1924-2024
中山大學 世纪华诞
100th ANNIVERSARY
SUN YAT-SEN UNIVERSITY

1924-2024

2024年“智能通信”暑期学校

主讲: 陈立

时间: 2024.8.11

世 纪 中 大

山 高 水 长

Some Basics of Information Theory

&

Channel Coding Theorem

I. Entropy – Mutual Information – Channel Capacity

Source Coding Theorem

Channel Coding Theorem

II. Information Inequalities

Some Basics of Information Theory

Information Theory, founded by Claude E. Shannon (1916-2001)

via "A Mathematical Theory of Communication," *Bell System Technical Journal*, 1948.

- What is information?
- How to measure information?
- How to represent information?
- How to transmit information and its limit?

Q: **为什么叫“信息”？**

《暮春怀故人》

李中（唐）

池馆寂寥三月尽，落花重叠盖莓苔。
惜春眷恋不忍扫，感物心情无计开。
梦断美人沈信息，目穿长路倚楼台。
琅玕绣段安可得，流水浮云共不回。

Some Basics of Information Theory

I. Entropy

Random Variable (discrete): X
Realizations of X: x; E.g., $x_1, \cdots, x_u$
Distribution of X: $P(x)$; E.g., $P(x_1), \cdots, P(x_u)$

$$H(X) = \mathbb{E}[\log_b P(x)^{-1}]$$
$$= \sum_x P(x) \log_b P(x)^{-1} \cdots \text{/ sym.}$$

- If $b = 2$, $H(X)$ is in bits/sym. {0,1} are utilized to form permutations to represent all possibilities
- Bit = binary + digit

I. Entropy

Example:

X:	A	B	C	D	E	
$P(x)$:	1	0	0	0	0	$H(X) = 0$
$P(x)$:	$1/3$	$1/3$	$1/3$	0	0	$H(X) = \log_2 3$ bits / sym.
$P(x)$:	$1/5$	$1/5$	$1/5$	$1/5$	$1/5$	$H(X) = \log_2 5$ bits / sym.

Some Basics of Information Theory

I. Entropy

Random variables (realizations): $X(x),\ Y(y)$
Distributions: $P(x),\ P(y),\ P(x,y),\ P(y|x),\ P(x|y)$

$$
\begin{aligned}
H(X,Y) &= \mathbb{E}[\log_2 P(x,y)^{-1}] \\
H(Y|X) &= \mathbb{E}[\log_2 P(y|x)^{-1}] \\
H(X|Y) &= \mathbb{E}[\log_2 P(x|y)^{-1}]
\end{aligned}
$$

