Lecture 3. Channel Capacity

Lin Zhang

Tsinghua-Berkeley Shenzhen Institute

Shenzhen, China, 2017

Outline

1. Channel Capacity and its Properties
   - Definition of Channel Capacity
   - Discrete Memoryless Channel
   - Channel Capacity and Its Properties
2. Discrete Memoryless Channel
3. Continuous and Analog Channel
4. Joint Typicality
5. Channel Coding
6. Summary

Communications and Channels

- Communication: Reliable transmission of messages from one peer to another.
- Shannon's channel model in 1948
- Types of channels, according to the continuity of values in both time and amplitude, e.g., discrete channel/digital channel, continuous channel and analog channel.

Discrete Memoryless Channel (DMC)

- **Definition 3.1** A discrete memoryless channel is defined with the set of input alphabets $\mathcal{X}$, the set of output alphabets $\mathcal{Y}$ and one-step transition probability matrix $\mathbf{Q} = \{q(y_j|x_i)\}_{i,j}$.
- Note that $q(y|x)$ is the conditional probability of y given x, and the channel matrix $\mathbf{Q}$ contains $|\mathcal{X}|$ rows and $|\mathcal{Y}|$ columns.
- **Theorem 3.1** Let the input of channel to be $\mathbf{X} = (X_1, X_2, \dots, X_n)$ and the output values $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)$. The discrete memoryless channel satisfies

$$I(\mathbf{X}; \mathbf{Y}) \leq \sum_{i=1}^{n} I(X_i; Y_j). \tag{1}$$

Channel Capacity

- **Definition 3.2** The capacity of a channel is defined as

$$C = \max_{p(x)} I(X;Y). \tag{2}$$

- The **information-theoretical** channel capacity is defined as the max mutual information between the input and output of the channel, over all possible input distribution $p(x)$ given a specific communication channel.
- For discrete memoryless channels, the equivalent formulation is

$$C = \max_{p(x)} I(\mathbf{p}, \mathbf{Q}), \tag{3}$$

where the mutual information is rewritten as the function of input distribution $\mathbf{p}$ and transition matrix $\mathbf{Q}$.

Properties of Channel Capacity

- **Theorem 3.2** The basic properties of C is listed as follows
  1. Non-negative: $C \geq 0$.
  2. Bounded: $C \leq \log |\mathcal{X}|$, and $C \leq \log |\mathcal{Y}|$.
- More importantly, $I(X;Y)$ is a continuously concave function over $p(x)$, implying that the maximum value, i.e, C, can be obtained using convex optimization.
- A classical example of channel models: noisy type-writer.
- **Definition 3.3** If the rows and columns of transition probability matrix are permutations of themselves, respectively, the channel is symmetric. If the rows are permutations of each other and the sum of columns are equal, the channel is weakly symmetric.

Outline

1. Channel Capacity and its Properties

2. Discrete Memoryless Channel
   - Simple Examples of DMC
   - Compute the DMC Capacity
   - The pmf to achieve the channel capacity
   - Cascading of Channels
   - Parallel Channels

3. Continuous and Analog Channel

4. Joint Typicality

5. Channel Coding

6. Summary

Simple examples of DMC

- **Theorem 3.3** The capacity of Binary Symmetric Channel (BSC) is

$$C = 1 - H(\epsilon), \tag{4}$$

where ϵ is the error rate, i.e., the flipping probability for each bit transmitted.

- **Theorem 3.4** The capacity of Binary Erasure Channel (BEC) is

$$C = 1 - \alpha, \tag{5}$$

where α is the erasure probability, i.e., the bit-loss rate.

- **Theorem 3.5** The capacity of weak symmetric channel $\mathbf{Q}$ equals

$$C = \log|\mathcal{Y}| - H(\text{the distribution w.r.t the row of } \mathbf{Q}), \tag{6}$$

achievable with equally distributed probabilities.

