Fundamentals of Information Theory

Basic Concepts

Yayu Gao
School of Electronic Information and Communications
Huazhong University of Science and Technology
Email: yayugao@hust.edu.cn

Outline

- Model of communication systems
- How to characterize the information source?
- How much information a message contains?
- What is entropy?
- Joint and conditional entropy
- Relative entropy and mutual information
- Entropies in communications
- Chain Rules
- Jensen’s Inequality and Log Sum Inequality
- Entropy rate: from single-outcome to sequence-outcome
- What is a Markov source?
- Differential Entropy: from discrete to continuous

本节学习目标

1. 写出定义与表达式，进行计算
   - ❑Joint entropy
   - ❑Conditional entropy
   - ❑Relative entropy
   - ❑Mutual information
2. 说出≥2条互信息的性质
3. 根据Venn Diagram，说出≥4个数学关系
4. 说出熵等相关概念在通信系统中的物理意义

重难点：

- ➢ 概念及其表达式
- ➢ 概念之间的关系
- ➢ 计算
- ➢ 物理意义

05

Joint and conditional entropy

Joint entropy: definition

- The joint entropy of **a pair** of discrete random variables (X, Y) with a joint distribution $p(x, y)$ is defined as

$$H(X,Y) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log[p(x,y)]$$
$$= -E\{\log[p(x,y)]\}$$

- Do we have $H(X, Y) = H(X) + H(Y)$?

Joint entropy: example

- There are 2 black balls and 1 white ball in the box.
- Case 1
  - X: Pick one ball and check the color, then put it back;
  - Y : Pick another ball and check the color.
- Case 2
  - X: Pick one ball and check the color, yet do not put it back;
  - Y : Pick another ball and check the color.
- What are the joint entropy of (X, Y) in these two cases?
- $\boldsymbol{H(X, Y) \leq H(X) + H(Y)}$**.**
- When does "=" hold?
- **Where is the missing information?**

Conditional entropy: definition

- Conditional entropy

$$
\begin{aligned}
H(Y|X) &= \sum_{x\in\mathcal{X}} p(x) H(Y|X=x) \\
&= -\sum_{x\in\mathcal{X}} p(x) \sum_{y\in\mathcal{Y}} p(y|x) \log\left[p(y|x)\right] \\
&= -\sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p(x,y) \log\left[p(y|x)\right] \\
&= -E\{\log[p(y|x)]\}
\end{aligned}
$$

- **上式被称为，联合集XY中，集Y相对于集X的条件熵**
- If X and Y are independent, $H(Y|X)$=? ***H(Y)***
- Can you prove it?

Conditional entropy: notes

- Do we have *H(Y|X)=H(X|Y)*? ***In general, NO.***
- 集*X*相对于集*Y*的条件熵

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

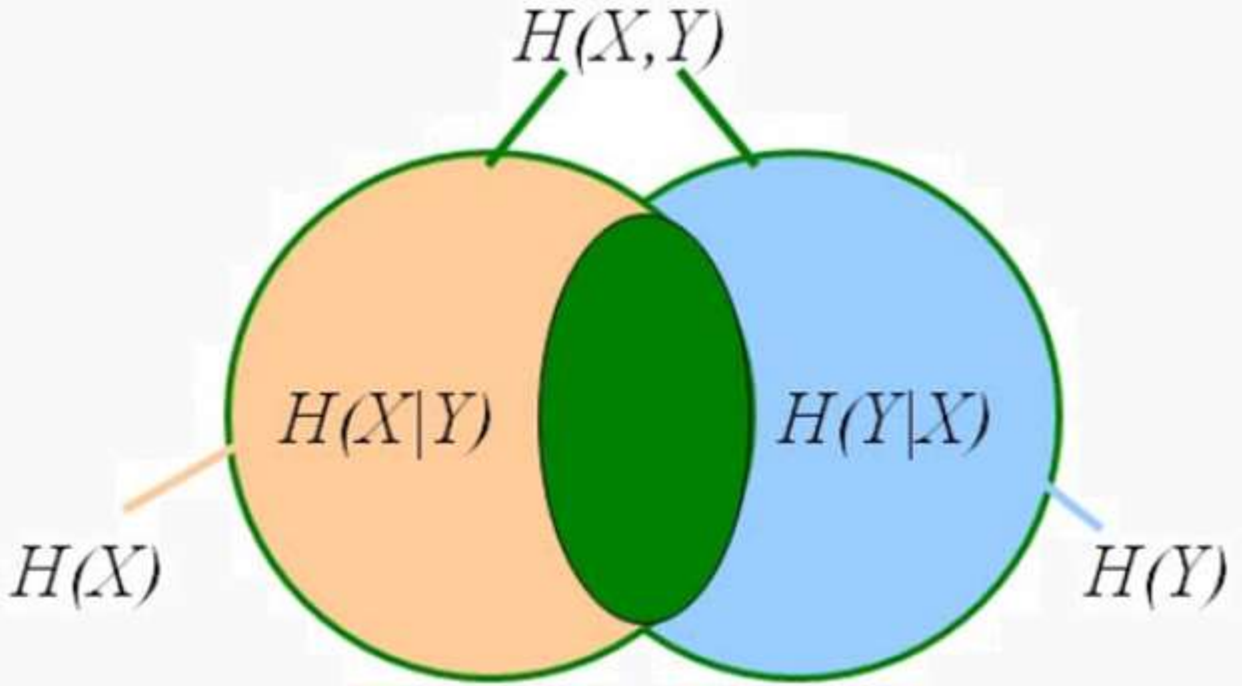
- 集*Y*相对于集*X*的条件熵

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

- Note: The average is taken over the joint distribution.
- **DO NOT** write it as

$$\boldsymbol{H(X|Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i|y_j)\cdot \log p(x_i|y_j)}$$

Venn Diagram

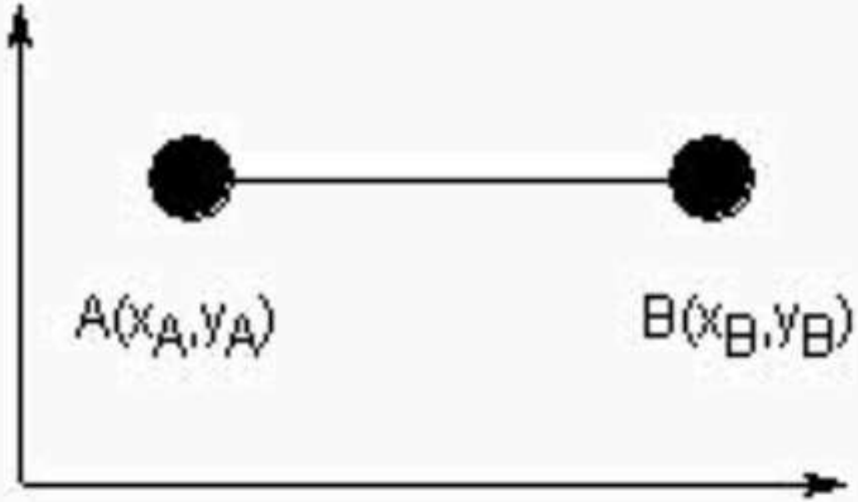


- What can you see?