Some Basics of Information Theory

I. Entropy

The Chain Rule

$$
\begin{aligned}
H(X,Y) &= H(X) + H(Y|X) \\
&= H(Y) + H(X|Y)
\end{aligned}
$$

$$H(X_1, \cdots, X_v) = \sum_{i=1}^{v} H(X_i|X_1 \cdots X_{i-1})$$

Inequality states to appear

$$H(X|Y) \leqslant H(X) \leqslant H(X,Y)$$

Some Basics of Information Theory

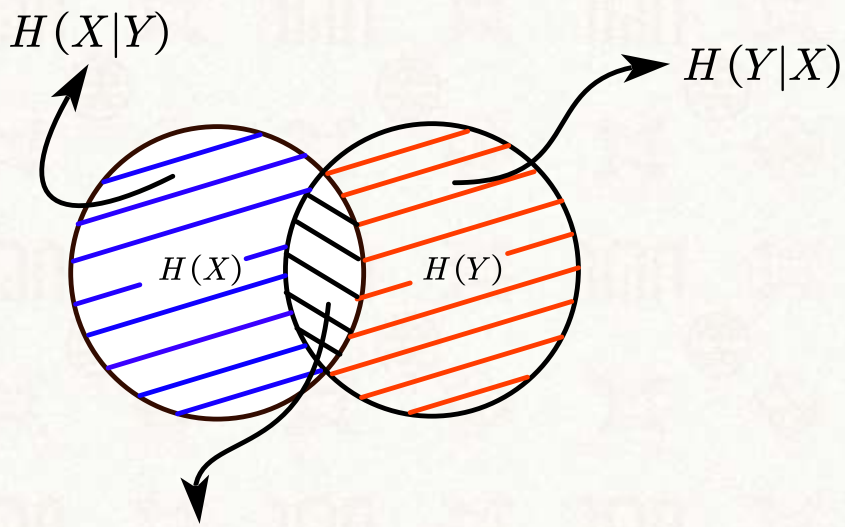
I. Entropy – Mutual Information

$$I(X;Y) = H(X) - H(X|Y)$$

$$= \mathbb{E}\left[\log_2 \frac{P(x|y)}{P(x)}\right]$$

Some Basics of Information Theory

I. Entropy – Mutual Information



$$I(X;Y) = H(Y) - H(Y|X)$$

$$I(X;Y) = H(X) + H(Y) - H(X,Y)$$

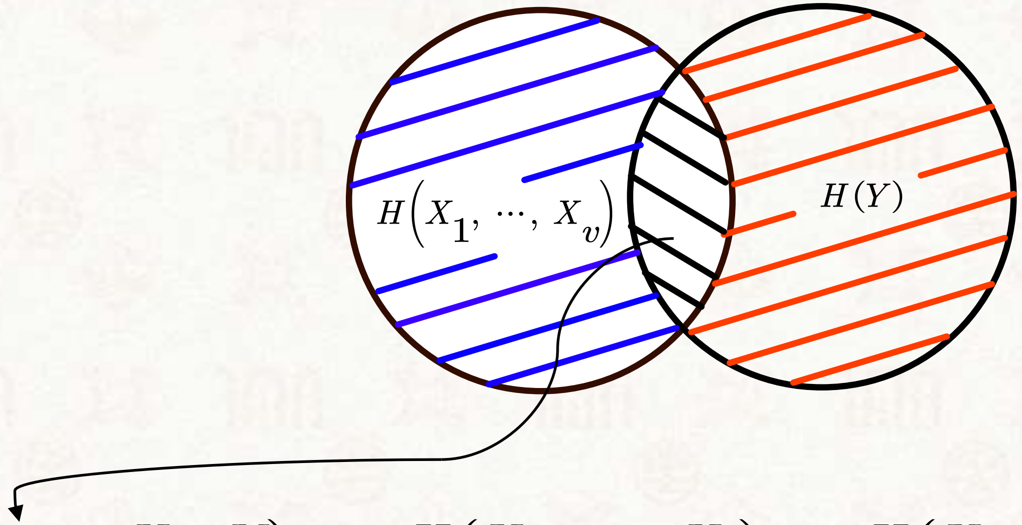
$$= \mathbb{E}\left[\log_2 \frac{P(x,y)}{P(x)P(y)}\right]$$

If X, Y are independent, $I(X;Y) = 0$

If $H(X) \subseteq H(Y)$ (or $H(Y) \subseteq H(X)$), $I(X;Y) = H(X)$ (or $= H(Y)$)

I. Entropy – Mutual Information

The Chain Rule



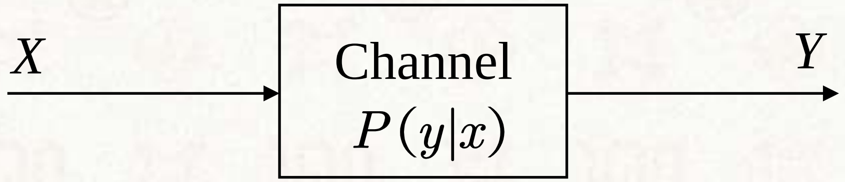
$$I(X_1, \cdots, X_v; Y) = H(X_1, \cdots, X_v) - H(X_1, \cdots, X_v|Y)$$

