# Principles of Communications

Meixia Tao

Shanghai Jiao Tong University

Chapter 9: Information Theory

Textbook: Chapter 12

《遶倕叻瑢》顧臭桧: 箈1.4苞

Communication Systems Engineering: Ch 6.1, Ch 9.1~9.2

# Information Theory

- Information theory is one of the key concepts in modern communications
- It deals with fundamental limits on communications
  - What is the highest rate at which information can be reliably transmitted over a communication channel?
  - What is the lowest rate at which information can be compressed and still be retrievable with small or no error?
  - What is the complexity of such optimal schemes?
- Topics to discuss
  - Modeling of information source
  - Source coding theorem
  - Modeling of communication channel
  - Channel capacity

# 9.1 Modeling of Information Source

- Information sources can be modeled by random processes
- The simplest model for information source is discrete memoryless source (DMS), a discrete-time, discrete-amplitude random process with i.i.d random variables
- A full description of DMS is given by:
  - Alphabet set $\mathcal{A} = \{a_1, a_1, \ldots, a_N\}$ where the random variable X takes its values
  - Probabilities $\{p_i\}_{i=1}^{N}$
- The information conveyed in different information sources can be different

# Information

- How to give a quantitative measure of information?
- Examples:
  - "the sun will rise" $\Rightarrow$ no information
  - "it will rain tomorrow" $\Rightarrow$ some information
  - "Final exam will be canceled" $\Rightarrow$ infinite information
- Information is connected with the elements of surprise, which is the result of uncertainty.
  - The smaller the probability of an event is, the more information the occurrence of that event will convey

# Measure of Information

- The information I that a source event x can will convey and the probability of the event P(x) satisfy:
  1. I=I[P(x)]
  2. P(x)↓→I↑, vice versa
     P(x)=1， I=0
  3. Consider multiple independent events x1, x2, …
     $I[P(x_1)P(x_2)\ldots]=I[P(x_1)]+I[P(x_2)]+\ldots$

- Definition (self information of symbol x):

$$I=\log_a \frac{1}{P(X)}=-\log_a P(X)$$

a=e nat a=2 bit

# Entropy (熵)

- Consider a discrete source with N possible symbols
- Entropy H(.) is defined as the average amount of information conveyed per symbol

$$H(X) \overset{\Delta}{=} E\left[I(x_j)\right] = \sum_{j=1}^{N} P(x_j)\log_2 \frac{1}{P(x_j)} \ (\text{bit/symbol})$$

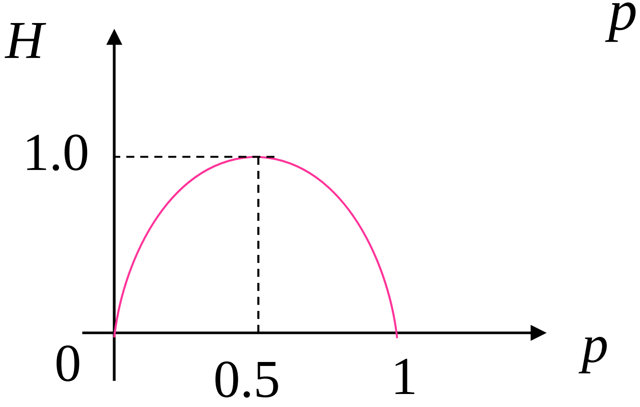
- Example: Consider a source having 3 symbols alphabet where $P(x_1)$ = ½, $P(x_2)$ =$P(x_3)$= ¼, and symbols are statically independent. Determine the entropy of the source.
- Solution: $H = p_1 \log_2 \frac{1}{p_1} + p_2 \log_2 \frac{1}{p_2} + p_3 \log_2 \frac{1}{p_3}$

$$= \frac{1}{2}\times 1 + \frac{1}{4}\times 2 + \frac{1}{4}\times 2 = 1.5\text{bit/Symbol}$$

# Entropy (Cont'd)

- How to maximize entropy?
- Consider binary case with two symbol alphabet {0, 1}, if we let P(1) = $p$, and P(0) = 1-$p$, then

$$H = p\log_2 \frac{1}{p} + (1-p)\log_2 \frac{1}{1-p}$$




Entropy is maximized when all the symbols are equiprobable

- N symbols: $$H = \sum_{n=1}^{N} \frac{1}{N}\log_2 N = \log_2 N \text{ bit/symbol}$$

# Exercise

- A source with bandwidth 4000Hz is sampled at the Nyquist rate. Assuming that the resulting sequence can be approximately modeled by a discrete memoryless source with alphabet {-2, -1, 0, 1, 2} and with corresponding probabilities {1/2, ¼, 1/8, 1/16, 1/16}, determine the rate of the source in bit/sec

# Solution

- We have

$$H(X) = \frac{1}{2}\log_2 2 + \frac{1}{4}\log_2 4 + \frac{1}{8}\log_2 8 + 2\times\frac{1}{16}\log_2 16$$
$$= \frac{15}{8} \text{ bits/sample}$$

- Since we have 8000 samples/sec the source produces information at a rate of 15kbits/sec.