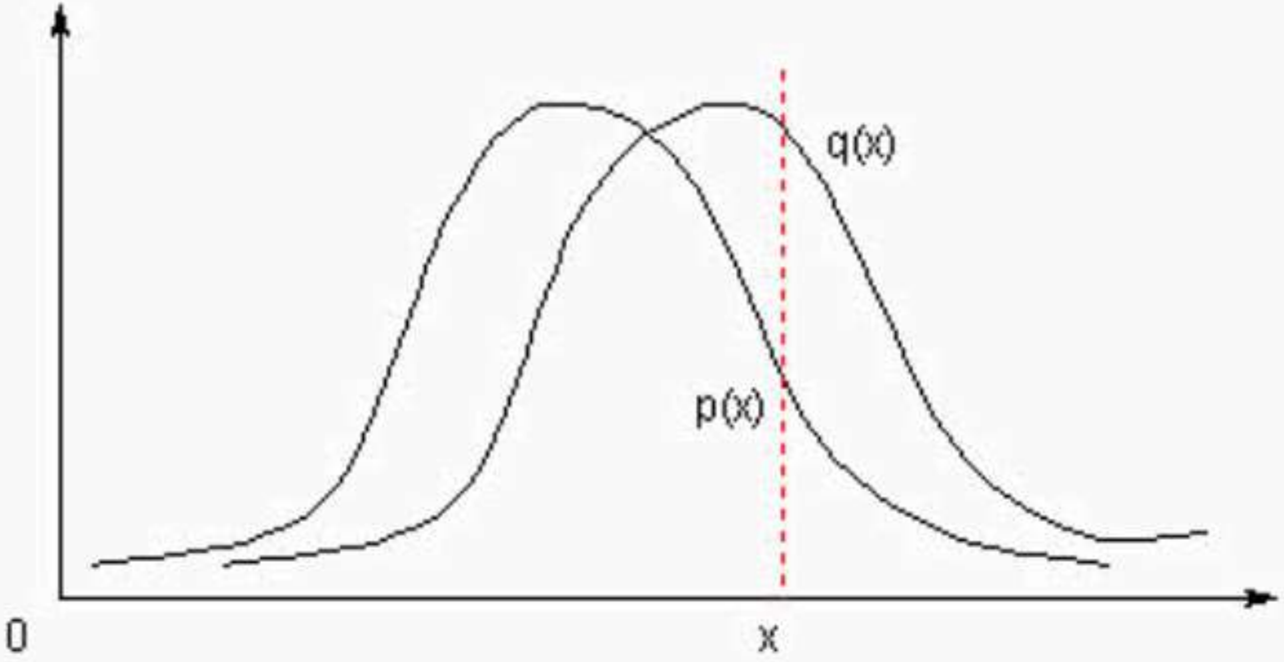
06

Relative Entropy

Relative entropy: Motivation



$$|A-B| = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2}$$



$$|I_B - I_A| = \left|\log\left[\frac{1}{q(x)}\right] - \log\left[\frac{1}{p(x)}\right]\right| = \log\left[\frac{p(x)}{q(x)}\right]$$

How to measure the **distance** between two *p.m.f.*?

$$\text{Average:} \quad \sum_{x\in\mathcal{X}} p(x)\log\left[\frac{p(x)}{q(x)}\right]$$

Relative entropy (Kullback-Leibler divergence): Definition

- Definition: a measure of the information distance or the information divergence between two *p.m.f.*, *p(x)* and *q(x).*

$$D(p(x)||q(x)) = \sum_{x\in\mathcal{X}} p(x) \log\left[\frac{p(x)}{q(x)}\right] = E_p\left\{\log\left[\frac{p(X)}{q(X)}\right]\right\}$$

- When *p(x)* is the true *p.m.f.* of *X*, this measures the inefficiency of assuming *q(x)* is the *p.m.f.* of *X*.
- It is “distance-like” in many respects.
- It is not a true distance, since it
  - is not symmetric $D(p||q)$ v.s. $D(q||p)$
  - does not satisfy the triangle inequality $D(p||q) + D(q||r)$ v.s. $D(p||r)$

Asymmetry of relative entropy: Example

Let $x \in \mathcal{X} = \{0, 1\}$, $p(0) = 1 - r$, $p(1) = r$, $q(0) = 1 - s$, $q(1) = s$.

$$\begin{aligned} D(p(x)||q(x)) &= p(0) \log \left[\frac{p(0)}{q(0)}\right] + p(1) \log \left[\frac{p(1)}{q(1)}\right] \\ &= (1-r) \log \left[\frac{1-r}{1-s}\right] + r \log \left[\frac{r}{s}\right] \end{aligned}$$

$$\begin{aligned} D(q(x)||p(x)) &= q(0) \log \left[\frac{q(0)}{p(0)}\right] + q(1) \log \left[\frac{q(1)}{p(1)}\right] \\ &= (1-s) \log \left[\frac{1-s}{1-r}\right] + s \log \left[\frac{s}{r}\right] \end{aligned}$$

- If $r = s$, $\Rightarrow D(p||q) = D(q||p)$
- If $r \neq s$, such as $r = 1/2$, $s = 1/4$
  $\Rightarrow D(p||q) = 0.2075$ bits, $D(q||p) = 0.1887$ bits
- Thus, in general $D(p||q) \neq D(q||p)$.

07

Mutual Information

Mutual information: Motivation

- Things are commonly related; two random variables are usually related.

How to characterize the relationship between two *r.v.*'s?

- Observe X alone, the information of X is $H(X)$.
- Knowing Y, the information of X becomes $H(X|Y)$.
- Knowing Y, the information of X is reduced by $\Delta = H(X) - H(X|Y)$.
- This reduced information Δ is the uncertainty reduction of X after knowing Y.