A general solution for the capacity of DMC

- As $I(\mathbf{p}, \mathbf{Q})$ is concave over $\mathbf{p}$, the channel capacity can be formulated as the maximization of the objective function

$$\text{maximize}_{\mathbf{p}} I(\mathbf{p}, \mathbf{Q}), \tag{7}$$

$$\begin{aligned} \text{subject to} \quad \sum_x p(x) &= 1 \qquad (8) \\ p(x) &\geq 0. \qquad (9) \end{aligned}$$

- The methods of Lagrangian multiplier, gradient-based search, iterative algorithms and KKT conditions.

Convex Programming using KKT conditions

- **Theorem 3.6** Karush-Kuhn-Tucker conditions
  Suppose the objective function $f(\mathbf{x})$ is defined over n-dimensional convex set $\mathcal{S}$, where $\mathcal{S} = \{\mathbf{x} = (x_1, x_2, \ldots, x_n) | x_i \geq 0, i = 1, 2, \ldots, n\}$, and it is differentiable and concave over the set. Let $\mathbf{x}^* = (x_1^*, x_2^*, \ldots, x_n^*) \in \mathcal{S}$ be the optimal solution. So the function $f(\mathbf{x})$ can achieve its maximum value over $\mathcal{S}$ at $\mathbf{x} = \mathbf{x}^*$ if and only if

$$\frac{\partial f(\mathbf{x})}{\partial x_i}|_{\mathbf{x}=\mathbf{x}^*} = 0, \qquad \text{if} \;\; \mathbf{x}_i^* > 0, \tag{10}$$

$$\frac{\partial f(\mathbf{x})}{\partial x_i}|_{\mathbf{x}=\mathbf{x}^*} \leq 0, \qquad \text{if} \;\; \mathbf{x}_i^* = 0. \tag{11}$$

- Note that $\mathbf{x}^* = (x_1^*, x_2^*, \ldots, x_n^*)$ is the interior point of $\mathcal{S}$ when $\mathbf{x}_i^* > 0$. If $\exists x_n^* = 0$, $\mathbf{x}^*$ is on the boundary of $\mathcal{S}$.

The capacity of DMC

- **Theorem 3.7** For the discrete memoryless channel with transition probability matrix $\mathbf{Q}$, the mutual information $I(\mathbf{p}, \mathbf{Q})$ is maximized to achieve the capacity C with respect to $\mathbf{p}^*$ if and only if

$$I(X = x_i; Y) = C, \qquad \text{for } x_i \text{ that } p^*(x_i) > 0 \tag{12}$$

$$I(X = x_i; Y) \leq C, \qquad \text{for } x_i \text{ that } p^*(x_i) = 0, \tag{13}$$

where $i \in \{1, 2, \ldots, n\}$ and

$$I(X = x_i; Y) = \sum_{j=1}^{J} q(y_j|x_i) \log \frac{q(y_j|x_i)}{p(y_j)} \tag{14}$$

indicates the average mutual information carried by x_i.

The uniqueness of the distribution

- **Theorem 3.8** The output distribution is unique when achieving the channel capacity, and all possible input distributions are optimal that maximizes the mutual information.
- The optimal input distribution $\mathbf{p}^*$ is not necessarily unique.
- **Theorem 3.9** The non-zero components of the optimal input distribution with the maximized number of zero components are uniquely determined, and the number of non-zeros components are no more than the number of output symbols.
- Note that the optimal input distribution with the most zero terms are not unique, but the non-zero components are permutations of each other.

The cascade of independent channels

- As a cascade of two independent channels

$$X \longrightarrow Y \longrightarrow Z, \tag{15}$$

  the data processing theorem tells $I(Y;Z) \geq I(X;Z)$.
- The capacity goes to zero with the increased number of cascades.
- The transition probability matrix of n cascades of channels is calculated as

$$\mathbf{Q} = \prod_{i=1}^{n} \mathbf{Q}_i, \tag{16}$$

  the product of all channel matrices.
- Data processing theorem tells us that the additional gain in channel capacity can be lost rather than obtained from the increased cascades.

