INFORMATION THEORY & CODING

Channel Code - 2

Dr. Rui Wang

Department of Electrical and Electronic Engineering
Southern Univ. of Science and Technology (SUSTech)

Email: wang.r@sustech.edu.cn

December 6, 2022

Review

- **Channel capacity.** The logarithm of the number of distinguishable inputs is given by

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

- **Examples**
  - Binary symmetric channel: $C = 1 - H(p)$
  - Binary erasure channel: $C = 1 - \alpha$
  - Symmetric channel: $C = \log|\mathcal{Y}| - H$ (row of trans. matrix)

Channel Code

Definition

An (M, n) code for the channel $(\mathcal{X}, p(y|x), \mathcal{Y})$ consists of :

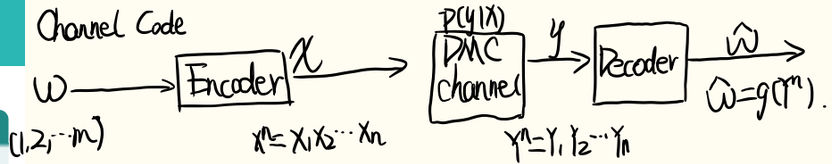
1. An index set $\{1, 2, \ldots, M\}$ representing messages.
2. An encoding function $X^n : \{1, 2, \ldots, M\} \to \mathcal{X}^n$, yielding codewords $x^n(1), x^n(2), \ldots, x^n(M)$. The set of codewords is called codebook.
3. A decoding function $g : \mathcal{Y}^n \to \{1, 2, \ldots, M\}$.

The rate R of an (M, n) code is

$$R = \frac{\log M}{n} \text{ bit per transmission}$$

On the other hand, we usually write

$$M = \lceil 2^{nR} \rceil$$



An (M,n) code consists of
① An index set representing messages $\{1, 2, \cdots, M\}$
② An encoder: $X^n(1), X^n(2), \cdots X^n(M)$.
③ A decoder $g: Y^n \to \{1, 2, 3, \cdots, M\}$

Code Rate: $R = \frac{\log_2 M}{n}$, 平均每个 中携带的信息量.

$M = \lceil 2^{nR} \rceil$ ⇒ 向上取整.

R与C之间的关系!!!

Channel Code

Definition

An (M, n) code for the channel $(\mathcal{X}, p(y|x), \mathcal{Y})$ consists of :

1. An index set $\{1, 2, \ldots, M\}$ representing messages.
2. An encoding function $X^n : \{1, 2, \ldots, M\} \to \mathcal{X}^n$, yielding codewords $x^n(1), x^n(2), \ldots, x^n(M)$. The set of codewords is called codebook.
3. A decoding function $g : \mathcal{Y}^n \to \{1, 2, \ldots, M\}$.

The rate R of an (M, n) code is

$$R = \frac{\log M}{n} \text{ bit per transmission}$$

On the other hand, we usually write

$$M = \lceil 2^{nR} \rceil$$

Performance Metric

- Conditional probability of error:

$$\lambda_i = \Pr[g(Y_n) \neq i | X^n = x^n(i)] = \sum_{y^n} p\Big(y^n | x^n(i)\Big) I\Big(g(y^n) \neq i\Big)$$

- Maximal probability of error: $\lambda^{(n)} = \max_{i \in \{1,2,\ldots,M\}} \lambda_i$
- Decoding error probability: $\Pr[W \neq g(Y^n)] = \sum_i \lambda_i \Pr[W = i]$
- Arithmetric average probability of error:

$$P_e^{(n)} = \frac{1}{M} \sum_{i=1}^{M} \lambda_i, \quad P_e^{(n)} \leq \lambda^{(n)}$$

If W is uniformly distributed:

$$P_e^{(n)} = \Pr[W \neq g(Y^n)] \text{ Decoding error probability}$$

Achievable Rate

- A rate R is achievable,

  if there exists a sequence of codes with rate R and codeword length n, denoted as $(\lceil 2^{nR} \rceil, n)$, such that the maximal probability of error $\lambda^{(n)} \to 0$ as $n \to \infty$.