Mutual information: Definition

- Definition: Mutual information is the **relative entropy** between the joint distribution and the product distribution of two random variables *X*, *Y*.

$$
\begin{aligned}
I(X;Y) &= D\left[p(x,y)||p(x)p(y)\right] \\
&= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log\left[\frac{p(x,y)}{p(x)p(y)}\right] \\
&= E_{(X,Y)}\left\{\log\left[\frac{p(X,Y)}{p(X)p(Y)}\right]\right\}
\end{aligned}
$$

- Measure of the information one random variable (say, *X*) contains in another (*Y*)
- Special cases
  - If *X* and *Y* are independent, $I(X;Y) = 0$.
  - If $Y = X$, $I(X;X) = H(X)$.

Some other concepts

- Conditional relative entropy

$$D(p(y|x)||q(y|x)) = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log\left[\frac{p(y|x)}{q(y|x)}\right]$$

- Conditional mutual information

$$\begin{aligned} I(X;Y|Z) &= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}}\sum_{z\in\mathcal{Z}} p(x,y,z)\log\left[\frac{p(x,y|z)}{p(x|z)p(y|z)}\right] \\ &= E_{p(x,y,z)}\left\{\log\left[\frac{p(X,Y|Z)}{p(X|Z)p(Y|Z)}\right]\right\} \end{aligned}$$

Mutual information: Properties

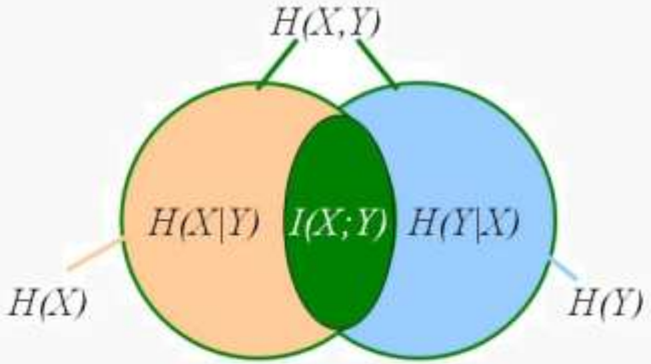
- Symmetry: $I(X;Y) = I(Y;X)$
  - It is indicated in “Mutual”.

- Non-negativity: $I(X;Y) \geq 0$

- Limits: $I(X;Y) \leq \min\,(H(X), H(Y))$

Mutual information vs. Entropy

$$\boxed{I(X;Y) = H(X) - H(X|Y)}$$

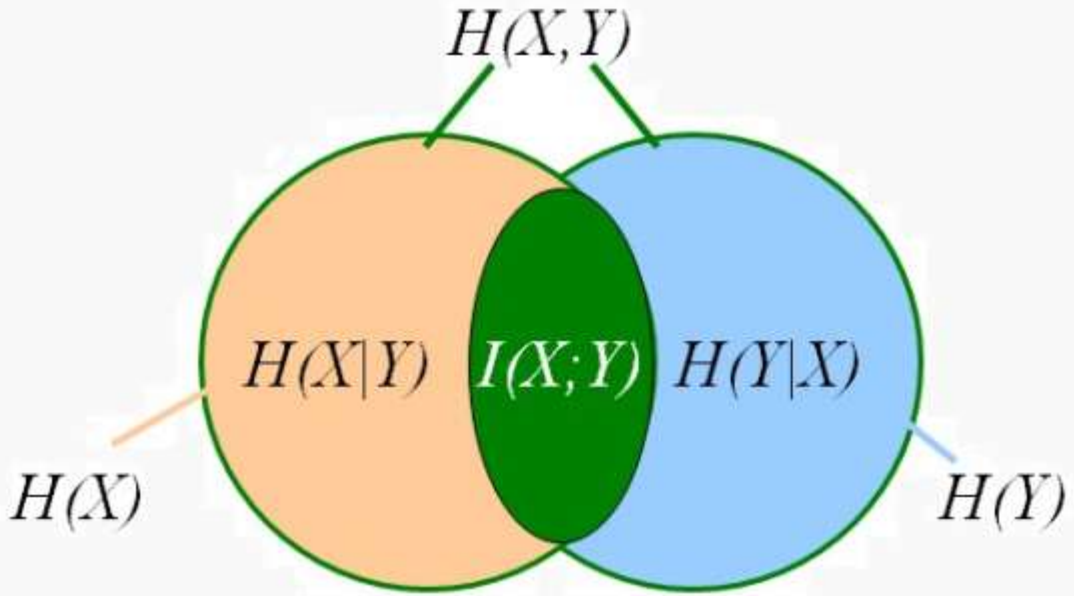


Proof:

$$\begin{aligned}
I(X;Y) &= \sum_x \sum_y p(x,y) \log\left[\frac{p(x,y)}{p(x)p(y)}\right] \\
&= \sum_x \sum_y p(x,y) \log\left[\frac{p(x|y)}{p(x)}\right] \\
&= \sum_x \sum_y p(x,y) \log[p(x|y)] - \sum_x \sum_y p(x,y) \log[p(x)] \\
&= -\sum_x p(x) \log[p(x)] - (-\sum_x \sum_y p(x,y) \log[p(x|y)]) \\
&= H(X) - H(X|Y)
\end{aligned}$$

Mutual information vs. Entropy

- Venn Diagram



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

Example #1

- Joint $p.m.f.$ is:

$Y \backslash X$	1	2	3	4	$p(y)$
1	$1/8$	$1/16$	$1/32$	$1/32$	$1/4$
2	$1/16$	$1/8$	$1/32$	$1/32$	$1/4$
3	$1/16$	$1/16$	$1/16$	$1/16$	$1/4$
4	$1/4$	0	0	0	$1/4$
$p(x)$	$1/2$	$1/4$	$1/8$	$1/8$	

- What is $H(X)$, $H(Y)$, $H(X|Y)$, $H(Y|X)$, $H(X, Y)$, $I(X; Y)$?

