《信道编码》
《Channel Coding》

Li Chen (陈立)
Professor, School of Electronics and Information Technology (SEIT)
Sun Yat-sen University
Office: 631C, SEIT Building
Email: chenli55@mail.sysu.edu.cn
Website: www.chencode.cn

《Channel Coding》

Textbooks:

1.《Elements of Information Theory》, by T. Cover and J. Thomas, Wiley (and introduced by Tsinghua University Press), 2003.

2.《Error control Coding》, by S. Lin and D. Costello, Prentice Hall, 2004.

3.《Non-binary error control coding for wireless communication and data storage》, by R. Carrasco and M. Johnston, Wiley, 2008.

4. 《信息论与编码理论》, 王育民、李晖著，高等教育出版社，2013.

Outlines

Evolution of Communications

Analogue comm.

Late 80s to early 90s

Digital comm.

Information theory and coding techniques

1G

2G

2.5G

EE+CS

3G

4G

Chapter 1 Fundamentals of Information Theory

- 1.1 An Introduction of Information
- 1.2 Measure of Information
- 1.3 Average Information (Entropy)
- 1.4 Channel Capacity

§ 1.1 An Introduction of Information

- What is information?
- How do we measure information?

Let us look at the following sentences:

1) I will be one year older next year.

   No information

   Boring!

2) I was born in 1993.

   Some information

   Being frank!

3) I was born in 1990s.

   More information

   Interesting, so which year?

The number of *possibilities* should be linked to the information!

§ 1.1 An Introduction of Information

Let us do the following game:

Throw a die once

You have 6 possible outcomes.
{1, 2, 3, 4, 5, 6}

Throw three dies

You have 6^3 possible outcomes.
{(1, 1, 1), (1, 1, 2), (1, 1, 3), (1, 1, 4)
……
(2, 1, 1), (2, 1, 2), (2, 1, 3), (2, 1, 4)
……
(6, 6, 3), (6, 6, 4), (6, 6, 5), (6, 6, 6)}

Information should be *‘additive’*.

§ 1.1 An Introduction of Information

Let us look at the following problem.

If there are 30 students in our class, and we would like to use binary bits to distinguish each of them, how many bits do we need?

Solution: 30 possibilities.

requires

$\log_2 30 = 4.907$ bits.

we need at least 5 bits to represent each of us.

Q: There are 7 billion people on our planet, how many bits do we need?

We can use *'logarithm'* to scale down the a huge amount of possibilities.

Number (binary bit) permutations are used to represent all possibilities.

§ 1.1 An Introduction of Information

Finally, let us look into the following game.

Pick one ball from the hat randomly,

The probability of picking up a white ball, $\frac{1}{4}$ (25%).

Representing the probability needs

$$\log_2 \frac{1}{1/4} = 2 \text{ bits.}$$

The probability of picking up a black ball, $\frac{3}{4}$ (75%).

Representing the probability needs

$$\log_2 \frac{1}{3/4} = 0.415 \text{ bits.}$$

§ 1.1 An Introduction of Information

- How do we measure the overall event? (On average, how many bits do we need to represent an outcome?)

$$\frac{1}{4} \cdot \log_2 \frac{1}{1/4} + \frac{3}{4} \log_2 \frac{1}{3/4} = 0.811 \text{ bits.}$$

- The measure of information should be

$$\sum_{i=1}^{N} P_i \log_2 P_i^{-1} = -\sum_{i=1}^{N} P_i \log_2 P_i$$

  - P_i: probability of the ith possible event.
  - N: Total number of possible events.

Measure of information should consider the *probabilities of various possible events.*

§ 1.2 Measure of Information

- Information: knowledge not precisely known by the recipient, as it is a measure of unexpectedness.
- Amount of information $\propto$ (probability of occurance)$^{-1}$
- Messages: $M_1 \quad M_2 \quad M_3 \ldots\ldots \quad M_q$

  Prob of occur: $P_1 \quad P_2 \quad P_3 \ldots\ldots \quad P_q \qquad (P_1 + P_2 + P_3 + \cdots + P_q = 1)$

Measure the amount of information carried by each message by

$$I(M_i) = log_x P_i^{-1}, \quad i = 1, 2, \ldots, q$$

$x = 2$, $I(M_i)$ in bits

$x = e$, $I(M_i)$ in nats

$x = 10$, $I(M_i)$ in Hartley.

- Observations: … …

§ 1.2 Measure of Information

1) $I(M_i) \to 0, \qquad if \qquad P_i \to 1;$
2) $I(M_i) \geq 0, \qquad when \quad 0 \leq P_i \leq 1;$
3) $I(M_i) > I(M_j), \quad if \qquad P_j > P_i$
4) Given M_i and M_j are statistically independent,
$$I(M_i \& M_j) = I(M_i) + I(M_j).$$

§ 1.2 Measure of Information

Example 1.1: A source outputs five possible messages. The probabilities of these messages are:

$$P_1 = \frac{1}{2} \quad P_2 = \frac{1}{4} \quad P_3 = \frac{1}{8} \quad P_4 = \frac{1}{16} \quad P_5 = \frac{1}{16}.$$

Determine the information contained in each of these messages.