  Recall that

  The rate R of an (M, n) code is

$$R = \frac{\log M}{n} \text{ bit per transmission.}$$

Error free → 极限情况下 Error probability → 0.

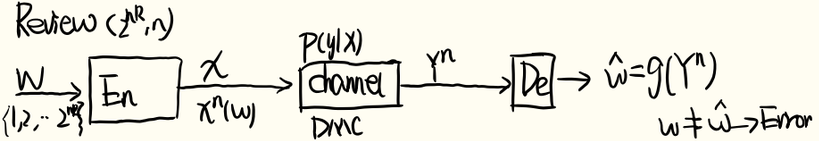
$\exists (\lceil 2^{nR} \rceil, n)$, $\forall n > 0$, integer

$\lambda^{(n)} \Downarrow$, $\forall n > 0$

$\Downarrow$

when $n \to \infty$, $\lambda^{(n)} \to 0$.

Review $(2^{nR}, n)$



$P(y^n|x^n) = P(y_n|x_n) \cdots P(y_1|x_1)$

codebook $C_n = \begin{pmatrix} x_c^n(1) \\ x_c^n(2) \\ \vdots \\ x_c^n(2^{nR}) \end{pmatrix} = \begin{pmatrix} x_1(1) & x_2(1) & \cdots & x_n(1) \\ x_1(2) & x_2(2) & \cdots & x_n(2) \\ \vdots & & & \\ x_1(2^{nR}) & x_2(2^{nR}) & \cdots & x_n(2^{nR}) \end{pmatrix}$

$\lambda_i = P_r[g(Y^n) \neq i | X_c^n(i)]$ $\quad \lambda^{(n)} = \max_i \lambda_i$

R is achievable $\Rightarrow \exists C_n, C_{n+1}, C_{n+2}, \cdots, \lambda^{(n)} \xrightarrow{n\to\infty} 0.$

$\begin{cases} R<C \Rightarrow R \text{ is achievable} \\ \lambda^{(n)} \to 0 \Rightarrow R \leq C. \end{cases}$

Joint Typical Set

- Joint typicality. Given two i.i.d. random variable sequences X^n and Y^n, the set of jointly typical sequences is

$$A_\epsilon^{(n)} = \Big\{ (x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n : \Big| -\frac{1}{n} \log p(x^n) - H(X) \Big| < \epsilon \Big| -\frac{1}{n} \log p(y^n) - H(Y) \Big| < \epsilon \Big| -\frac{1}{n} \log p(x^n, y^n) - H(X,Y) \Big| < \epsilon \Big\}$$

where $p(x^n, y^n) = \prod_{i=1}^n p(x_i, y_i)$. → $x_1, x_2 \cdots x_n$ 之间相互独立
$y_1, y_2 \cdots y_n$ 之间相互独立
y_i 由 x_i 决定，所以成对

Typical set.

$X^n = X_1 X_2 \cdots X_n \sim P(X^n) = P(X_1)P(X_2)\cdots P(X_n)$

$A_\varepsilon^n = \{ x^n : \left| -\frac{1}{n} (\log P(X^n) - H(X) \right| < \varepsilon \}$

$P_r[X^n \in A_\varepsilon^{(n)}] \to 1, \ n \to \infty$

Joint Typical Set $(X^n, Y^n) \sim P(X^n, Y^n) = P(x_1, y_1)P(x_2, y_2) \cdots$

$A_\varepsilon^{(n)} = \{ (x^n, y^n) : \left| -\frac{1}{n} \log P(x^n) - H(X) \right| < \varepsilon, \left| -\frac{1}{n} \log P(y^n) - H(Y) \right| < \varepsilon, \left| -\frac{1}{n} \log P(x^n, y^n) - H(X,Y) \right| < \varepsilon \}$