Solution of example #1

$$
\begin{aligned}
H(X) &= -\sum_{x\in\mathcal{X}} p(x)\log[p(x)] \\
&= H\left(\frac{1}{2},\frac{1}{4},\frac{1}{8},\frac{1}{8}\right) \\
&= -\left[\frac{1}{2}\log(\frac{1}{2}) + \frac{1}{4}\log(\frac{1}{4}) + \frac{1}{8}\log(\frac{1}{8}) + \frac{1}{8}\log(\frac{1}{8})\right] \\
&= 1.75 \text{ bits}
\end{aligned}
$$

$$
\begin{aligned}
H(Y) &= -\sum_{y\in\mathcal{Y}} p(y)\log[p(y)] \\
&= H\left(\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4}\right) \\
&= -\left[\frac{1}{4}\log(\frac{1}{4}) + \frac{1}{4}\log(\frac{1}{4}) + \frac{1}{4}\log(\frac{1}{4}) + \frac{1}{4}\log(\frac{1}{4})\right] \\
&= 2 \text{ bits}
\end{aligned}
$$

Solution of example #1

$$
\begin{aligned}
H(X|Y) &= \sum_{y\in\mathcal{Y}} p(y) H(X|Y=y) \\
&= \sum_{y\in\mathcal{Y}} p(y) \sum_{x\in\mathcal{X}} p(x|y) \log\left[\frac{1}{p(x|y)}\right] \\
&= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log\left[p(x|y)\right] \\
&= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log\left[\frac{p(x,y)}{p(y)}\right] \\
&= -\left[\begin{array}{c}
\frac{1}{8}\log\frac{\frac{1}{8}}{\frac{1}{4}} + \frac{1}{16}\log\frac{\frac{1}{16}}{\frac{1}{4}} + \frac{1}{32}\log\frac{\frac{1}{32}}{\frac{1}{4}} + \frac{1}{32}\log\frac{\frac{1}{32}}{\frac{1}{4}} \\
+\frac{1}{16}\log\frac{\frac{1}{16}}{\frac{1}{4}} + \frac{1}{8}\log\frac{\frac{1}{8}}{\frac{1}{4}} + \frac{1}{32}\log\frac{\frac{1}{32}}{\frac{1}{4}} + \frac{1}{32}\log\frac{\frac{1}{32}}{\frac{1}{4}} \\
+\frac{1}{16}\log\frac{\frac{1}{16}}{\frac{1}{4}} + \frac{1}{16}\log\frac{\frac{1}{16}}{\frac{1}{4}} + \frac{1}{16}\log\frac{\frac{1}{16}}{\frac{1}{4}} + \frac{1}{16}\log\frac{\frac{1}{16}}{\frac{1}{4}} \\
+\frac{1}{4}\log\frac{\frac{1}{4}}{\frac{1}{4}} + 0\log\frac{0}{\frac{1}{4}} + 0\log\frac{0}{\frac{1}{4}} + 0\log\frac{0}{\frac{1}{4}}
\end{array}\right] \\
&= 1.375 \text{ bits}
\end{aligned}
$$

Solution of example #1

$$
\begin{aligned}
H(Y|X) &= \sum_{x\in\mathcal{X}} p(x) H(Y|X=x) \\
&= \sum_{x\in\mathcal{X}} p(x) \sum_{y\in\mathcal{Y}} p(y|x) \log\left[\frac{1}{p(y|x)}\right] \\
&= -\sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p(x,y) \log\left[p(y|x)\right] \\
&= -\sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p(x,y) \log\left[\frac{p(x,y)}{p(x)}\right] \\
&= -\left[\begin{array}{c}
\frac{1}{8}\log\frac{\frac{1}{8}}{\frac{1}{2}} + \frac{1}{16}\log\frac{\frac{1}{16}}{\frac{1}{4}} + \frac{1}{32}\log\frac{\frac{1}{32}}{\frac{1}{8}} + \frac{1}{32}\log\frac{\frac{1}{32}}{\frac{1}{8}} \\
+\frac{1}{16}\log\frac{\frac{1}{16}}{\frac{1}{2}} + \frac{1}{8}\log\frac{\frac{1}{8}}{\frac{1}{4}} + \frac{1}{32}\log\frac{\frac{1}{32}}{\frac{1}{8}} + \frac{1}{32}\log\frac{\frac{1}{32}}{\frac{1}{8}} \\
+\frac{1}{16}\log\frac{\frac{1}{16}}{\frac{1}{2}} + \frac{1}{16}\log\frac{\frac{1}{16}}{\frac{1}{4}} + \frac{1}{16}\log\frac{\frac{1}{16}}{\frac{1}{8}} + \frac{1}{16}\log\frac{\frac{1}{16}}{\frac{1}{8}} \\
+\frac{1}{4}\log\frac{\frac{1}{4}}{\frac{1}{2}} + 0\log\frac{0}{\frac{1}{4}} + 0\log\frac{0}{\frac{1}{8}} + 0\log\frac{0}{\frac{1}{8}}
\end{array}\right] \\
&= 1.625 \text{ bits}
\end{aligned}
$$

Solution of example #1

$$
\begin{aligned}
H(X,Y) &= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log[p(x,y)] \\
&= -\left[\begin{array}{c}
\frac{1}{8}\log\frac{1}{8}+\frac{1}{16}\log\frac{1}{16}+\frac{1}{32}\log\frac{1}{32}+\frac{1}{32}\log\frac{1}{32} \\
+\frac{1}{16}\log\frac{1}{16}+\frac{1}{8}\log\frac{1}{8}+\frac{1}{32}\log\frac{1}{32}+\frac{1}{32}\log\frac{1}{32} \\
+\frac{1}{16}\log\frac{1}{16}+\frac{1}{16}\log\frac{1}{16}+\frac{1}{16}\log\frac{1}{16}+\frac{1}{16}\log\frac{1}{16} \\
+\frac{1}{4}\log\frac{1}{4}+0\log 0+0\log 0+0\log 0
\end{array}\right] \\
&= 3.375 \text{ bits}
\end{aligned}
$$

$H(X) = 1.75$ bits, $H(Y) = 2$ bits, $H(X|Y) = 1.375$ bits, $H(Y|X) = 1.625$ bits

Note that $H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$ by observation in this example.

Solution of example #1

- Method 1:

$$I(X;Y) = H(X) - H(X|Y) = 1.75 - 1.375 = 0.375 \text{ bit}$$

$$I(X;Y) = H(Y) - H(Y|X) = 2 - 1.625 = 0.375 \text{ bit}$$

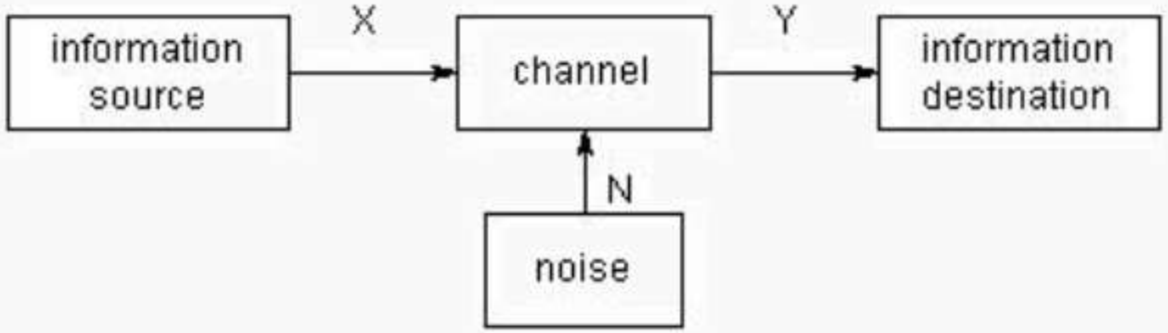
- Method 2:

$$\begin{aligned}
I(X;Y) &= D\left[p(x,y) \| p(x)p(y)\right] \\
&= \sum_{x\in\mathcal{X}} \sum_{y\in\in\mathcal{Y}} p(x,y) \log\left[\frac{p(x,y)}{p(x)p(y)}\right] \\
&= \left[\begin{array}{c}
\frac{1}{8}\log\frac{\frac{1}{8}}{\frac{1}{2}\cdot\frac{1}{4}} + \frac{1}{16}\log\frac{\frac{1}{16}}{\frac{1}{4}\cdot\frac{1}{4}} + \frac{1}{32}\log\frac{\frac{1}{32}}{\frac{1}{8}\cdot\frac{1}{4}} + \frac{1}{32}\log\frac{\frac{1}{32}}{\frac{1}{8}\cdot\frac{1}{4}} \\
+\frac{1}{16}\log\frac{\frac{1}{16}}{\frac{1}{2}\cdot\frac{1}{4}} + \frac{1}{8}\log\frac{\frac{1}{8}}{\frac{1}{4}\cdot\frac{1}{4}} + \frac{1}{32}\log\frac{\frac{1}{32}}{\frac{1}{8}\cdot\frac{1}{4}} + \frac{1}{32}\log\frac{\frac{1}{32}}{\frac{1}{8}\cdot\frac{1}{4}} \\
+\frac{1}{16}\log\frac{\frac{1}{16}}{\frac{1}{2}\cdot\frac{1}{4}} + \frac{1}{16}\log\frac{\frac{1}{16}}{\frac{1}{4}\cdot\frac{1}{4}} + \frac{1}{16}\log\frac{\frac{1}{16}}{\frac{1}{8}\cdot\frac{1}{4}} + \frac{1}{16}\log\frac{\frac{1}{16}}{\frac{1}{8}\cdot\frac{1}{4}} \\
+\frac{1}{4}\log\frac{\frac{1}{4}}{\frac{1}{2}\cdot\frac{1}{4}} + 0\log 0 + 0\log 0 + 0\log 0
\end{array}\right] \\
&= 0.375 \text{ bit}
\end{aligned}$$

08

Entropies in Communications

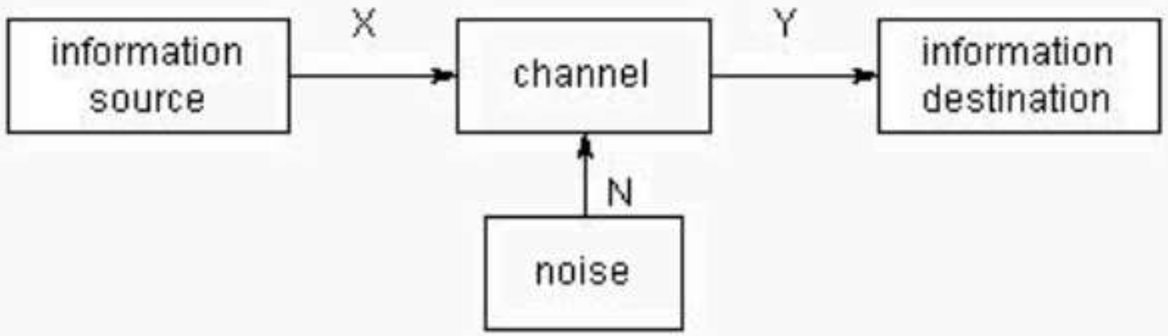
Entropies in Communications

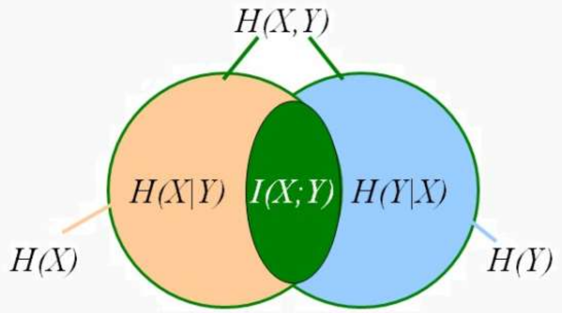


- System model
  - Source sends *r.v. X,* destination receives *r.v. Y.*
  - Realization of X (or Y) is x_i (or y_i)

- **How much information transmitted from source to information?**
- Options: H(X), H(Y), H(X, Y), H(X|Y), H(Y|X), I(X; Y)

How much information transmitted from source to information?

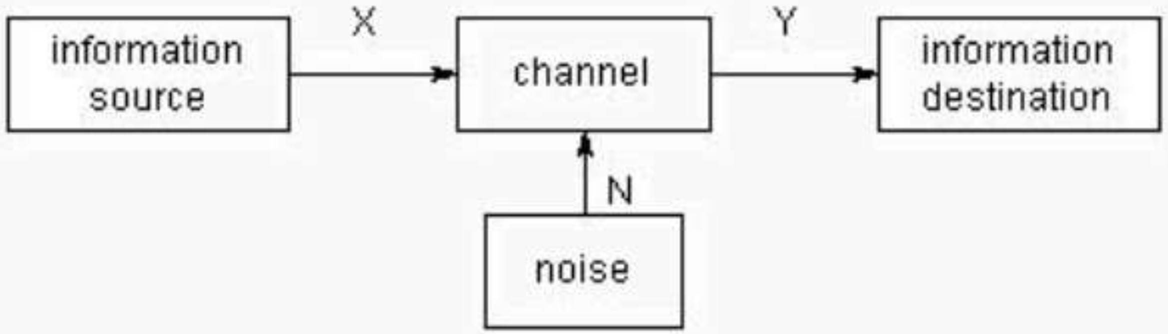


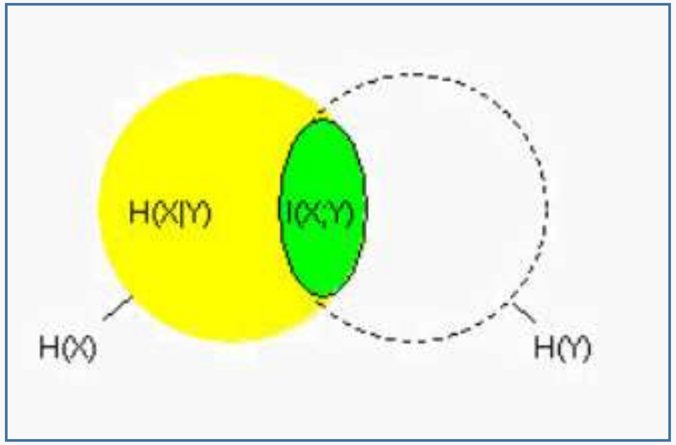


- System model
  - Source sends *r.v. X,* destination receives *r.v. Y.*
- ***I(X;Y)***: information successfully transmitted from the source to the destination.