Parallel channels

- **Theorem 3.10** Parallel channels sharing the same input X with different output values $Y_1, Y_2, \ldots, Y_n$.
  1. The capacity of parallel channels with the same input is greater than that of any single channel and less than $\max H(X)$.
  2. Diversity in wireless communications is the case.
- **Theorem 3.11** The channels used in parallel with multiple input and multiple output (MIMO).
  1. Sub-channels are independent on each other and $\mathbf{X} = \{X_1, X_2, \ldots, X_n\}$ are transmitted through these sub-channels to get $\mathbf{Y} = \{Y_1, Y_2, \ldots, Y_n\}$, simultaneously. The equivalent capacity is the sum of capacities of the parallel sub-channels.
  2. Multiplexing in communications is the case.
- **Theorem 3.12** The probabilistic utilization of n channels leads to the equivalent capacity $C = \log \sum_{i=1}^{n} 2^{C_n}$ with probability $p_i(C) = 2^{(C_i - C)}$. Opportunistic communication is the case.

Outline

1. Channel Capacity and its Properties

2. Discrete Memoryless Channel

3. Continuous and Analog Channel
   - Capacity-cost Function
   - Additive Noise Continuous Channel
   - Water Filling
   - The notion of analog channel capacity
   - Shannon Formula
   - Insights of Shannon Formula

4. Joint Typicality

5. Channel Coding

The Capacity of Continuous Channel

- Continuous channel that is discrete in time and continuous in values with power constraint.
- **Definition 3.4** Define the function $b(\cdot) > 0$ as the cost for a sequence $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ for continuous memoryless channel $\{X, q(y|x), Y\}$. Let the joint probability distribution of random vector $\mathbf{X} = \{X_1, X_2, \ldots, X_n\}$ be $p(\mathbf{x})$. Then, the average cost is

$$E(b(\mathbf{X})) = \sum_{\mathbf{x}} p(\mathbf{x}) b(\mathbf{x}). \tag{17}$$

- **Definition 3.5** Let $\mathbf{X}$ and $\mathbf{Y}$ be the n-dimensional vectors as the input and output of a continuous channel. The capacity-cost function is defined as

$$C(\beta) = \lim_{n\to\infty} \frac{1}{n} \max_{p(\mathbf{x}):E(b(\mathbf{X}))\leq n\beta} I(\mathbf{X};\mathbf{Y}). \tag{18}$$

Continuous memoryless channel with Additive Noise

- The mutual information is simplified to $nI(X,Y)$ with stationary memoryless input, and the capacity-utility function becomes

$$C(\beta) = \max_{p(x):E(b(X))\leq\beta} I(X;Y). \tag{19}$$

- The channel output equals $Y = X + Z$, where Z is memoryless stationary noise added on the input X.
- **Theorem 3.13** The capacity-cost function of continuous memoryless additive noisy channel is

$$C(P_S) = \max_{p(\mathbf{x}):E(X^2)\leq P_S} h(Y) - h(Z). \tag{20}$$

Additive White Gaussian Noise channel (AWGN)

- **Theorem 3.14** The capacity-cost function for memoryless additive gaussian noise channel is

$$C(P_S) = \frac{1}{2}\log(1+\frac{P_S}{P_N}), \tag{21}$$

  where $\frac{P_S}{P_N}$ is the signal-to-noise ratio.
- AWGN channel can always be fully exploited using gaussian input at the costs of equal transmission power.
- **Theorem 3.15** If the input signal X is gaussian, the mutual information is minimized with gaussian noise under power constraint P_N.
- It indicates that gaussian noise is the worst when the input is given as gaussian distribution.

Continuous parallel channels and water-filling

- Input vector $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ and allocated power P_{S_i} of the i-th element; channel noise vector $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_n)$ and noise power P_{N_i}, $i = 1, 2, \ldots, n$.
- The capacity of the parallel channels is

$$C(P_S) = \max_{p(\mathbf{X}):\sum_{i=1} P_{S_i} \le P_0} I(\mathbf{X}; \mathbf{Y}), \tag{22}$$

  where P_0 is the budget for the total transmitting power.
- The capacity is achievable by allocating power as water-filling via Lagrange multiplier.