# Joint and Conditional Entropy

- When dealing with two or more random sources, exactly in the same way that joint and conditional probabilities are introduced, one can introduce joint and conditional entropies.
- The **joint entropy** of (X, Y) is defined as

$$H(X,Y) = -\sum_{x,y} p(x,y) \log p(x,y)$$

- The **conditional entropy** of X given Y is defined as

$$H(X|Y) = -\sum_{x,y} p(x,y) \log p(x|y)$$

- Using chain rule, it can be shown that

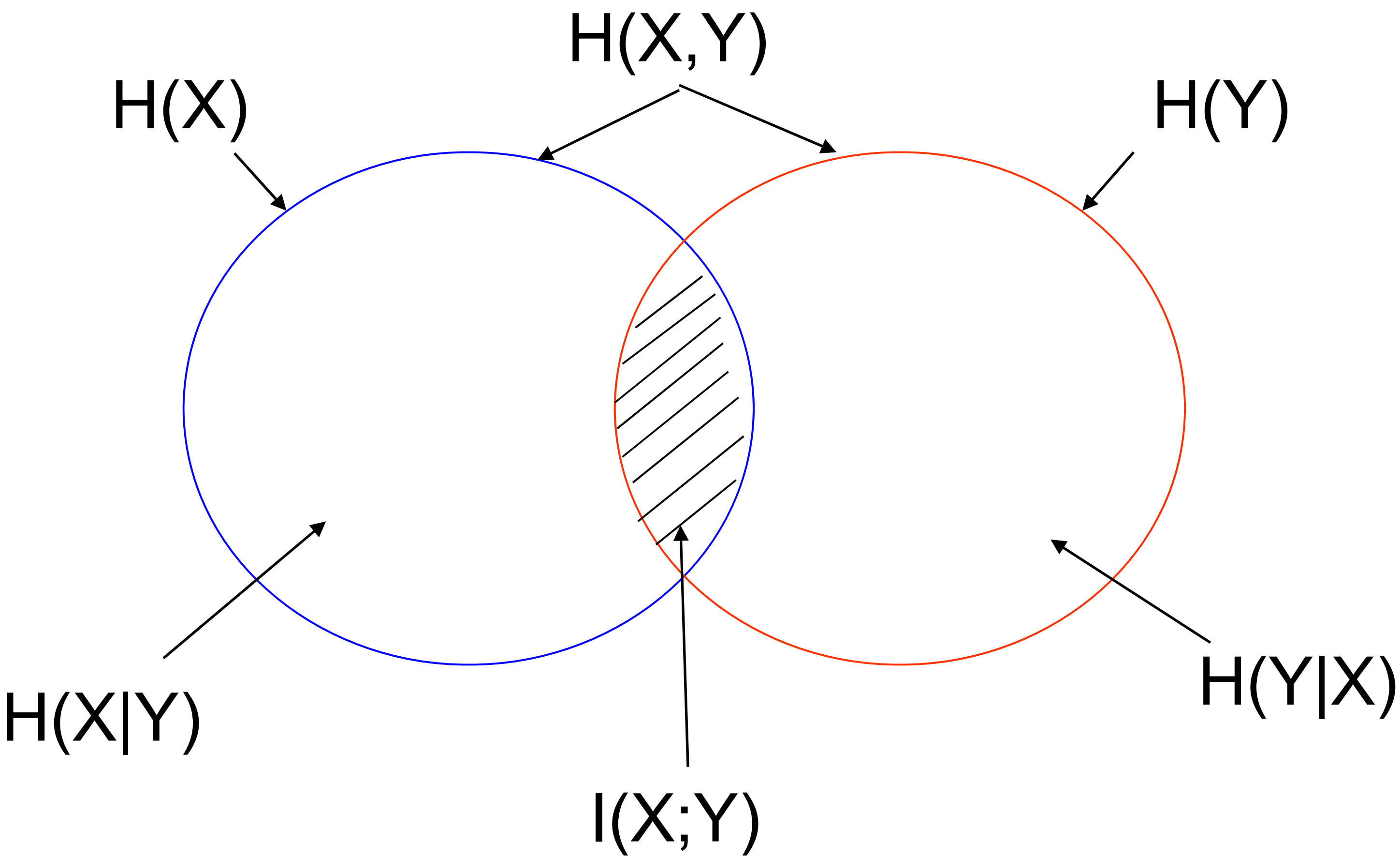
$$H(X,Y) = H(X|Y) + H(Y)$$

# Mutual Information

- Given by
  - H(X) denotes the uncertainty of the random varaible X
  - H(X|Y) denotes the uncertainty of random variable X after random variable Y is known.
- Then, H(X)-H(X|Y)
  - Denotes the amount of uncertainty of X that has been removed given Y is known
  - In other words, it is the amount of information provided by random variable Y about random variable X
- Definition of **mutual information**