$$I(X;Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) \log \left[\frac{p(x,y)}{p(x)p(y)} \right] = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x)p(y|x) \log \left[\frac{p(y|x)}{p(y)} \right]$$

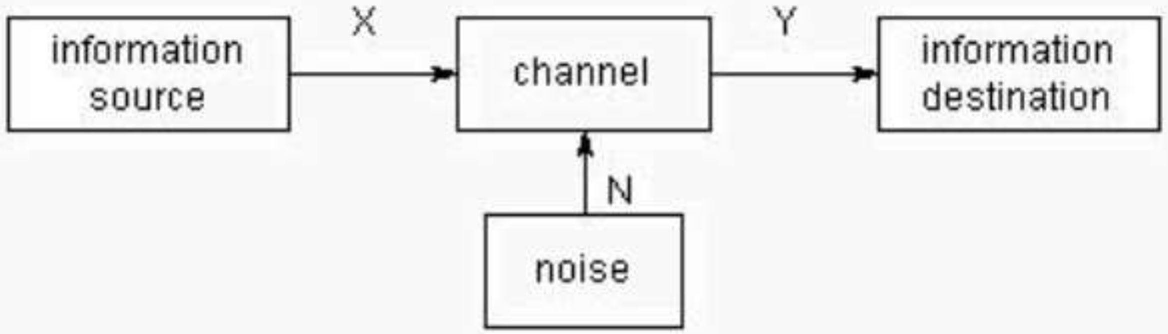
How much information is lost after the transmission?

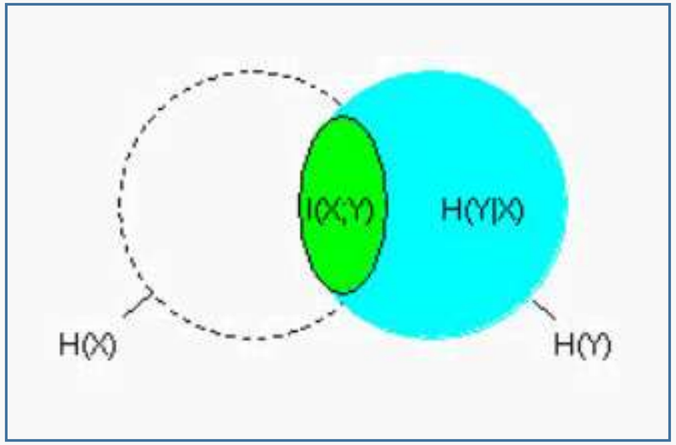




- Ideally, $H(X)$ should be transmitted from the source to the destination.
- $H(X) = H(X|Y) + I(X;Y)$
- $I(X;Y) = H(X) - H(X|Y)$
- At Destination, after Y is received, there still exists average uncertainty about source X due to the transmission distortion in the channel.
- **H(X|Y): loss entropy**

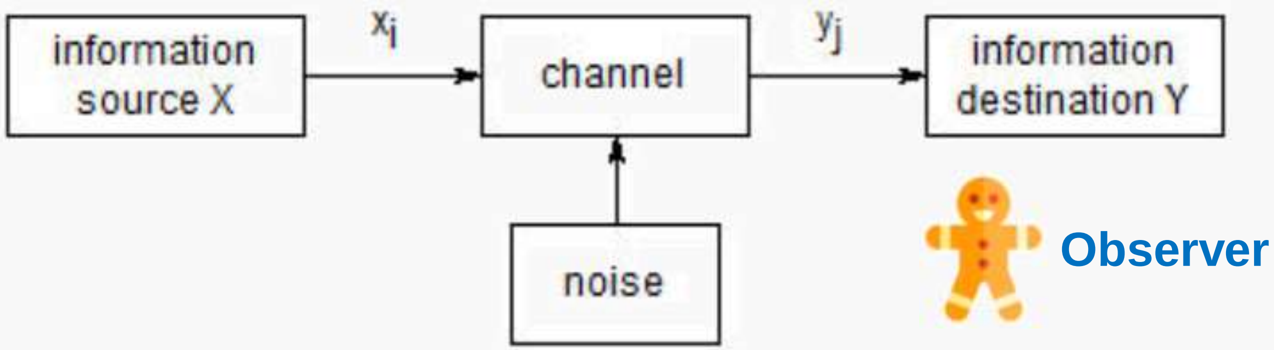
How much uncertainty due to channel noise?





- Ideally, if there is no noise in the channel, there should exist deterministic relationship between the sender and the receiver.
- $H(Y) = I(Y; X) + H(Y|X)$
- $I(Y; X) = H(Y) - H(Y|X)$
- At Source, after X is sent, there still exists average uncertainty about destination Y due to the channel noise.
- **H(Y|X): noise entropy**

Mutual information of realization: at destination



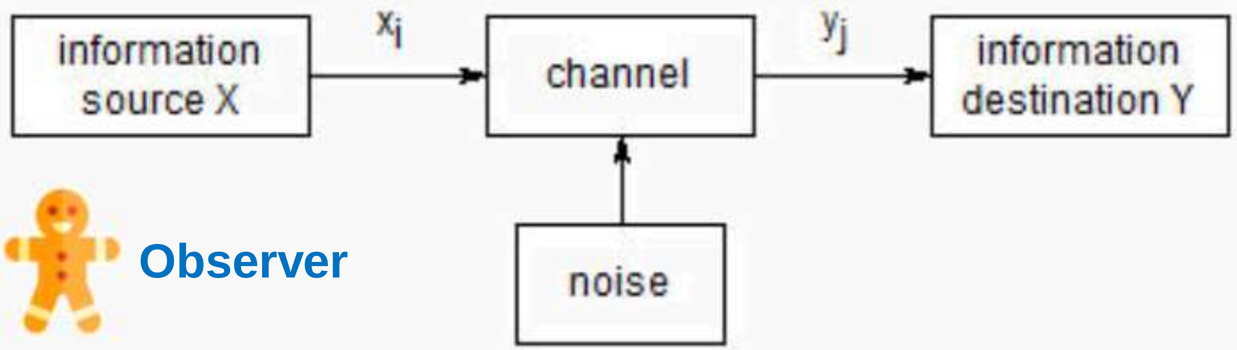
观察者站在信宿端。
互信息为通信后，从y_j获得的关于x_i的信息量。

- **Before** communication, is the source sending x_i ?
  - Priori probability p(x_i): uncertainty on x_i without receiving y_j
- **After** communication (received y_j), is the source sending x_i ?
  - Posteriori probability p(x_i| y_j): uncertainty on x_i with receiving y_j

