

Chapter 5

Channel Coding Theorems and Approximations of Output Statistics for Arbitrary Channels

Po-Ning Chen

Department of Communications Engineering

National Chiao-Tung University

Hsin Chu, Taiwan 30050

Preface

II: 5-1

- The Shannon channel capacity in Volume I of the lecture notes is derived under the assumption that the channel is memoryless.
- With moderate modification of the proof, this result was extended to stationary-ergodic channels for which the capacity formula becomes the maximization of the mutual information rate:

$$\lim_{n \rightarrow \infty} \sup_{X^n} \frac{1}{n} I(X^n; Y^n).$$

- Yet, for more general channels, such as non-stationary or non-ergodic channels, a more general expression for channel capacity needs to be derived.

General models for channels

II: 5-2

- The channel transition probability in its most general form is denoted by $\{W^n = P_{Y^n|X^n}\}_{n=1}^{\infty}$, which is abbreviated by \mathbf{W} for convenience.
- Similarly, the input and output random processes are respectively denoted by \mathbf{X} and \mathbf{Y} .

Notations

II: 5-3

- The sup- and inf- (mutual-)information rates are respectively defined by

$$\bar{I}(\mathbf{X}; \mathbf{Y}) \triangleq \sup\{\theta : \underline{i}(\theta) < 1\}$$

and

$$\underline{I}(\mathbf{X}; \mathbf{Y}) \triangleq \sup\{\theta : \bar{i}(\theta) \leq 0\},$$

where

$$\underline{i}(\theta) \triangleq \liminf_{n \rightarrow \infty} Pr \left\{ \frac{1}{n} i_{X^n W^n}(X^n; Y^n) \leq \theta \right\}$$

is the inf-spectrum of the normalized information density,

$$\bar{i}(\theta) \triangleq \limsup_{n \rightarrow \infty} Pr \left\{ \frac{1}{n} i_{X^n W^n}(X^n; Y^n) \leq \theta \right\}$$

is the sup-spectrum of the normalized information density, and

$$i_{X^n W^n}(x^n; y^n) \triangleq \log \frac{P_{Y^n|X^n}(y^n|x^n)}{P_{Y^n}(y^n)}$$

is the *information density*.

Historical background

II: 5-4

- In 1994, Verdú and Han have shown that the channel capacity in its most general form is

$$C \triangleq \sup_{\mathbf{X}} I(\mathbf{X}; \mathbf{Y}).$$

- In their proof, they showed the achievability part in terms of Feinstein's Lemma, and provide a new proof for the converse part. In this section, we will not adopt the original proof of Verdú and Han in the converse theorem. Instead, we will use a new and tighter bound established by Poor and Verdú in 1995.

Notations and definitions

II: 5-5

Definition 5.1 (fixed-length data transmission code) An (n, M) fixed-length data transmission code for channel input alphabet \mathcal{X}^n and output alphabet \mathcal{Y}^n consists of

1. M informational messages intended for transmission;
2. an encoding function

$$f : \{1, 2, \dots, M\} \rightarrow \mathcal{X}^n;$$

3. a decoding function

$$g : \mathcal{Y}^n \rightarrow \{1, 2, \dots, M\},$$

which is (usually) a deterministic rule that assigns a guess to each possible received vector.

The channel inputs in $\{x^n \in \mathcal{X}^n : x^n = f(m) \text{ for some } 1 \leq m \leq M\}$ are the codewords of the data transmission code.

Notations and definitions

II: 5-6

Definition 5.2 (average probability of error) The average probability of error for a $\mathcal{C}_n = (n, M)$ code with encoder $f(\cdot)$ and decoder $g(\cdot)$ transmitted over channel $Q_{Y^n|X^n}$ is defined as

$$P_e(\mathcal{C}_n) = \frac{1}{M} \sum_{i=1}^M \lambda_i,$$

where

$$\lambda_i \triangleq \sum_{\{y^n \in \mathcal{Y}^n : g(y^n) \neq i\}} Q_{Y^n|X^n}(y^n | f(i)).$$

Under the criterion of average probability of error, all of the codewords are treated equally, namely the prior probability of the selected M codewords are uniformly distributed.

Feinstein's Lemma

II: 5-7

Lemma 5.3 (Feinstein's Lemma) Fix a positive n . For every $\gamma > 0$ and input distribution P_{X^n} on \mathcal{X}^n , there exists an (n, M) block code for the transition probability $P_{W^n} = P_{Y^n|X^n}$ that its average error probability $P_e(\mathcal{C}_n)$ satisfies

$$P_e(\mathcal{C}_n) < Pr \left[\frac{1}{n} i_{X^n W^n}(X^n; Y^n) < \frac{1}{n} \log M + \gamma \right] + e^{-n\gamma}.$$

Proof:

Step 1: Notations. Define

$$\mathcal{G} \triangleq \left\{ (x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n : \frac{1}{n} i_{X^n W^n}(x^n; y^n) \geq \frac{1}{n} \log M + \gamma \right\}.$$

Let $\nu \triangleq e^{-n\gamma} + P_{X^n W^n}(\mathcal{G}^c)$.