$$= \sum_{i=1}^{v} H(X_i|X_1, \cdots, X_{i-1}) - \sum_{i=1}^{v} H(X_i|X_1, \cdots, X_{i-1}, Y)$$

$$= \sum_{i=1}^{v} I(X_i; Y|X_1, \cdots, X_{i-1})$$

1924-2024 中山大學 世纪华诞 100th ANNIVERSARY SUN YAT-SEN UNIVERSITY

I. Mutual Information – Channel Capacity



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

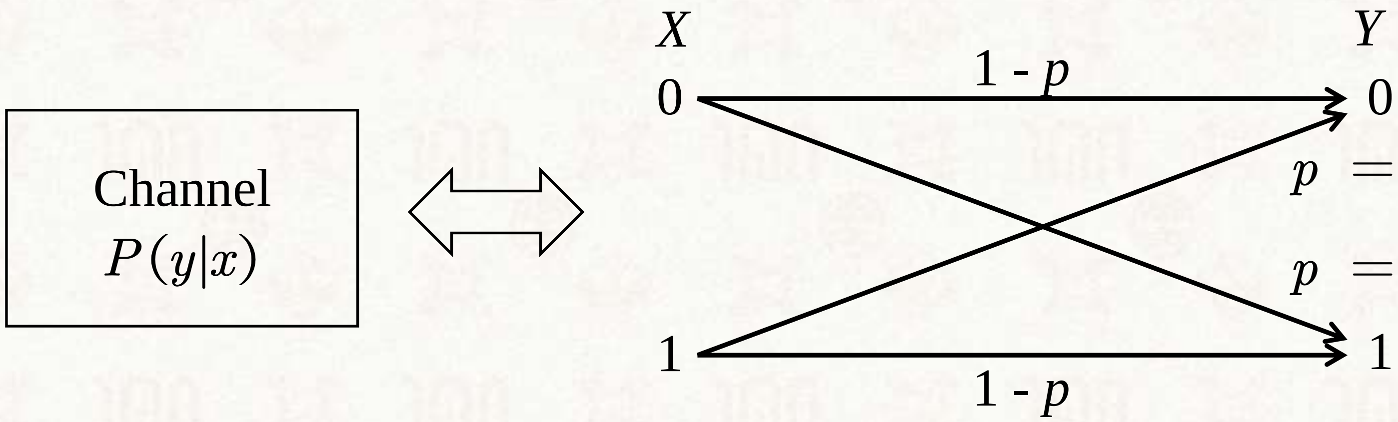
$$= \max_{P(x)} \left\{ \mathbb{E}\left[\log_2 \frac{P(y|x)}{\sum_x P(y|x)P(x)} \right] \right\}$$

The maximum information conveying capability of the channel (characterized by $P(y|x)$) can be reached by adjusting the input distribution $P(x)$.

Some Basics of Information Theory

I. Mutual Information – Channel Capacity

BSC:



$$p = P(y = 0 \mid x = 1)$$

$$p = P(y = 1 \mid x = 0)$$

When $P(x = 0) = P(x = 1) = \frac{1}{2}$

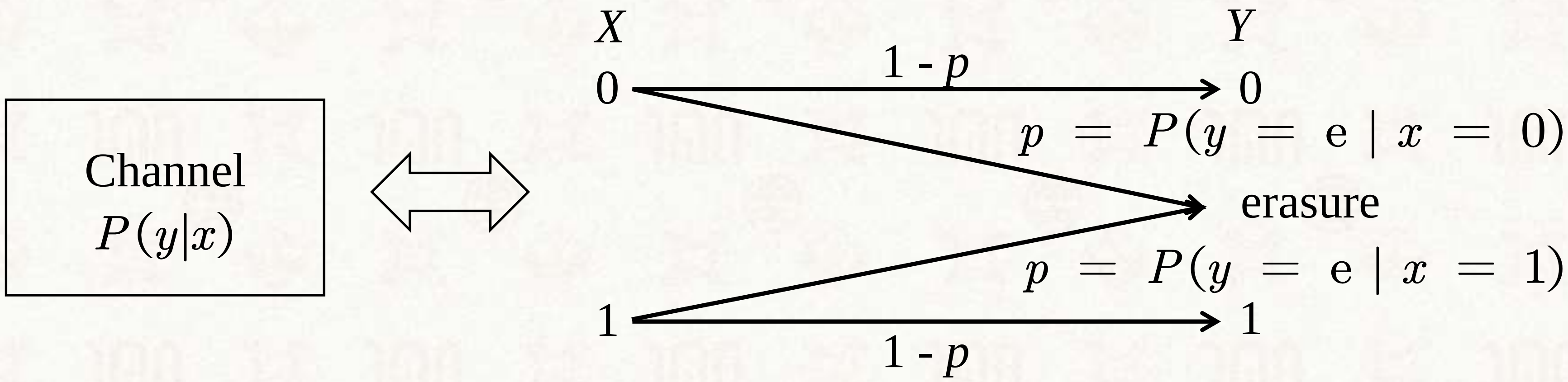
$$C = 1 - H_2(p)$$

$$= 1 - H(Y|X) \text{ bits / sym.}$$

Some Basics of Information Theory

I. Mutual Information – Channel Capacity

BEC:



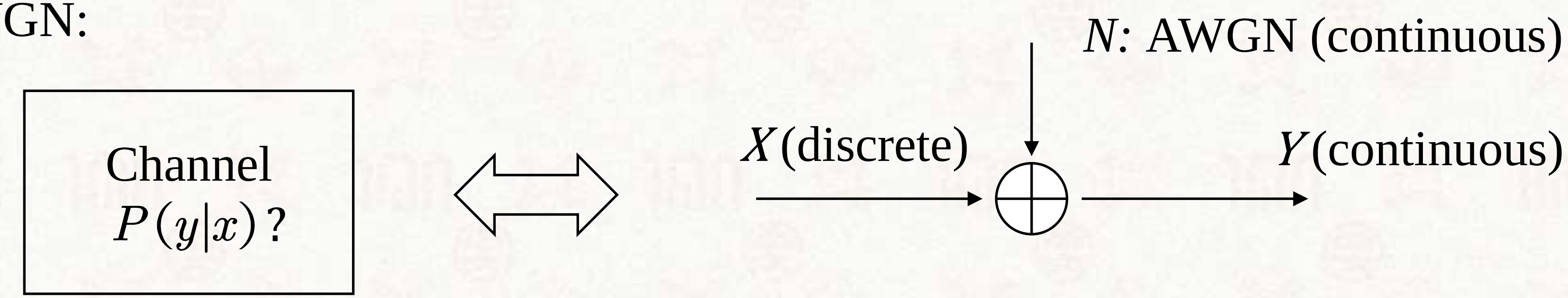
When $P(x = 0) = P(x = 1) = \dfrac{1}{2}$

$C = 1 - p$ bits / sym.

Some Basics of Information Theory

I. Mutual Information – Channel Capacity

AWGN:



1) Baseband, and when X is Gaussian distributed

$$C = \frac{1}{2}\log_2\left(1+\frac{\sigma_X^2}{\sigma_N^2}\right) \text{ bits / sym.}$$