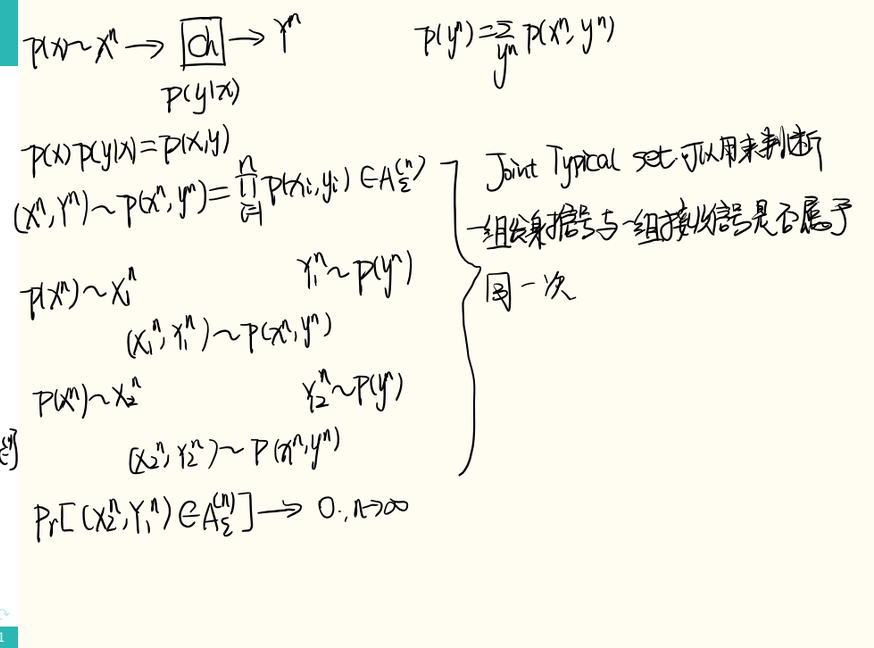
→ Ch

Joint AEP

- **Joint AEP** Let (X^n, Y^n) be the sequences of length n drawn i.i.d. according to $p(x^n, y^n) = \prod_{i=1}^{n} p(x_i, y_i)$, then:
  1. $\Pr\left[(X^n, Y^n) \in A_\epsilon^{(n)}\right] \to 1$ as $n \to \infty$.
  2. $\left|A_\epsilon^{(n)}\right| \leq 2^{n(H(X,Y)+\epsilon)}$.
  3. If $(\tilde{X}^n, \tilde{Y}^n) \sim p(x^n)p(y^n)$, then

$$\Pr\left[\left(\tilde{X}^n, \tilde{Y}^n\right) \in A_\epsilon^{(n)}\right] \leq 2^{-n(I(X;Y)-3\epsilon)}.$$

Please refer to p196 for the proof (proof of Theorem 7.6.1)



$p(y|x) \Rightarrow C = \max_{p(x)} I(X;Y)$, let $p(x) = \arg\max I(X;Y)$ $p(x,y) = p(x)p(y|x)$

$\Rightarrow p(x^n, y^n) = \prod_{i=1}^{n} p(x_i, y_i) \Rightarrow A_\varepsilon^{(n)}$

Proof: Encoder $\Rightarrow p(x)$. i.i.d generate C.

$\Rightarrow |\mathcal{X}|^{2^{nR} \times n}$ codebooks

Transmit message w with C.

① $X_c^n(w)$ ② receive $Y^n \sim p[y^n | x^n = x_c^n(w)]$

③ g: find the only $\hat{w}$, $(x_c^n(\hat{w}), Y^n) \in A_\varepsilon^{(n)}$

Error: (1) $(x_c^{(n)}(w), Y^n) \notin A_\varepsilon^{(n)}$ (2) $(x_c^{(n)}(i), Y^n) \in A_\varepsilon^{(n)}$, $i \neq w$.

let $\mathcal{E}_r = \{\hat{w}(y^n) \neq w\}$ $P_r(\mathcal{E}_r) = \sum_C P_r(C) \cdot P_e^n(C)$ → 使用C码出错的概率

使用C的概率

$= \sum_C P_r(C) \sum_{w=1}^{2^{nR}} \lambda_w(C) \frac{1}{2^{nR}} = \frac{1}{2^{nR}} \sum_{w=1}^{2^{nR}} \sum_C P_r(C) \lambda_1(C) = \sum_C P_r(C) \lambda_1(C)$

平均错误概率

假设w=1

Channel Coding Theorem

Theorem (Channel coding theorem)

For a discrete memoryless channel, all rates below capacity C are achievable. Specifically, for every rate $R < C$, there exists a sequence of $(2^{nR}, n)$ codes with maximum probability of error $\lambda^{(n)} \to 0$.

