right]\right)^2 = \left(\sum_t t\sqrt{P_{T_n}(t)}\,\frac{\left(P_{T_n|H}(t|1) - P_{T_n|H}(t|0)\right)}{\sqrt{P_{T_n}(t)}}\right)^2$$

Second Moment Method for TV Distance

Lemma (2nd Moment Method [Evans-Kenyon-Peres-Schulman 2000])

$$\left\| P_{T_n|H=1} - P_{T_n|H=0} \right\|_{\mathsf{TV}} \geq \frac{\left(\mathbb{E}[T_n|H=1] - \mathbb{E}[T_n|H=0]\right)^2}{4\,\mathbb{VAR}(T_n)}$$

where $\|P - Q\|_{\mathsf{TV}} = \frac{1}{2}\|P - Q\|_{\ell^1}$ is the *total variation (TV) distance* between the distributions P and Q.

Proof: Cauchy-Schwarz inequality

$$\left(\mathbb{E}\left[T_n^+\right] - \mathbb{E}\left[T_n^-\right]\right)^2 = \left(\sum_t t\sqrt{P_{T_n}(t)}\,\frac{\left(P_{T_n|H}(t|1) - P_{T_n|H}(t|0)\right)}{\sqrt{P_{T_n}(t)}}\right)^2$$

$$\leq \left(\sum_t t^2 P_{T_n}(t)\right)\left(\sum_t \frac{\left(P_{T_n|H}(t|1) - P_{T_n|H}(t|0)\right)^2}{P_{T_n}(t)}\right)$$

Second Moment Method for TV Distance

Lemma (2nd Moment Method [Evans-Kenyon-Peres-Schulman 2000])

$$\left\| P_{T_n|H=1} - P_{T_n|H=0} \right\|_{\mathsf{TV}} \geq \frac{\left(\mathbb{E}[T_n|H=1] - \mathbb{E}[T_n|H=0]\right)^2}{4\,\mathbb{VAR}(T_n)}$$

where $\|P - Q\|_{\mathsf{TV}} = \frac{1}{2}\|P - Q\|_{\ell^1}$ is the *total variation (TV) distance* between the distributions P and Q.

Proof: Recall that T_n is *zero-mean*

$$\left(\mathbb{E}\left[T_n^+\right] - \mathbb{E}\left[T_n^-\right]\right)^2 = \left(\sum_t t\sqrt{P_{T_n}(t)}\,\frac{\left(P_{T_n|H}(t|1) - P_{T_n|H}(t|0)\right)}{\sqrt{P_{T_n}(t)}}\right)^2$$

$$\leq \mathbb{VAR}(T_n)\left(\sum_t \frac{\left(P_{T_n|H}(t|1) - P_{T_n|H}(t|0)\right)^2}{P_{T_n}(t)}\right)$$

Second Moment Method for TV Distance

Lemma (2nd Moment Method [Evans-Kenyon-Peres-Schulman 2000])

$$\left\| P_{T_n|H=1} - P_{T_n|H=0} \right\|_{\mathsf{TV}} \geq \frac{\left(\mathbb{E}[T_n|H=1] - \mathbb{E}[T_n|H=0]\right)^2}{4\,\mathbb{VAR}(T_n)}$$

where $\|P - Q\|_{\mathsf{TV}} = \frac{1}{2}\|P - Q\|_{\ell^1}$ is the *total variation (TV) distance* between the distributions P and Q.