The Feinstein's Lemma obviously holds if $\nu \geq 1$, because then

$$P_e(\mathcal{C}_n) \leq 1 \leq \nu \triangleq Pr \left[\frac{1}{n} i_{X^n W^n}(X^n; Y^n) < \frac{1}{n} \log M + \gamma \right] + e^{-n\gamma}.$$

So we assume $\nu < 1$, which immediately results in

$$P_{X^n W^n}(\mathcal{G}^c) < \nu < 1,$$

or equivalently,

$$P_{X^n W^n}(\mathcal{G}) > 1 - \nu > 0.$$

Feinstein's Lemma

II: 5-8

Therefore,

$$P_{X^n}(\mathcal{A} \triangleq \{x^n \in \mathcal{X}^n : P_{Y^n|X^n}(\mathcal{G}_{x^n}|x^n) > 1 - \nu\}) > 0,$$

where $\mathcal{G}_{x^n} \triangleq \{y^n \in \mathcal{Y}^n : (x^n, y^n) \in \mathcal{G}\}$, because if $P_{X^n}(\mathcal{A}) = 0$,

$$(\forall x^n \text{ with } P_{X^n}(x^n) > 0) P_{Y^n|X^n}(\mathcal{G}_{x^n}|x^n) \leq 1 - \nu$$

$$\Rightarrow \sum_{x^n \in \mathcal{X}^n} P_{X^n}(x^n) P_{Y^n|X^n}(\mathcal{G}_{x^n}|x^n) = P_{X^n Y^n}(\mathcal{G}) \leq 1 - \nu.$$

Step 2: Encoder. Choose an x_1^n in \mathcal{A} (Recall that $P_{X^n}(\mathcal{A}) > 0$.) Define $\Gamma_1 = \mathcal{G}_{x_1^n}$. (Then $P_{Y^n|X^n}(\Gamma_1|x_1^n) > 1 - \nu$.)

Next choose, if possible, a point $x_2^n \in \mathcal{X}^n$ without replacement (i.e., x_2^n cannot be identical to x_1^n) for which

$$P_{Y^n|X^n}(\mathcal{G}_{x_2^n} - \Gamma_1 | x_2^n) > 1 - \nu,$$

and define $\Gamma_2 \triangleq \mathcal{G}_{x_2^n} - \Gamma_1$.

Continue in the following way as for codeword i : choose x_i^n to satisfy

$$P_{Y^n|X^n} \left(\mathcal{G}_{x_i^n} - \bigcup_{j=1}^{i-1} \Gamma_j \middle| x_i^n \right) > 1 - \nu,$$

Feinstein's Lemma

II: 5-9

and define $\Gamma_i \triangleq \mathcal{G}_{x_i^n} - \bigcup_{j=1}^{i-1} \Gamma_j$.

Repeat the above codeword selecting procedure until either M codewords have been selected or all the points in \mathcal{A} have been exhausted.

Step 3: Decoder. Define the decoding rule as

$$\phi(\mathbf{y}^n) = \begin{cases} i, & \text{if } \mathbf{y}^n \in \Gamma_i \\ \text{arbitrary } \mathbf{y}, & \text{otherwise.} \end{cases}$$

Step 4: Probability of error. For all selected codewords, the error probability given codeword i is transmitted, $\lambda_{e|i}$, satisfies

$$\lambda_{e|i} \leq P_{Y^n|X^n}(\Gamma_i^c|x_i^n) < \nu.$$

(Note that $(\forall i) P_{X^n|X^n}(\Gamma_i|x_i^n) \geq 1 - \nu$ by step 2.) Therefore, if we can show that the above codeword selecting procedures will not be terminated before M , then

$$P_e(\mathcal{E}_n) = \frac{1}{M} \sum_{i=1}^M \lambda_{e|i} < \nu.$$

Step 5: Claim. The codeword selecting procedure in step 2 will not be terminated before M .

Feinstein's Lemma

II: 5-10

proof: We will prove it by contradiction.

Suppose the above procedure terminated before M , say at $N < M$. Define the set

$$\mathcal{F} \triangleq \bigcup_{i=1}^N \Gamma_i \in \mathcal{Y}^n.$$

Consider the probability

$$P_{X^n W^n}(\mathcal{G}) = P_{X^n W^n}[\mathcal{G} \cap (\mathcal{X}^n \times \mathcal{F})] + P_{X^n W^n}[\mathcal{G} \cap (\mathcal{X}^n \times \mathcal{F}^c)]. \quad (5.2.1)$$

Since for any $y^n \in \mathcal{G}_{x_i^n}$,

$$P_{Y^n}(y^n) \leq \frac{P_{Y^n|X^n}(y^n|x_i^n)}{M \cdot e^{n\gamma}},$$

we have

$$\begin{aligned} P_{Y^n}(\Gamma_i) &\leq P_{Y^n}(\mathcal{G}_{x_i^n}) \\ &\leq \frac{1}{M} e^{-n\gamma} P_{Y^n|X^n}(\mathcal{G}_{x_i^n}) \\ &\leq \frac{1}{M} e^{-n\gamma}. \end{aligned}$$

Feinstein's Lemma

II: 5-11

So the first term of the right hand side in (5.2.1) can be upper bounded by

$$\begin{aligned} P_{X^n W^n}[\mathcal{G} \cap (\mathcal{X}^n \times \mathcal{F})] &\leq P_{X^n W^n}(\mathcal{X}^n \times \mathcal{F}) \\ &= P_{Y^n}(\mathcal{F}) \\ &= \sum_{i=1}^N P_{Y^n}(\Gamma_i) \\ &\leq N \times \frac{1}{M} e^{-n\gamma} = \frac{N}{M} e^{-n\gamma}. \end{aligned}$$

As for the second term of the right hand side in (5.2.1), we can upper bound it by

$$\begin{aligned} P_{X^n W^n}[\mathcal{G} \cap (\mathcal{X}^n \times \mathcal{F}^c)] &= \sum_{x^n \in \mathcal{X}^n} P_{X^n}(x^n) P_{Y^n|X^n}(\mathcal{G}_{x^n} \cap \mathcal{F}^c | x^n) \\ &= \sum_{x^n \in \mathcal{X}^n} P_{X^n}(x^n) P_{Y^n|X^n} \left(\mathcal{G}_{x^n} - \bigcup_{i=1}^N \Gamma_i \middle| x^n \right) \\ &\leq \sum_{x^n \in \mathcal{X}^n} P_{X^n}(x^n) (1 - \nu) \leq (1 - \nu), \end{aligned}$$

Feinstein's Lemma

II: 5-12

where the last step follows since for all $x^n \in \mathcal{X}^n$,

$$P_{Y^n|X^n} \left(\mathcal{G}_{x^n} - \bigcup_{i=1}^N \Gamma_i \middle| x^n \right) \leq 1 - \nu.$$

(Because otherwise we could find the $(N + 1)$ -th codeword.)