Conversely, any sequence of $(2^{nR}, n)$ codes with $\lambda^{(n)} \to 0$ must have $R \leq C$.

proof.

最大的Error probability.

$n \to \infty$, $\lambda^{(n)} \to 0$, $P_e^{(n)} \to 0$

Achievability: when $R < C$, there exists zero-error code.

Converse: zero-error codes must have $R \leq C$.

construct $(2^{nR}, n)$ $x_c^n(1)$ 第i个message的第n个symbol.

Encoder $(x_1(w), x_2(w) \cdots x_n(w))$. $w = 1, 2, \cdots, 2^{nR}$

codebook $C =$

$2^{nR} \times n$

$x_1(2^{nR})$ $x_2(2^{nR})$ $\cdots$ $x_n(2^{nR})$

$x_c^n(2^{nR})$

让p(x)成独立同分布每个symbol.

Channel $p(y|x)$ let $p(x)$ be the capacity-achieving source distribution

$I(X;Y) = C$

w, i都是在同一个codebook上，求均会使 $\lambda_w(C)=\lambda_i(C)$.

let $e_i(C)=\{(X_C^n(i), Y_i^n)\in A_\varepsilon^{(n)}\}$

$\lambda_i(C)=P_r[\bar{e}_i(C)\cup e_2(C)\cup e_3(C)\cdots e_{2^{nR}}(C)\,|\,w=i]$

$\le P_r[\bar{e}_i(C)|w=i]+\sum_{i=2}^{2^{nR}} P_r[e_i(C)|w=i]$

$P_r(\xi_r)\le \sum_C P_r[C]P_r[\bar{e}_i(C)|w=i]+\sum_{i=2}^{2^{nR}}\sum_C P_r[C]P_r[e_i(C)|w=i]$

$\sum_C P_r[C]\,P_r[\bar{e}_i(C)|w=i]=\sum_C\prod_{i=1}^{2^{nR}} P_r[X_C^n(i)]P_r[\bar{e}_i(C)]$

$=\sum_{X_1^n} P_r[X_1^n]\sum_{C:\lambda_C^n(1)=X_1^n}\prod_{i=1}^{2^{nR}} P_r[X_C^n(i)]P_r[\bar{e}_i(C)]$

$=\sum_{X_1^n} P_r[X_1^n]\,P_r[\bar{e}_i(C)]\sum_{C:X_C^n(1)=X_1^n}\prod_{i=1}^{2^{nR}} P_r[\bar{X}_C^n(i)]$ (the last factor $=1$)

$=\sum_{X_1^n} P_r[X_1^n]\,P_r\left[(X_C^n(1), Y_1^n)\notin A_\varepsilon^{(n)}\right]$

Random Codebook

- Generate a $(2^{nR}, n)$ code at random according to $p(x)$, where $p(x)$ is the capacity achieving distribution. The 2^{nR} are the rows of a matrix:

$$\mathcal{C}=\begin{bmatrix} x_1(1) & x_2(1) & \dots & x_n(1) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(2^{nR}) & x_2(2^{nR}) & \dots & x_n(2^{nR}) \end{bmatrix}.$$

  Each entry is generated i.i.d. according to $p(x)$.
- Encoding: map the message $w=\{1,2,3,\dots,2^{nR}\}$ to codeword $[x_1(w),x_2(w),\dots,x_n(w)]$, i.e.

$$\mathcal{C}\to[x_1(w),x_2(w),\dots,x_n(w)]=x_{\mathcal{C}}^n(w), w=1,2,\dots,2^{nR}$$

- We shall prove the average detection error probability (over all codebooks) tends to zero as n increase, which implies that there must exists one good codebook whose detection error probability tends to zero

Decoder. $w\to X^n(w)\xrightarrow{\text{channel}} Y^n\to\hat{w}=g(Y^n)$

- $P(y|x)$

$P(x,y)=P(x)\cdot P(y|x)$

$P(x^n,y^n)=\prod_{i=1}^{n} P(x_i,y_i)$

$A_\varepsilon^{(n)}$

g: find the only $\hat{w}$ such that $(X_C^n(\hat{w}), Y^n)$ is jointly typical.