$$I(X;Y) = H(X) - H(X \mid Y)$$

# Entropy, Conditional Entropy and Mutual Information



# Differential Entropy

- The **differential entropy** of a discrete-time continuous alphabet source X with pdf f(x) is defined as:

$$h(X) = -\int_{-\infty}^{\infty} f_X(x) \log f_X(x) dx$$

- Example: the differential entropy of $X \sim N(0, \sigma^2)$ is

$$h(X) = \frac{1}{2}\log_2\left(2\pi e\sigma^2\right) \text{ bits}$$

- Mutual information between two continuous random variables X and Y:

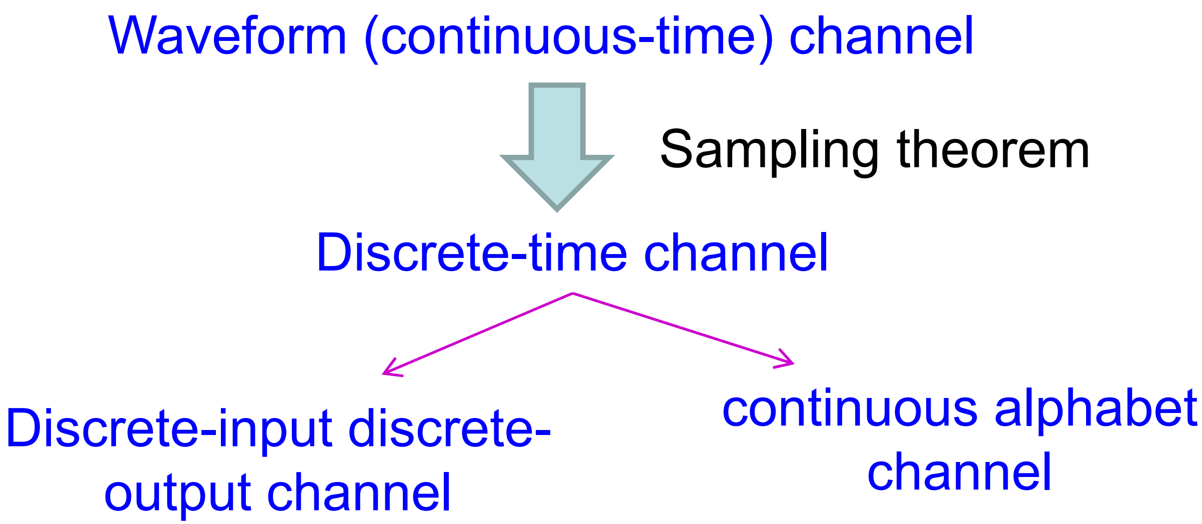
$$I(X;Y) = h(X) - h(X \mid Y)$$

# 9.2 Source Coding Theorem

- **Source coding theorem**:
  - A source with entropy (or entropy rate) H can be encoded with an arbitrarily small error probability at any rate R (bits/source output) as long as R > H.
  - Conversely, if R < H, the error probability will be bounded away from zero, independent of the complexity of the encoder and decoder employed

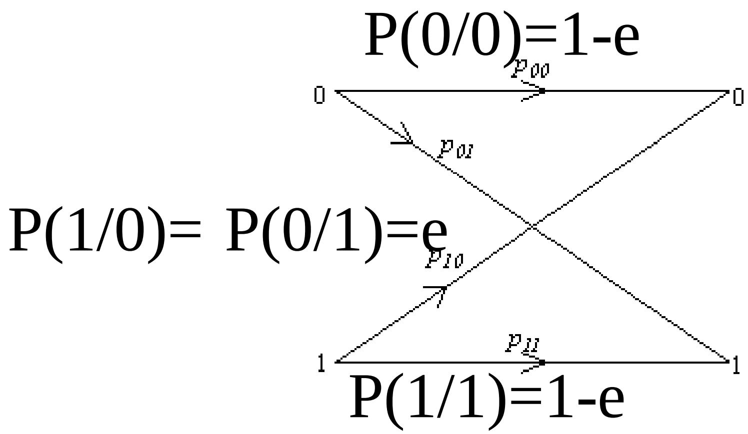
# 9.3 Modeling of Communication Channel

- Recall that a communication channel is any medium over which information can be transmitted
- It is characterized by a relationship between its input and output, which is generally a stochastic relation due to the presence of fading and noise