Proof: Hammersley-Chapman-Robbins bound

$$\left(\mathbb{E}\left[T_n^+\right] - \mathbb{E}\left[T_n^-\right]\right)^2 = \left(\sum_t t\sqrt{P_{T_n}(t)}\,\frac{\left(P_{T_n|H}(t|1) - P_{T_n|H}(t|0)\right)}{\sqrt{P_{T_n}(t)}}\right)^2$$

$$\leq 4\,\mathbb{VAR}(T_n)\underbrace{\left(\frac{1}{4}\sum_t \frac{\left(P_{T_n|H}(t|1) - P_{T_n|H}(t|0)\right)^2}{P_{T_n}(t)}\right)}_{\text{Vincze-Le Cam distance}}$$

Second Moment Method for TV Distance

Lemma (2nd Moment Method [Evans-Kenyon-Peres-Schulman 2000])

$$\left\| P_{T_n|H=1} - P_{T_n|H=0} \right\|_{\mathsf{TV}} \geq \frac{\left(\mathbb{E}[T_n|H=1] - \mathbb{E}[T_n|H=0]\right)^2}{4\,\mathbb{VAR}(T_n)}$$

where $\|P - Q\|_{\mathsf{TV}} = \frac{1}{2}\|P - Q\|_{\ell^1}$ is the *total variation (TV) distance* between the distributions P and Q.

Proof:

$$\left(\mathbb{E}\left[T_n^+\right] - \mathbb{E}\left[T_n^-\right]\right)^2 = \left(\sum_t t\sqrt{P_{T_n}(t)}\,\frac{\left(P_{T_n|H}(t|1) - P_{T_n|H}(t|0)\right)}{\sqrt{P_{T_n}(t)}}\right)^2$$

$$\leq 4\,\mathbb{VAR}(T_n)\left(\frac{1}{4}\sum_t \frac{\left(P_{T_n|H}(t|1) - P_{T_n|H}(t|0)\right)^2}{P_{T_n}(t)}\right)$$

$$\leq 4\,\mathbb{VAR}(T_n)\left\| P_{T_n|H=1} - P_{T_n|H=0} \right\|_{\mathsf{TV}}$$

■

Achievability Proof

Theorem (Achievability)

For any $0 < R < 1/2$, consider the binary hypothesis testing problem with $H \sim \text{Ber}\left(\frac{1}{2}\right)$, and $X_1^n \stackrel{\text{i.i.d.}}{\sim} \text{Ber}\left(q + \frac{h}{n^R}\right)$ given $H = h \in \{0, 1\}$.

Proof: Start with Le Cam's relation

$$P_{\text{ML}}^n = \frac{1}{2}\left(1 - \left\|P_{T_n|H=1} - P_{T_n|H=0}\right\|_{\text{TV}}\right)$$

Theorem (Achievability)

For any $0 < R < 1/2$, consider the binary hypothesis testing problem with $H \sim \text{Ber}\left(\frac{1}{2}\right)$, and $X_1^n \overset{\text{i.i.d.}}{\sim} \text{Ber}\left(q + \frac{h}{n^R}\right)$ given $H = h \in \{0, 1\}$.

Proof: Apply second moment method lemma

$$P_{\text{ML}}^n = \frac{1}{2}\left(1 - \left\|P_{T_n|H=1} - P_{T_n|H=0}\right\|_{\text{TV}}\right)$$
$$\leq \frac{1}{2}\left(1 - \frac{(\mathbb{E}[T_n|H=1] - \mathbb{E}[T_n|H=0])^2}{4\,\mathbb{VAR}(T_n)}\right)$$

Achievability Proof

Theorem (Achievability)

For any $0 < R < 1/2$, consider the binary hypothesis testing problem with $H \sim \text{Ber}\left(\frac{1}{2}\right)$, and $X_1^n \overset{\text{i.i.d.}}{\sim} \text{Ber}\left(q + \frac{h}{n^R}\right)$ given $H = h \in \{0, 1\}$.

Proof: After explicit computation and simplification...

$$P_{\text{ML}}^n = \frac{1}{2}\left(1 - \left\|P_{T_n|H=1} - P_{T_n|H=0}\right\|_{\text{TV}}\right)$$
$$\leq \frac{1}{2}\left(1 - \frac{\left(\mathbb{E}[T_n|H=1] - \mathbb{E}[T_n|H=0]\right)^2}{4\,\mathbb{VAR}(T_n)}\right)$$

Achievability Proof

Theorem (Achievability)

For any $0 < R < 1/2$, consider the binary hypothesis testing problem with $H \sim \text{Ber}\left(\frac{1}{2}\right)$, and $X_1^n \stackrel{\text{i.i.d.}}{\sim} \text{Ber}\left(q + \frac{h}{n^R}\right)$ given $H = h \in \{0, 1\}$.