Consequently, $P_{X^n W^n}(\mathcal{G}) \leq (N/M)e^{-n\gamma} + (1 - \nu)$. By definition of \mathcal{G} ,

$$P_{X^n W^n}(\mathcal{G}) = 1 - \nu + e^{-n\gamma} \leq \frac{N}{M}e^{-n\gamma} + (1 - \nu),$$

which implies $N \geq M$, a contradiction. □

Poor-Verdú lemma

II: 5-13

Lemma 5.4 (Poor-Verdú Lemma) Suppose X and Y are random variables, where X taking values on a finite (or countably infinite) set

$$\mathcal{X} = \{x_1, x_2, x_3, \dots\}.$$

The minimum probability of error P_e in estimating X from Y satisfies

$$P_e \geq (1 - \alpha) P_{X,Y} \{(x, y) \in \mathcal{X} \times \mathcal{Y} : P_{X|Y}(x|y) \leq \alpha\}$$

for each $\alpha \in [0, 1]$.

Proof: It is known that the minimum-error-probability estimate $e(y)$ of X when receiving y is

$$e(y) = \arg \max_{x \in \mathcal{X}} P_{X|Y}(x|y). \quad (5.2.2)$$

Therefore, the error probability incurred in testing among the values of X is given

Poor-Verdú lemma

II: 5-14

by

$$\begin{aligned} 1 - P_e &= Pr\{X = e(Y)\} \\ &= \int_{\mathcal{Y}} \left[\sum_{\{x : x=e(y)\}} P_{X|Y}(x|y) \right] dP_Y(y) \\ &= \int_{\mathcal{Y}} \left(\max_{x \in \mathcal{X}} P_{X|Y}(x|y) \right) dP_Y(y) \\ &= \int_{\mathcal{Y}} \left(\max_{x \in \mathcal{X}} f_x(y) \right) dP_Y(y) \\ &= E \left[\max_{x \in \mathcal{X}} f_x(Y) \right], \end{aligned}$$

where $f_x(y) \triangleq P_{X|Y}(x|y)$. Let $h_j(y)$ be the j -th element in the re-ordering set of $\{f_{x_1}(y), f_{x_2}(y), f_{x_3}(y), \dots\}$ according to ascending element values. In other words,

$$h_1(y) \geq h_2(y) \geq h_3(y) \geq \dots$$

and

$$\{h_1(y), h_2(y), h_3(y), \dots\} = \{f_{x_1}(y), f_{x_2}(y), f_{x_3}(y), \dots\}.$$

Then

$$1 - P_e = E[h_1(Y)]. \tag{5.2.3}$$

Poor-Verdú lemma

II: 5-15

For any $\alpha \in [0, 1]$, we can write

$$P_{X,Y} \{(x, y) \in \mathcal{X} \times \mathcal{Y} : f_x(y) > \alpha\} = \int_{\mathcal{Y}} P_{X|Y} \{x \in \mathcal{X} : f_x(y) > \alpha\} dP_Y(y).$$

Observe that

$$\begin{aligned} P_{X|Y} \{x \in \mathcal{X} : f_x(y) > \alpha\} &= \sum_{x \in \mathcal{X}} P_{X|Y}(x|y) \cdot \mathbf{1}[f_x(y) > \alpha] \\ &= \sum_{x \in \mathcal{X}} f_x(y) \cdot \mathbf{1}[f_x(y) > \alpha] \\ &= \sum_{j=1}^{\infty} h_j(y) \cdot \mathbf{1}[h_j(y) > \alpha], \end{aligned}$$

where $\mathbf{1}(\cdot)$ is the indicator function. Therefore,

$$\begin{aligned} P_{XY} \{(x, y) \in \mathcal{X} \times \mathcal{Y} : f_x(y) > \alpha\} &= \int_{\mathcal{Y}} \left(\sum_{j=1}^{\infty} h_j(y) \cdot \mathbf{1}[h_j(y) > \alpha] \right) dP_Y(y) \\ &\geq \int_{\mathcal{Y}} h_1(y) \cdot \mathbf{1}(h_1(y) > \alpha) dP_Y(y) \\ &= E[h_1(Y) \cdot \mathbf{1}(h_1(Y) > \alpha)]. \end{aligned}$$

Poor-Verdú lemma

II: 5-16

It remains to relate $E[h_1(Y) \cdot \mathbf{1}(h_1(Y) > \alpha)]$ with $E[h_1(Y)]$, which is exactly $1 - P_e$. For any $\alpha \in [0, 1]$ and any random variable U with $Pr\{0 \leq U \leq 1\} = 1$,

$$U \leq \alpha + (1 - \alpha) \cdot U \cdot \mathbf{1}(U > \alpha)$$

holds with probability one. This can be easily proved by upper-bounding U in terms of α when $0 \leq U \leq \alpha$, and $\alpha + (1 - \alpha)U$, otherwise. Thus

$$E[U] \leq \alpha + (1 - \alpha)E[U \cdot \mathbf{1}(U > \alpha)].$$

By letting $U = h_1(Y)$, together with (5.2.3), we finally obtain

$$\begin{aligned} (1 - \alpha)P_{XY} \{(x, y) \in \mathcal{X} \times \mathcal{Y} : f_x(y) > \alpha\} &\geq (1 - \alpha)E[h_1(Y) \cdot \mathbf{1}[h_1(Y) > \alpha]] \\ &\geq E[h_1(Y)] - \alpha \\ &= (1 - P_e) - \alpha \\ &= (1 - \alpha) - P_e. \end{aligned}$$