The capacity of analog channel

- Analog channel is continuous in both time and values, e.g., fiber, wire, electromagnetic wave.
- A special case: Additive white gaussian noise (AWGN) channel.
  1. limited bandwidth: W
  2. additive noise: $y(t) = x(t) + z(t)$
  3. white noise: stationary ergodic stochastic process and a uniform distribution in spectral density.
  4. Gaussian noise: The noise is gaussian at any time.
- Problem formulation
  1. Input signal $x(t)$ with limited bandwidth $[-W, W]$, duration T and power constrain P.
  2. Output signal $y(t) = x(t) + z(t)$
  3. Noise $z(t)$ is additive white gaussian with zero-mean, and its double-side spectral density is

$$N(f) = \frac{N_0}{2}, \text{ if } |f| \leq W \tag{23}$$

and zero otherwise.

The capacity of analog channel

- Sampling the analog signal in time is equivalent to the parallel combination of addictive gaussian channels. So the total capacity is written as

$$C_T(P_S) = \frac{1}{2}\sum_{i=1}^{2WT}\log(1+\frac{P_{S_i}}{P_{N_i}}) \tag{24}$$

under the noise power constraint $P_N = 2WTP_{N_i} = 2WT\frac{N_0}{2} = WTN_0$.

- The transmission power is allocated as

$$P_{S_i} = \frac{P_S T + P_N}{2WT} - \frac{N_0}{2} = \frac{P_S}{2W} \tag{25}$$

via water-filling. Then we can obtain

$$\frac{P_{S_i}}{P_{N_i}} = \frac{P_S}{WN_0}, \tag{26}$$

which is used to get Shannon formula.

Shannon Formula

- **Theorem 3.16** The capacity of AWGN channel is

$$C = W \log(1 + \frac{P_S}{N_0 W}). \tag{27}$$

- Remarks: the physical dimesions of terms in Shannon formula
  1. capacity C $\quad bps$ bit-per-second
  2. bandwidth W $\quad Hz$ or s^{-1}
  3. signal power P_S $\quad Watt$
  4. noise spectral density N_0 $\quad Watt/Hz$
- Two ways to increase the channel capacity
  1. expand the bandwidth W, bounded by

$$\lim_{W\to\infty} C(P_S) = \lim_{W\to\infty} \frac{P_S}{N_0} \log(1 + \frac{P_S}{N_0 W})^{\frac{N_0 W}{P_S}} = \frac{P_S}{N_0} \log e \approx 1.44 \frac{P_S}{N}. \tag{28}$$

  2. increase the transmission power P_S, approximated as

$$C(P_S) \approx W \log \frac{P_S}{N_0 W} \tag{29}$$

at high SNR regime.

Information and thermal-dynamics

- When approaching the capacity, the energy per bit is

$$E_b = \frac{P_S}{C}(J/bit). \tag{30}$$

- The normalized SNR ratio $\frac{E_b}{N_0}$ is

$$\frac{E_b}{N_0} = \frac{P_S}{N_0 C} = \frac{P_S}{N_0 W \log(1 + \frac{P_S}{N_0 W})} = \frac{SNR}{\log(1 + SNR)}, \tag{31}$$

  lower bounded by Shannon-limit $\ln 2 \approx 0.693$ or $-1.59dB$ if SNR approaches zero.
- Remarks:
  1. Shannon limit indicates the minimum required energy for 1-bit transmission.
  2. Thermal dynamics shows the noise power spectral density $N_0 = kT$ in W/Hz, where k is Boltzmann constant and T is Kelvin temperature. Therefore, the minimum energy required per bit is $E_b \geq \ln^2 kT \approx 0.693kT$ from the point of view of physics.

