Permutation Channels

Anuran Makur

Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology

10 July 2020

Outline

Three Motivations

- **Coding theory:** [DG01], [Mit06], [Met09], [KV15], [KT18], . . .

Three Motivations

- **Coding theory:** [DG01], [Mit06], [Met09], [KV15], [KT18], . . .
  - *Random deletion channel*: LDPC codes nearly achieve capacity for large alphabets
  - Codes correct for transpositions of symbols

Three Motivations

- **Coding theory:** [DG01], [Mit06], [Met09], [KV15], [KT18], ...
  - *Random deletion channel*: LDPC codes nearly achieve capacity for large alphabets
  - Codes correct for transpositions of symbols
  - Permutation channels with insertions, deletions, substitutions, or erasures
  - Construction and analysis of *multiset codes*

Three Motivations

- **Coding theory:** [DG01], [Mit06], [Met09], [KV15], [KT18], ...
  - *Random deletion channel*: LDPC codes nearly achieve capacity for large alphabets
  - Codes correct for transpositions of symbols
  - Permutation channels with insertions, deletions, substitutions, or erasures
  - Construction and analysis of *multiset codes*
- **Communication networks:** [XZ02], [WWM09], [GG10], [KV13], ...
  - Mobile ad hoc networks, multipath routed networks, etc.

Three Motivations

- **Coding theory:** [DG01], [Mit06], [Met09], [KV15], [KT18], ...
  - *Random deletion channel*: LDPC codes nearly achieve capacity for large alphabets
  - Codes correct for transpositions of symbols
  - Permutation channels with insertions, deletions, substitutions, or erasures
  - Construction and analysis of *multiset codes*
- **Communication networks:** [XZ02], [WWM09], [GG10], [KV13], ...
  - Mobile ad hoc networks, multipath routed networks, etc.
  - *Out-of-order delivery* of packets
  - Correct for packet errors/losses when packets *do not have sequence numbers*

Three Motivations

- **Coding theory:** [DG01], [Mit06], [Met09], [KV15], [KT18], ...
  - *Random deletion channel*: LDPC codes nearly achieve capacity for large alphabets
  - Codes correct for transpositions of symbols
  - Permutation channels with insertions, deletions, substitutions, or erasures
  - Construction and analysis of *multiset codes*
- **Communication networks:** [XZ02], [WWM09], [GG10], [KV13], ...
  - Mobile ad hoc networks, multipath routed networks, etc.
  - *Out-of-order delivery* of packets
  - Correct for packet errors/losses when packets *do not have sequence numbers*
- **Molecular/Biological Communications:** [YKGR+15], [KPM16], [HSRT17], [SH19], ...

Three Motivations

- **Coding theory:** [DG01], [Mit06], [Met09], [KV15], [KT18], ...
  - *Random deletion channel*: LDPC codes nearly achieve capacity for large alphabets
  - Codes correct for transpositions of symbols
  - Permutation channels with insertions, deletions, substitutions, or erasures
  - Construction and analysis of *multiset codes*
- **Communication networks:** [XZ02], [WWM09], [GG10], [KV13], ...
  - Mobile ad hoc networks, multipath routed networks, etc.
  - *Out-of-order delivery* of packets
  - Correct for packet errors/losses when packets *do not have sequence numbers*
- **Molecular/Biological Communications:** [YKGR+15], [KPM16], [HSRT17], [SH19], ...
  - *DNA based storage systems*
  - Source data encoded into DNA molecules

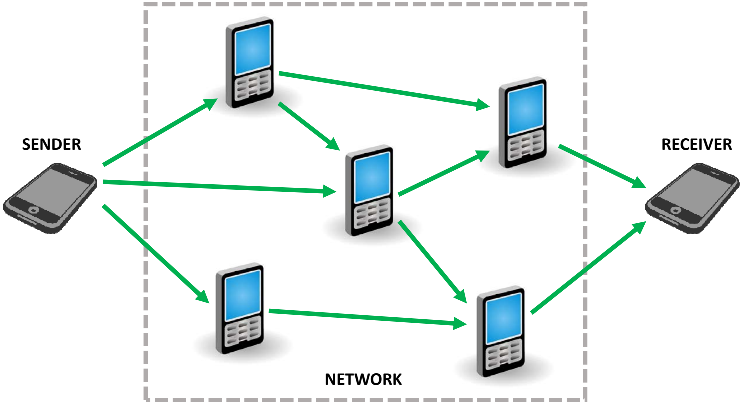
Three Motivations

- **Coding theory:** [DG01], [Mit06], [Met09], [KV15], [KT18], ...
  - *Random deletion channel*: LDPC codes nearly achieve capacity for large alphabets
  - Codes correct for transpositions of symbols
  - Permutation channels with insertions, deletions, substitutions, or erasures
  - Construction and analysis of *multiset codes*
- **Communication networks:** [XZ02], [WWM09], [GG10], [KV13], ...
  - Mobile ad hoc networks, multipath routed networks, etc.
  - *Out-of-order delivery* of packets
  - Correct for packet errors/losses when packets *do not have sequence numbers*
- **Molecular/Biological Communications:** [YKGR$^+$15], [KPM16], [HSRT17], [SH19], ...
  - *DNA based storage systems*
  - Source data encoded into DNA molecules
  - Fragments of DNA molecules cached
  - Receiver reads encoded data by *shotgun sequencing* (i.e., random sampling)

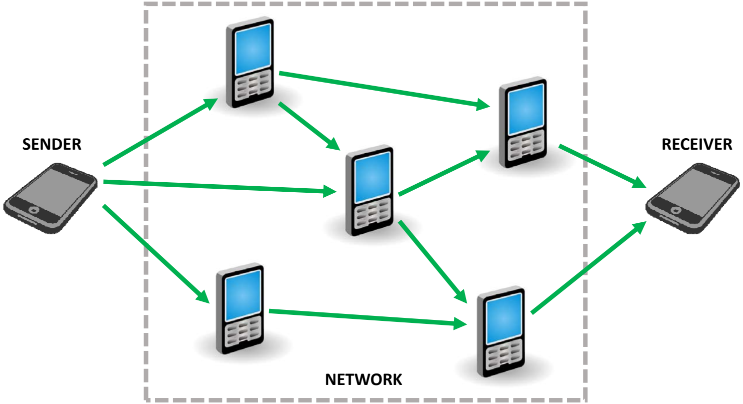
Three Motivations

- **Coding theory:** [DG01], [Mit06], [Met09], [KV15], [KT18], ...
  - *Random deletion channel*: LDPC codes nearly achieve capacity for large alphabets
  - Codes correct for transpositions of symbols
  - Permutation channels with insertions, deletions, substitutions, or erasures
  - Construction and analysis of *multiset codes*
- **Communication networks:** [XZ02], [WWM09], [GG10], [KV13], ...
  - Mobile ad hoc networks, multipath routed networks, etc.
  - *Out-of-order delivery* of packets
  - Correct for packet errors/losses when packets *do not have sequence numbers*
- **Molecular/Biological Communications:** [YKGR$^+$15], [KPM16], [HSRT17], [SH19], ...
  - *DNA based storage systems*
  - Source data encoded into DNA molecules
  - Fragments of DNA molecules cached
  - Receiver reads encoded data by *shotgun sequencing* (i.e., random sampling)

Motivation: Point-to-point Communication in Packet Networks

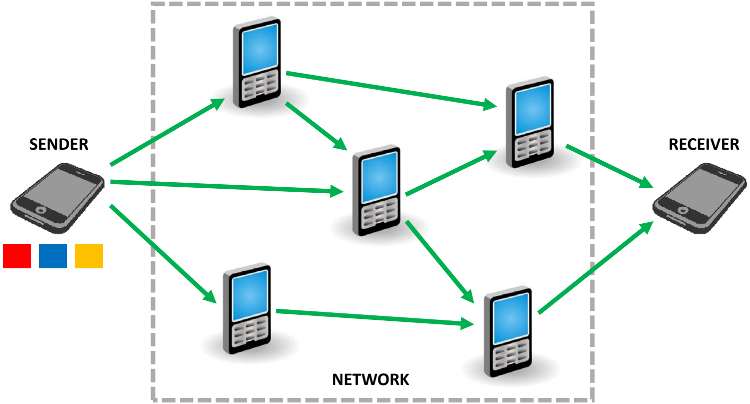


Motivation: Point-to-point Communication in Packet Networks



Model communication network as a channel

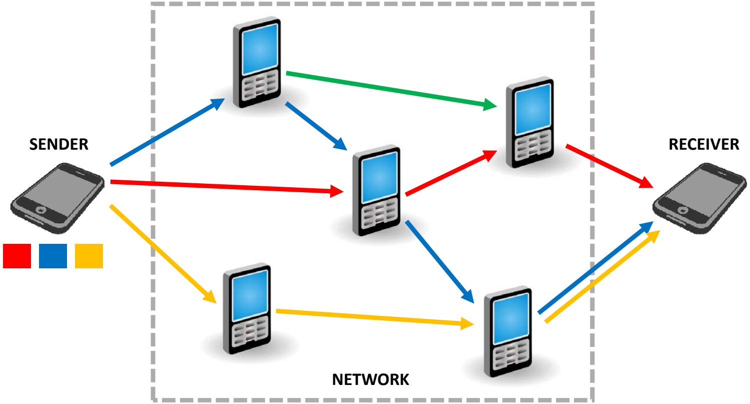
Motivation: Point-to-point Communication in Packet Networks



Model communication network as a channel:

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

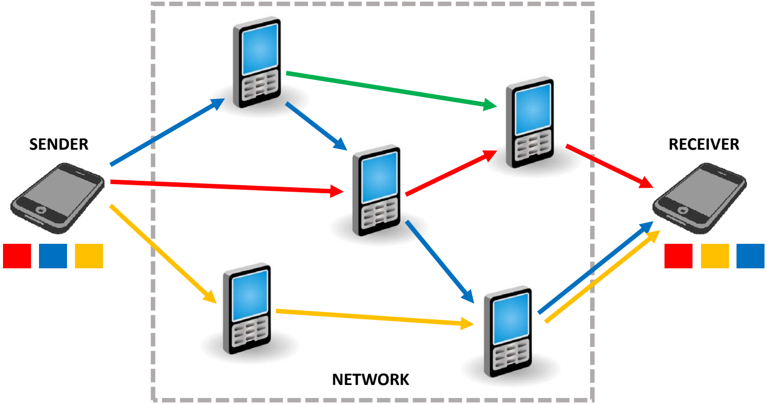
Motivation: Point-to-point Communication in Packet Networks



Model communication network as a channel:

- Alphabet symbols = all possible b-bit packets
- Multipath routed network or evolving network topology

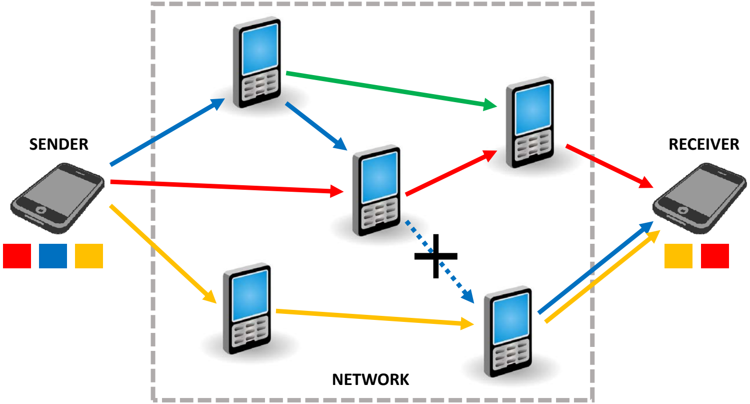
Motivation: Point-to-point Communication in Packet Networks



Model communication network as a channel:

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

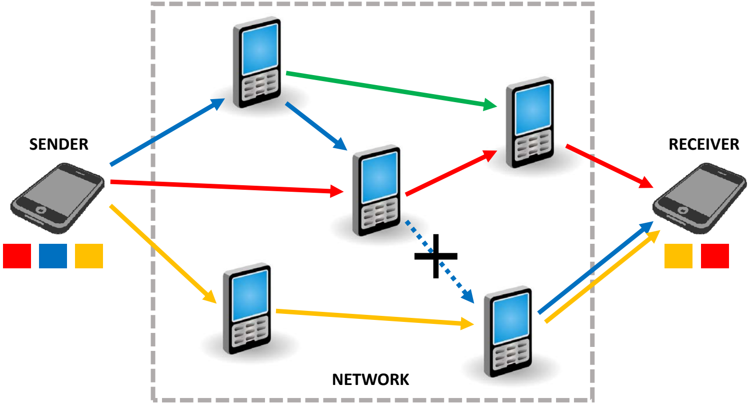
Motivation: Point-to-point Communication in Packet Networks



Model communication network as a channel:

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

Motivation: Point-to-point Communication in Packet Networks

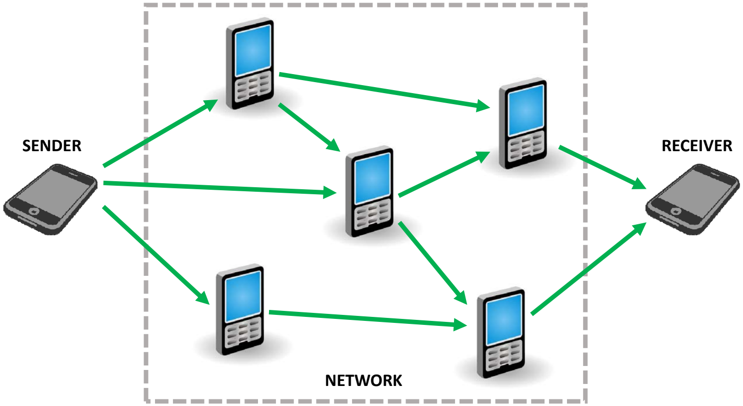


Model communication network as a channel:

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

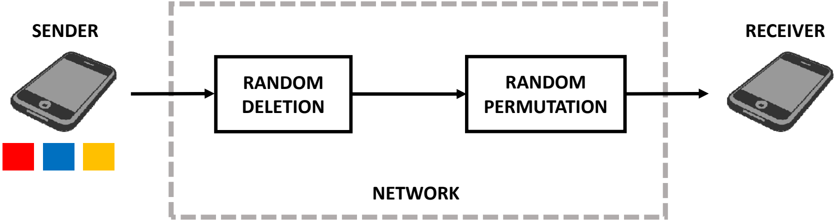
Example: Coding for Random Deletion Network

Consider a communication network where packets can be dropped:



Example: Coding for Random Deletion Network

Consider a communication network where packets can be dropped:

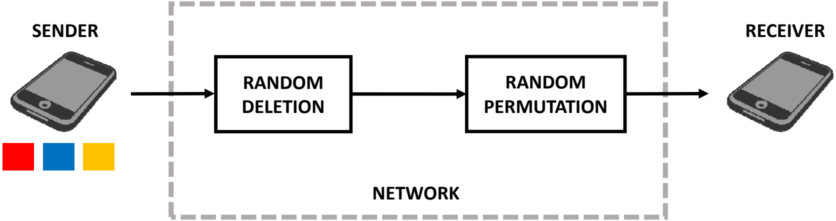


Abstraction:

- n-length codeword = sequence of n packets

Example: Coding for Random Deletion Network

Consider a communication network where packets can be dropped:

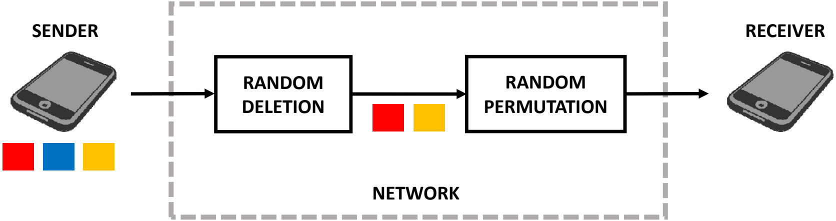


Abstraction:

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

Example: Coding for Random Deletion Network

Consider a communication network where packets can be dropped:

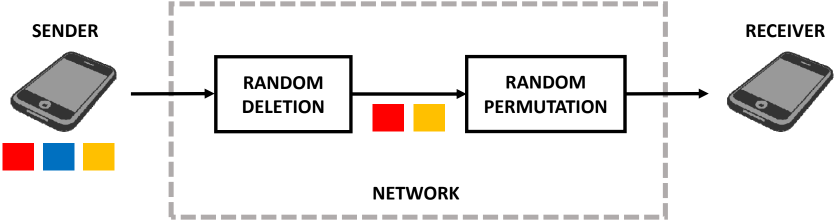


Abstraction:

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

Example: Coding for Random Deletion Network

Consider a communication network where packets can be dropped:

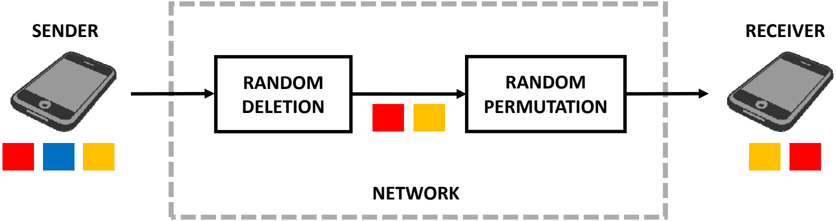


Abstraction:

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

Example: Coding for Random Deletion Network

Consider a communication network where packets can be dropped:

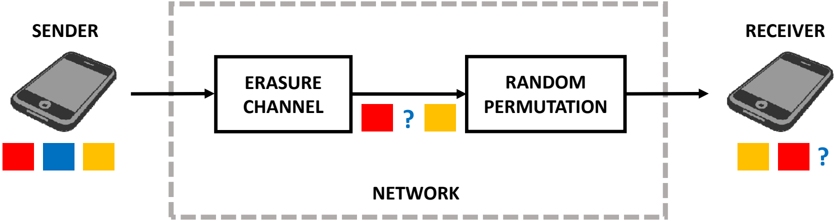


Abstraction:

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

Example: Coding for Random Deletion Network

Consider a communication network where packets can be dropped:

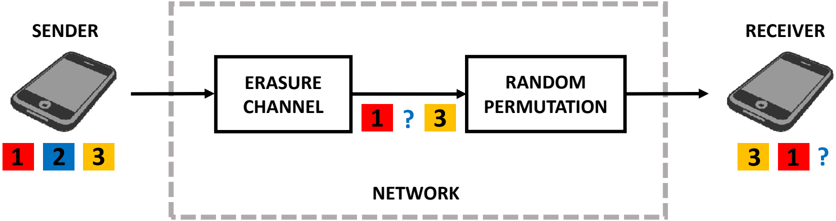


Abstraction:

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

Example: Coding for Random Deletion Network

Consider a communication network where packets can be dropped:

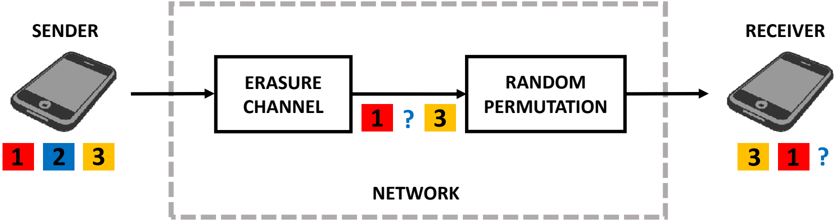


Abstraction:

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

Example: Coding for Random Deletion Network

Consider a communication network where packets can be dropped:

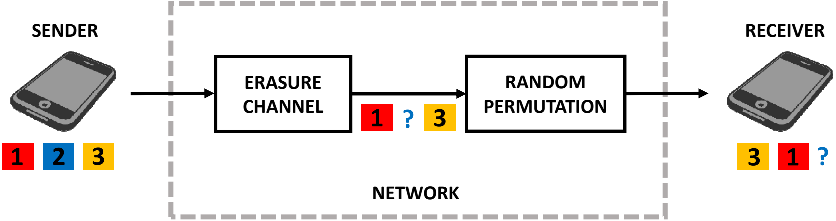


Abstraction:

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

Example: Coding for Random Deletion Network

Consider a communication network where packets can be dropped:

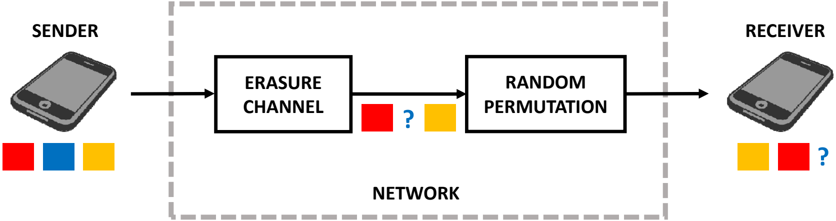


Abstraction:

- n-length codeword = sequence of n packets
- Erasure channel: Erase each symbol/packet independently with prob $p \in (0, 1)$
- Random permutation block: Randomly permute packets of codeword
- Coding: Add sequence numbers and use standard coding techniques
- More refined coding techniques *simulate* sequence numbers [Mit06], [Met09]

Example: Coding for Random Deletion Network

Consider a communication network where packets can be dropped:

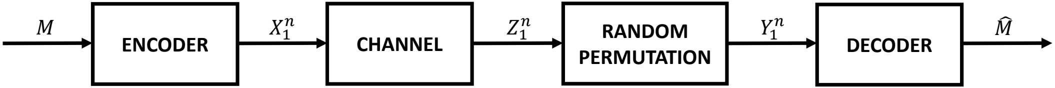


Abstraction:

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

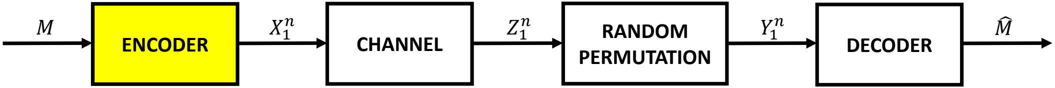
How do you code in such channels without increasing alphabet size?

Permutation Channel Model



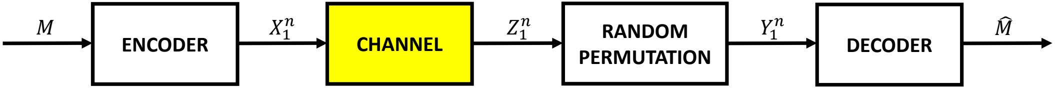
- Sender sends message $M \sim \text{Uniform}(\mathcal{M})$
- $n =$ blocklength

Permutation Channel Model



- Sender sends message $M \sim \text{Uniform}(\mathcal{M})$
- $n =$ blocklength
- Randomized encoder $f_n : \mathcal{M} \to \mathcal{X}^n$ produces codeword $X_1^n = (X_1, \ldots, X_n) = f_n(M)$

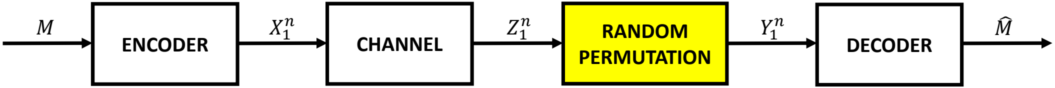
Permutation Channel Model



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

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

Permutation Channel Model

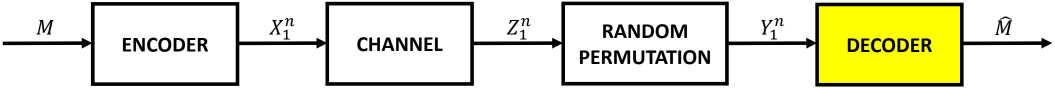


- Sender sends message $M \sim \text{Uniform}(\mathcal{M})$
- $n =$ blocklength
- Randomized encoder $f_n : \mathcal{M} \to \mathcal{X}^n$ produces codeword $X_1^n = (X_1, \ldots, X_n) = f_n(M)$
- Discrete memoryless channel $P_{Z|X}$ with input & output alphabets $\mathcal{X}$ & $\mathcal{Y}$ produces 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 π generates Y_1^n from Z_1^n: $Y_{\pi(i)} = Z_i$ for $i \in \{1, \ldots, n\}$

Permutation Channel Model



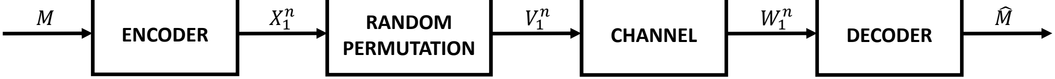
- Sender sends message $M \sim \text{Uniform}(\mathcal{M})$
- $n =$ blocklength
- Randomized encoder $f_n : \mathcal{M} \to \mathcal{X}^n$ produces codeword $X_1^n = (X_1, \dots, X_n) = f_n(M)$
- Discrete memoryless channel $P_{Z|X}$ with input & output alphabets $\mathcal{X}$ & $\mathcal{Y}$ produces 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 π generates Y_1^n from Z_1^n: $Y_{\pi(i)} = Z_i$ for $i \in \{1, \dots, n\}$
- Randomized decoder $g_n : \mathcal{Y}^n \to \mathcal{M} \cup \{\text{error}\}$ produces estimate $\hat{M} = g_n(Y_1^n)$ at receiver

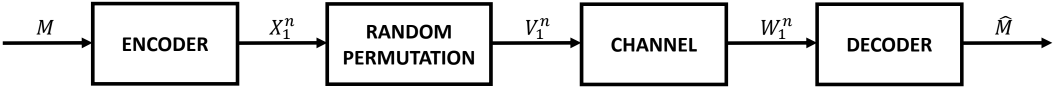
Permutation Channel Model

What if we analyze the "swapped" model?