□

Poor-Verdú lemma

II: 5-17

Corollary 5.5 Every $\mathcal{C}_n = (n, M)$ code satisfies

$$P_e(\mathcal{C}_n) \geq (1 - e^{-n\gamma}) Pr \left[\frac{1}{n} i_{X^n W^n}(X^n; Y^n) \leq \frac{1}{n} \log M - \gamma \right]$$

for every $\gamma > 0$, where X^n places probability mass $1/M$ on each codeword, and $P_e(\mathcal{C}_n)$ denotes the error probability of the code.

Proof: Taking $\alpha = e^{-n\gamma}$ in Lemma 5.4, and replacing X and Y in Lemma 5.4 by its n -fold counterparts, i.e., X^n and Y^n , we obtain

$$\begin{aligned} P_e(\mathcal{C}_n) &\geq (1 - e^{-n\gamma}) P_{X^n W^n} \left[(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n : P_{X^n|Y^n}(x^n|y^n) \leq e^{-n\gamma} \right] \\ &= (1 - e^{-n\gamma}) P_{X^n W^n} \left[(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n : \frac{P_{X^n|Y^n}(x^n|y^n)}{1/M} \leq \frac{e^{-n\gamma}}{1/M} \right] \\ &= (1 - e^{-n\gamma}) P_{X^n W^n} \left[(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n : \frac{P_{X^n|Y^n}(x^n|y^n)}{P_{X^n}(x^n)} \leq \frac{e^{-n\gamma}}{1/M} \right] \\ &= (1 - e^{-n\gamma}) P_{X^n W^n} \left[(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n : \right. \\ &\quad \left. \frac{1}{n} \log \frac{P_{X^n|Y^n}(x^n|y^n)}{P_{X^n}(x^n)} \leq \frac{1}{n} \log M - \gamma \right] \\ &= (1 - e^{-n\gamma}) Pr \left[\frac{1}{n} i_{X^n W^n}(X^n; Y^n) \leq \frac{1}{n} \log M - \gamma \right]. \end{aligned}$$

□

Definitions of general Shannon capacity

II: 5-18

Definition 5.6 (ε -achievable rate) Fix $\varepsilon \in [0, 1]$. $R \geq 0$ is an ε -achievable rate if there exists a sequence of $\mathcal{C}_n = (n, M_n)$ channel block codes such that

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log M_n \geq R$$

and

$$\limsup_{n \rightarrow \infty} P_e(\mathcal{C}_n) \leq \varepsilon.$$

Definition 5.7 (ε -capacity C_ε) Fix $\varepsilon \in [0, 1]$. The supremum of ε -achievable rates is called the ε -capacity, C_ε .

It is straightforward for the definition that C_ε is non-decreasing in ε , and $C_1 = \log |\mathcal{X}|$.

Definition 5.8 (capacity C) The channel capacity C is defined as the supremum of the rates that are ε -achievable for all $\varepsilon \in [0, 1]$. It follows immediately from the definition that $C = \inf_{0 \leq \varepsilon \leq 1} C_\varepsilon = \lim_{\varepsilon \downarrow 0} C_\varepsilon = C_0$ and that C is the supremum of all the rates R for which there exists a sequence of $\mathcal{C}_n = (n, M_n)$ channel block codes such that

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log M_n \geq R,$$

and

$$\limsup_{n \rightarrow \infty} P_e(\mathcal{C}_n) = 0.$$

ε -capacity

II: 5-19

Theorem 5.9 (ε -capacity) For $0 < \varepsilon < 1$, the ε -capacity C_ε for arbitrary channels satisfies

$$C_\varepsilon = \sup_{\mathbf{X}} \underline{I}_\varepsilon(\mathbf{X}; \mathbf{Y}).$$

Proof:

1. $C_\varepsilon \geq \sup_{\mathbf{X}} \underline{I}_\varepsilon(\mathbf{X}; \mathbf{Y})$.

Fix input \mathbf{X} . It suffices to show the existence of $\mathcal{C}_n = (n, M_n)$ data transmission code with rate

$$\underline{I}_\varepsilon(\mathbf{X}; \mathbf{Y}) - \gamma < \frac{1}{n} \log M_n < \underline{I}_\varepsilon(\mathbf{X}; \mathbf{Y}) - \frac{\gamma}{2}$$

and probability of decoding error satisfying

$$\limsup_{n \rightarrow \infty} P_e(\mathcal{C}_n) \leq \varepsilon$$

for every $\gamma > 0$. (Because if such code exists, then $\limsup_{n \rightarrow \infty} P_e(\mathcal{C}_n) \leq \varepsilon$ and $\liminf_{n \rightarrow \infty} (1/n) \log M_n \geq \underline{I}_\varepsilon(\mathbf{X}; \mathbf{Y}) - \gamma$, which implies $C_\varepsilon \geq \underline{I}_\varepsilon(\mathbf{X}; \mathbf{Y})$.)