Outline

1. Channel Capacity and its Properties
2. Discrete Memoryless Channel
3. Continuous and Analog Channel
4. Joint Typicality
   - Strong Typicality
   - Strong Typical Sequences
   - Jointly Typical Sequences
   - Joint Typicality Lemma
5. Channel Coding
6. Summary

Strong Typicality

- Let x^n be a sequence with elements drawn from a finite alphabet set $\mathcal{X}$. Define the empirical pmf of x^n as

$$\pi(x|x^n) := \frac{|\{i : x_i = x\}|}{n} \text{ for } x \in \mathcal{X} \tag{32}$$

- Let $X_1, X_2, \cdots$ be i.i.d. with $X_i \sim p_X(x_i)$. For each $x \in \mathcal{X}$

$$\pi(x|X^n) \to p(x) \text{ in probability} \tag{33}$$

Strong Typical Set

- **Definition 3.6** Typical Set: For $X \sim p(x)$ and $\epsilon \in (0,1)$, define the set $\mathcal{T}_\epsilon^{(n)}(X)$ of typical sequences x^n as

$$\mathcal{T}_\epsilon^{(n)}(X) := \{x^n : |\pi(x|x^n) - p(x)| \leq \epsilon \cdot p(x) \text{ for all } x \in \mathcal{X}\} \tag{34}$$

- **Theorem 3.17** Typical Average Lemma: Let $x^n \in \mathcal{T}_\epsilon^{(n)}(X)$. Then for any nonnegative function $g(x)$ on $\mathcal{X}$,

$$(1-\epsilon)E(g(X)) \leq \frac{1}{n}\sum_{i=1}^{n} g(x_i) \leq (1+\epsilon)E(g(X)) \tag{35}$$

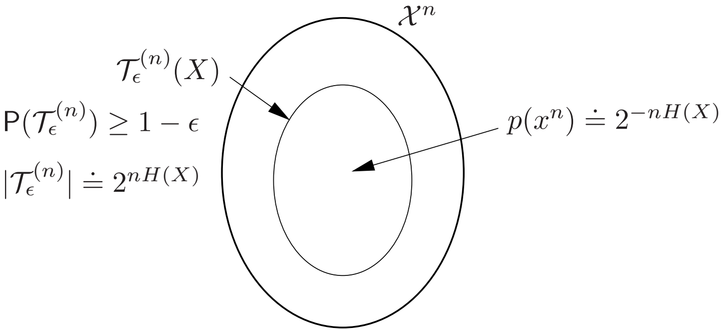
Properties of Strong Typical Set

- Let $p(x^n) = \prod_{i=1}^{n} p_X(x_i)$. Then, for each $x^n \in \mathcal{T}_\epsilon^{(n)}(X)$

$$2^{-n(H(X)+\delta(\epsilon))} \leq p(x^n) \leq 2^{-n(H(X)-\delta(\epsilon))} \tag{36}$$

  where $\delta(\epsilon) = \epsilon \cdot H(X) \to 0$ as $\epsilon \to 0$. This follows from the typical average lemma by taking $g(x) = -\log p_X(x)$.
- The cardinality of the typical set $|\mathcal{T}_\epsilon^{(n)}| \leq 2^{n(H(X)+\delta(\epsilon))}$.
- If $X_1, X_2, \ldots$ are i.i.d. with $X_i \sim p_X(x_i)$, then by the LLN $P\{X^n \in \mathcal{T}_\epsilon^{(n)}\} \to 1$.
- The cardinality of the typical set $|\mathcal{T}_\epsilon^{(n)}| \geq (1-\epsilon)2^{n(H(X)-\delta(\epsilon))}$ for n sufficiently large.

Illustration of Typical Set Properties



Jointly Typical Sequences

- We generalize the notion of typicality to the multiple random variables.
- Let (x^n, y^n) be a pair of sequences with elements drawn from a pair of finite alphabets $(\mathcal{X}, \mathcal{Y})$. Define the joint empirical pmf (joint type) as

$$\pi(x, y|x^n, y^n) = \frac{|\{i : (x_i, y_i) = (x, y)\}|}{n} \text{ for } (x, y) \in \mathcal{X} \times \mathcal{Y} \tag{37}$$