Solution:

$$I(M_1) = \log_2 \frac{1}{1/2} = 1 \text{ bit}$$

$$I(M_2) = \log_2 \frac{1}{1/4} = 2 \text{ bit}$$

$$I(M_3) = \log_2 \frac{1}{1/8} = 3 \text{ bit}$$

$$I(M_4) = \log_2 \frac{1}{1/16} = 4 \text{ bit}$$

$$I(M_5) = \log_2 \frac{1}{1/16} = 4 \text{ bit}$$

Total amount of information = 14 bits. Is it right?

§ 1.3 Average Information (Entropy (熵))

Given a source vector of length N, and it has U possible symbols $S_1, S_2, \ldots S_U$, each of which has probability of $P_1, P_2, \ldots P_U$ of occurrence.

To represent the source vector, we need

$$I = \sum_{i=1}^{U} N\, P_i log_2 P_i^{-1} \text{ bits.}$$

So on average, how many information bits do we need for a source symbol?

$$H = \frac{I}{N} = \sum_{i=1}^{U} P_i log_2 P_i^{-1} \text{ bits/symbol}$$

H is called the source entropy – average number of information per source symbol.

§ 1.3 Average Information (Entropy)

Example 1.2: A source vector contains symbols of four possible outcomes $A,\ B,\ C,\ D$. They occur with probabilities of $\frac{1}{4},\ \frac{1}{3},\ \frac{1}{3}$ and $\frac{1}{12}$, respectively. Determine the entropy of the source vector.

$$H = \frac{1}{4}\log_2\frac{1}{1/4} + \frac{2}{3}\log_2\frac{1}{1/3} + \frac{1}{12}\log_2\frac{1}{1/12}$$

$$= 1.856 \text{ bits/symbol}$$

§ 1.3 Average Information (Entropy)

Entropy of a binary source: The source vector has only two possible symbols, i.e., 0 and 1. Let $P(0)$ denote the probability of a source symbol being 0, and $P(1)$ denote the probability of a source symbol being 1, we have

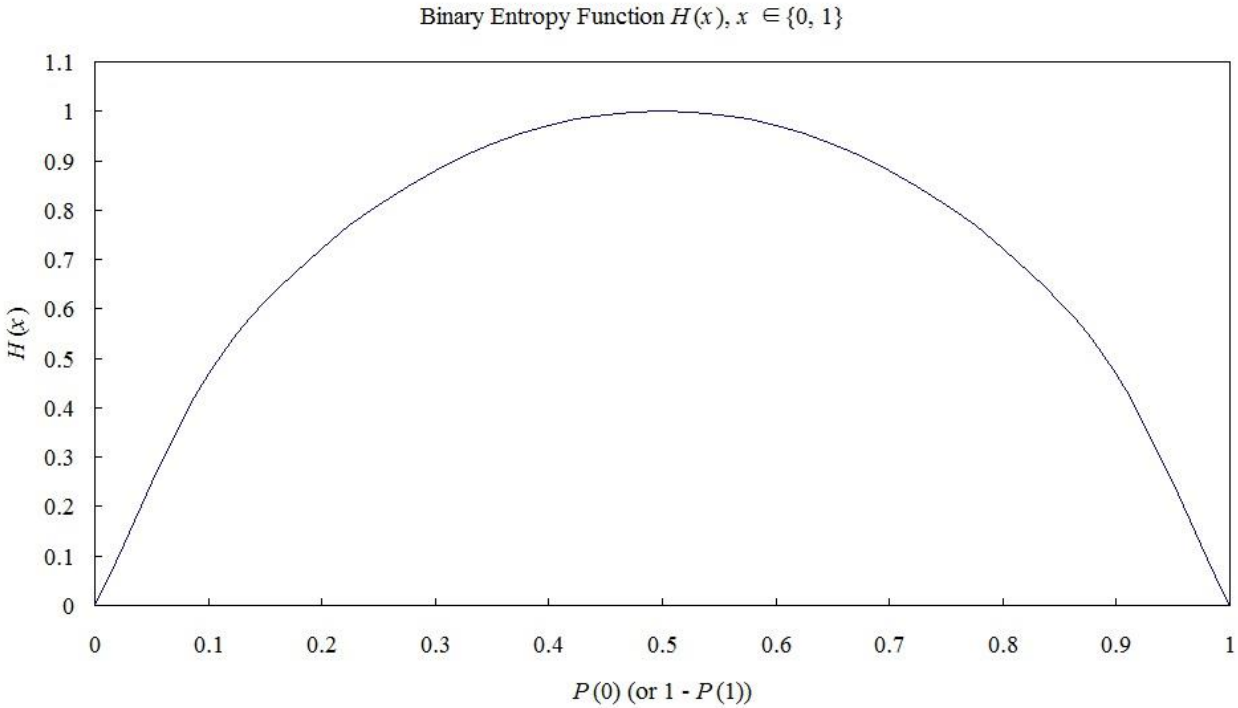
$$H = P(0) \cdot \log_2 P(0)^{-1} + P(1) \log_2 (1)^{-1}$$

or

$$H = P(0) \cdot \log_2 P(0)^{-1} + (1 - P(0)) \cdot \log_2 (1 - P(0))^{-1}$$

Binary Entropy Function

§ 1.3 Average Information (Entropy)