ε -capacity

II: 5-20

From Lemma 5.3, there exists an $\mathcal{C}_n = (n, M_n)$ code whose error probability satisfies

$$\begin{aligned} P_e(\mathcal{C}_n) &< Pr \left[\frac{1}{n} i_{X^n W^n}(X^n; Y^n) < \frac{1}{n} \log M_n + \frac{\gamma}{4} \right] + e^{-n\gamma/4} \\ &\leq Pr \left[\frac{1}{n} i_{X^n W^n}(X^n; Y^n) < \left(\underline{I}_\varepsilon(\mathbf{X}; \mathbf{Y}) - \frac{\gamma}{2} \right) + \frac{\gamma}{4} \right] + e^{-n\gamma/4} \\ &\leq Pr \left[\frac{1}{n} i_{X^n W^n}(X^n; Y^n) < \underline{I}_\varepsilon(\mathbf{X}; \mathbf{Y}) - \frac{\gamma}{4} \right] + e^{-n\gamma/4}. \end{aligned}$$

Since

$$\underline{I}_\varepsilon(\mathbf{X}; \mathbf{Y}) \triangleq \sup \left\{ R : \limsup_{n \rightarrow \infty} Pr \left[\frac{1}{n} i_{W^n W^n}(X^n; Y^n) \leq R \right] \leq \varepsilon \right\},$$

we obtain

$$\limsup_{n \rightarrow \infty} Pr \left[\frac{1}{n} i_{X^n W^n}(X^n; Y^n) < \underline{I}_\varepsilon(\mathbf{X}; \mathbf{Y}) - \frac{\gamma}{4} \right] \leq \varepsilon.$$

Hence, the proof of the direct part is completed by noting that

$$\begin{aligned} \limsup_{n \rightarrow \infty} P_e(\mathcal{C}_n) &\leq \limsup_{n \rightarrow \infty} Pr \left[\frac{1}{n} i_{X^n W^n}(X^n; Y^n) < \underline{I}_\varepsilon(\mathbf{X}; \mathbf{Y}) - \frac{\gamma}{4} \right] \\ &\quad + \limsup_{n \rightarrow \infty} e^{-n\gamma/4} = \varepsilon. \end{aligned}$$

ε -capacity

II: 5-21

$$2. C_\varepsilon \leq \sup_{\mathbf{X}} \underline{I}_\varepsilon(\mathbf{X}; \mathbf{Y}).$$

Suppose that there exists a sequence of $\mathcal{C}_n = (n, M_n)$ codes with rate strictly larger than $\sup_{\mathbf{X}} \underline{I}_\varepsilon(\mathbf{X}; \mathbf{Y})$ and $\limsup_{n \rightarrow \infty} P_e(\mathcal{C}_n) \leq \varepsilon$. Let the ultimate code rate for this code be $\sup_{\mathbf{X}} \underline{I}_\varepsilon(\mathbf{X}; \mathbf{Y}) + 3\rho$ for some $\rho > 0$. Then for sufficiently large n ,

$$\frac{1}{n} \log M_n > \sup_{\mathbf{X}} \underline{I}_\varepsilon(\mathbf{X}; \mathbf{Y}) + 2\rho.$$

Since the above inequality holds for every \mathbf{X} , it certainly holds if taking input \hat{X}^n which places probability mass $1/M_n$ on each codeword, i.e.,

$$\frac{1}{n} \log M_n > \underline{I}_\varepsilon(\hat{\mathbf{X}}; \hat{\mathbf{Y}}) + 2\rho, \quad (5.2.4)$$

where $\hat{\mathbf{Y}}$ is the channel output due to channel input $\hat{\mathbf{X}}$. Then from Corollary 5.5, the error probability of the code satisfies

$$\begin{aligned} P_e(\mathcal{C}_n) &\geq (1 - e^{-n\rho}) Pr \left[\frac{1}{n} i_{\hat{X}^n W^n}(\hat{X}^n; \hat{Y}^n) \leq \frac{1}{n} \log M_n - \rho \right] \\ &\geq (1 - e^{-n\rho}) Pr \left[\frac{1}{n} i_{\hat{X}^n W^n}(\hat{X}^n; \hat{Y}^n) \leq \underline{I}_\varepsilon(\hat{\mathbf{X}}; \hat{\mathbf{Y}}) + \rho \right], \end{aligned}$$

where the last inequality follows from (5.2.4), which by taking the limsup of

ε -capacity

II: 5-22

both sides, we have

$$\varepsilon \geq \limsup_{n \rightarrow \infty} P_e(\mathcal{C}_n) \geq \limsup_{n \rightarrow \infty} Pr \left[\frac{1}{n} i_{\hat{X}^n W^n}(\hat{X}^n; \hat{Y}^n) \leq I_\varepsilon(\hat{\mathbf{X}}; \hat{\mathbf{Y}}) + \rho \right] > \varepsilon,$$

and a desired contradiction is obtained. □

General Shannon capacity and strong capacity

II: 5-23

Theorem 5.10 (general Shannon capacity) The channel capacity C for arbitrary channel satisfies

$$C = \sup_{\mathbf{X}} \underline{I}(\mathbf{X}; \mathbf{Y}).$$

Definition 5.11 (strong capacity) Define the *strong converse capacity* (or *strong capacity*) C_{SC} as the infimum of the rates R such that for all channel block codes $\mathcal{C}_n = (n, M_n)$ with

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log M_n \geq R,$$

we have

$$\liminf_{n \rightarrow \infty} P_e(\mathcal{C}_n) = 1.$$

Theorem 5.12 (general strong capacity)

$$C_{SC} \triangleq \sup_{\mathbf{X}} \bar{I}(\mathbf{X}; \mathbf{Y}).$$

Examples

II: 5-24

Example 5.13 (Shannon capacity) Let the input and output alphabets be $\{0, 1\}$, and let every output Y_i be given by:

$$Y_i = X_i \oplus N_i.$$

Assume the input process \mathbf{X} and the noise process \mathbf{N} are independent.