$= \sum_{x_C^n(1)} P_r[x_C^n(1)] P_r[(x_C^n(1), Y_i^n) \notin A_\varepsilon^{(n)}] = P_r[(X_1^n, Y_i^n) \notin A_\varepsilon^{(n)}] \to 0, \ n \to \infty.$

$= P_r[(X_1^n, Y_i^n) \notin A_\varepsilon^{(n)}] \le \varepsilon$ for sufficiently large n.

$\sum_C P_r[C] P_r[e_C(C) | w = 1] \qquad i = 2 \cdots 2^{nR}$

$= P_r[(X_i^n, Y_i^n) \in A_\varepsilon^n | w = 1] \le 2^{-n[I(X;Y) - 3\varepsilon]}$

$P_r[\varepsilon_r] \le \varepsilon + \sum_{i=2}^{2^{nR}} 2^{-n[I(X;Y)-3\varepsilon]} = \varepsilon + 2^{-n[I(X;Y)-3\varepsilon]} (2^{nR} - 1)$

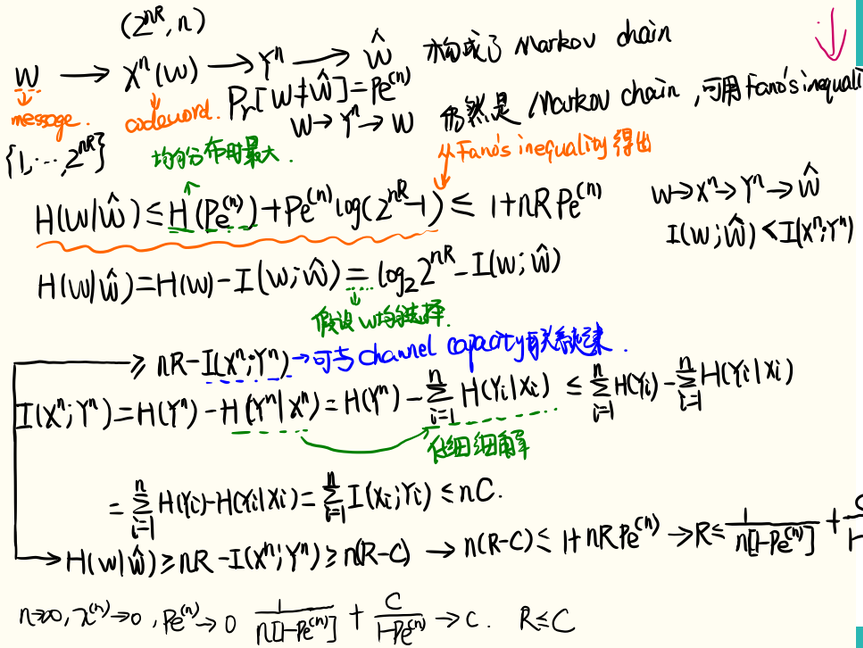
$< \varepsilon + 2^{-n[I(X;Y)-3\varepsilon]} 2^{nR} = \varepsilon + 2^{-n[I(X;Y)-R-3\varepsilon]} = \varepsilon + 2^{-n[C-R-3\varepsilon]}$

If $C - R - 3\varepsilon > 0$. $P_r[\varepsilon_r] \to 0$. $\Rightarrow$ If $R < C - 3\varepsilon$, $P_r[\varepsilon_r] \to 0$

$\Rightarrow$ If $R < C$, $P_r[\varepsilon_r] \to 0$.