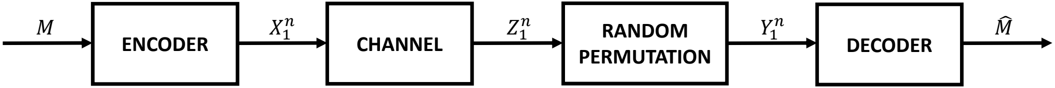
Permutation Channel Model

What if we analyze the "swapped" model?



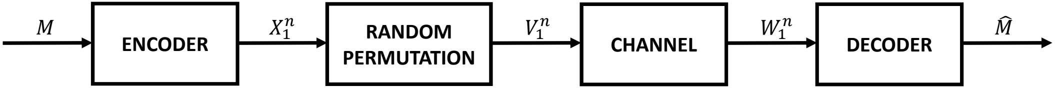
Proposition (Equivalent Models)

If channel $P_{W|V}$ is equal to channel $P_{Z|X}$, then channel $P_{W_1^n|X_1^n}$ is equal to channel $P_{Y_1^n|X_1^n}$.



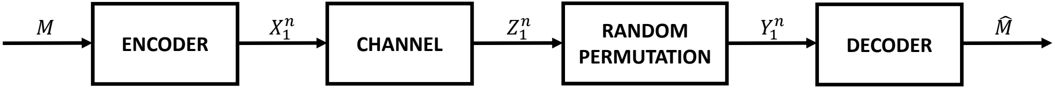
Permutation Channel Model

What if we analyze the "swapped" model?



Proposition (Equivalent Models)

If channel $P_{W|V}$ is equal to channel $P_{Z|X}$, then channel $P_{W_1^n|X_1^n}$ is equal to channel $P_{Y_1^n|X_1^n}$.

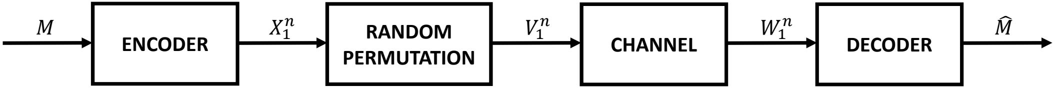


Remarks:

- Proof follows from direct calculation.

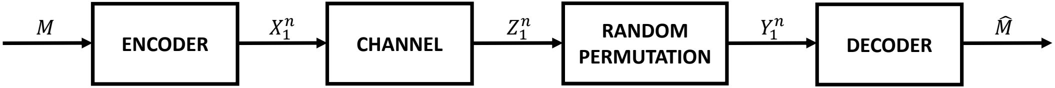
Permutation Channel Model

What if we analyze the "swapped" model?



Proposition (Equivalent Models)

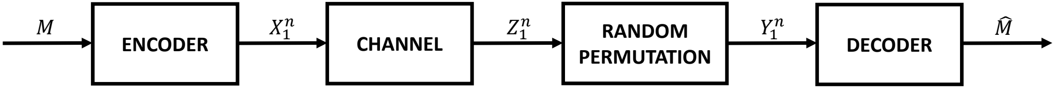
If channel $P_{W|V}$ is equal to channel $P_{Z|X}$, then channel $P_{W_1^n|X_1^n}$ is equal to channel $P_{Y_1^n|X_1^n}$.



Remarks:

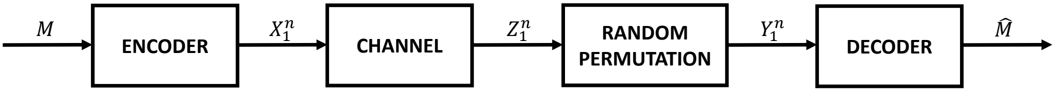
- Proof follows from direct calculation.
- Can analyze *either* model!

Coding for the Permutation Channel

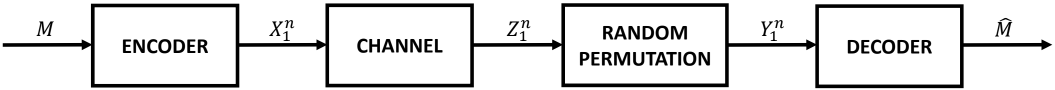


- **General Principle:**
  "Encode the information in an object that is invariant under the [permutation] transformation." [KV13]

Coding for the Permutation Channel



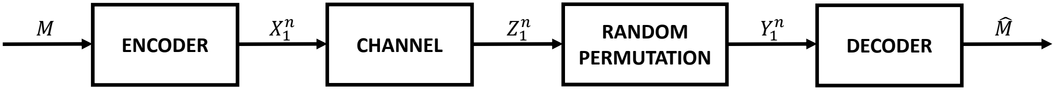
- **General Principle:**
  "Encode the information in an object that is invariant under the [permutation] transformation." [KV13]
- Multiset codes are studied in [KV13], [KV15], and [KT18].



- **General Principle:**
  "Encode the information in an object that is invariant under the [permutation] transformation." [KV13]
- Multiset codes are studied in [KV13], [KV15], and [KT18].

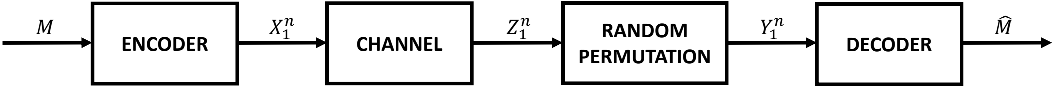
What are the fundamental information theoretic limits of this model?

Information Capacity of the Permutation Channel



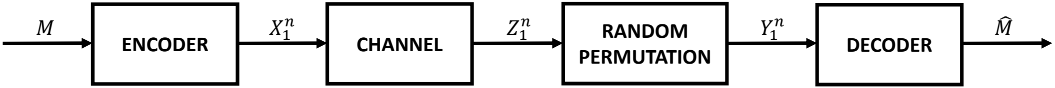
- Average probability of error $P_{\text{error}}^n \triangleq \mathbb{P}(M \neq \hat{M})$

Information Capacity of the Permutation Channel



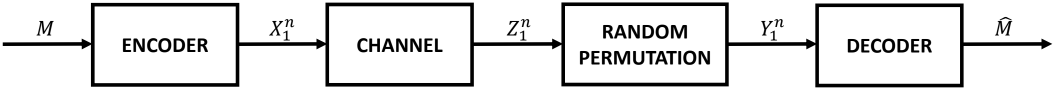
- Average probability of error $P_{\text{error}}^n \triangleq \mathbb{P}(M \neq \hat{M})$
- "Rate" of coding scheme (f_n, g_n) is $R \triangleq \dfrac{\log(|\mathcal{M}|)}{\log(n)}$

Information Capacity of the Permutation Channel



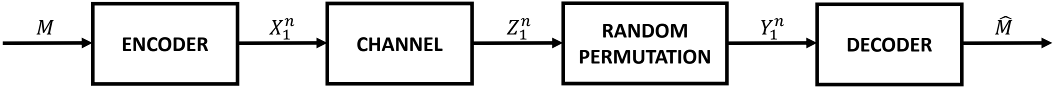
- Average probability of error $P_{\text{error}}^n \triangleq \mathbb{P}(M \neq \hat{M})$
- "Rate" of coding scheme (f_n, g_n) is $R \triangleq \dfrac{\log(|\mathcal{M}|)}{\log(n)}$
- $|\mathcal{M}| = n^R$

Information Capacity of the Permutation Channel



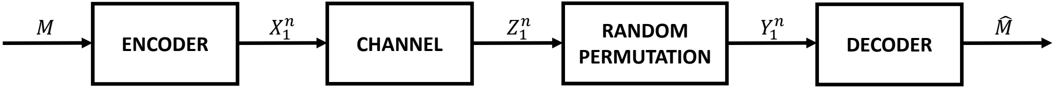
- Average probability of error $P^n_{\text{error}} \triangleq \mathbb{P}(M \neq \hat{M})$
- "Rate" of coding scheme (f_n, g_n) is $R \triangleq \dfrac{\log(|\mathcal{M}|)}{\log(n)}$
- $|\mathcal{M}| = n^R$ because number of empirical distributions of Y_1^n is $poly(n)$

Information Capacity of the Permutation Channel



- Average probability of error $P_{\text{error}}^n \triangleq \mathbb{P}(M \neq \hat{M})$
- "Rate" of coding scheme (f_n, g_n) is $R \triangleq \dfrac{\log(|\mathcal{M}|)}{\log(n)}$
- $|\mathcal{M}| = n^R$
- 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

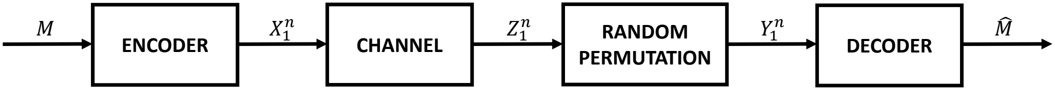


- Average probability of error $P_{\text{error}}^n \triangleq \mathbb{P}(M \neq \hat{M})$
- "Rate" of coding scheme (f_n, g_n) is $R \triangleq \dfrac{\log(|\mathcal{M}|)}{\log(n)}$
- $|\mathcal{M}| = n^R$
- Rate $R \geq 0$ is achievable $\Leftrightarrow \exists \{(f_n, g_n)\}_{n \in \mathbb{N}}$ such that $\lim_{n \to \infty} P_{\text{error}}^n = 0$

Definition (Permutation Channel Capacity)

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

Information Capacity of the Permutation Channel



- Average probability of error $P^n_{\text{error}} \triangleq \mathbb{P}(M \neq \hat{M})$
- "Rate" of coding scheme (f_n, g_n) is $R \triangleq \dfrac{\log(|\mathcal{M}|)}{\log(n)}$
- $|\mathcal{M}| = n^R$
- Rate $R \geq 0$ is achievable $\Leftrightarrow \exists \{(f_n, g_n)\}_{n \in \mathbb{N}}$ such that $\lim_{n \to \infty} P^n_{\text{error}} = 0$

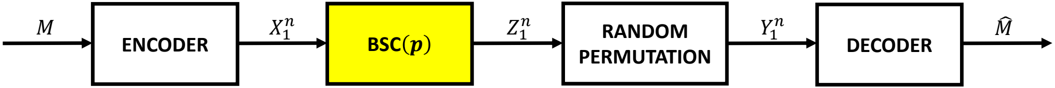
Definition (Permutation Channel Capacity)

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

Main Question

What is the permutation channel capacity of a general $P_{Z|X}$?