Then

$$\underline{H}(\mathbf{Y}) - \bar{H}(\mathbf{Y}|\mathbf{X}) \leq \underline{I}(\mathbf{X}; \mathbf{Y}) \leq \bar{H}(\mathbf{Y}) - \bar{H}(\mathbf{Y}|\mathbf{X})$$

or equivalently,

$$\underline{H}(\mathbf{Y}) - \bar{H}(\mathbf{N}) \leq \underline{I}(\mathbf{X}; \mathbf{Y}) \leq \bar{H}(\mathbf{Y}) - \bar{H}(\mathbf{N}).$$

By the channel symmetry, we obtain:

$$C = \log(2) - \bar{H}(\mathbf{N}) \text{ nats.}$$

Case A) If \mathbf{N} is a non-stationary binary independent sequence with

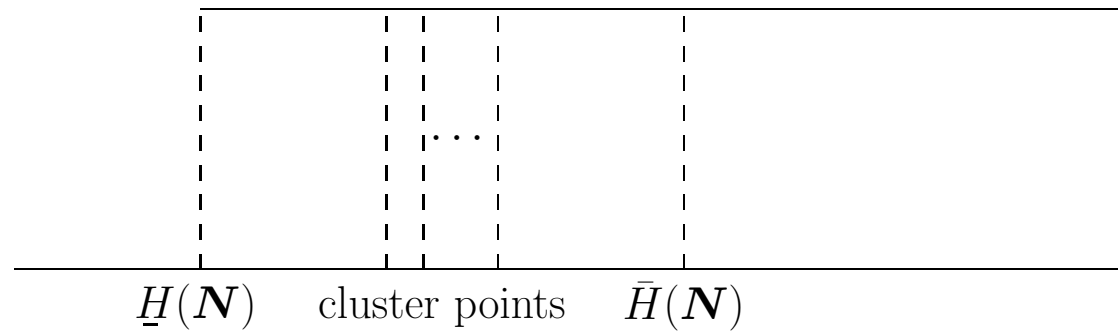
$$Pr\{N_i = 1\} = p_i,$$

then

$$C = \log(2) - \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n h_b(p_i) \quad \text{nats/channel usage.}$$

Examples

II: 5-25



The ultimate CDFs of $(1/n) \log P_{N^n}(N^n)$.

Case B) If \mathbf{N} has the same distribution as the source process in Example 4.23, then $\bar{H}(\mathbf{N}) = \log(2)$ nats, which yields zero channel capacity.

Examples

II: 5-26

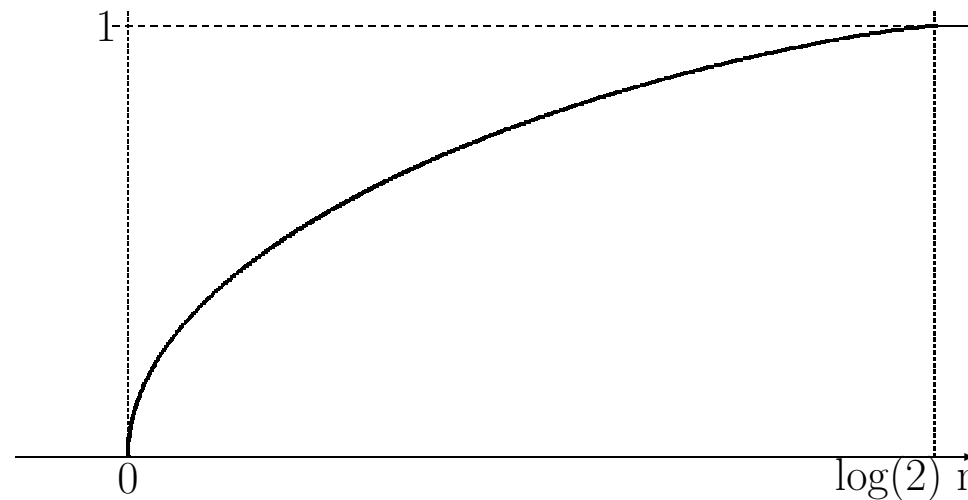
Example 5.14 (strong capacity)

Case A)

$$C_{SC} = 1 - \liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n h_b(p_i).$$

Case B)

$$C_{SC} = \log(2) \text{ nats/channel usage.}$$



The ultimate CDF $i(\theta)$ of the normalized information density for Example 5.14-Case B).

Examples

II: 5-27

Example 5.15 (ε -capacity) Continue from case B of Example 5.13.

$$C_\varepsilon = i^{-1}(\theta).$$

Structures of good data transmission codes

II: 5-28

- The channel capacity for discrete memoryless channel is shown to be:

$$C \triangleq \max_{P_X} I(P_X, Q_{Y|X}).$$

- Let $P_{\bar{X}}$ be the optimizer of the above maximization operation. Then

$$C \triangleq \max_{P_X} I(P_X, Q_{Y|X}) = I(P_{\bar{X}}, Q_{Y|X}).$$

- Here, the performance of the code is assumed to be the average error probability, namely

$$P_e(\mathcal{C}_n) = \frac{1}{M} \sum_{i=1}^M P_e(\mathcal{C}_n | x_i^n),$$

if the code book is $\mathcal{C}_n \triangleq \{x_1^n, x_2^n, \dots, x_M^n\}$.