Jointly Typical Decoding

- Decoding: finds the only $\hat{w}$ such that $(x_{\mathcal{C}}^n(\hat{w}), Y_{\mathcal{C}}^n)$ is jointly typical.
- Decoding error: Suppose message 1 is sent to via codeword $x_{\mathcal{C}}^n(1)$ and $Y_{\mathcal{C}}^n$ is the received signal, the possible decoding error events include:
  - $(x_{\mathcal{C}}^n(1), Y_{\mathcal{C}}^n)$ is not joint typical.
  - $(x_{\mathcal{C}}^n(i), Y_{\mathcal{C}}^n)$ is joint typical $(i = 2, 3, \ldots, 2^{nR})$.
- Idea of proof: According to joint AEP, since $x_{\mathcal{C}}^n(1)$ and $Y_{\mathcal{C}}^n$ are generated according to joint distribution $p(x^n, y^n)$, the chance of the first event is small. Moreover, since $Y_{\mathcal{C}}^n$ is generated independently of $x_{\mathcal{C}}^n(i)$, the total chance of the second event is also small.



Proof for achievability

- A message W is chosen according to a uniform distribution

$$\Pr[W = w] = 2^{-nR},$$

for $w = 1, 2, \ldots, 2^{nR}$. The w-th codeword $x_{\mathcal{C}}^n(w)$, corresponding to the w-th row of $\mathcal{C}$, is sent over the channel.

- The receiver receives a sequence $Y_{\mathcal{C}}^n$ according to the distribution according to the distribution

$$\Pr\left(y_{\mathcal{C}}^n | x_{\mathcal{C}}^n(w)\right) = \prod_{i=1}^{n} \Pr\left(y_{i,\mathcal{C}} | x_{i,\mathcal{C}}(w)\right),$$

and guesses which message was sent using jointly typical decoding.

Proof for achievability

- Let $\varepsilon = \{\hat{W}(Y^n) \neq W\}$ denote the error event, $\lambda_w(\mathcal{C})$ be the error probability of the w-th codeword of code $\mathcal{C}$. The average probability of error, over all codewords and all codebooks, is:

$$\begin{aligned}\Pr(\varepsilon) &= \sum_{\mathcal{C}} \Pr(\mathcal{C}) P_e^{(n)}(\mathcal{C}) = \sum_{\mathcal{C}} \Pr(\mathcal{C}) \frac{1}{2^{nR}} \sum_{w=1}^{2^{nR}} \lambda_w(\mathcal{C}) \\ &= \frac{1}{2^{nR}} \sum_{w=1}^{2^{nR}} \sum_{\mathcal{C}} \Pr(\mathcal{C}) \lambda_w(\mathcal{C}) = \sum_{\mathcal{C}} \Pr(\mathcal{C}) \lambda_1(\mathcal{C}),\end{aligned}$$

where $\sum_{\mathcal{C}} \Pr(\mathcal{C}) \lambda_1(\mathcal{C}) = \sum_{\mathcal{C}} \Pr(\mathcal{C}) \lambda_w(\mathcal{C})$, $\forall w \neq 1$.

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Proof for achievability

- Let $Y_{\mathcal{C}}^n$ be the received signal for $x_{\mathcal{C}}^n(1)$

$$e_i(\mathcal{C}) = \{(x_{\mathcal{C}}^n(i), Y_{\mathcal{C}}^n) \in A_\epsilon^{(n)}\}, i \in \{1, 2, \ldots, 2^{nR}\},$$

and $e_i^c(\mathcal{C}) =! e_i(C)$. Thus,

$$\Pr[\varepsilon] = \sum_{\mathcal{C}} \Pr(\mathcal{C})\lambda_1(\mathcal{C}) = \sum_{\mathcal{C}} \Pr(\mathcal{C}) \Pr\left[e_1^c(\mathcal{C}) \cup (\cup_{i=2}^{2^{nR}} e_i(\mathcal{C}))\middle| W = 1\right]$$

$$\leq \sum_{\mathcal{C}} \Pr(\mathcal{C}) \Pr[e_1^c(\mathcal{C})|W = 1] + \sum_{\mathcal{C}} \Pr(\mathcal{C}) \sum_{i=2}^{2^{nR}} \Pr[e_i(\mathcal{C})|W = 1]$$

$$= \sum_{\mathcal{C}} \Pr(\mathcal{C}) \Pr[e_1^c(\mathcal{C})|W = 1] + \sum_{i=2}^{2^{nR}} \sum_{\mathcal{C}} \Pr(\mathcal{C}) \Pr[e_i(\mathcal{C})|W = 1]$$