Proof: For any $0 < R < \frac{1}{2}$,

$$P_{\text{ML}}^n = \frac{1}{2}\left(1 - \left\|P_{T_n|H=1} - P_{T_n|H=0}\right\|_{\text{TV}}\right)$$

$$\leq \frac{1}{2}\left(1 - \frac{\left(\mathbb{E}[T_n|H=1] - \mathbb{E}[T_n|H=0]\right)^2}{4\,\mathbb{VAR}(T_n)}\right)$$

$$\leq \frac{3}{2n^{1-2R}}$$

Achievability Proof

Theorem (Achievability)

For any $0 < R < 1/2$, consider the binary hypothesis testing problem with $H \sim \text{Ber}\left(\frac{1}{2}\right)$, and $X_1^n \overset{\text{i.i.d.}}{\sim} \text{Ber}\left(q + \frac{h}{n^R}\right)$ given $H = h \in \{0, 1\}$.
Then, $\lim_{n \to \infty} P_{\text{ML}}^n = 0$.

Proof: For any $0 < R < \frac{1}{2}$,

$$\begin{aligned} P_{\text{ML}}^n &= \frac{1}{2}\left(1 - \left\| P_{T_n|H=1} - P_{T_n|H=0} \right\|_{\text{TV}}\right) \\ &\leq \frac{1}{2}\left(1 - \frac{\left(\mathbb{E}[T_n|H=1] - \mathbb{E}[T_n|H=0]\right)^2}{4\,\mathbb{VAR}(T_n)}\right) \\ &\leq \frac{3}{2n^{1-2R}} \to 0 \text{ as } n \to \infty \end{aligned}$$

Achievability Proof

Theorem (Achievability)

For any $0 < R < 1/2$, consider the binary hypothesis testing problem with $H \sim \text{Ber}\left(\frac{1}{2}\right)$, and $X_1^n \stackrel{\text{i.i.d.}}{\sim} \text{Ber}\left(q + \frac{h}{n^R}\right)$ given $H = h \in \{0,1\}$.
Then, $\lim_{n\to\infty} P_{\text{ML}}^n = 0$. This implies that:

$$C_{\text{perm}}(\text{BSC}(p)) \geq \frac{1}{2}.$$

Proof: For any $0 < R < \frac{1}{2}$,

$$\begin{aligned} P_{\text{ML}}^n &= \frac{1}{2}\left(1 - \left\|P_{T_n|H=1} - P_{T_n|H=0}\right\|_{\text{TV}}\right) \\ &\leq \frac{1}{2}\left(1 - \frac{\left(\mathbb{E}[T_n|H=1] - \mathbb{E}[T_n|H=0]\right)^2}{4\,\mathbb{VAR}(T_n)}\right) \\ &\leq \frac{3}{2n^{1-2R}} \to 0 \text{ as } n \to \infty \end{aligned}$$

Outline

1. Introduction
2. Achievability
3. Converse
   - Fano's Inequality Argument
   - CLT Approximation
4. Conclusion

Converse: Fano's Inequality Argument

- Consider the Markov chain $M \rightarrow X_1^n \rightarrow Z_1^n \rightarrow Y_1^n \rightarrow S_n \triangleq \sum_{i=1}^n Y_i$, and a sequence of encoder-decoder pairs $\{(f_n, g_n)\}_{n \in \mathbb{N}}$ such that $|\mathcal{M}| = n^R$ and $\lim_{n \to \infty} P_{\text{error}}^n = 0$