$$I(x_i; y_j) = I(x_i) - I(x_i|y_j)$$

$$= \log\left[\frac{1}{p(x_i)}\right] - \log\left[\frac{1}{p(x_i|y_j)}\right] = \log\left[\frac{p(x_i|y_j)}{p(x_i)}\right]$$

Mutual information of realization: at source



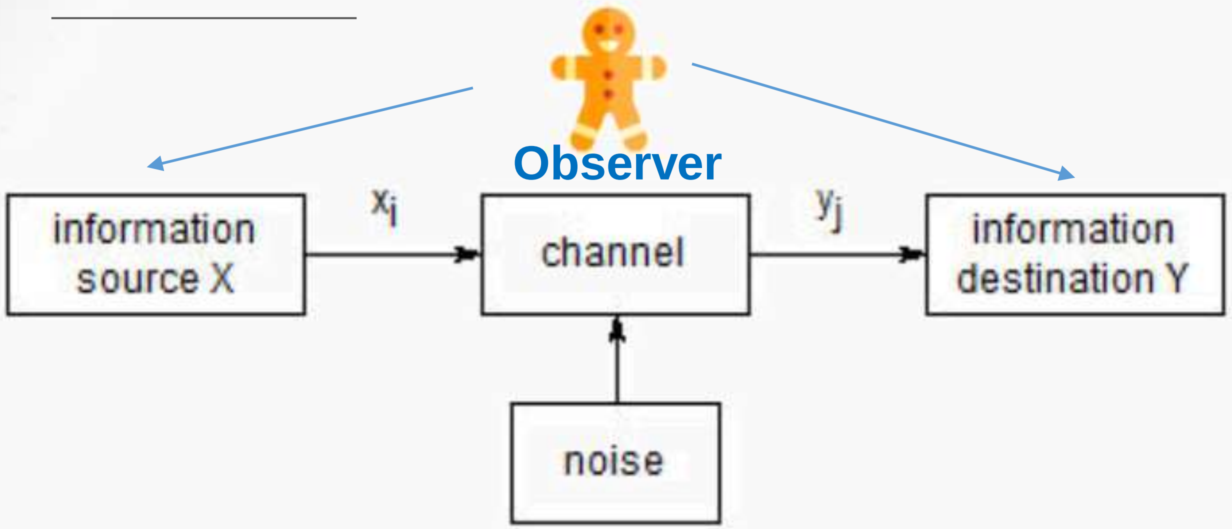
观察者站在信源端。
互信息为通信后，从x_i获得的关于y_j的信息量。

- **Before** communication, is the destination receiving y_j ?
  - Priori probability p(y_j): uncertainty on y_j without sending x_i
- **After** communication (sent x_i), is the destination receiving y_j ?
  - Posteriori probability p(y_j | x_i): uncertainty on y_j with sending x_i

$$I(y_j; x_i) = I(y_j) - I(y_j|x_i)$$

$$= \log\left[\frac{1}{p(y_j)}\right] - \log\left[\frac{1}{p(y_j|x_i)}\right] = \log\left[\frac{p(y_j|x_i)}{p(y_j)}\right]$$

Mutual information of realization: at system

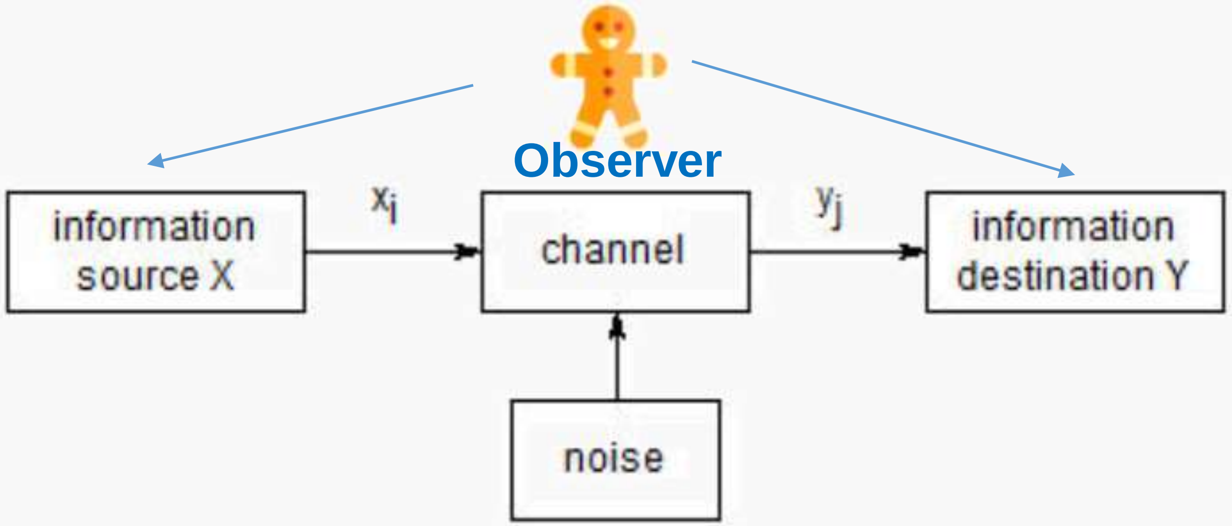


- **Before** communication, is the source sending x_i , destination receiving y_j ?
  - X and Y are considered to be statistically independent.

$$p(x_i, y_j) = p(x_i)p(y_j)$$

$$I_{before}(x_i, y_j) = \log\left[\frac{1}{p(x_i, y_j)}\right] = \log\left[\frac{1}{p(x_i)p(y_j)}\right] = \log\left[\frac{1}{p(x_i)}\right] + \log\left[\frac{1}{p(y_j)}\right] = I(x_i) + I(y_j)$$