Proof for achievability

$$\sum_{\mathcal{C}} \Pr(\mathcal{C}) \Pr[e_1^c(\mathcal{C})|W=1]$$

$$= \sum_{\mathcal{C}} \left(\prod_{i=1}^{2^{nR}} \Pr(x_{\mathcal{C}}^n(i)) \right) \Pr[e_1^c(\mathcal{C})|W=1]$$

$$= \sum_{x_1^n} \sum_{\mathcal{C}: x_{\mathcal{C}}^n(1)=x_1^n} \prod_{i=1}^{2^{nR}} \Pr(x_{\mathcal{C}}^n(i)) \Pr(x_1^n \text{ and } Y^n \text{ are not joint typical}|W=1)$$

$$= \sum_{x_1^n} \Pr(x_1^n) \Pr(x_1^n \text{ and } Y^n \text{ are not joint typical}|W=1)$$

$$\times \sum_{\mathcal{C}: x_{\mathcal{C}}^n(1)=x_1^n} \prod_{i=2}^{2^{nR}} \Pr(x_{\mathcal{C}}^n(i))$$

$$= \sum_{x_1^n} \Pr(x_1^n) \Pr(x_1^n \text{ and } Y^n \text{ are not joint typical}|W=1)$$

$$= \Pr(X_1^n \text{ and } Y^n \text{ are not joint typical}|W=1) = \Pr(E_1^c|W=1)$$

Proof for achievability

- Similarly,

$$\sum_{\mathcal{C}} \Pr(\mathcal{C}) \Pr[e_1(\mathcal{C})|W=1] = \Pr(X_i^n \text{ and } Y^n \text{ are joint typical}|W=1)$$

$$= \Pr(E_i|W=1)$$

- As a result,

$$\Pr[\varepsilon] \leq \Pr[E_1^c|W=1] + \sum_{i=2}^{2^{nR}} \Pr[E_i|W=1]$$

Proof for achievability

- By the joint AEP, $\Pr[E_1^c|W=1] \leq \epsilon$ for n sufficiently large. By the code generation process, $X^n(1)$ and $X^n(i)$ are independent for $i \neq 1$, so are Y^n and $X^n(i)$. Hence the probability that $X^n(i)$ and Y^n are jointly typical is $\leq 2^{-n(I(X;Y)-3\epsilon)}$ by the joint AEP.

$$\begin{aligned}
\Pr[\varepsilon] &\leq \epsilon + \sum_{i=2}^{2^{nR}} 2^{-n(I(X;Y)-3\epsilon)} \\
&= \epsilon + (2^{nR}-1)2^{-n(I(X;Y)-3\epsilon)} \\
&\leq \epsilon + 2^{3n\epsilon}2^{-n(I(X;Y)-R)} \\
&\leq 2\epsilon \quad \text{for } R \leq I(X;Y) - 4\epsilon \text{ and sufficiently large n}
\end{aligned}$$

  Hence, if $R < I(X;Y)$, we can choose ϵ and n so that the average probability of error, over codebooks and codewords, is less than 2ϵ.

- Since $p(x)$ is the capacity achieving distribution, $R < I(X;Y)$ beacomes $R < C$.

Proof for achievability

- Get rid of the average over codebooks. Since the average probability of error is $\leq 2\epsilon$, there exists at least one codebook $\mathcal{C}^*$ with a small average probability of error ($\Pr(\varepsilon|\mathcal{C}^*) \leq 2\epsilon$). Since we have chosen $\hat{W}$ according to a uniform distribution, we have

$$\Pr(\varepsilon|\mathcal{C}^*) = \frac{1}{2^{nR}} \sum_{i=1}^{2^{nR}} \lambda_i(\mathcal{C}^*).$$

- Throw away the worst half of the codewords in the best codebook $\mathcal{C}^*$. We have $\Pr(\varepsilon|\mathcal{C}^*) \leq \frac{1}{2^{nR}} \sum \lambda_i(\mathcal{C}^*) \leq 2\epsilon$. This implies that at least half the indices i and their associated codewords $X^n(I)$ must have conditional probability of error $\lambda_i \leq 4\epsilon$. If we reindex the codewords, we have 2^{nR-1} codewords. The rate now is $R' = R - \frac{1}{n}$ with maximal probability of error $\lambda^{(n)} \leq 4\epsilon$.