Converse: Fano's Inequality Argument

- Consider the Markov chain $M \to X_1^n \to Z_1^n \to Y_1^n \to S_n \triangleq \sum_{i=1}^n Y_i$, and a sequence of encoder-decoder pairs $\{(f_n, g_n)\}_{n \in \mathbb{N}}$ such that $|\mathcal{M}| = n^R$ and $\lim_{n \to \infty} P_{\text{error}}^n = 0$
- Standard argument, cf. [Cover-Thomas 2006]: M is uniform

$$R \log(n) = H(M)$$

Converse: Fano's Inequality Argument

- Consider the Markov chain $M \rightarrow X_1^n \rightarrow Z_1^n \rightarrow Y_1^n \rightarrow S_n \triangleq \sum_{i=1}^n Y_i$, and a sequence of encoder-decoder pairs $\{(f_n, g_n)\}_{n \in \mathbb{N}}$ such that $|\mathcal{M}| = n^R$ and $\lim_{n \rightarrow \infty} P_{\text{error}}^n = 0$
- Standard argument, cf. [Cover-Thomas 2006]: Fano's inequality, DPI

$$\begin{aligned} R \log(n) &= H(M|\hat{M}) + I(M; \hat{M}) \\ &\leq 1 + P_{\text{error}}^n R \log(n) + I(M; Y_1^n) \end{aligned}$$

Converse: Fano's Inequality Argument

- Consider the Markov chain $M \to X_1^n \to Z_1^n \to Y_1^n \to S_n \triangleq \sum_{i=1}^n Y_i$, and a sequence of encoder-decoder pairs $\{(f_n, g_n)\}_{n\in\mathbb{N}}$ such that $|\mathcal{M}| = n^R$ and $\lim_{n\to\infty} P_{\text{error}}^n = 0$
- Standard argument, cf. [Cover-Thomas 2006]: sufficiency

$$\begin{aligned} R\log(n) &= H(M|\hat{M}) + I(M;\hat{M}) \\ &\leq 1 + P_{\text{error}}^n R\log(n) + I(M;Y_1^n) \\ &= 1 + P_{\text{error}}^n R\log(n) + I(M;S_n) \end{aligned}$$

Converse: Fano's Inequality Argument

- Consider the Markov chain $M \to X_1^n \to Z_1^n \to Y_1^n \to S_n \triangleq \sum_{i=1}^n Y_i$, and a sequence of encoder-decoder pairs $\{(f_n, g_n)\}_{n\in\mathbb{N}}$ such that $|\mathcal{M}| = n^R$ and $\lim_{n\to\infty} P_{\text{error}}^n = 0$
- Standard argument, cf. [Cover-Thomas 2006]: DPI

$$\begin{aligned} R\log(n) &= H(M|\hat{M}) + I(M;\hat{M}) \\ &\leq 1 + P_{\text{error}}^n R\log(n) + I(M;Y_1^n) \\ &= 1 + P_{\text{error}}^n R\log(n) + I(M;S_n) \\ &\leq 1 + P_{\text{error}}^n R\log(n) + I(X_1^n;S_n) \end{aligned}$$

Converse: Fano's Inequality Argument

- Consider the Markov chain $M \to X_1^n \to Z_1^n \to Y_1^n \to S_n \triangleq \sum_{i=1}^n Y_i$, and a sequence of encoder-decoder pairs $\{(f_n, g_n)\}_{n \in \mathbb{N}}$ such that $|\mathcal{M}| = n^R$ and $\lim_{n \to \infty} P_{\text{error}}^n = 0$
- Standard argument, cf. [Cover-Thomas 2006]:

$$\begin{aligned} R \log(n) &= H(M|\hat{M}) + I(M; \hat{M}) \\ &\leq 1 + P_{\text{error}}^n R \log(n) + I(M; Y_1^n) \\ &= 1 + P_{\text{error}}^n R \log(n) + I(M; S_n) \\ &\leq 1 + P_{\text{error}}^n R \log(n) + I(X_1^n; S_n) \end{aligned}$$

- Divide by $\log(n)$

$$R \leq \frac{1}{\log(n)} + P_{\text{error}}^n R + \frac{I(X_1^n; S_n)}{\log(n)}$$