§ 1.3 Average Information (Entropy)

<u>Mutual Information of a channel</u>

$$\boxed{\text{Source}} \xrightarrow{X} \boxed{\text{Channel}} \xrightarrow{Y} \boxed{\text{Sink}}$$

Y is a noisy (corrupted) version of X

Given the observation of Y, how much uncertainty about X is left at the sink?

- Let $P(x_i|y_k)$ denote the probability of $X = x_i$ given $Y = y_k$ is observed, $i = 1, 2, \ldots$, and $k = 1, 2, \ldots$;
- The conditional entropy of X is

$$H(X|Y = y_k) = \sum_i P(x_i|y_k) \log_2[\frac{1}{P(x_i|y_k)}]$$

- Since the observations $y_1, y_2, \ldots$, happen with probabilities of $P(y_1), P(y_2), \ldots$, the average conditional entropy should be

$$\begin{aligned} H(X|Y) &= \sum_k H(x|y = y_k) \cdot P(y_k) \\ &= \sum_k \sum_i P(x_i|y_k) \cdot P(y_k) \cdot \log_2[\frac{1}{P(x_i|y_k)}] \\ &= \sum_k \sum_i P(x_i, y_k) \cdot \log_2[\frac{1}{P(x_i|y_k)}]. \end{aligned}$$

§ 1.3 Average Information (Entropy)

Look at the difference between

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

How much we don't know BEFORE the channel observations.

How much we still don't know AFTER the channel observations.

How much information is carried by the channel, and this is called the **Mutual Information** of the channel, denotes $I(X, Y)$.

§ 1.3 Average Information (Entropy)

Properties of mutual information

Property 1: $I(X,\ Y) \geq 0$

Property 2: $I(X,\ Y) = I(Y,\ X)$

Property 3: $I(X,Y) = H(X) + H(Y) - H(X,\ Y)$

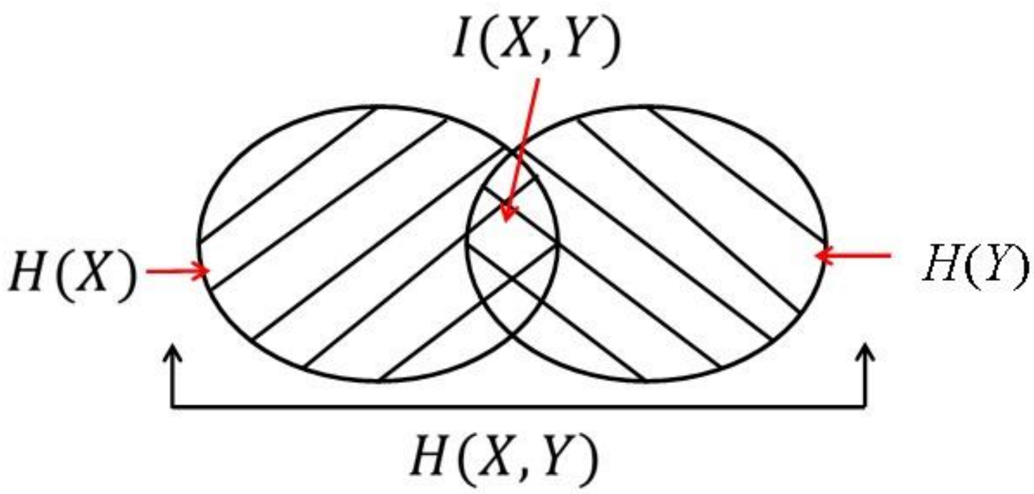
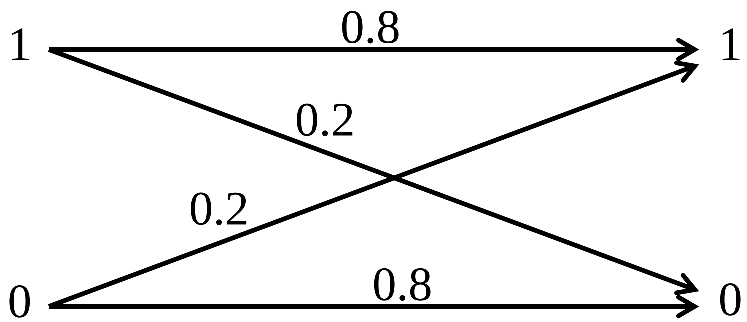


Fig. A Venn diagram

Remark: mutual information $I(X,\ Y)$ describes the amount of information one variable X contains about the other Y, or vice versa as in $I(Y,\ X)$.