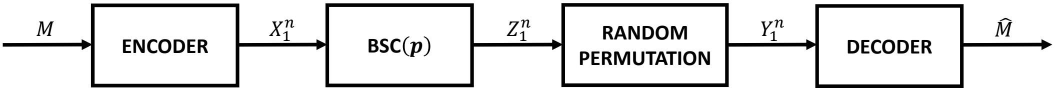
Example: Binary Symmetric Channel



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

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

Example: Binary Symmetric Channel

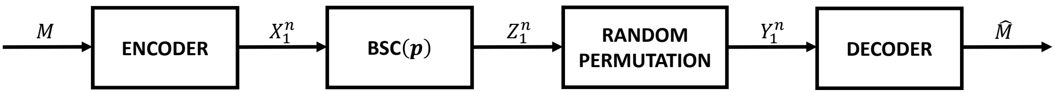


- Channel is binary symmetric channel, denoted $\mathsf{BSC}(p)$:

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

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

Example: Binary Symmetric Channel

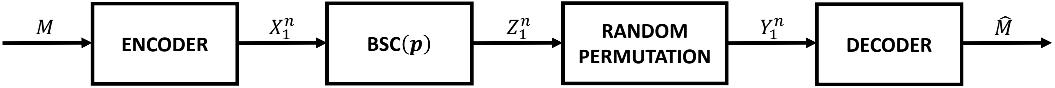


- Channel is binary symmetric channel, denoted BSC(p):

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

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

Example: Binary Symmetric Channel



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

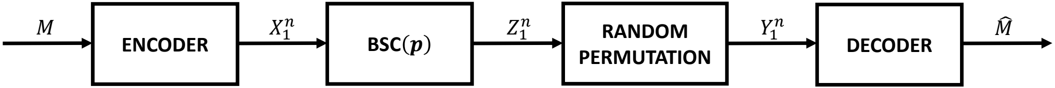
$$\forall z, x \in \{0,1\}, \quad P_{Z|X}(z|x) = \begin{cases} 1-p, & \text{for } z = x \\ p, & \text{for } z \neq x \end{cases}$$

- Alphabets are $\mathcal{X} = \mathcal{Y} = \{0,1\}$
- Assume crossover probability $p \in (0,1)$ and $p \neq \frac{1}{2}$
- **Question:** What is the permutation channel capacity of the BSC?

Outline

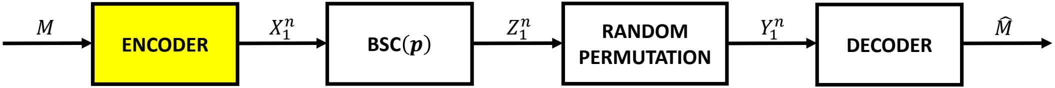
1. Introduction
2. Achievability and Converse for the BSC
   - Encoder and Decoder
   - Testing between Converging Hypotheses
   - Second Moment Method for TV Distance
   - Fano's Inequality and CLT Approximation
3. General Achievability Bound
4. General Converse Bounds
5. Conclusion

Warm-up: Sending Two Messages



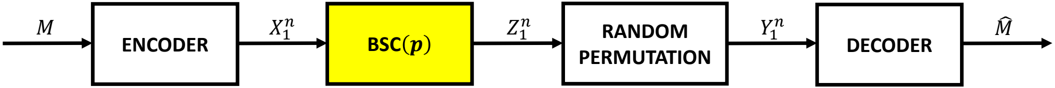
- Fix a message $m \in \{0, 1\}$

Warm-up: Sending Two Messages



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

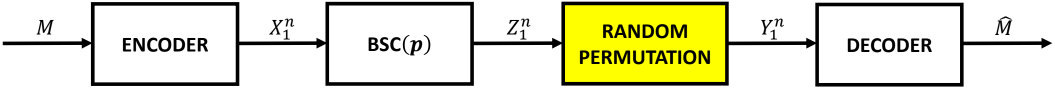
Warm-up: Sending Two Messages



- 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

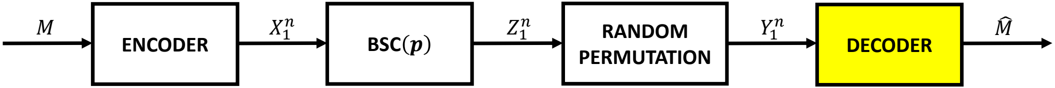
Warm-up: Sending Two Messages



- 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 $Y_1^n \overset{\text{i.i.d.}}{\sim} \text{Ber}(p * q_m)$

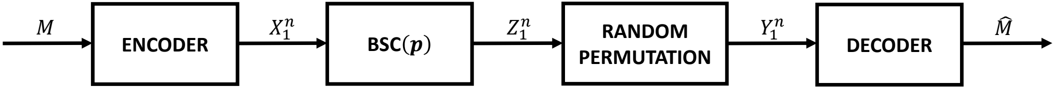
Warm-up: Sending Two Messages



- 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: $\hat{M} = \mathbb{1}\{\frac{1}{n}\sum_{i=1}^{n} Y_i \geq \frac{1}{2}\}$ (for $p < \frac{1}{2}$)

Warm-up: Sending Two Messages



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

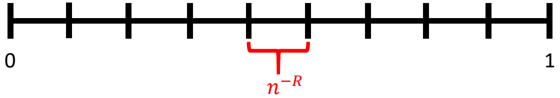
- Memoryless BSC(p) outputs $Z_1^n \overset{\text{i.i.d.}}{\sim} \text{Ber}(p * q_m)$, where $p * q_m \triangleq p(1-q_m) + q_m(1-p)$ is the convolution of p and q_m
- Random permutation generates $Y_1^n \overset{\text{i.i.d.}}{\sim} \text{Ber}(p * q_m)$
- Maximum Likelihood (ML) decoder: $\hat{M} =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

Encoder and Decoder

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

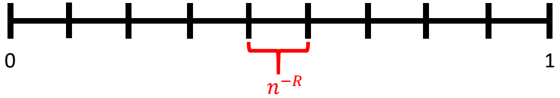
Encoder and Decoder

- 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

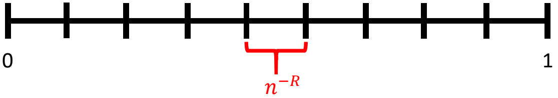
- Suppose $\mathcal{M} = \{1, \ldots, n^R\}$ for some $R > 0$
- **Randomized encoder:** Given $m \in \mathcal{M}$, $f_n(m) = X_1^n \overset{\text{i.i.d.}}{\sim} \text{Ber}\left(\frac{m}{n^R}\right)$



- 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

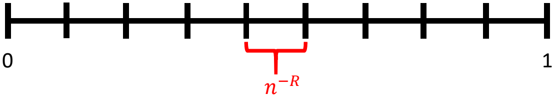
- Suppose $\mathcal{M} = \{1, \ldots, n^R\}$ for some $R > 0$
- **Randomized encoder:** Given $m \in \mathcal{M}$, $f_n(m) = X_1^n \overset{\text{i.i.d.}}{\sim} \text{Ber}\left(\frac{m}{n^R}\right)$



- 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 $y_1^n \in \{0,1\}^n$, $g_n(y_1^n) = \arg\max_{m \in \mathcal{M}} P_{Y_1^n|M}(y_1^n|m)$

Encoder and Decoder

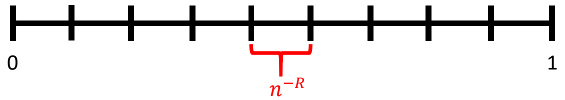
- Suppose $\mathcal{M} = \{1, \ldots, n^R\}$ for some $R > 0$
- **Randomized encoder:** Given $m \in \mathcal{M}$, $f_n(m) = X_1^n \overset{\text{i.i.d.}}{\sim} \text{Ber}\left(\frac{m}{n^R}\right)$



- 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 $y_1^n \in \{0,1\}^n$, $g_n(y_1^n) = \arg\max_{m \in \mathcal{M}} P_{Y_1^n|M}(y_1^n|m)$
- **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. $\frac{m}{n^R} - \frac{m+1}{n^R} \to 0$

Encoder and Decoder

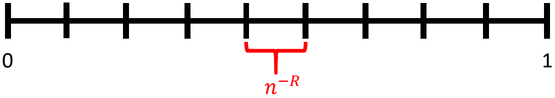
- Suppose $\mathcal{M} = \{1, \ldots, n^R\}$ for some $R > 0$
- **Randomized encoder:** Given $m \in \mathcal{M}$, $f_n(m) = X_1^n \overset{\text{i.i.d.}}{\sim} \text{Ber}\left(\frac{m}{n^R}\right)$



- 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 $\Rightarrow \lim_{n\to\infty} P_{\text{error}}^n = 0$

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 $P_{\text{ML}}^n \triangleq \mathbb{P}(\hat{H}_{\text{ML}}^n(T_n) \neq H)$

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:

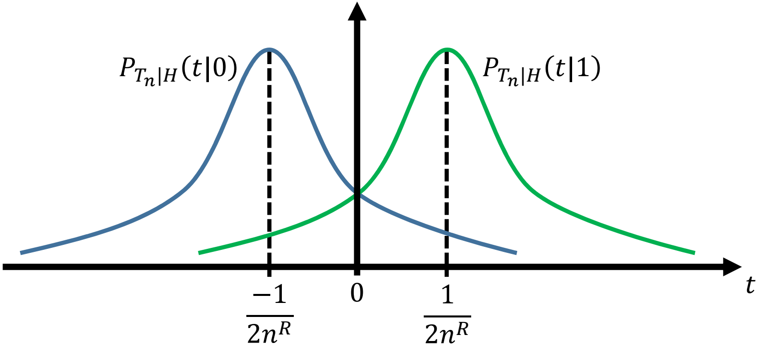
$$T_n \triangleq \frac{1}{n}\sum_{i=1}^{n} X_i - q - \frac{1}{2n^R}$$

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

Intuition via Central Limit Theorem

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

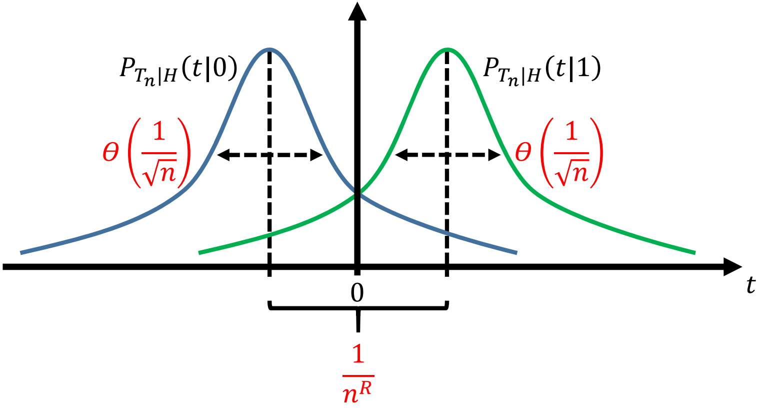
Figure:



Intuition via Central Limit Theorem

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

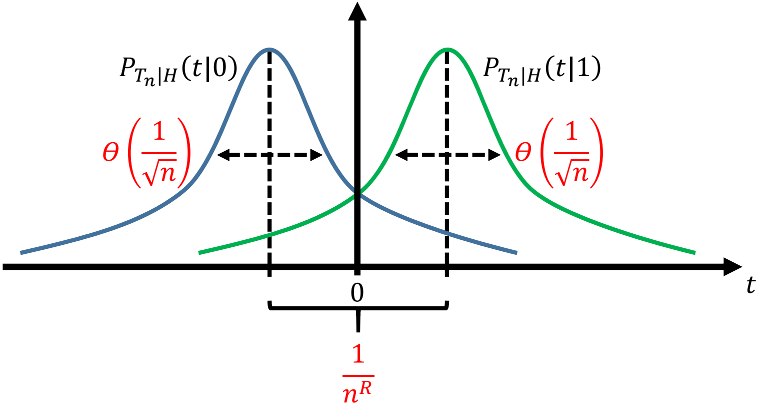
Figure:



Intuition via Central Limit Theorem

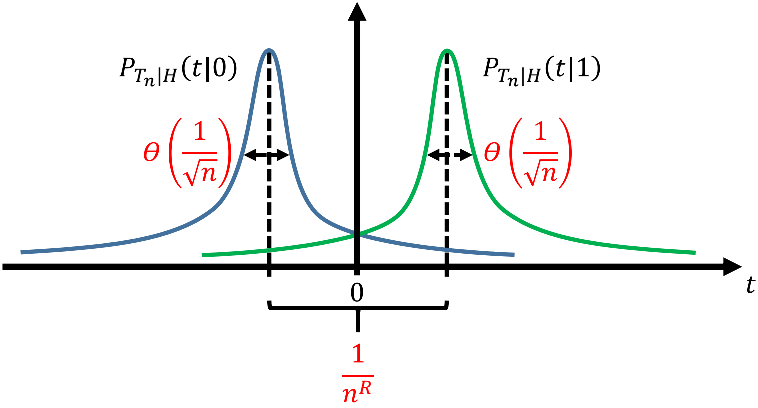
- For large n, $P_{T_n|H}(\cdot|0)$ and $P_{T_n|H}(\cdot|1)$ are Gaussian distributions
- $|\mathbb{E}[T_n|H=0] - \mathbb{E}[T_n|H=1]| = 1/n^R$
- Standard deviations are $\Theta(1/\sqrt{n})$

Figure:



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

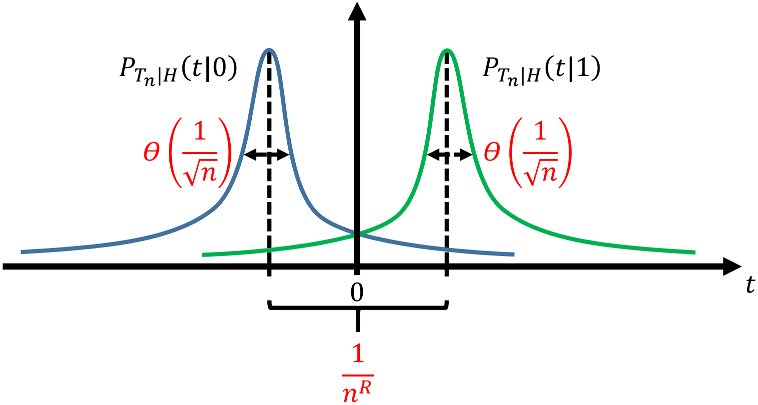
Case $R < \frac{1}{2}$**:**



Intuition via Central Limit Theorem

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

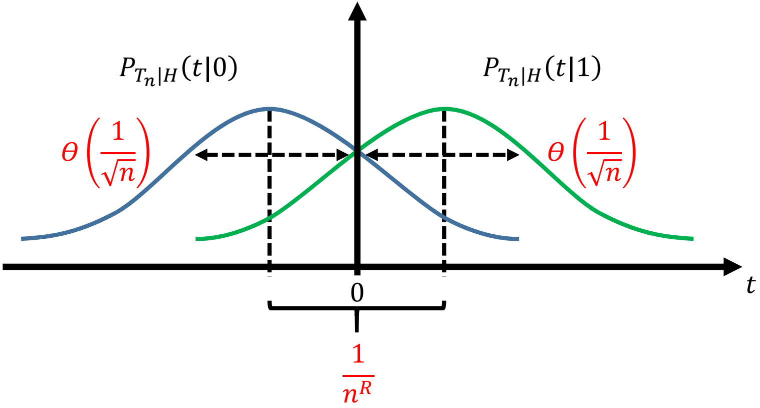
Case $R < \frac{1}{2}$**:** Decoding is possible ☺



Intuition via Central Limit Theorem

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

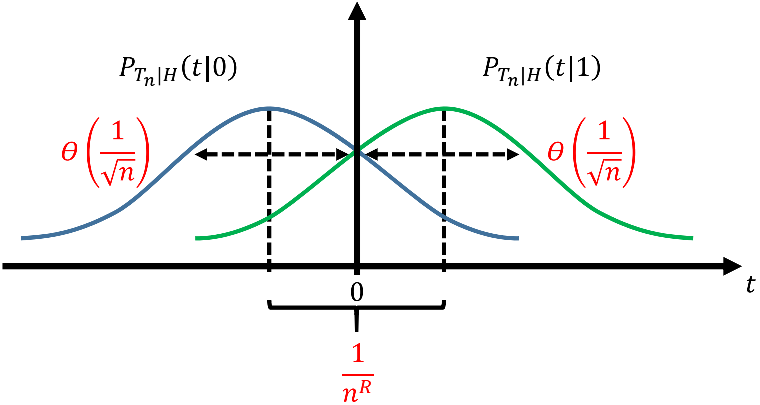
Case $R > \frac{1}{2}$**:**



Intuition via Central Limit Theorem

- For large n, $P_{T_n|H}(\cdot|0)$ and $P_{T_n|H}(\cdot|1)$ are Gaussian distributions
- $|\mathbb{E}[T_n|H=0] - \mathbb{E}[T_n|H=1]| = 1/n^R$
- 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 [EKPS00])

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

where $\|P - Q\|_{\mathsf{TV}} = \frac{1}{2}\|P - Q\|_1$ denotes the *total variation (TV) distance* between the distributions P and Q.

Second Moment Method for TV Distance

Lemma (2nd Moment Method [EKPS00])

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

where $\|P - Q\|_{\mathsf{TV}} = \frac{1}{2}\|P - Q\|_1$ denotes 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 [EKPS00])

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

where $\|P - Q\|_{\mathsf{TV}} = \frac{1}{2}\|P - Q\|_1$ denotes 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 [EKPS00])

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

where $\|P - Q\|_{\mathsf{TV}} = \frac{1}{2}\|P - Q\|_1$ denotes 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 [EKPS00])

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

where $\|P - Q\|_{\mathsf{TV}} = \frac{1}{2}\|P - Q\|_1$ denotes 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 [EKPS00])

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

where $\|P - Q\|_{\mathsf{TV}} = \frac{1}{2}\|P - Q\|_1$ denotes the *total variation (TV) distance* between the distributions P and Q.

Proof: Hammersley-Chapman-Robbins bound

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

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

Second Moment Method for TV Distance

Lemma (2nd Moment Method [EKPS00])

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

where $\|P - Q\|_{\mathsf{TV}} = \frac{1}{2}\|P - Q\|_1$ denotes the *total variation (TV) distance* between the distributions P and Q.

Proof:

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

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

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

BSC Achievability Proof

Proposition (BSC 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: Start with Le Cam's relation

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

BSC Achievability Proof

Proposition (BSC 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

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

BSC Achievability Proof

Proposition (BSC 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...

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

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

BSC Achievability Proof

Proposition (BSC 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: For any $0 < R < \frac{1}{2}$,

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

BSC Achievability Proof

Proposition (BSC 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 $0 < R < \frac{1}{2}$,

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

BSC Achievability Proof

Proposition (BSC 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, $\lim_{n\to\infty} P_{\text{ML}}^n = 0$. 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

Recall: Basic Definitions of Information Measures

- Consider discrete random variables $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ with joint distribution $P_{X,Y}$.

Recall: Basic Definitions of Information Measures

- Consider discrete random variables $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ with joint distribution $P_{X,Y}$.
- **Shannon Entropy:**

$$H(X) \triangleq -\sum_{x \in \mathcal{X}} P_X(x) \log(P_X(x))$$

Recall: Basic Definitions of Information Measures

- Consider discrete random variables $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ with joint distribution $P_{X,Y}$.
- **Shannon Entropy:**

$$H(X) \triangleq -\sum_{x \in \mathcal{X}} P_X(x) \log(P_X(x))$$

- **Conditional Shannon Entropy:**

$$H(X|Y) \triangleq -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P_{X,Y}(x,y) \log\left(P_{X|Y}(x|y)\right)$$

Recall: Basic Definitions of Information Measures

- Consider discrete random variables $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ with joint distribution $P_{X,Y}$.
- **Shannon Entropy:**

$$H(X) \triangleq - \sum_{x \in \mathcal{X}} P_X(x) \log(P_X(x))$$

- **Conditional Shannon Entropy:**

$$H(X|Y) \triangleq - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P_{X,Y}(x,y) \log\big(P_{X|Y}(x|y)\big)$$

- **Mutual Information:**

$$I(X;Y) \triangleq \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P_{X,Y}(x,y) \log\left(\frac{P_{X,Y}(x,y)}{P_X(y)P_Y(y)}\right)$$

Recall: Basic Definitions of Information Measures

- Consider discrete random variables $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ with joint distribution $P_{X,Y}$.
- **Shannon Entropy:**

$$H(X) \triangleq - \sum_{x \in \mathcal{X}} P_X(x) \log(P_X(x))$$

- **Conditional Shannon Entropy:**

$$H(X|Y) \triangleq - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P_{X,Y}(x,y) \log\left(P_{X|Y}(x|y)\right)$$

- **Mutual Information:**

$$I(X;Y) \triangleq \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P_{X,Y}(x,y) \log\left(\frac{P_{X,Y}(x,y)}{P_X(y)P_Y(y)}\right)$$

$$= H(X) - H(X|Y)$$

Recall: Two Information Inequalities

Consider discrete random variables X, Y, Z that form a Markov chain $X \to Y \to Z$.

Recall: Two Information Inequalities

Consider discrete random variables X, Y, Z that form a Markov chain $X \to Y \to Z$.

Lemma (Data Processing Inequality [CT06])

$$I(X;Z) \leq I(X;Y)$$

with equality if and only if Z is a *sufficient statistic* of Y for X, i.e., $X \to Z \to Y$ also forms a Markov chain.

Recall: Two Information Inequalities

Consider discrete random variables X, Y, Z that form a Markov chain $X \to Y \to Z$.

Lemma (Data Processing Inequality [CT06])

$$I(X;Z) \leq I(X;Y)$$

with equality if and only if Z is a *sufficient statistic* of Y for X, i.e., $X \to Z \to Y$ also forms a Markov chain.

Lemma (Fano's Inequality [CT06])

If X takes values in the finite alphabet $\mathcal{X}$, then

$$H(X|Z) \leq 1 + \mathbb{P}(X \neq Z) \log(|\mathcal{X}|)$$

where we perceive Z as an estimator for X based on Y.

BSC Converse Proof: Fano's Inequality Argument

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

BSC Converse Proof: Fano's Inequality Argument

- Consider the Markov chain $M \to X_1^n \to Z_1^n \to Y_1^n \to S_n \triangleq \sum_{i=1}^n Y_i \to \hat{M}$, 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 [CT06]: M is uniform

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

BSC Converse Proof: Fano's Inequality Argument

- Consider the Markov chain $M \to X_1^n \to Z_1^n \to Y_1^n \to S_n \triangleq \sum_{i=1}^n Y_i \to \hat{M}$, 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 [CT06]: Fano's inequality, data processing inequality

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

BSC Converse Proof: Fano's Inequality Argument

- Consider the Markov chain $M \to X_1^n \to Z_1^n \to Y_1^n \to S_n \triangleq \sum_{i=1}^n Y_i \to \hat{M}$, 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 [CT06]: sufficiency

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

BSC Converse Proof: Fano's Inequality Argument

- Consider the Markov chain $M \to X_1^n \to Z_1^n \to Y_1^n \to S_n \triangleq \sum_{i=1}^n Y_i \to \hat{M}$, 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 [CT06]: data processing inequality

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

BSC Converse Proof: Fano's Inequality Argument

- Consider the Markov chain $M \rightarrow X_1^n \rightarrow Z_1^n \rightarrow Y_1^n \rightarrow S_n \triangleq \sum_{i=1}^n Y_i \rightarrow \hat{M}$, 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 [CT06]:

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

- Divide by $\log(n)$:

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

BSC Converse Proof: Fano's Inequality Argument

- Consider the Markov chain $M \to X_1^n \to Z_1^n \to Y_1^n \to S_n \triangleq \sum_{i=1}^n Y_i \to \hat{M}$, 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 [CT06]:

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

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

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

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

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

BSC Converse Proof: CLT Approximation

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

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

BSC Converse Proof: CLT Approximation

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

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

BSC Converse Proof: CLT Approximation

Using [CT06, Problem 2.14], i.e., $\max\{H(X), H(Y)\} \leq H(X+Y)$ for $X \perp\!\!\!\perp Y$,

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

BSC Converse Proof: CLT Approximation

Approximate binomial entropy using CLT [ALY10]:

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

BSC Converse Proof: CLT Approximation

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

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

BSC Converse Proof: CLT Approximation

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

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

Hence, we have $R \leq \lim_{n\to\infty} I(X_1^n; S_n)/\log(n) = \frac{1}{2}$.