Converse: Fano's Inequality Argument

- Consider the Markov chain $M \to X_1^n \to Z_1^n \to Y_1^n \to S_n \triangleq \sum_{i=1}^n Y_i$, and a sequence of encoder-decoder pairs $\{(f_n, g_n)\}_{n \in \mathbb{N}}$ such that $|\mathcal{M}| = n^R$ and $\lim_{n \to \infty} P_{\text{error}}^n = 0$
- Standard argument, cf. [Cover-Thomas 2006]:

$$\begin{aligned} R\log(n) &= H(M|\hat{M}) + I(M;\hat{M}) \\ &\leq 1 + P_{\text{error}}^n R\log(n) + I(M;Y_1^n) \\ &= 1 + P_{\text{error}}^n R\log(n) + I(M;S_n) \\ &\leq 1 + P_{\text{error}}^n R\log(n) + I(X_1^n;S_n) \end{aligned}$$

- Divide by $\log(n)$ and let $n \to \infty$:

$$R \leq \lim_{n \to \infty} \frac{I(X_1^n;S_n)}{\log(n)}$$

Converse: CLT Approximation

Upper bound on $I(X_1^n; S_n)$:

$I(X_1^n; S_n) = H(S_n) - H(S_n | X_1^n)$

Converse: CLT Approximation

Since $S_n \in \{0, \ldots, n\}$,

$$
\begin{aligned}
I(X_1^n; S_n) &= H(S_n) - H(S_n | X_1^n) \\
&\leq \log(n+1) - \sum_{x_1^n \in \{0,1\}^n} P_{X_1^n}(x_1^n) \, H(S_n | X_1^n = x_1^n)
\end{aligned}
$$

Converse: CLT Approximation

Given $X_1^n = x_1^n$ with $\sum_{i=1}^n x_i = k$, $S_n = \text{bin}(k, 1-p) + \text{bin}(n-k, p)$:

$$
\begin{aligned}
I(X_1^n; S_n) &= H(S_n) - H(S_n|X_1^n) \\
&\leq \log(n+1) - \sum_{x_1^n \in \{0,1\}^n} P_{X_1^n}(x_1^n)\, H(\text{bin}(k, 1-p) + \text{bin}(n-k, p))
\end{aligned}
$$

Converse: CLT Approximation

Using Problem 2.14 in [Cover-Thomas 2006],

$$
\begin{aligned}
I(X_1^n; S_n) &= H(S_n) - H(S_n | X_1^n) \\
&\leq \log(n+1) - \sum_{x_1^n \in \{0,1\}^n} P_{X_1^n}(x_1^n)\, H(\text{bin}(k, 1-p) + \text{bin}(n-k, p)) \\
&\leq \log(n+1) - \sum_{x_1^n \in \{0,1\}^n} P_{X_1^n}(x_1^n)\, H\left(\text{bin}\left(\frac{n}{2}, p\right)\right)
\end{aligned}
$$

Converse: CLT Approximation

Approximate binomial entropy using CLT, cf. [Adell-Lekuona-Yu 2010]:

$$
\begin{aligned}
I(X_1^n; S_n) &= H(S_n) - H(S_n | X_1^n) \\
&\leq \log(n+1) - \sum_{x_1^n \in \{0,1\}^n} P_{X_1^n}(x_1^n)\, H(\mathrm{bin}(k, 1-p) + \mathrm{bin}(n-k, p)) \\
&\leq \log(n+1) - \sum_{x_1^n \in \{0,1\}^n} P_{X_1^n}(x_1^n)\, H\left(\mathrm{bin}\left(\frac{n}{2}, p\right)\right) \\
&= \log(n+1) - \sum_{x_1^n \in \{0,1\}^n} P_{X_1^n}(x_1^n) \left(\frac{1}{2}\log(\pi e p(1-p)n) + O\left(\frac{1}{n}\right)\right)
\end{aligned}
$$