§ 1.3 Average Information (Entropy)

Example 1.3: Given the binary symmetric channel shown as



We know $P(x=0)=0.3$, $P(x=1)=0.7$, $P(y=1|x=1)=0.8$,
$P(y=1|x=0)=0.2$, $P(y=0|x=1)=0.2$ and $P(y=0|x=0)=0.8$.
Please determine the mutual information of such a channel.

Solution:

- Entropy of the binary source is

$$\begin{aligned} H(x) &= -P(x=0)\log_2 P(x=0) - P(x=1)\log_2 P(x=1) \\ &= 0.3\cdot\log_2\frac{1}{0.3} + 0.7\cdot\log_2\frac{1}{0.7} \\ &= 0.881 \text{ bits} \end{aligned}$$

§ 1.3 Average Information (Entropy)

- With $P(x)$ and $P(y|x)$, we know

$$P(y=1)=P(y=1|x=1)P(x=1)+P(y=1|x=0)P(x=0)$$
$$=0.62$$
$$P(y=0)=P(y=0|x=1)P(x=1)+P(y=0|x=0)P(X=0)$$
$$=0.38$$

$$P(x=0,y=0)=P(y=0|x=0)\cdot P(x=0)=0.24$$
$$P(x=0|y=0)=\frac{P(x=0,y=0)}{P(y=0)}=0.63$$
$$P(x=1,y=0)=P(y=0|x=1)\cdot P(x=1)=0.14$$
$$P(x=1|y=0)=\frac{P(x=1,y=0)}{P(y=0)}=0.37$$
$$P(x=0,y=1)=P(y=1|x=0)P(x=0)=0.06$$
$$P(x=0|y=1)=\frac{P(x=0,y=1)}{P(y=1)}=0.10$$
$$P(x=1,y=1)=P(y=1|x=1)P(x=1)=0.56$$
$$P(x=1|y=1)=\frac{P(x=1,y=1)}{P(y=1)}=0.90$$

§ 1.3 Average Information (Entropy)

- Hence, the conditional entropy is:

$$H(X\,|\,Y)=P(x=0,y=0)\log_2\frac{1}{P(x=0\,|\,y=0)}+P(x=1,y=0)\log_2\frac{1}{P(x=1\,|\,y=0)}$$

$$+P(x=0,y=1)\log_2\frac{1}{P(x=0\,|\,y=1)}+P(x=1,y=1)\log_2\frac{1}{P(x=1\,|\,y=1)}$$

$$=0.24\log_2\frac{1}{0.63}+0.14\log_2\frac{1}{0.37}+0.06\log_2\frac{1}{0.10}+0.56\log_2\frac{1}{0.90}$$

$$=0.644\text{bits/sym}$$

- The mutual information is:

$$I(X,Y)=H(X)-H(X\,|\,Y)=0.237\text{bits/sym}$$

§ 1.4 Channel Capacity

$$\boxed{\text{Source}} \xrightarrow{X} \boxed{\text{Channel}} \xrightarrow{Y} \boxed{\text{Sink}}$$

- Channel is a medium that conveys the information X of source to the sink, and X is now being observed as Y.
- Channel imposes signal impairments into X, e.g.,
  for additive white Gaussian noise (AWGN) channel

$$y_i = x_i + \underbrace{n_i}_{\text{noise}}$$

  for fading channel,

$$y_i = \underbrace{\alpha}_{\text{fading coeff.}} \cdot x_i + \underbrace{n_i}_{\text{noise}}$$

- Recall: Mutual Information of a channel $I(X,Y)$ defines how much information can be carried by the channel as

$$I(X,Y) = H(X) - H(X|Y).$$

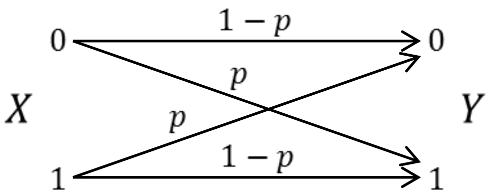
- Channel Capacity: the maximum mutual information taken over all distribution of X as

$$C = \max_{P(x_i)} I(X,Y) \quad \text{bits/symbol.}$$

$\boxed{C \text{ defines the best capacity that a channel can convey the unknown information.}}$

§ 1.4 Channel Capacity

<u>Channel Capacity for Binary Symmetric Channel (BSC)</u>



- Assume that $P(x=0)=P(x=1)=\frac{1}{2}$, hence $H(X)=1$.
- It is known that $P(y=0|x=0)=P(y=1|x=1)=1-p$

  $P(y=0|x=1)=P(y=1|x=0)=p$.
- The conditional entropy $H(X|Y)$ is

$$
\begin{aligned}
H(X|Y) &= \sum_k \sum_i P(x_i, y_k) \cdot \log_2 \frac{1}{P(x_i|y_k)} \\
&= P(y=0|x=0) \cdot P(x=0) \cdot \log_2 \frac{1}{P(x=0|y=0)} \\
&\quad + P(y=0|x=1) \cdot P(x=1) \cdot \log_2 \frac{1}{P(x=1|y=0)} \\
&\quad + P(y=1|x=0) \cdot P(x=0) \cdot \log_2 \frac{1}{P(x=0|y=1)} \\
&\quad + P(y=1|x=1) \cdot P(x=1) \cdot \log_2 \frac{1}{P(x=1|y=1)}
\end{aligned}
$$