BSC Converse Proof: CLT Approximation

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

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

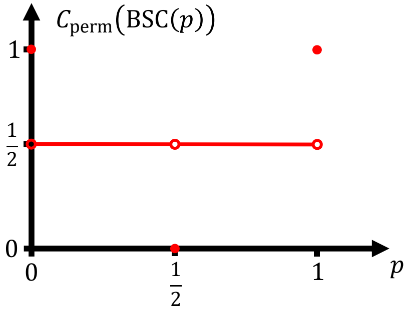
Hence, we have $R \leq \lim_{n\to\infty} I(X_1^n; S_n)/\log(n) = \frac{1}{2}$.

Proposition (BSC Converse)

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

Proposition (Pemutation Channel Capacity of BSC)

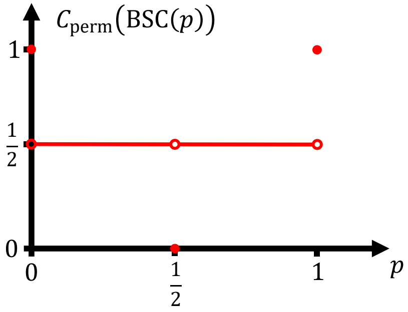
$$C_{\text{perm}}(\mathsf{BSC}(p)) = \begin{cases} 1, & \text{for } p = 0, 1 \\ \frac{1}{2}, & \text{for } p \in \left(0, \frac{1}{2}\right) \cup \left(\frac{1}{2}, 1\right) \\ 0, & \text{for } p = \frac{1}{2} \end{cases}$$



Information Capacity of the BSC Permutation Channel

Proposition (Pemutation Channel Capacity of BSC)

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



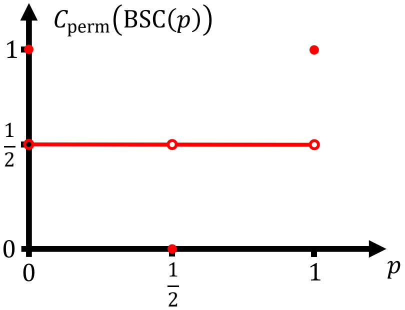
Remarks:

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

Information Capacity of the BSC Permutation Channel

Proposition (Pemutation Channel Capacity of BSC)

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



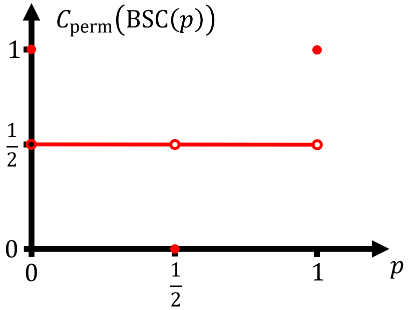
Remarks:

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

Information Capacity of the BSC Permutation Channel

Proposition (Pemutation Channel Capacity of BSC)

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

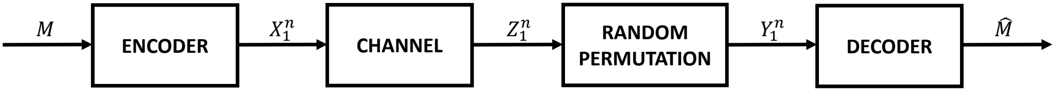


Remarks:

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

Outline

Recall General Problem



- Average probability of error $P_{\text{error}}^n \triangleq \mathbb{P}(M \neq \hat{M})$
- "Rate" of coding scheme (f_n, g_n) is $R \triangleq \dfrac{\log(|\mathcal{M}|)}{\log(n)}$
- Rate $R \geq 0$ is achievable $\Leftrightarrow \exists \{(f_n, g_n)\}_{n \in \mathbb{N}}$ such that $\lim_{n\to\infty} P_{\text{error}}^n = 0$

Definition (Permutation Channel Capacity)

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

Main Question

What is the permutation channel capacity of a general $P_{Z|X}$?

Achievability: Coding Scheme

- Let $r = \text{rank}(P_{Z|X})$ and $k = \lfloor \sqrt{n} \rfloor$

Achievability: Coding Scheme

- Let $r = \text{rank}(P_{Z|X})$ and $k = \lfloor \sqrt{n} \rfloor$
- Consider $\mathcal{X}' \subseteq \mathcal{X}$ with $|\mathcal{X}'| = r$ such that $\{P_{Z|X}(\cdot|x) : x \in \mathcal{X}'\}$ are linearly independent

- Let $r = \text{rank}(P_{Z|X})$ and $k = \lfloor \sqrt{n} \rfloor$
- Consider $\mathcal{X}' \subseteq \mathcal{X}$ with $|\mathcal{X}'| = r$ such that $\{P_{Z|X}(\cdot|x) : x \in \mathcal{X}'\}$ are linearly independent
- **Message set:**

$$\mathcal{M} \triangleq \left\{ p = (p(x) : x \in \mathcal{X}') \in (\mathbb{Z}_+)^{\mathcal{X}'} : \sum_{x \in \mathcal{X}'} p(x) = k \right\}$$

Achievability: Coding Scheme

- Let $r = \text{rank}(P_{Z|X})$ and $k = \lfloor \sqrt{n} \rfloor$
- Consider $\mathcal{X}' \subseteq \mathcal{X}$ with $|\mathcal{X}'| = r$ such that $\{P_{Z|X}(\cdot|x) : x \in \mathcal{X}'\}$ are linearly independent
- **Message set:**

$$\mathcal{M} \triangleq \left\{ p = (p(x) : x \in \mathcal{X}') \in (\mathbb{Z}_+)^{\mathcal{X}'} : \sum_{x \in \mathcal{X}'} p(x) = k \right\}$$

where $|\mathcal{M}| = \binom{k+r-1}{r-1} = \Theta\left(n^{\frac{r-1}{2}}\right)$

Achievability: Coding Scheme

- Let $r = \text{rank}(P_{Z|X})$ and $k = \lfloor \sqrt{n} \rfloor$
- Consider $\mathcal{X}' \subseteq \mathcal{X}$ with $|\mathcal{X}'| = r$ such that $\{P_{Z|X}(\cdot|x) : x \in \mathcal{X}'\}$ are linearly independent
- **Message set:**

$$\mathcal{M} \triangleq \left\{ p = (p(x) : x \in \mathcal{X}') \in (\mathbb{Z}_+)^{\mathcal{X}'} : \sum_{x \in \mathcal{X}'} p(x) = k \right\}$$

where $|\mathcal{M}| = \binom{k+r-1}{r-1} = \Theta\left(n^{\frac{r-1}{2}}\right)$

- **Randomized Encoder:**

$$\forall p \in \mathcal{M}, \quad f_n(p) = X_1^n \overset{\text{i.i.d.}}{\sim} P_X \quad \text{where} \quad P_X(x) = \begin{cases} \frac{p(x)}{k}, & \text{for } x \in \mathcal{X}' \\ 0, & \text{for } x \in \mathcal{X} \backslash \mathcal{X}' \end{cases}$$

Achievability: Coding Scheme

- Let stochastic matrix $\tilde{P}_{Z|X} \in \mathbb{R}^{r \times |\mathcal{Y}|}$ have rows $\{P_{Z|X}(\cdot|x) : x \in \mathcal{X}'\}$
- Let $\tilde{P}_{Z|X}^{\dagger}$ denote its *Moore-Penrose pseudoinverse*

Achievability: Coding Scheme

- Let stochastic matrix $\tilde{P}_{Z|X} \in \mathbb{R}^{r \times |\mathcal{Y}|}$ have rows $\{P_{Z|X}(\cdot|x) : x \in \mathcal{X}'\}$
- Let $\tilde{P}^{\dagger}_{Z|X}$ denote its *Moore-Penrose pseudoinverse*
- **(Sub-optimal) Thresholding Decoder:** For any $y_1^n \in \mathcal{Y}^n$,
  Step 1: Construct its type/empirical distribution/histogram

$$\forall y \in \mathcal{Y}, \quad \hat{P}_{y_1^n}(y) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{y_i = y\}$$

Achievability: Coding Scheme

- Let stochastic matrix $\tilde{P}_{Z|X} \in \mathbb{R}^{r \times |\mathcal{Y}|}$ have rows $\{P_{Z|X}(\cdot|x) : x \in \mathcal{X}'\}$
- Let $\tilde{P}^{\dagger}_{Z|X}$ denote its *Moore-Penrose pseudoinverse*
- **(Sub-optimal) Thresholding Decoder:** For any $y_1^n \in \mathcal{Y}^n$,

  Step 1: Construct its type/empirical distribution/histogram

$$\forall y \in \mathcal{Y}, \quad \hat{P}_{y_1^n}(y) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{y_i = y\}$$

  Step 2: Generate estimate $\hat{p} \in (\mathbb{Z}_+)^{\mathcal{X}'}$ with components

$$\forall x \in \mathcal{X}', \quad \hat{p}(x) = \underset{j \in \{0,\ldots,k\}}{\arg\min} \left| \sum_{y \in \mathcal{Y}} \hat{P}_{y_1^n}(y) \left[\tilde{P}^{\dagger}_{Z|X}\right]_{y,x} - \frac{j}{k} \right|$$

Achievability: Coding Scheme

- Let stochastic matrix $\tilde{P}_{Z|X} \in \mathbb{R}^{r \times |\mathcal{Y}|}$ have rows $\{P_{Z|X}(\cdot|x) : x \in \mathcal{X}'\}$
- Let $\tilde{P}^{\dagger}_{Z|X}$ denote its *Moore-Penrose pseudoinverse*
- **(Sub-optimal) Thresholding Decoder:** For any $y_1^n \in \mathcal{Y}^n$,

  Step 1: Construct its type/empirical distribution/histogram

$$\forall y \in \mathcal{Y}, \quad \hat{P}_{y_1^n}(y) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{y_i = y\}$$

  Step 2: Generate estimate $\hat{p} \in (\mathbb{Z}_+)^{\mathcal{X}'}$ with components

$$\forall x \in \mathcal{X}', \quad \hat{p}(x) = \operatorname*{arg\,min}_{j \in \{0,\ldots,k\}} \left| \sum_{y \in \mathcal{Y}} \hat{P}_{y_1^n}(y) \left[\tilde{P}^{\dagger}_{Z|X}\right]_{y,x} - \frac{j}{k} \right|$$

  Step 3: Output decoded message

$$g_n(y_1^n) = \begin{cases} \hat{p}, & \text{if } \hat{p} \in \mathcal{M} \\ \text{error}, & \text{otherwise} \end{cases}$$

Achievability: Rank Bound

Theorem (Rank Bound)

For any channel $P_{Z|X}$:

$$C_{\text{perm}}(P_{Z|X}) \geq \frac{\text{rank}(P_{Z|X}) - 1}{2} .$$

Remarks about Coding Scheme:

- Showing $\lim_{n\to\infty} P^n_{\text{error}} = 0$ proves theorem.

Achievability: Rank Bound

Theorem (Rank Bound)

For any channel $P_{Z|X}$:

$$C_{\text{perm}}(P_{Z|X}) \geq \frac{\text{rank}(P_{Z|X}) - 1}{2} .$$

Remarks about Coding Scheme:

- Showing $\lim_{n\to\infty} P^n_{\text{error}} = 0$ proves theorem.
- *Intuition*: Conditioned on $M = p$, $\hat{P}_{Y_1^n} \approx P_Z$ with high probability as $n \to \infty$.

Achievability: Rank Bound

Theorem (Rank Bound)

For any channel $P_{Z|X}$:

$$C_{\text{perm}}(P_{Z|X}) \geq \frac{\text{rank}(P_{Z|X}) - 1}{2} .$$

Remarks about Coding Scheme:

- Showing $\lim_{n\to\infty} P_{\text{error}}^n = 0$ proves theorem.
- *Intuition*: Conditioned on $M = p$, $\hat{P}_{Y_1^n} \approx P_Z$ with high probability as $n \to \infty$. Hence, $\sum_{y\in\mathcal{Y}} \hat{P}_{Y_1^n}(y)\big[\tilde{P}_{Z|X}^{\dagger}\big]_{y,x} \approx P_X(x)$ for all $x \in \mathcal{X}'$ with high probability.