Converse: CLT Approximation

Upper bound on $I(X_1^n; S_n)$:

$$
\begin{aligned}
I(X_1^n; S_n) &= H(S_n) - H(S_n | X_1^n) \\
&\leq \log(n+1) - \sum_{x_1^n \in \{0,1\}^n} P_{X_1^n}(x_1^n)\, H(\text{bin}(k, 1-p) + \text{bin}(n-k, p)) \\
&\leq \log(n+1) - \sum_{x_1^n \in \{0,1\}^n} P_{X_1^n}(x_1^n)\, H\left(\text{bin}\left(\frac{n}{2}, p\right)\right) \\
&= \log(n+1) - \frac{1}{2}\log(\pi e p(1-p)n) + O\left(\frac{1}{n}\right)
\end{aligned}
$$

Converse: CLT Approximation

Upper bound on $I(X_1^n; S_n)$:

$$\begin{aligned}
I(X_1^n; S_n) &= H(S_n) - H(S_n|X_1^n) \\
&\leq \log(n+1) - \sum_{x_1^n \in \{0,1\}^n} P_{X_1^n}(x_1^n)\, H(\text{bin}(k, 1-p) + \text{bin}(n-k, p)) \\
&\leq \log(n+1) - \sum_{x_1^n \in \{0,1\}^n} P_{X_1^n}(x_1^n)\, H\left(\text{bin}\left(\frac{n}{2}, p\right)\right) \\
&= \log(n+1) - \frac{1}{2}\log(\pi e p(1-p)n) + O\left(\frac{1}{n}\right)
\end{aligned}$$

Hence, we have:

$$R \leq \lim_{n\to\infty} \frac{I(X_1^n; S_n)}{\log(n)} = \frac{1}{2}$$

Converse: CLT Approximation

Upper bound on $I(X_1^n; S_n)$:

$$\begin{aligned} I(X_1^n; S_n) &= H(S_n) - H(S_n|X_1^n) \\ &\leq \log(n+1) - \sum_{x_1^n \in \{0,1\}^n} P_{X_1^n}(x_1^n)\, H(\text{bin}(k, 1-p) + \text{bin}(n-k, p)) \\ &\leq \log(n+1) - \sum_{x_1^n \in \{0,1\}^n} P_{X_1^n}(x_1^n)\, H\left(\text{bin}\left(\frac{n}{2}, p\right)\right) \\ &= \log(n+1) - \frac{1}{2}\log(\pi e p(1-p)n) + O\left(\frac{1}{n}\right) \end{aligned}$$

Hence, we have:

$$R \leq \lim_{n \to \infty} \frac{I(X_1^n; S_n)}{\log(n)} = \frac{1}{2}$$

Theorem (Converse)

$$C_{\text{perm}}(\text{BSC}(p)) \leq \frac{1}{2}$$

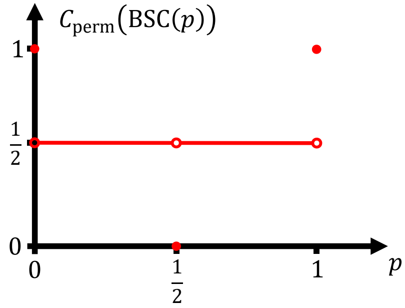
Outline

1. Introduction
2. Achievability
3. Converse
4. Conclusion

Conclusion

Theorem (Pemutation Channel Capacity of BSC)

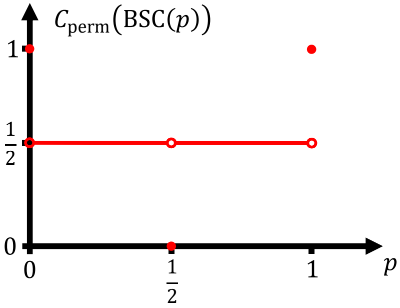
$$C_{\text{perm}}(\text{BSC}(p)) = \begin{cases} 1, & \text{for } p = 0, 1 \\ \frac{1}{2}, & \text{for } p \in \left(0, \frac{1}{2}\right) \cup \left(\frac{1}{2}, 1\right) \\ 0, & \text{for } p = \frac{1}{2} \end{cases}$$