- Due to the random coding argument, a deterministic good code with arbitrarily small error probability and rate less than channel capacity must exist.
- Question is what is the relationship between the good code and the optimizer $P_{\bar{X}}$? It is widely believed that if the code is good (with rate close to capacity and low error probability), then the output statistics $P_{\tilde{Y}^n}$ — due to the equally-likely code—must approximate the output distribution, denoted by $P_{\bar{Y}^n}$, due to the input distribution achieving the channel capacity.

Structures of good data transmission codes

II: 5-29

Theorem 5.16 For any channel $W^n = (Y^n|X^n)$ with finite input alphabet and capacity C that satisfies the strong converse (i.e., $C = C_{SC}$), the following statement holds.

Fixed any $\gamma > 0$ and any sequence of $\{\mathcal{C}_n = (n, M_n)\}_{n=1}^{\infty}$ block codes with

$$\frac{1}{n} \log M_n \geq C - \gamma/2,$$

and vanishing error probability,

$$\frac{1}{n} \|\tilde{Y}^n - \bar{Y}^n\| \leq \gamma \quad \text{for all sufficiently large } n,$$

where \tilde{Y}^n is the output due to the block code and \bar{Y}^n is the output due the \bar{X}^n that satisfies

$$I(\bar{X}^n; \bar{Y}^n) = \max_{X^n} I(X^n; Y^n).$$

To be specific,

$$P_{\tilde{Y}^n}(y^n) = \sum_{x^n \in \mathcal{C}_n} P_{\tilde{X}^n}(x^n) P_{W^n}(y^n|x^n) = \sum_{x^n \in \mathcal{C}_n} \frac{1}{M} P_{W^n}(y^n|x^n)$$

and

$$P_{\bar{Y}^n}(y^n) = \sum_{x^n \in \mathcal{X}^n} P_{\bar{X}^n}(x^n) P_{W^n}(y^n|x^n).$$

Structures of good data transmission codes

II: 5-30

- Note that the above theorem holds for arbitrary channels, not restricted to only discrete memoryless channels.
- One may query that “can a result in the spirit of the above theorem be proved for the input statistics rather than the output statistics?”
- The answer is negative.
- Hence, the statement that *the statistics of any good code must approximate those that maximize the mutual information* is erroneously taken for granted.
 - However, we do not rule out the possibility of the existence of good codes that approximate those that maximize the mutual information.

Structures of good data transmission codes

II: 5-31

- To see this, simply consider the normalized entropy of \overline{X}^n versus that of \tilde{X}^n (which is uniformly distributed over the codewords) for discrete memoryless channels:

$$\begin{aligned}\frac{1}{n}H(\overline{X}^n) - \frac{1}{n}H(\tilde{X}^n) &= \left[\frac{1}{n}H(\overline{X}^n|\overline{Y}^n) + \frac{1}{n}I(\overline{X}^n; \overline{Y}^n) \right] - \frac{1}{n}\log(M_n) \\ &= [H(\overline{X}|\overline{Y}) + I(\overline{X}; \overline{Y})] - \frac{1}{n}\log(M_n) \\ &= [H(\overline{X}|\overline{Y}) + C] - \frac{1}{n}\log(M_n).\end{aligned}$$

A good code with vanishing error probability exists for $(1/n)\log(M_n)$ arbitrarily close to C ; hence, we can find a good code sequence to satisfy

$$\lim_{n \rightarrow \infty} \left[\frac{1}{n}H(\overline{X}^n) - \frac{1}{n}H(\tilde{X}^n) \right] = H(\overline{X}|\overline{Y}).$$

Since the term $H(\overline{X}|\overline{Y})$ is in general positive, where a quick example is the BSC with crossover probability p , which yields $H(\overline{X}|\overline{Y}) = -p\log(p) - (1-p)\log(1-p)$, the two input distributions can by no means resemble to each other.

Resolvability for channels

II: 5-32

- The discussion of the previous section somewhat motivates the necessity to find a equally-distributed (over a subset of input alphabet) input distribution that generates the output statistics, which is close to the output due to the input that maximizes the mutual information.
- Since such approximations are usually performed by computers, it may be natural to connect approximations of the input and output statistics with the concept of *resolvability*.

Notations and definitions

II: 5-33

- In a data transmission system as shown in Figure 5.3, suppose that the source, channel and output are respectively denoted by

$$X^n \triangleq (X_1, \dots, X_n),$$

$$W^n \triangleq (W_1, \dots, W_n),$$

and

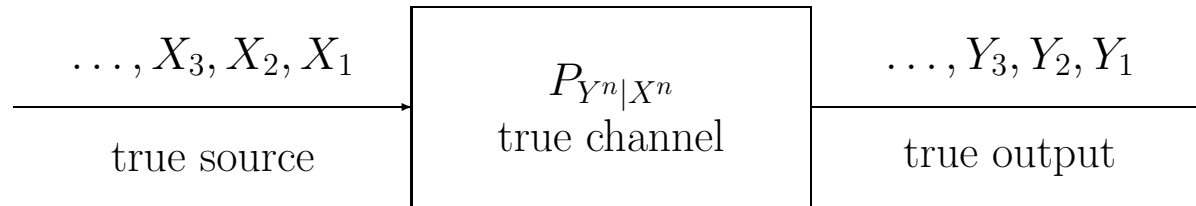
$$Y^n \triangleq (Y_1, \dots, Y_n),$$

where W_i has distribution $P_{Y_i|X_i}$.