§ 1.4 Channel Capacity

- $$P(y=0)=\frac{1}{2}(1-p)+\frac{1}{2}p=\frac{1}{2}=P(y=1)$$

  $$P(x=0\mid y=0)=\frac{P(x=0,y=0)}{P(y=0)}=1-p$$

  $$P(x=0\mid y=1)=\frac{P(x=0,y=1)}{P(y=1)}=p$$

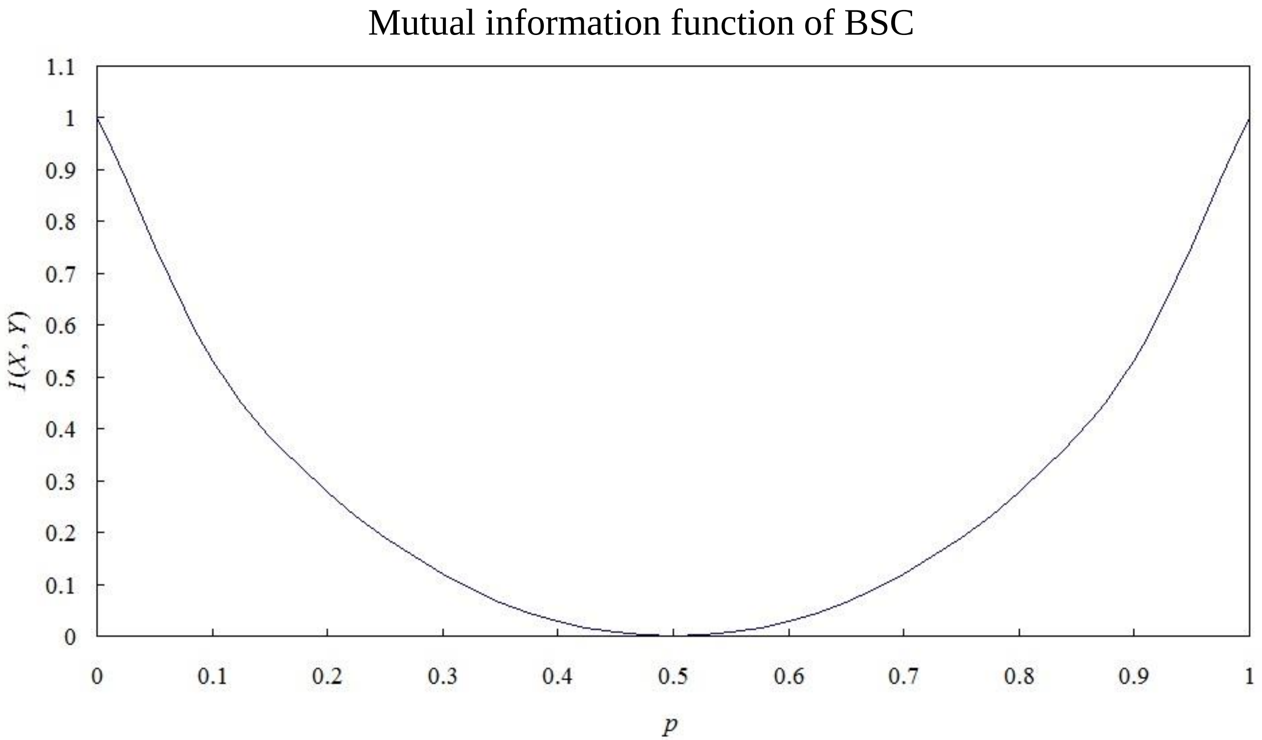
- $$P(x=1|y=0)=\frac{P(x=1,y=0)}{P(y=0)}=p$$

  $$P(x=1|y=1)=\frac{P(x=1,y=1)}{P(y=1)}=1-p$$

- $$H(X|Y)=-[\tfrac{1}{2}(1-p)\log_2(1-p)+\tfrac{1}{2}p\log_2 p+\tfrac{1}{2}p\log_2 p+\tfrac{1}{2}(1-p)\log_2(1-p)]$$
  $$=-p\log_2 p-(1-p)\log_2(1-p)$$

- $$I(X,Y)=H(X)-H(X|Y)$$
  $$=1+p\log_2 p+(1-p)\log_2(1-p).$$

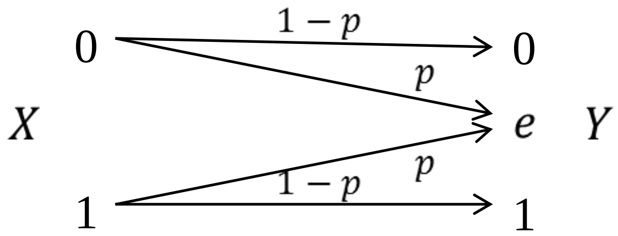
§ 1.4 Channel Capacity



$$I(x, y) = 1 + p\log_2 p + (1-p)\log_2(1-p)$$

§ 1.4 Channel Capacity

Channel Capacity for Binary Erasure Channel (BEC)