Conclusion

Theorem (Pemutation Channel Capacity of BSC)

$$C_{\text{perm}}(\text{BSC}(p)) = \begin{cases} 1, & \text{for } p = 0, 1 \\ \frac{1}{2}, & \text{for } p \in \left(0, \frac{1}{2}\right) \cup \left(\frac{1}{2}, 1\right) \\ 0, & \text{for } p = \frac{1}{2} \end{cases}$$



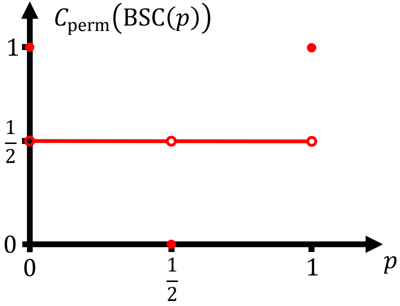
Remarks:

- $C_{\text{perm}}(\cdot)$ is discontinuous and non-convex

Conclusion

Theorem (Pemutation Channel Capacity of BSC)

$$C_{\text{perm}}(\text{BSC}(p)) = \begin{cases} 1, & \text{for } p = 0, 1 \\ \frac{1}{2}, & \text{for } p \in \left(0, \frac{1}{2}\right) \cup \left(\frac{1}{2}, 1\right) \\ 0, & \text{for } p = \frac{1}{2} \end{cases}$$



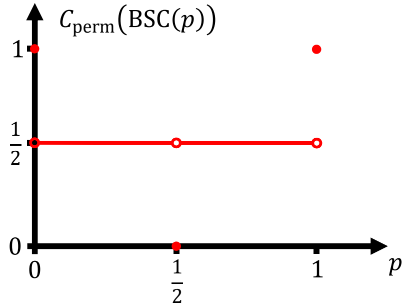
Remarks:

- $C_{\text{perm}}(\cdot)$ is discontinuous and non-convex
- $C_{\text{perm}}(\cdot)$ is generally agnostic to parameters of channel

Conclusion

Theorem (Pemutation Channel Capacity of BSC)

$$C_{\text{perm}}(\text{BSC}(p)) = \begin{cases} 1, & \text{for } p = 0, 1 \\ \frac{1}{2}, & \text{for } p \in \left(0, \frac{1}{2}\right) \cup \left(\frac{1}{2}, 1\right) \\ 0, & \text{for } p = \frac{1}{2} \end{cases}$$



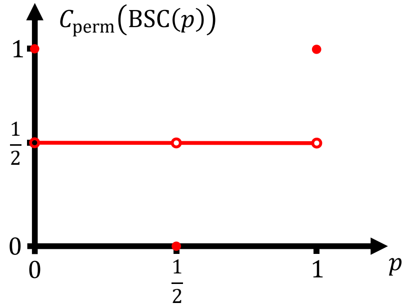
Remarks:

- $C_{\text{perm}}(\cdot)$ is discontinuous and non-convex
- $C_{\text{perm}}(\cdot)$ is generally agnostic to parameters of channel
- Computationally tractable coding scheme in proof

Conclusion

Theorem (Pemutation Channel Capacity of BSC)

$$C_{\text{perm}}(\text{BSC}(p)) = \begin{cases} 1, & \text{for } p = 0, 1 \\ \frac{1}{2}, & \text{for } p \in \left(0, \frac{1}{2}\right) \cup \left(\frac{1}{2}, 1\right) \\ 0, & \text{for } p = \frac{1}{2} \end{cases}$$



Remarks:

- $C_{\text{perm}}(\cdot)$ is discontinuous and non-convex
- $C_{\text{perm}}(\cdot)$ is generally agnostic to parameters of channel
- Computationally tractable coding scheme in proof
- Proof technique yields more general results

Thank You!