- **Definition 3.7** Jointly Typical Set. Let $(X, Y) \sim p(x, y)$. The set $\mathcal{T}_\epsilon^{(n)}(X, Y)$ of joint ϵ-typical n-sequences (x^n, y^n) is defined as

$$\mathcal{T}_\epsilon^{(n)} := \{(x^n, y^n) : |\pi(x, y|x^n, y^n) - p(x, y)| \leq \epsilon \cdot p(x, y) \text{ for all } (x, y) \in \mathcal{X} \times \mathcal{Y}\} \tag{38}$$

In other words, $\mathcal{T}_\epsilon^{(n)}(X, Y) = \mathcal{T}_\epsilon^{(n)}((X, Y))$.

Properties of Jointly Typical Sequences

- Let $p(x^n, y^n) = \prod_{i=1}^{n} p_{X,Y}(x_i, y_i)$. If $(x^n, y^n) \in \mathcal{T}_\epsilon^{(n)}(X, Y)$, then
  1. $x^n \in \mathcal{T}_\epsilon^{(n)}(X)$ and $y^n \in \mathcal{T}_\epsilon^{(n)}(Y)$.
  2. $p(x^n, y^n) \approx 2^{-nH(X,Y)}$.
  3. $p(x^n) \approx 2^{-nH(X)}$, and $p(y^n) \approx 2^{-nH(Y)}$.
  4. $p(x^n|y^n) \approx 2^{-nH(X|Y)}$ and $p(y^n|x^n) \approx 2^{-nH(Y|X)}$.
- As in the single random variable case, for n sufficiently large

$$|\mathcal{T}_\epsilon^{(n)}(X, Y)| \approx 2^{nH(X,Y)} \tag{39}$$

- Let $\mathcal{T}_\epsilon^{(n)}(Y|x^n) := \{y^n : (x^n, y^n) \in \mathcal{T}_\epsilon^{(n)}(X, Y)\}$. Then

$$|\mathcal{T}_\epsilon^{(n)}(Y|x^n)| \leq 2^{n(H(Y|X)+\delta(\epsilon))}, \tag{40}$$

  where $\delta(\epsilon) = \epsilon \cdot H(Y|X)$.

Conditional Typicality Lemma

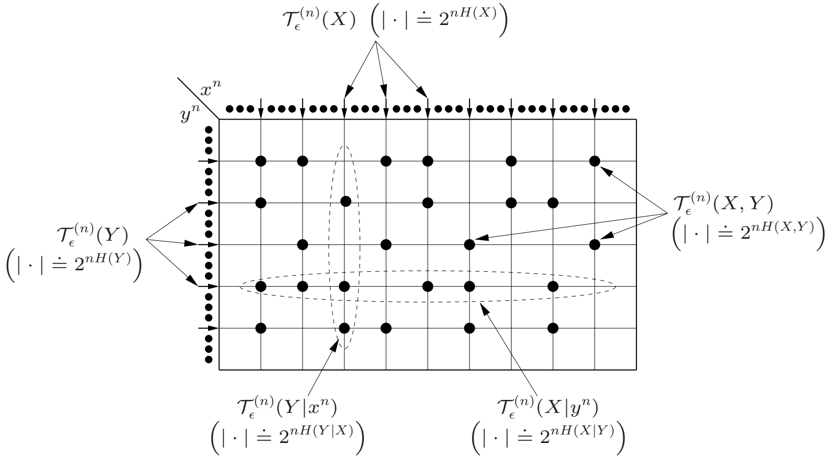
- **Lemma 3.18** Conditional Typicality Lemma. Let $x^n \in \mathcal{T}_\epsilon^{(n)}(X)$ and $Y^n \sim \prod_{i=1}^n p_{Y|X}(y_i|x_i)$. Then for every $\epsilon > \epsilon'$,

$$P\{(x^n, Y^n) \in \mathcal{T}_\epsilon^{(n)}(X, Y)\} \to 1 \text{ as } n \to \infty. \tag{41}$$