Achievability: Rank Bound

Theorem (Rank Bound)

For any channel $P_{Z|X}$:

$$C_{\text{perm}}(P_{Z|X}) \geq \frac{\text{rank}(P_{Z|X}) - 1}{2} .$$

Remarks about Coding Scheme:

- Showing $\lim_{n\to\infty} P_{\text{error}}^n = 0$ proves theorem.
- *Intuition*: Conditioned on $M = p$, $\hat{P}_{Y_1^n} \approx P_Z$ with high probability as $n \to \infty$.
  Hence, $\sum_{y\in\mathcal{Y}} \hat{P}_{Y_1^n}(y)\big[\tilde{P}_{Z|X}^{\dagger}\big]_{y,x} \approx P_X(x)$ for all $x \in \mathcal{X}'$ with high probability.
- *Computational complexity*: Decoder has $O(n)$ running time.

Achievability: Rank Bound

Theorem (Rank Bound)

For any channel $P_{Z|X}$:

$$C_{\text{perm}}(P_{Z|X}) \geq \frac{\text{rank}(P_{Z|X}) - 1}{2} \, .$$

Remarks about Coding Scheme:

- Showing $\lim_{n\to\infty} P^n_{\text{error}} = 0$ proves theorem.
- *Intuition*: Conditioned on $M = p$, $\hat{P}_{Y_1^n} \approx P_Z$ with high probability as $n \to \infty$. Hence, $\sum_{y\in\mathcal{Y}} \hat{P}_{Y_1^n}(y) \big[\tilde{P}^{\dagger}_{Z|X}\big]_{y,x} \approx P_X(x)$ for all $x \in \mathcal{X}'$ with high probability.
- *Computational complexity*: Decoder has $O(n)$ running time.
- *Probabilistic method*: Good deterministic codes exist.

Achievability: Rank Bound

Theorem (Rank Bound)

For any channel $P_{Z|X}$:

$$C_{\text{perm}}(P_{Z|X}) \geq \frac{\text{rank}(P_{Z|X}) - 1}{2} .$$

Remarks about Coding Scheme:

- Showing $\lim_{n\rightarrow\infty} P^n_{\text{error}} = 0$ proves theorem.
- *Intuition*: Conditioned on $M = p$, $\hat{P}_{Y_1^n} \approx P_Z$ with high probability as $n \rightarrow \infty$. Hence, $\sum_{y\in\mathcal{Y}} \hat{P}_{Y_1^n}(y)\left[\tilde{P}^{\dagger}_{Z|X}\right]_{y,x} \approx P_X(x)$ for all $x \in \mathcal{X}'$ with high probability.
- *Computational complexity*: Decoder has $O(n)$ running time.
- *Probabilistic method*: Good deterministic codes exist.
- *Expurgation*: Achievability bound holds under maximal probability of error criterion.

Outline

Converse: Output Alphabet Bound

Theorem (Output Alphabet Bound)

For any entry-wise *strictly positive* channel $P_{Z|X} > 0$:

$$C_{\text{perm}}(P_{Z|X}) \leq \frac{|\mathcal{Y}| - 1}{2}.$$

Converse: Output Alphabet Bound

Theorem (Output Alphabet Bound)

For any entry-wise *strictly positive* channel $P_{Z|X} > 0$:

$$C_{\mathsf{perm}}(P_{Z|X}) \leq \frac{|\mathcal{Y}| - 1}{2} .$$

Remarks:

- Proof hinges on Fano's inequality and CLT approximation of binomial entropy.

Converse: Output Alphabet Bound

Theorem (Output Alphabet Bound)

For any entry-wise *strictly positive* channel $P_{Z|X} > 0$:

$$C_{\text{perm}}(P_{Z|X}) \leq \frac{|\mathcal{Y}| - 1}{2}.$$

Remarks:

- Proof hinges on Fano's inequality and CLT approximation of binomial entropy.
- What if $|\mathcal{X}|$ is much smaller than $|\mathcal{Y}|$?

Converse: Output Alphabet Bound

Theorem (Output Alphabet Bound)

For any entry-wise *strictly positive* channel $P_{Z|X} > 0$:

$$C_{\text{perm}}(P_{Z|X}) \leq \frac{|\mathcal{Y}| - 1}{2}.$$

Remarks:

- Proof hinges on Fano's inequality and CLT approximation of binomial entropy.
- What if $|\mathcal{X}|$ is much smaller than $|\mathcal{Y}|$?
- **Want:** Converse bound in terms of input alphabet size.

Converse: Effective Input Alphabet Bound

Theorem (Effective Input Alphabet Bound)

For any entry-wise *strictly positive* channel $P_{Z|X} > 0$:

$$C_{\text{perm}}(P_{Z|X}) \leq \frac{\text{ext}(P_{Z|X}) - 1}{2}$$

where $\text{ext}(P_{Z|X})$ denotes the number of *extreme points* of $\text{conv}\{P_{Z|X}(\cdot|x) : x \in \mathcal{X}\}$.

Converse: Effective Input Alphabet Bound

Theorem (Effective Input Alphabet Bound)

For any entry-wise *strictly positive* channel $P_{Z|X} > 0$:

$$C_{\text{perm}}(P_{Z|X}) \leq \frac{\text{ext}(P_{Z|X}) - 1}{2}$$

where $\text{ext}(P_{Z|X})$ denotes the number of *extreme points* of $\text{conv}\{P_{Z|X}(\cdot|x) : x \in \mathcal{X}\}$.

Remarks:

- *Effective input alphabet size*: $\text{rank}(P_{Z|X}) \leq \text{ext}(P_{Z|X}) \leq |\mathcal{X}|$.

Converse: Effective Input Alphabet Bound

Theorem (Effective Input Alphabet Bound)

For any entry-wise *strictly positive* channel $P_{Z|X} > 0$:

$$C_{\text{perm}}(P_{Z|X}) \leq \frac{\text{ext}(P_{Z|X}) - 1}{2}$$

where $\text{ext}(P_{Z|X})$ denotes the number of *extreme points* of $\text{conv}\{P_{Z|X}(\cdot|x) : x \in \mathcal{X}\}$.

Remarks:

- *Effective input alphabet size*: $\text{rank}(P_{Z|X}) \leq \text{ext}(P_{Z|X}) \leq |\mathcal{X}|$.
- For any channel $P_{Z|X} > 0$, $C_{\text{perm}}(P_{Z|X}) \leq \left(\min\{\text{ext}(P_{Z|X}), |\mathcal{Y}|\} - 1\right)/2$.

Converse: Effective Input Alphabet Bound

Theorem (Effective Input Alphabet Bound)

For any entry-wise *strictly positive* channel $P_{Z|X} > 0$:

$$C_{\text{perm}}(P_{Z|X}) \leq \frac{\text{ext}(P_{Z|X}) - 1}{2}$$

where $\text{ext}(P_{Z|X})$ denotes the number of *extreme points* of $\text{conv}\{P_{Z|X}(\cdot|x) : x \in \mathcal{X}\}$.

Remarks:

- *Effective input alphabet size*: $\text{rank}(P_{Z|X}) \leq \text{ext}(P_{Z|X}) \leq |\mathcal{X}|$.
- For any channel $P_{Z|X} > 0$, $C_{\text{perm}}(P_{Z|X}) \leq \big(\min\{\text{ext}(P_{Z|X}), |\mathcal{Y}|\} - 1\big)/2$.
- For any *general* channel $P_{Z|X}$, $C_{\text{perm}}(P_{Z|X}) \leq \min\{\text{ext}(P_{Z|X}), |\mathcal{Y}|\} - 1$.

Converse: Effective Input Alphabet Bound

Theorem (Effective Input Alphabet Bound)

For any entry-wise *strictly positive* channel $P_{Z|X} > 0$:

$$C_{\text{perm}}(P_{Z|X}) \leq \frac{\text{ext}(P_{Z|X}) - 1}{2}$$

where $\text{ext}(P_{Z|X})$ denotes the number of *extreme points* of $\text{conv}\{P_{Z|X}(\cdot|x) : x \in \mathcal{X}\}$.

Remarks:

- *Effective input alphabet size*: $\text{rank}(P_{Z|X}) \leq \text{ext}(P_{Z|X}) \leq |\mathcal{X}|$.
- For any channel $P_{Z|X} > 0$, $C_{\text{perm}}(P_{Z|X}) \leq \big(\min\{\text{ext}(P_{Z|X}), |\mathcal{Y}|\} - 1\big)/2$.
- For any *general* channel $P_{Z|X}$, $C_{\text{perm}}(P_{Z|X}) \leq \min\{\text{ext}(P_{Z|X}), |\mathcal{Y}|\} - 1$.
- **How do we prove above theorem?**

Brief Digression: Degradation

Definition (Degradation/Blackwell Order [Bla51], [She51], [Ste51], [Cov72], [Ber73])

Given channels $P_{Z_1|X}$ and $P_{Z_2|X}$ with common input alphabet $\mathcal{X}$, $P_{Z_2|X}$ is a degraded version of $P_{Z_1|X}$ if $P_{Z_2|X} = P_{Z_1|X} P_{Z_2|Z_1}$ for some channel $P_{Z_2|Z_1}$.

Brief Digression: Degradation

Definition (Degradation/Blackwell Order [Bla51], [She51], [Ste51], [Cov72], [Ber73])

Given channels $P_{Z_1|X}$ and $P_{Z_2|X}$ with common input alphabet $\mathcal{X}$, $P_{Z_2|X}$ is a degraded version of $P_{Z_1|X}$ if $P_{Z_2|X} = P_{Z_1|X} P_{Z_2|Z_1}$ for some channel $P_{Z_2|Z_1}$.

Theorem (Blackwell-Sherman-Stein [Bla51], [She51], [Ste51])

The *observation model* $P_{Z_2|X}$ is a degraded version of $P_{Z_1|X}$ if and only if for every prior distribution P_X, and every loss function $L : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, the *Bayes risks* satisfy:

$$\min_{f(\cdot)} \mathbb{E}\left[L(X, f(Z_1))\right] \leq \min_{g(\cdot)} \mathbb{E}\left[L(X, g(Z_2))\right]$$

where the minima are over all randomized estimators of X.

Brief Digression: Symmetric Channels

Definition (q-ary Symmetric Channel)

A q-ary symmetric channel, denoted q-SC(δ), with total crossover probability $\delta \in [0, 1]$ and alphabet $\mathcal{X}$ where $|\mathcal{X}| = q$, is given by the doubly stochastic matrix:

$$W_\delta \triangleq \begin{bmatrix} 1-\delta & \frac{\delta}{q-1} & \cdots & \frac{\delta}{q-1} \\ \frac{\delta}{q-1} & 1-\delta & \cdots & \frac{\delta}{q-1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\delta}{q-1} & \frac{\delta}{q-1} & \cdots & 1-\delta \end{bmatrix}.$$

Brief Digression: Symmetric Channels

Definition (q-ary Symmetric Channel)

A q-ary symmetric channel, denoted q-SC(δ), with total crossover probability $\delta \in [0,1]$ and alphabet $\mathcal{X}$ where $|\mathcal{X}| = q$, is given by the doubly stochastic matrix:

$$W_\delta \triangleq \begin{bmatrix} 1-\delta & \frac{\delta}{q-1} & \cdots & \frac{\delta}{q-1} \\ \frac{\delta}{q-1} & 1-\delta & \cdots & \frac{\delta}{q-1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\delta}{q-1} & \frac{\delta}{q-1} & \cdots & 1-\delta \end{bmatrix}.$$

Proposition (Degradation by Symmetric Channels)

Given channel $P_{Z|X}$ with $\nu = \min_{x \in \mathcal{X}, y \in \mathcal{Y}} P_{Z|X}(y|x)$,

if $0 \leq \delta \leq \dfrac{\nu}{1 - \nu + \frac{\nu}{q-1}}$, then $P_{Z|X}$ is a degraded version of q-SC(δ).