2) Band limited (carrier frequency: W Hz), and when X is Gaussian distributed

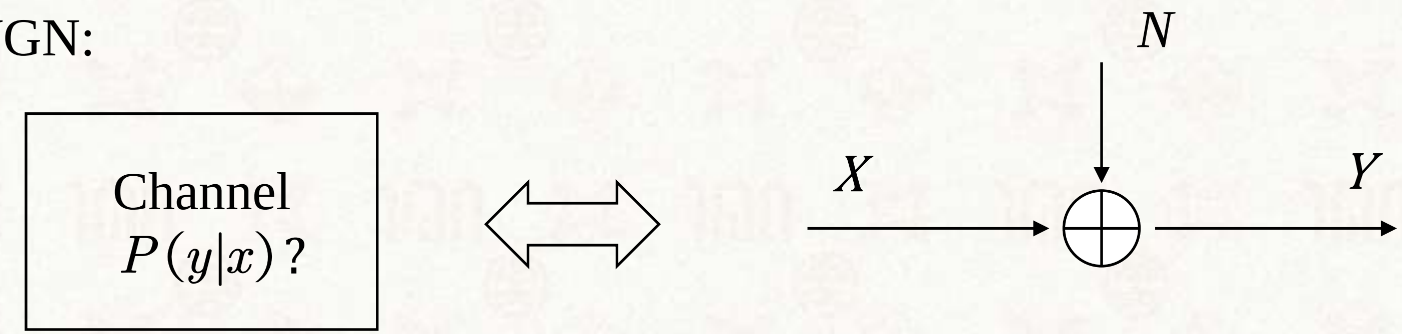
$$C = W\log_2\left(1+\frac{E}{WN_0}\right) \text{ bits / sec.}$$

Note: $E = 2W\sigma_X^2,\ N_0 = 2\sigma_N^2$

Some Basics of Information Theory

I. Mutual Information – Channel Capacity

AWGN:



3) Finite modulation alphabet, and Baseband

$$X:\ \{x_1,\ x_2,\ \cdots,\ x_u\}$$

When $P(x_1) = \cdots = P(x_u) = \dfrac{1}{u}$

$$C = \log_2 u - \frac{1}{u}\sum_{i=1}^{u} \mathbb{E}\left[\log_2 \frac{\sum_{i'=1}^{u} P(y|x_{i'})}{P(y|x_i)}\right]$$