- Erasure e represents the data loss phenomenon.
- Assume that $P(x=0)=P(x=1)=\frac{1}{2}$. Hence, $H(X)=1$.
- $P(Y=0|X=0)=1-p=P(Y=1|X=1)$

  $P(Y=0)=P(Y=0|X=0)P(X=0)=\frac{1}{2}(1-p)=P(Y=1)$
- $P(Y=e)=P(Y=e|X=0)P(X=0)+P(Y=e|X=1)P(X=1)$

  $=\frac{1}{2}p+\frac{1}{2}p=p.$
- $P(X=0,Y=0)=P(X=0\,|\,Y=0)P(Y=0)=\dfrac{1}{2}(1-p)$ $\quad P(X=0,Y=e)=P(X=0\,|\,Y=e)P(Y=e)=\dfrac{1}{2}p$

  $P(X=1,Y=1)=P(X=1\,|\,Y=1)P(Y=1)=\dfrac{1}{2}(1-p)$ $\quad P(X=1,Y=e)=P(X=1\,|\,Y=e)P(Y=e)=\dfrac{1}{2}p$

§ 1.4 Channel Capacity

$$P(x=0|y=e)=\frac{P(x=0,y=e)}{P(y=e)}=\frac{1}{2} \qquad P(x=0|y=0)=\frac{P(x=0,y=0)}{P(y=0)}=1$$

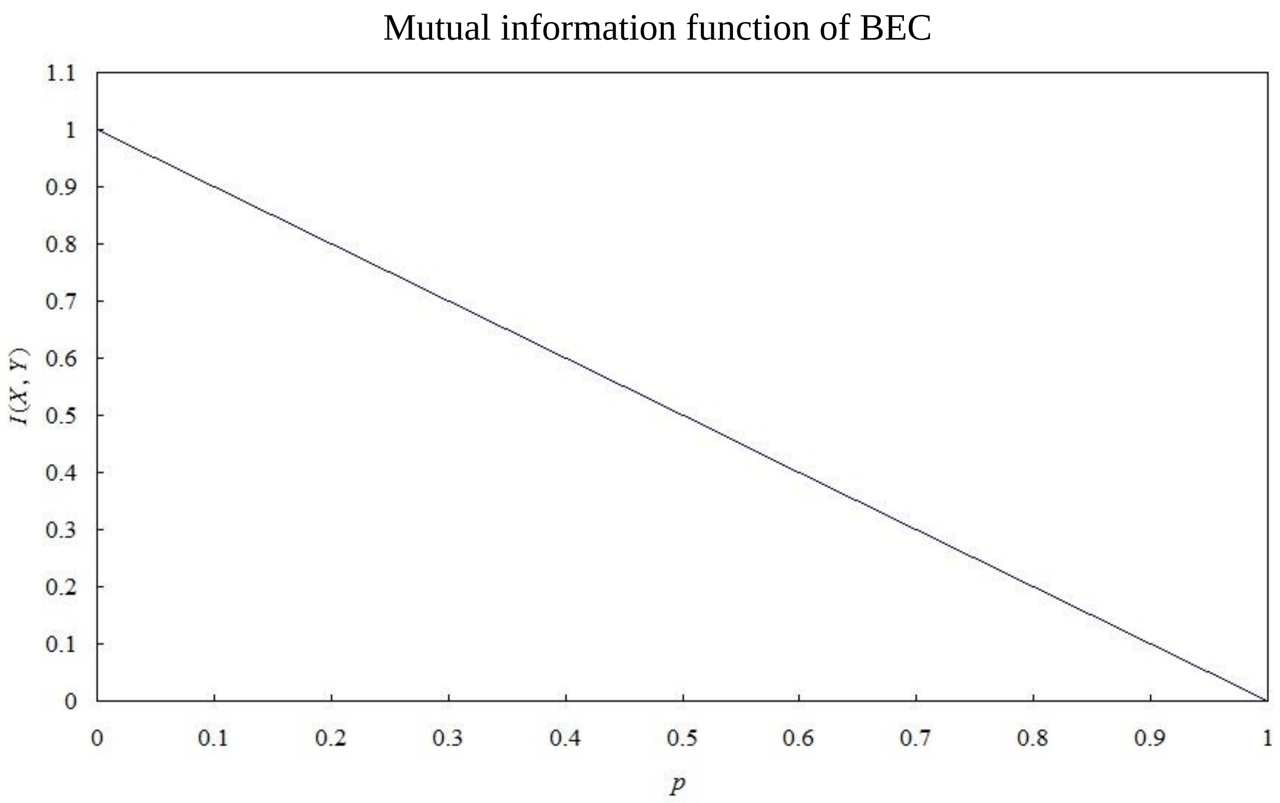
$$P(x=1|y=e)=\frac{P(x=1,y=e)}{P(y=e)}=\frac{1}{2} \qquad P(x=1|y=1)=\frac{P(x=1,y=1)}{P(y=1)}=1$$