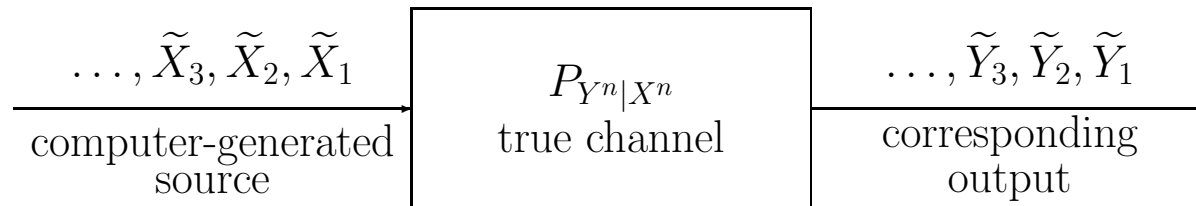
- To simulate the behavior of the channel, a computer-generated input may be necessary as shown in Figure 5.4.
- As stated in Chapter 4, such computer-generated input is based on an algorithm formed by a few basic uniform random experiments, which has finite resolution.
- Our goal is to find a good computer-generated input \tilde{X}^n such that the corresponding output \tilde{Y}^n is very close to the true output Y^n .
- Then according to the AEP theorem, we can form a good block code by adopting the typical set w.r.t. $P_{\tilde{X}^n}$ and have each codeword be (almost) equally-distributed with probability mass $e^{-H(\tilde{X}^n)}$.

Notations and definitions

II: 5-34



The communication system.



The simulated communication system.

Notations and definitions

II: 5-35

Definition 5.17 (ε -resolvability for input \mathbf{X} and channel \mathbf{W}) Fix $\varepsilon > 0$, and suppose that the (true) input random variable and (true) channel statistics are \mathbf{X} and $\mathbf{W} = (\mathbf{Y}|\mathbf{X})$, respectively.

Then the ε -resolvability $S_\varepsilon(\mathbf{X}, \mathbf{W})$ for input \mathbf{X} and channel \mathbf{W} is defined by:

$$S_\varepsilon(\mathbf{X}, \mathbf{W}) \triangleq \min \left\{ R : (\forall \gamma > 0)(\exists \tilde{\mathbf{X}} \text{ and } N)(\forall n > N) \right. \\ \left. \frac{1}{n}R(\tilde{X}^n) < R + \gamma \text{ and } \|Y^n - \tilde{Y}^n\| < \varepsilon \right\},$$

where $P_{\tilde{Y}^n} = P_{\tilde{X}^n}P_{W^n}$.

- Note that if we take the channel W^n to be an identity channel for all n , namely $\mathcal{X}^n = \mathcal{Y}^n$ and $P_{Y^n|X^n}(y^n|x^n)$ is either 1 or 0, then the ε -resolvability for input \mathbf{X} and channel \mathbf{W} is reduced to source ε -resolvability for source \mathbf{X} only:

$$S_\varepsilon(\mathbf{X}, \mathbf{W}_{\text{Identity}}) = S_\varepsilon(\mathbf{X}).$$

Similar reductions can be applied to all the following definitions.

Notations and definitions

II: 5-36

Definition 5.18 (ε -mean-resolvability for input \mathbf{X} and channel \mathbf{W})

Fix $\varepsilon > 0$, and suppose that the (true) input random variable and (true) channel statistics are respectively \mathbf{X} and \mathbf{W} .

Then the ε -mean-resolvability $\bar{S}_\varepsilon(\mathbf{X}, \mathbf{W})$ for input \mathbf{X} and channel \mathbf{W} is defined by:

$$\bar{S}_\varepsilon(\mathbf{X}, \mathbf{W}) \triangleq \min \left\{ R : (\forall \gamma > 0)(\exists \tilde{\mathbf{X}} \text{ and } N)(\forall n > N) \right. \\ \left. \frac{1}{n} H(\tilde{X}^n) < R + \gamma \text{ and } \|Y^n, \tilde{Y}^n\|_1 < \varepsilon \right\},$$

where $P_{\tilde{Y}^n} = P_{\tilde{X}^n} P_{W^n}$ and $P_{Y^n} = P_{X^n} P_{W^n}$.

Definition 5.19 (resolvability and mean resolvability for input \mathbf{X} and channel \mathbf{W})

The *resolvability* and *mean-resolvability* for input \mathbf{X} and \mathbf{W} are defined respectively as:

$$S(\mathbf{X}, \mathbf{W}) \triangleq \sup_{\varepsilon > 0} S_\varepsilon(\mathbf{X}, \mathbf{W}) \quad \text{and} \quad \bar{S}(\mathbf{X}, \mathbf{W}) \triangleq \sup_{\varepsilon > 0} \bar{S}_\varepsilon(\mathbf{X}, \mathbf{W}).$$

Notations and definitions

II: 5-37

Definition 5.20 (resolvability and mean resolvability for channel \mathbf{W})

The *resolvability* and *mean-resolvability* for channel \mathbf{W} are defined respectively as:

$$S(\mathbf{W}) \triangleq \sup_{\mathbf{X}} S(\mathbf{X}, \mathbf{W}),$$

and

$$\bar{S}(\mathbf{W}) \triangleq \sup_{\mathbf{X}} \bar{S}(\mathbf{X}, \mathbf{W}).$$

Theorem 5.21

$$S(\mathbf{W}) = C_{SC} = \sup_{\mathbf{X}} \bar{I}(\mathbf{X}; \mathbf{Y}).$$

It is somewhat a reasonable inference that if no computer algorithms can produce a desired good output statistics under the number of random nats specified, then all codes should be bad codes for this rate.

Theorem 5.22

$$\bar{S}(\mathbf{W}) = C = \sup_{\mathbf{X}} I(\mathbf{X}; \mathbf{Y}).$$

It is not yet clear that the operational relation between the resolvability (or mean-resolvability) and capacities for channels. This could be some interesting problems to think of.