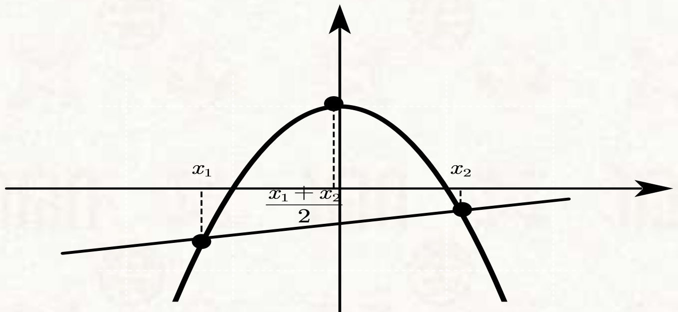
Some Basics of Information Theory

II. Information Inequalities

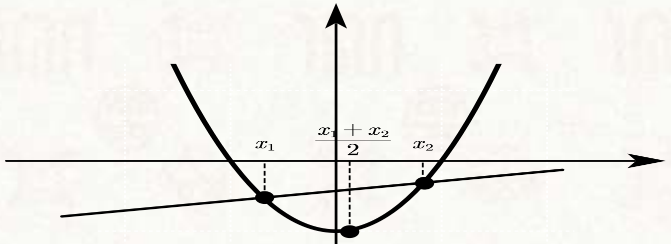
Jensen's Inequality

Concave $\mathbb{E}[f(x)] \leqslant f(\mathbb{E}[x])$

Convex $f(\mathbb{E}[x]) \leqslant \mathbb{E}[f(x)]$

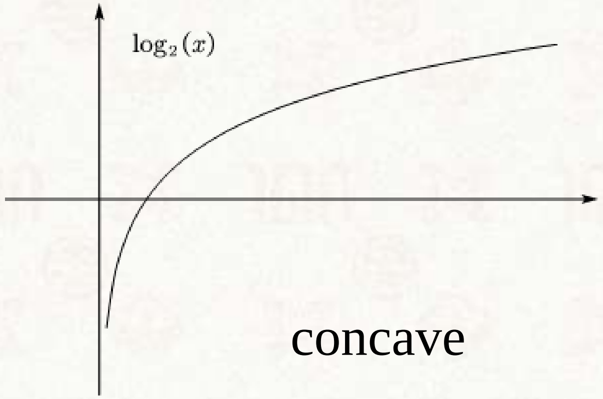


$$\frac{f(x_1)+f(x_2)}{2} < f\left(\frac{x_1+x_2}{2}\right)$$



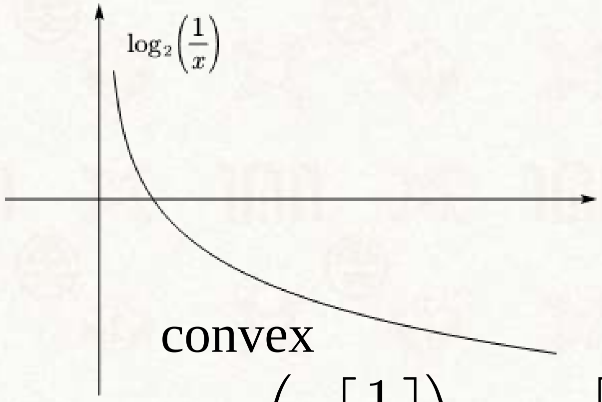
$$\frac{f(x_1)+f(x_2)}{2} > f\left(\frac{x_1+x_2}{2}\right)$$

The logarithm function,



concave

$$\mathbb{E}[\log_2 x] \leqslant \log_2(\mathbb{E}[x])$$



convex

$$\log_2\left(\mathbb{E}\left[\frac{1}{x}\right]\right) \leqslant \mathbb{E}\left[\log_2\frac{1}{x}\right]$$

1924-2024 中山大學 世纪华诞 100th ANNIVERSARY SUN YAT-SEN UNIVERSITY

II. Information Inequalities

Jensen's Inequality => Maximum Entropy Distribution

$$\begin{aligned} H(X) &= \mathbb{E}[\log_2 P(x)^{-1}] \\ &\leqslant \log_2(\mathbb{E}[P(x)^{-1}]) \\ &= \log_2\left(\sum_{i=1}^{u} P(x_i)P(x_i)^{-1}\right) \\ &\overset{(a)}{=} \log_2 u. \quad \text{bits / sym.} \end{aligned}$$

(a): when $P(x_1) = \cdots = P(x_u) = \frac{1}{u}$

Some Basics of Information Theory

II. Information Inequalities

Fano's Inequality*

Given $X,\ Y \in \{X_1,\ \cdots,\ X_u\}$,

$$P_e = P(X \neq Y).$$

$$H(X|Y) \leqslant H_2(P_e) + P_e \log_2(u - 1)$$

A casino setup: $H_2(P_e)$ - # Information for eye blinking

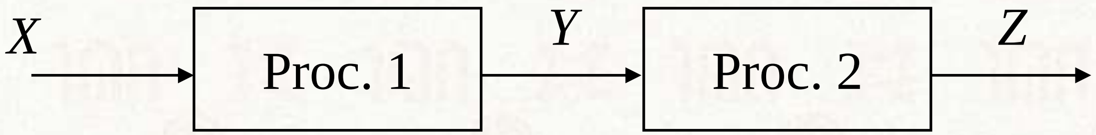
$P_e \log_2(u - 1)$ - # Information for blinking left eye

$*$: used to prove Shannon's channel coding theorem

Some Basics of Information Theory

II. Information Inequalities

Data Processing Inequality*



$X - Y - Z$ form a Markov chain

$$P(z|xy) = P(z|y)$$
$$P(x|yz) = P(x|y)$$

$$I(X;Z) \leqslant I(X;Y)$$
$$I(X;Z) \leqslant I(Y;Z)$$

* : used to prove Shannon's channel coding theorem