- Therefor, the conditional entropy $H(X|Y)$ can be determined as

$$\begin{aligned} H(X|Y) &= -[\sum_k \sum_i P(x_i, y_k) \cdot \log_2 P(x_i|y_k)] \\ &= -[P(x=0, y=0) \cdot \log_2 P(x=0|y=0) \\ &\quad +P(x=1, y=1) \cdot \log_2 P(x=1|y=e) \\ &\quad +P(x=0, y=e) \cdot \log_2 P(x=0|y=e) \\ &\quad +P(x=1, y=e) \cdot \log_2 P(x=1|y=e)] \\ &= -[\tfrac{1}{2}p \cdot 0 + \tfrac{1}{2}p \cdot 0 + \tfrac{1}{2}p \cdot \log_2 \tfrac{1}{2} + \tfrac{1}{2}p \cdot \log_2 \tfrac{1}{2}] \\ &= p \end{aligned}$$

- $I(X,Y) = H(X) - H(X|Y) = 1 - p.$

§ 1.4 Channel Capacity



$$I(x, y) = 1 - p$$

§ 1.4 Channel Capacity

Channel Capacity for AWGN Channel

- $y_i = x_i + n_i$
- The channel is discrete memoryless, and its output is continuous.
- $I(X,Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$
- $C = \max[\sum_k \sum_i P(x_i, y_k) \log_2 \frac{1}{P(x_i, y_k)}]$,

  $i = 1,2,\cdots,m,$ and m is the order of a modulation, $k \to \infty$.
- Given input signal x and output signal y are Gaussian distributed, with

$$P(x) = \frac{1}{\sqrt{2\pi}\sigma_x} \cdot e^{-\frac{x^2}{2\sigma_x^2}} \text{ and } P(y) = \frac{1}{\sqrt{2\pi}\sigma_y} \cdot e^{-\frac{y^2}{2\sigma_y^2}}$$

- $\sigma_y^2 = \sigma_x^2 + \sigma_n^2 \leftarrow$ variance (power) of the noise.

  variance (power) of the transmitted signal

  variance (power) of the received signal
- We have

$$H(Y|X) = \log_2 \sqrt{2\pi e \sigma_n^2}$$
$$H(Y) = \log_2 \sqrt{2\pi e(\sigma_x^2 + \sigma_n^2)}$$

§ 1.4 Channel Capacity

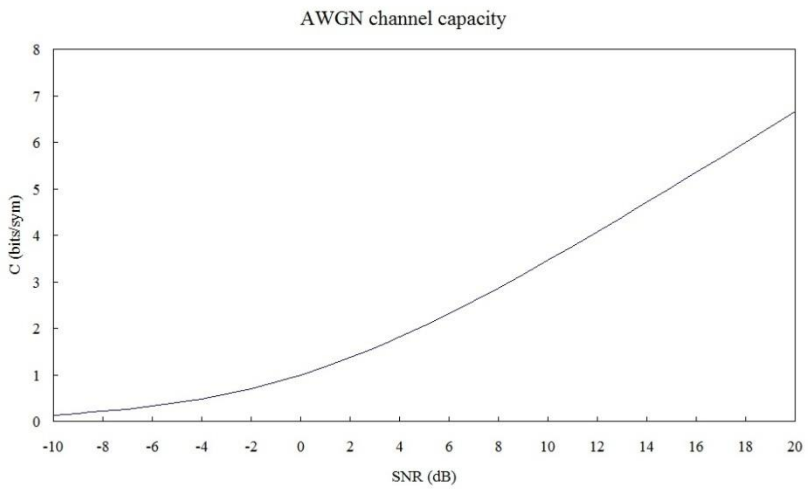
- $$I(X,Y) = H(Y) - H(Y|X)$$
$$= \log_2 \sqrt{\frac{2\pi e(\sigma_x^2+\sigma_n^2)}{2\pi e \sigma_n^2}}$$
$$= \log_2 \sqrt{1 + \frac{\sigma_x^2}{\sigma_n^2}}$$
$$= \log_2 \sqrt{1 + SNR} \quad \text{bits/sym}$$
- Here given the channel bandwidth as B (sym/sec), the AWGN channel can transmit

  $\frac{1}{2}B\log_2(1+SNR)$ bits/sec (For 1-dim. signal, y_i, x_i and n_i are real numbers)

  $B\log_2(1+SNR)$ bits/sec (For 2-dim. signal, y_i, x_i and n_i are complex numbers)
- **Example 1.6:** Determine the capacity of a low-pass channel with usual bandwidth of 3000 Hz and $\frac{S}{N} = 10$ dB (signal/noise) at the channel output. Assume the channel noise to be Gaussian and white.