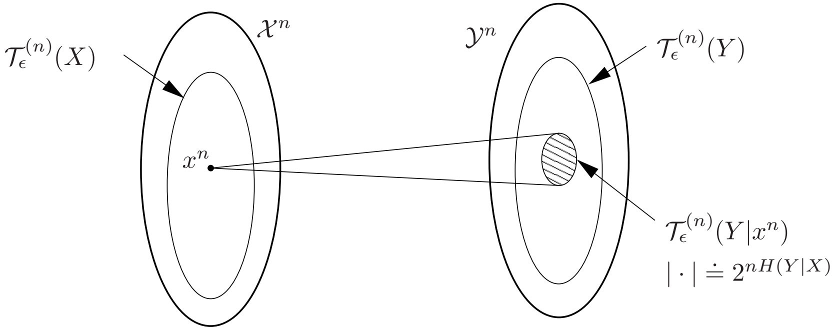
- The conditional typicality lemma implies that for all $x^n \in \mathcal{T}_{\epsilon'}^{(n)}(X)$

$$|\mathcal{T}_\epsilon^{(n)}(Y|x^n)| \geq (1-\epsilon)2^{n(H(Y|X)-\delta(\epsilon))} \text{ for } n \text{ sufficiently large.} \tag{42}$$

Useful Picture



Another Useful Picture



Joint Typicality Lemma

- **Theorem 3.19** Joint Typicality Lemma.

  Let $(X, Y, Z) \sim p(x, y, z)$. Then there exists $\delta(\epsilon) \to 0$ as $\epsilon \to 0$ such that the following statements hold:

  1. Given $(x^n, y^n) \in \mathcal{T}_\epsilon^{(n)}(X, Y)$, let $\tilde{Z}^n$ be distributed according to $\prod_{i=1}^n p_{Z|X}(\tilde{z}_i|x_i)$ (instead of $p_{Z|X,Y}(\tilde{z}_i|x_i, y_i)$). Then,
     (a) $\mathsf{P}\{(x^n, y^n, \tilde{Z}^n) \in \mathcal{T}_\epsilon^{(n)}(X, Y, Z)\} \le 2^{-n(I(Y;Z|X)-\delta(\epsilon))}$, and
     (b) for sufficiently large n,
     $\mathsf{P}\{(x^n, y^n, \tilde{Z}^n) \in \mathcal{T}_\epsilon^{(n)}(X, Y, Z)\} \ge (1-\epsilon)2^{-n(I(Y;Z|X)+\delta(\epsilon))}$
  2. If $(\tilde{X}^n, \tilde{Y}^n)$ is distributed according to an arbitrary pmf $p(\tilde{x}^n, \tilde{y}^n)$ and $\tilde{Z}^n \sim \prod_{i=1}^n p_{Z|X}(\tilde{z}_i|\tilde{x}_i)$, conditionally independent of $\tilde{Y}^n$ given $\tilde{X}^n$, then

  $$\mathsf{P}\{(\tilde{X}^n, \tilde{Y}^n, \tilde{Z}^n) \in \mathcal{T}_\epsilon^{(n)}(X, Y, Z)\} \le 2^{-n(I(Y;Z|X)-\delta(\epsilon))}$$

Outline

1. Channel Capacity and its Properties

2. Discrete Memoryless Channel

3. Continuous and Analog Channel

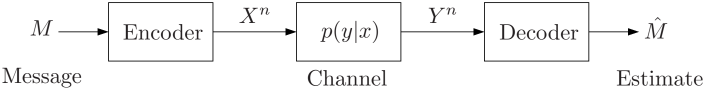
4. Joint Typicality

5. Channel Coding
   - Channel Coding and Operational Capacity
   - Channel Coding Theorem

6. Summary

DMC revisited

- Sender X wishes to send a message reliably to a receiver Y over a communication channel.
- A DMC $(\mathcal{X}, p(y|x), \mathcal{Y})$ consists of two finit sets $\mathcal{X}$, $\mathcal{Y}$, and a collection of conditional pmfs $p(y|x)$ on $\mathcal{Y}$.
- By memoryless, we mean that when the DMC $(\mathcal{X}, p(y|x), \mathcal{Y})$ is used over n transmissions with message M and input X^n, the output Y_i at time $i \in [1:n]$ is distributed according to $p(y_i|x_i, y_{i-1}, m) = p(y_i|x_i)$