Proof for the converse

- The index W is uniformly distributed on the set $\mathcal{W} = \{1, 2, \ldots, 2^{nR}\}$, and the sequence Y^n is related to W. From Y^n, we estimate the index W as $\hat{W} = g(Y^n)$. Thus, $W \to X^n(W) \to Y^n \to \hat{W}$ forms a Markov chain.

Data processing inequality: $I(W; \hat{W}) \leq I(X^n(W); Y^n)$

Lemma (Fano's inequality)

For a discrete memoryless channel with a codebook $\mathcal{C}$ and the input message W uniformly distributed over 2^{nR}, we have

$$H(W|\hat{W}) \leq 1 + P_e^{(n)} nR.$$

Proof for the converse

- The index W is uniformly distributed on the set $\mathcal{W} = \{1, 2, \ldots, 2^{nR}\}$, and the sequence Y^n is related to W. From Y^n, we estimate the index W as $\hat{W} = g(Y^n)$. Thus, $W \to X^n(W) \to Y^n \to \hat{W}$ forms a Markov chain.

$$\text{Data processing inequality:} \quad I(W; \hat{W}) \leq I(X^n(W); Y^n)$$

Lemma (Fano's inequality)

For a discrete memoryless channel with a codebook $\mathcal{C}$ and the input message W uniformly distributed over 2^{nR}, we have

$$H(W|\hat{W}) \leq 1 + P_e^{(n)} nR.$$

Lemma

Let Y^n be the result of passing X^n through a discrete memoryless channel of capacity C. Then

$$I(X^n;Y^n) \leq nC, \quad \textit{for all} \quad p(x^n).$$

Proof.

$$I(X^n;Y^n) = H(Y^n) - H(Y^n|X^n) = H(Y^n) - \sum_{i=1}^{n} H(Y_i|Y_1,\ldots,Y_{i-1},X^n)$$

$$= H(Y^n) - \sum_{i=1}^{n} H(Y_i|X_i) \quad \text{memoryless}$$

$$\leq \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|X_i) \quad \text{independence bound}$$

$$= \sum_{i=1}^{n} I(X_i|Y_i) \leq nC$$

Lemma

Let Y^n be the result of passing X^n through a discrete memoryless channel of capacity C. Then

$$I(X^n;Y^n) \le nC, \quad \textit{for all} \quad p(x^n).$$

Proof.

$$I(X^n;Y^n) = H(Y^n) - H(Y^n|X^n) = H(Y^n) - \sum_{i=1}^{n} H(Y_i|Y_1,\ldots,Y_{i-1},X^n)$$

$$= H(Y^n) - \sum_{i=1}^{n} H(Y_i|X_i) \quad \text{memoryless}$$

$$\le \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|X_i) \quad \text{independence bound}$$

$$= \sum_{i=1}^{n} I(X_i|Y_i) \le nC$$

Proof for the converse

Proof.

Converse to channel coding theorem: Since W has a uniform distribution, we have

$$\begin{aligned} nR = H(W) &= H(W|\hat{W}) + I(W;\hat{W}) \\ &\leq 1 + P_e^{(n)} nR + I(W;\hat{W}) \quad \text{Fano's inequality} \\ &\leq 1 + P_e^{(n)} nR + I(X^n;Y^n) \quad \text{data-processing inequality} \\ &\leq 1 + P_e^{(n)} nR + nC \quad \text{Lemma 7.9.2} \end{aligned}$$

We obtain $R \leq \frac{1}{n(1+P_e^{(n)})} + \frac{C}{1+P_e^{(n)}} \to \frac{1}{n} + C$.

Letting $n \to \infty$, we have $R \leq C$.

□

Reading & Homework

- Reading: Chapter 7: 7.6-7.10
- Homework: Problems 7.15, 7.31.