  Solution:
$$C = B\log_2(1 + \frac{S}{N})$$
$$= 3000 \log_2(1 + 10) \cong 10378 \quad \text{bits/sec}$$

§ 1.4 Channel Capacity



– How to realize the channel capacity in a practical communication system?

– With the existence of channel impairments, how can we secure the recovery of the transmitted data?

– The use of channel codes: map an arbitrary k bits information into n bits codeword and $n > k$. The introduced redundancy ($n - k$ bits) can correct the error brought by the channel. We call $r = k/n$ as the **code rate**.

§ 1.4 Channel Capacity

- With a binary signalling modulation, e.g., BPSK, reliable communication can be achieved if the code rate does not exceed the channel capacity, i.e.,

$$r < C$$

- With $SNR > 2^r - 1$, reliable communication is possible. The better code you use, the closer your error performance will approach to the SNR threshold value.

- In a general communication system that employs a channel code of rate *r* and a modulation scheme of order *m*, we need

$$r \cdot m < C$$

§ 1.4 Channel Capacity

- E_s – average symbol energy, E_c – average coded bit energy, E_b – average information bit energy, their relationships are

$$E_s = m\,E_c, \qquad E_c = E_b\,r$$

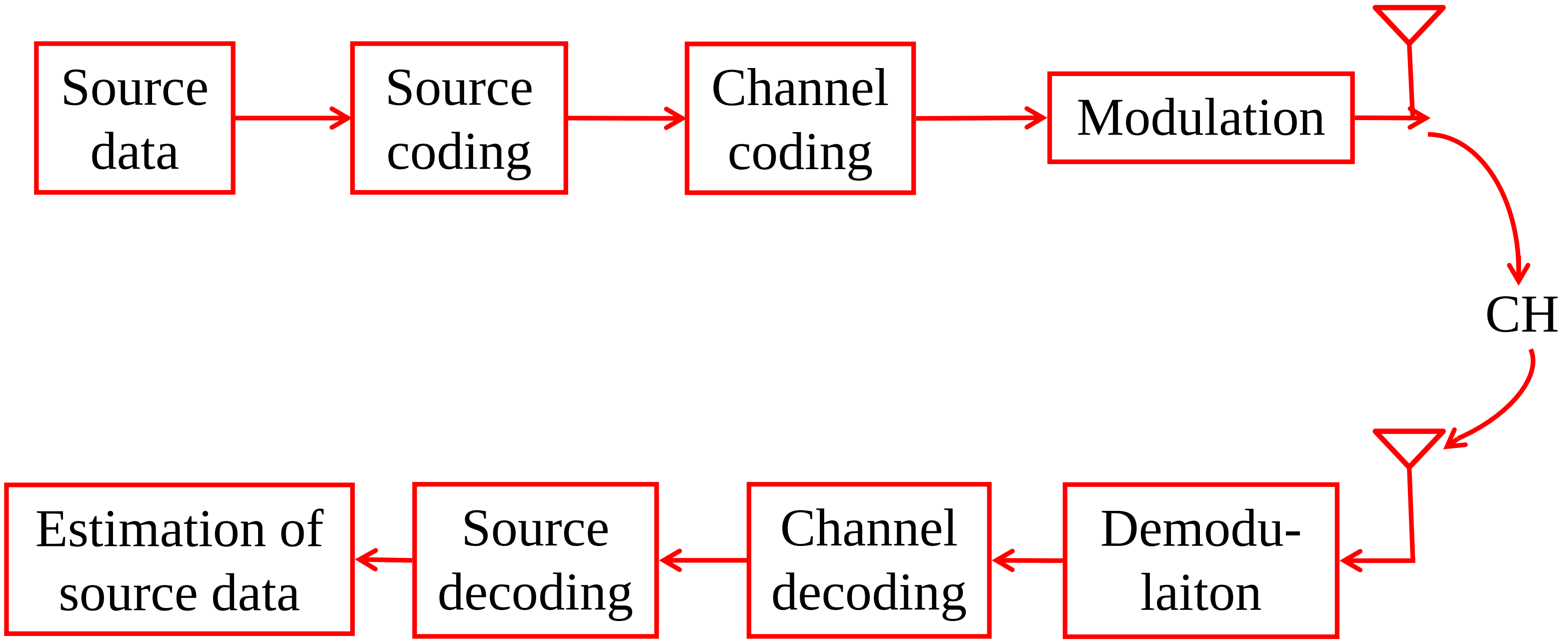
- To secure a reliable communication, we need

$$\log_2(1+\frac{E_s}{N_0}) > rm \quad \Longrightarrow \quad \log_2(1+\frac{E_b}{N_0}mr) > rm$$

- Hence, with $\frac{E_b}{N_0} > \frac{2^{mr}-1}{mr}$, reliable communication is possible.

§ 1.4 Channel Capacity

Now, we are ready to materialize a communication system:



§ 1.4 Channel Capacity

- Let us examine the performance of various coded communication systems using a rate half (r = 0.5) channel code and QPSK (m =2).