Brief Digression: Symmetric Channels

Proposition (Degradation by Symmetric Channels)

Given channel $P_{Z|X}$ with $\nu = \min_{x \in \mathcal{X}, y \in \mathcal{Y}} P_{Z|X}(y|x)$,

if $0 \leq \delta \leq \dfrac{\nu}{1 - \nu + \frac{\nu}{q-1}}$, then $P_{Z|X}$ is a degraded version of q-SC(δ).

Remarks:

- Prop follows from computing extremal δ such that $W_\delta^{-1} P_{Z|X}$ is row stochastic.

Brief Digression: Symmetric Channels

Proposition (Degradation by Symmetric Channels)

Given channel $P_{Z|X}$ with $\nu = \min_{x \in \mathcal{X},\, y \in \mathcal{Y}} P_{Z|X}(y|x)$,

if $0 \leq \delta \leq \dfrac{\nu}{1 - \nu + \frac{\nu}{q-1}}$, then $P_{Z|X}$ is a degraded version of q-SC(δ).

Remarks:

- Prop follows from computing extremal δ such that $W_\delta^{-1} P_{Z|X}$ is row stochastic.
- Bound on δ can be improved when more is known about $P_{Z|X}$:

Brief Digression: Symmetric Channels

Proposition (Degradation by Symmetric Channels)

Given channel $P_{Z|X}$ with $\nu = \min_{x \in \mathcal{X}, y \in \mathcal{Y}} P_{Z|X}(y|x)$,

if $0 \leq \delta \leq \dfrac{\nu}{1 - \nu + \frac{\nu}{q-1}}$, then $P_{Z|X}$ is a degraded version of q-SC(δ).

Remarks:

- Prop follows from computing extremal δ such that $W_\delta^{-1} P_{Z|X}$ is row stochastic.
- Bound on δ can be improved when more is known about $P_{Z|X}$:
  - Markov chain [MP18]: $\delta \leq \nu / \left(1 - (q-1)\nu + \frac{\nu}{q-1}\right)$.

Brief Digression: Symmetric Channels

Proposition (Degradation by Symmetric Channels)

Given channel $P_{Z|X}$ with $\nu = \min_{x \in \mathcal{X}, y \in \mathcal{Y}} P_{Z|X}(y|x)$,

if $0 \leq \delta \leq \dfrac{\nu}{1 - \nu + \frac{\nu}{q-1}}$, then $P_{Z|X}$ is a degraded version of q-SC(δ).

Remarks:

- Prop follows from computing extremal δ such that $W_\delta^{-1} P_{Z|X}$ is row stochastic.
- Bound on δ can be improved when more is known about $P_{Z|X}$:
  - Markov chain [MP18]: $\delta \leq \nu / \left(1 - (q-1)\nu + \frac{\nu}{q-1}\right)$.
  - Additive noise channel on Abelian group $\mathcal{X}$ [MP18]: $\delta \leq (q-1)\nu$.

Brief Digression: Symmetric Channels

Proposition (Degradation by Symmetric Channels)

Given channel $P_{Z|X}$ with $\nu = \min_{x \in \mathcal{X}, y \in \mathcal{Y}} P_{Z|X}(y|x)$,

if $0 \leq \delta \leq \dfrac{\nu}{1 - \nu + \frac{\nu}{q-1}}$, then $P_{Z|X}$ is a degraded version of q-SC(δ).

Remarks:

- Prop follows from computing extremal δ such that $W_\delta^{-1} P_{Z|X}$ is row stochastic.
- Bound on δ can be improved when more is known about $P_{Z|X}$:
  - Markov chain [MP18]: $\delta \leq \nu / \left(1 - (q-1)\nu + \frac{\nu}{q-1}\right)$.
  - Additive noise channel on Abelian group $\mathcal{X}$ [MP18]: $\delta \leq (q-1)\nu$.
  - Alternative bounds for Markov chains [MOS13].

Brief Digression: Symmetric Channels

Proposition (Degradation by Symmetric Channels)

Given channel $P_{Z|X}$ with $\nu = \min_{x \in \mathcal{X},\, y \in \mathcal{Y}} P_{Z|X}(y|x)$,

if $0 \leq \delta \leq \dfrac{\nu}{1 - \nu + \frac{\nu}{q-1}}$, then $P_{Z|X}$ is a degraded version of q-SC(δ).

Remarks:

- Prop follows from computing extremal δ such that $W_\delta^{-1} P_{Z|X}$ is row stochastic.
- Bound on δ can be improved when more is known about $P_{Z|X}$:
  - Markov chain [MP18]: $\delta \leq \nu / \left(1 - (q-1)\nu + \frac{\nu}{q-1}\right)$.
  - Additive noise channel on Abelian group $\mathcal{X}$ [MP18]: $\delta \leq (q-1)\nu$.
  - Alternative bounds for Markov chains [MOS13].
- Many applications in information theory, statistics, and probability [MP18], [MOS13].

Proof Idea: Degradation by Symmetric Channels

Theorem (Effective Input Alphabet Bound)

For any entry-wise strictly positive channel $P_{Z|X} > 0$:

$$C_{\text{perm}}(P_{Z|X}) \leq \frac{\text{ext}(P_{Z|X}) - 1}{2}.$$

Proof Sketch:

- *Degradation by symmetric channels* + *tensorization* of degradation + data processing

$$\Rightarrow \quad I(X_1^n; Y_1^n) \leq I(X_1^n; \tilde{Y}_1^n)$$

where Y_1^n and $\tilde{Y}_1^n$ are outputs of permutation channels with $P_{Z|X}$ and q-SC(δ), resp.

Proof Idea: Degradation by Symmetric Channels

Theorem (Effective Input Alphabet Bound)

For any entry-wise strictly positive channel $P_{Z|X} > 0$:

$$C_{\text{perm}}(P_{Z|X}) \leq \frac{\text{ext}(P_{Z|X}) - 1}{2} .$$

Proof Sketch:

- *Degradation by symmetric channels* + *tensorization* of degradation + data processing

$$\Rightarrow \quad I(X_1^n; Y_1^n) \leq I(X_1^n; \tilde{Y}_1^n)$$

  where Y_1^n and $\tilde{Y}_1^n$ are outputs of permutation channels with $P_{Z|X}$ and q-SC(δ), resp.
- *Convexity* of KL divergence $\Rightarrow$ Reduce $|\mathcal{X}|$ to $\text{ext}(P_{Z|X})$.

Proof Idea: Degradation by Symmetric Channels

Theorem (Effective Input Alphabet Bound)

For any entry-wise strictly positive channel $P_{Z|X} > 0$:

$$C_{\text{perm}}(P_{Z|X}) \leq \frac{\text{ext}(P_{Z|X}) - 1}{2}.$$

Proof Sketch:

- *Degradation by symmetric channels* + *tensorization* of degradation + data processing

$$\Rightarrow \quad I(X_1^n; Y_1^n) \leq I(X_1^n; \tilde{Y}_1^n)$$

  where Y_1^n and $\tilde{Y}_1^n$ are outputs of permutation channels with $P_{Z|X}$ and q-SC(δ), resp.
- *Convexity* of KL divergence $\Rightarrow$ Reduce $|\mathcal{X}|$ to $\text{ext}(P_{Z|X})$.
- *Fano argument* of output alphabet bound $\Rightarrow$ effective input alphabet bound.

Outline

Strictly Positive and "Full Rank" Channels

Achievability and converse bounds yield:

Theorem (Strictly Positive and "Full Rank" Channels)

For any entry-wise *strictly positive* channel $P_{Z|X} > 0$ that is *"full rank"* in the sense that $r \triangleq \text{rank}(P_{Z|X}) = \min\{\text{ext}(P_{Z|X}), |\mathcal{Y}|\}$:

$$C_{\text{perm}}(P_{Z|X}) = \frac{r-1}{2}\,.$$

Strictly Positive and "Full Rank" Channels

Achievability and converse bounds yield:

Theorem (Strictly Positive and "Full Rank" Channels)

For any entry-wise *strictly positive* channel $P_{Z|X} > 0$ that is *"full rank"* in the sense that $r \triangleq \text{rank}(P_{Z|X}) = \min\{\text{ext}(P_{Z|X}), |\mathcal{Y}|\}$:

$$C_{\text{perm}}(P_{Z|X}) = \frac{r-1}{2} .$$

Recall Example: C_{perm} of non-trivial binary symmetric channel is $\frac{1}{2}$.

Main Result:
For any entry-wise *strictly positive* channel $P_{Z|X} > 0$:

$$\frac{\operatorname{rank}(P_{Z|X}) - 1}{2} \leq C_{\text{perm}}(P_{Z|X}) \leq \frac{\min\{\operatorname{ext}(P_{Z|X}), |\mathcal{Y}|\} - 1}{2}.$$

Conclusion

Main Result:

For any entry-wise *strictly positive* channel $P_{Z|X} > 0$:

$$\frac{\text{rank}(P_{Z|X}) - 1}{2} \leq C_{\text{perm}}(P_{Z|X}) \leq \frac{\min\{\text{ext}(P_{Z|X}), |\mathcal{Y}|\} - 1}{2}.$$

Future Directions:

- Characterize C_{perm} of all (entry-wise strictly positive) channels.

Conclusion

Main Result:

For any entry-wise *strictly positive* channel $P_{Z|X} > 0$:

$$\frac{\operatorname{rank}(P_{Z|X}) - 1}{2} \leq C_{\text{perm}}(P_{Z|X}) \leq \frac{\min\{\operatorname{ext}(P_{Z|X}), |\mathcal{Y}|\} - 1}{2}.$$

Future Directions:

- Characterize C_{perm} of all (entry-wise strictly positive) channels.
- Perform error exponent analysis (i.e., tight bounds on P_{error}^n).

Conclusion

Main Result:

For any entry-wise *strictly positive* channel $P_{Z|X} > 0$:

$$\frac{\text{rank}(P_{Z|X}) - 1}{2} \le C_{\text{perm}}(P_{Z|X}) \le \frac{\min\{\text{ext}(P_{Z|X}), |\mathcal{Y}|\} - 1}{2}.$$

Future Directions:

- Characterize C_{perm} of all (entry-wise strictly positive) channels.
- Perform error exponent analysis (i.e., tight bounds on P^n_{error}).
- Prove strong converse results (i.e., phase transition for P^n_{error}).

Conclusion

Main Result:

For any entry-wise *strictly positive* channel $P_{Z|X} > 0$:

$$\frac{\text{rank}(P_{Z|X}) - 1}{2} \leq C_{\text{perm}}(P_{Z|X}) \leq \frac{\min\{\text{ext}(P_{Z|X}), |\mathcal{Y}|\} - 1}{2}.$$

Future Directions:

- Characterize C_{perm} of all (entry-wise strictly positive) channels.
- Perform error exponent analysis (i.e., tight bounds on P^n_{error}).
- Prove strong converse results (i.e., phase transition for P^n_{error}).
- Perform finite blocklength analysis (i.e., exact asymptotics for maximum achievable $|\mathcal{M}|$).

Conclusion

Main Result:

For any entry-wise *strictly positive* channel $P_{Z|X} > 0$:

$$\frac{\text{rank}(P_{Z|X}) - 1}{2} \leq C_{\text{perm}}(P_{Z|X}) \leq \frac{\min\{\text{ext}(P_{Z|X}), |\mathcal{Y}|\} - 1}{2} .$$

Future Directions:

- Characterize C_{perm} of all (entry-wise strictly positive) channels.
- Perform error exponent analysis (i.e., tight bounds on P^n_{error}).
- Prove strong converse results (i.e., phase transition for P^n_{error}).
- Perform finite blocklength analysis (i.e., exact asymptotics for maximum achievable $|\mathcal{M}|$).
- Analyze permutation channels with more complex probability models in the random permutation block.

References

This talk was based on:

- A. Makur, "**Information capacity of BSC and BEC permutation channels**," in *Proceedings of the 56th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, USA, October 2-5 2018, pp. 1112–1119.
- A. Makur, "**Bounds on permutation channel capacity**," in *Proceedings of the IEEE International Symposium on Information Theory* (ISIT), Los Angeles, CA, USA, June 21-26 2020.
- A. Makur, "**Coding theorems for noisy permutation channels**," accepted to *IEEE Transactions on Information Theory*, July 2020.

Thank You!