Channel Coding

- **Definition 3.8** Channel Code. A $(2^{nR}, n)$ code with rate $R \geq 0$ bits/transmission for the DMC $(\mathcal{X}, p(y|x), \mathcal{Y})$ consists of:
  1. A message set $[1 : 2^{nR}] = \{1, 2, \ldots, 2^{\lceil nR \rceil}\}$
  2. An encoding function (encoder) $x^n : [1 : 2^{nR}] \to \mathcal{X}^n$ that assigns a codeword $x^n(m)$ to each message $m \in [1 : 2^{nR}]$. The set $\mathcal{C} := \{x^n(1), \ldots, x^n(2^{\lceil nR \rceil})\}$ is referred to as the codebook.
  3. A decoding function (decoder) $\hat{m} : \mathcal{Y}^n \to [1 : 2^{nR}] \cup \{e\}$ that assigns a message $\hat{m} \in [1 : 2^{nR}]$ or an error message e to each received sequence y^n.
- We assume the message $M \sim$ Unif$[1 : 2^{nR}]$, i.e., it is chosen uniformly at random from the message set.

Probability of Error and Operational Capacity

- **Definition 3.9** Probability of error: Let $\lambda_m(\mathcal{C}) = P\{\hat{M} \neq m | M = m\}$ be the conditional probability of error given that message m is sent. The average probability of error of a $(2^{nR}, n)$ code is

$$P_e^{(n)} = P\{\hat{M} \neq M\} = \frac{1}{2^{\lceil nR \rceil}} \sum_{m=1}^{2^{\lceil nR \rceil}} \lambda_m(\mathcal{C}). \tag{43}$$

- **Definition 3.10** Achievability of rate. A rate R is said to be achievable if there exists a sequence of $(2^{nR}, n)$ codes with probability of error $P_e^{(n)} \to 0$ as $n \to \infty$.
- **Definition 3.11** Operational Capacity. The capacity C of a discrete memoryless channel is the supremum of all achievable rates.

Channel Coding Theorem

- **Theorem 3.20** Shannon's Channel Coding Theorem. The (operational) capacity of the DMC $(\mathcal{X}, p(y|x), \mathcal{Y})$ is given by $C = \max_{p(x)} I(X;Y)$.
- Proof of Channel Coding Theorem:
  - Achievability: We show that any rate $R < C$ is achievable, i.e., there exists a sequence of $(2^{nR}, n)$ codes with average probability of error $P_e(n) \to 0$. The proof of achievability uses random coding and joint typicality decoding
  - Weak converse: We show that given any sequence of $(2^{nR}, n)$ codes with $P_e(n) \to 0$, then $R \leq C$. The proof of converse uses Fano?s inequality and properties of mutual information.

Outline

1. Channel Capacity and its Properties
2. Discrete Memoryless Channel
3. Continuous and Analog Channel
4. Joint Typicality
5. Channel Coding
6. Summary

Summary

- Definition of capacity and Discrete Memoryless Channel.
- The calculation of the capacity of a few typical channels.
- A generalized calculation of the capacity of DMC.
- The capacity of continuous channel and capacity-cost function
- Continuous Gaussian Channel
- AWGN channel
- Combinations of channels: cascade, parallel, water-filling and power allocation.
- Strong Typicality, Jointly Typical Sequences
- Shannon Channel Coding Theorem

Reference

- *Lin Zhang*, "Lecture notes on Fundamentals of applied information theory", *2014-spring, in Chinese.*
- *Claude E. Shannon*, "A Mathematical Theory of Communication", *Bell System Technical Journal, 1948.*
- *Cover T M, Thomas J A.*, "Elements of information theory", *John Wiley and Sons, 2012.*
- *Xuelong Zhu*, "Fundamentals of applied information theory", *Tsinghua Univ. Press, 2001, in Chinese.*