# Binary-Symmetric Channel

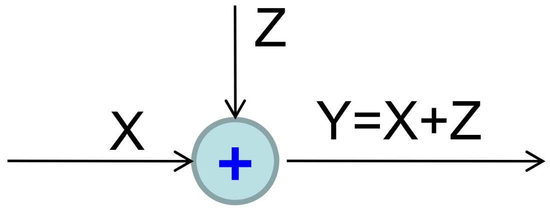
- BSC channel is characterized by the crossover probability e=P(0|1)=P(1|0)
- For instance, $e = Q\left(\sqrt{\frac{2E_b}{N_0}}\right)$



# AWGN Channel

- Both input and output are real numbers
- The input satisfy some power constraint

$$\sum_{i=1}^{n} x_i^2 \leq P$$



# 9.4 Channel Capacity

- In 1948, Shannon proved that
  - there exists a maximum rate, called **channel capacity** and denoted as C in bits/sec, at which one can communicate over a channel with arbitrarily small error probability
  - one can theoretically transmit over a channel at a rate R ≤ C with almost error free
  - Otherwise, if R > C, then reliable transmission is not possible
  - The capacity of a discrete-memoryless channel is given by

$$C = \max_{p(x)} I(X;Y) \quad \text{(max over all possible input distribution)}$$

⇨ **The Noisy Channel Coding Theorem**

(one of the fundamental results in information theory)

Claude E. Shannon
(1916-2001)

# Binary Symmetric Channel Capacity

- Since
$$\begin{aligned} I(X;Y) &= H(Y) - H(Y \mid X) \\ &= H(Y) - \sum p(x) H(Y \mid X = x) \\ &= H(Y) - \sum p(x) H(P_e) \\ &= H(Y) - H(P_e) \\ &\le 1 - H(P_e) \end{aligned}$$

- Here, $H(p) = -p\log_2 p - (1-p)\log_2(1-p)$

  $H(Y) \le 1$ Equality holds when X is equal probably

- Thus, the capacity of a BSC is

$$C = 1 - H(P_e)$$

# Gaussian Channel Capacity

- Consider a discrete-time Gaussian channel with

$$Y = X + Z$$

  - Input power constraint: $\sum_{i=1}^{n} x_i^2 \le P$
  - $Z \sim N(0, P_N)$

- Its capacity is given by (proof?)

$$C = \frac{1}{2}\log\left(1 + \frac{P}{P_N}\right)$$

- Now, consider a continuous-time, bandlimited AWGN channel with noise PSD=N0/2, input power constraint P, bandwidth W.
- Sample it at Nyquist rate and obtain a discrete-time channel. The power/sample will be P and the noise power/sample will be

$$P_N = \int_{-W}^{W} \frac{N_0}{2} df = WN_0$$

- Thus,

$$C = \frac{1}{2}\log\left(1 + \frac{P}{N_0 W}\right) \text{ bits/transmission}$$

- Since the number of transmissions/sec is 2W, we obtain the channel capacity in bits/sec

$$C = W\log\left(1 + \frac{P}{N_0 W}\right) \text{ bits/sec}$$

**(Shannon Formula)**

# Example

- Find the capacity of a telephone channel with bandwidth W=3000Hz, and SNR of 39dB

- Solution:
  - The SNR of 39 dB is equivalent to 7943. Using Shannon formula, we have

$$C = 3000 \log(1 + 7943) \approx \sim 38{,}867 \text{ bits/sec}$$

# Insights from Shannon Formula

1. Increasing signal power P increases the capacity C
   - When SNR is high enough, every doubling of P adds additional B bits/s in capacity
   - When P approaches infinity, so is C
2. Increasing channel bandwidth W can increase C, but cannot increase infinitely (as noise power also increases)

$$\lim_{W\to\infty} C = \lim_{W\to\infty}\left[\frac{WN_0}{P}\log\left(1+\frac{P}{N_0W}\right)\right]\frac{P}{N_0}$$

$$=\frac{P}{N_0}\log e = 1.44\frac{P}{N_0}$$

3. Bandwidth efficiency – energy efficiency tradeoff

- In any practical system, we must have

$$R \le W \log_2\left(1+\frac{P}{N_0 W}\right)$$

- Defining r=R/W, the spectral bit rate

$$r = \frac{R}{W} \le \log_2(1+\frac{P}{N_0 W})$$

- Let Eb be the energy per bit, $E_b = \frac{P}{R}$

- Then, $$r \le \log_2\left(1+r\frac{E_b}{N_0}\right)$$

$E_b/N_0$ = SNR per bit
r = spectral efficiency

- As r=R/B -> 0

$$\left.\frac{E_b}{N_0}\right|_{r\to 0} = \lim_{r\to 0}\frac{1}{r}(2^r - 1)$$
$$= \ln 2$$
$$= 0.693$$
$$= -1.59\text{dB}$$

**Shannon Limit**, an absolute minimum for reliable communication