Mutual information of realization: at system

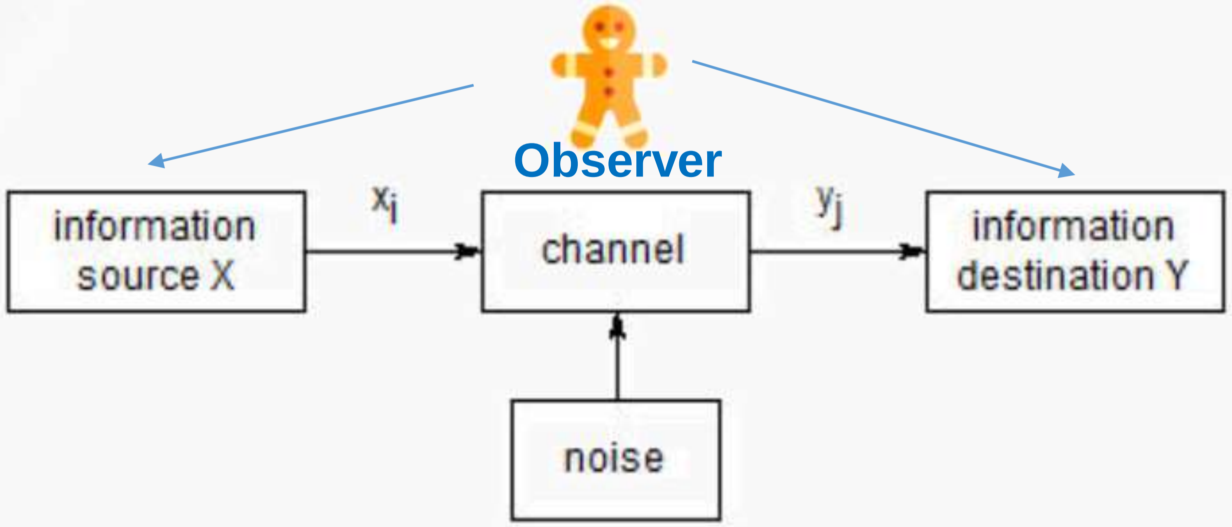


- **After** communication, is the source sending x_i , destination receiving y_j ?
  - X and Y are related due to channel characteristics.

$$p(x_i, y_j) = p(x_i)p(y_j|x_i) = p(y_j)p(x_i|y_j)$$

$$I_{after}(x_i, y_j) = \log\left[\frac{1}{p(x_i, y_j)}\right] = I(x_i, y_j)$$

Mutual information of realization: at system



观察者站在系统整体角度。
互信息为通信后整体不确定度的减少。

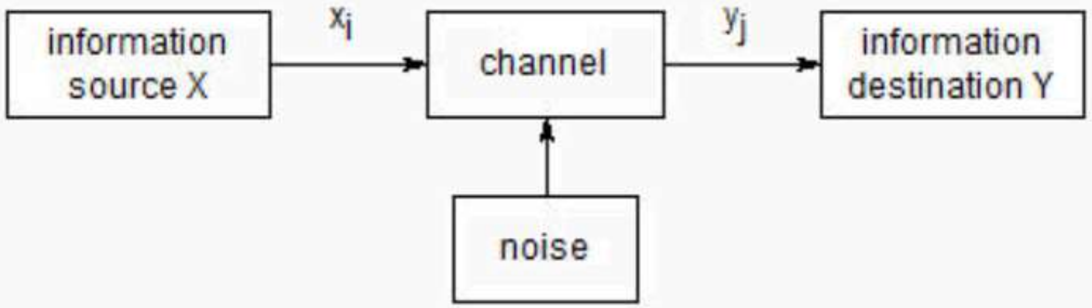
$$I(x_i;y_j)=[I(x_i)+I(y_j)]-I(x_i,y_j)$$

$$=[-\log p(x_i)-\log p(y_j)]-[-\log p(x_i,y_j)]$$

$$=\log\frac{p(x_i,y_j)}{p(x_i)p(y_j)}$$

Mutual information of realization: equivalency

- At destination, $I(x_i; y_j) = I(x_i) - I(x_i|y_j)$.
- At source, $I(y_j; x_i) = I(y_j) - I(y_j|x_i)$.
- From system, $I(x_i; y_j) = I(x_i) + I(y_j) - I(x_i, y_j)$



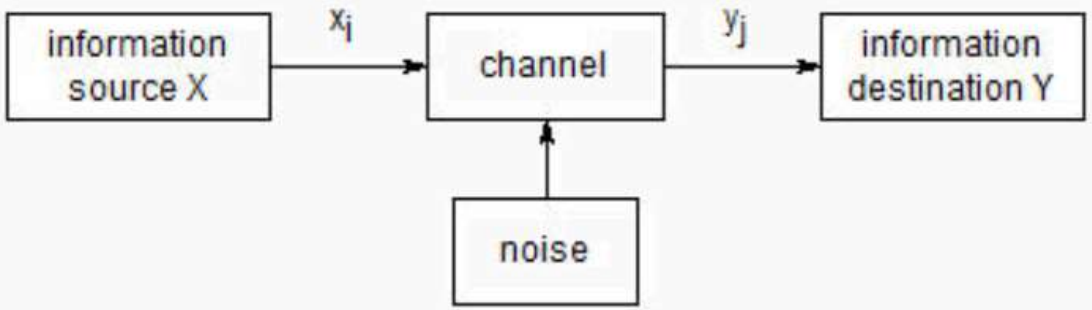
$$I(x_i, y_j) = \log\left[\frac{1}{p(x_i, y_j)}\right] = \log\left[\frac{1}{p(x_i)p(y_j|x_i)}\right] = \log\left[\frac{1}{p(x_i)}\right] + \log\left[\frac{1}{p(y_j|x_i)}\right]$$
$$= I(x_i) + I(y_j|x_i)$$

$$I(x_i) + I(y_j) - I(x_i, y_j) = I(x_i) + I(y_j) - [I(x_i) + I(y_j|x_i)] = I(y_j) - I(y_j|x_i)$$

$$I(y_j, x_i) = \log\left[\frac{1}{p(y_j, x_i)}\right] = \log\left[\frac{1}{p(y_j)p(x_i|y_j)}\right] = \log\left[\frac{1}{p(y_j)}\right] + \log\left[\frac{1}{p(x_i|y_j)}\right]$$
$$= I(y_j) + I(x_i|y_j)$$

$$I(x_i) + I(y_j) - I(y_j, x_i) = I(x_i) + I(y_j) - [I(y_j) + I(x_i|y_j)] = I(x_i) - I(x_i|y_j)$$

Mutual information: micro-level vs. macro-level

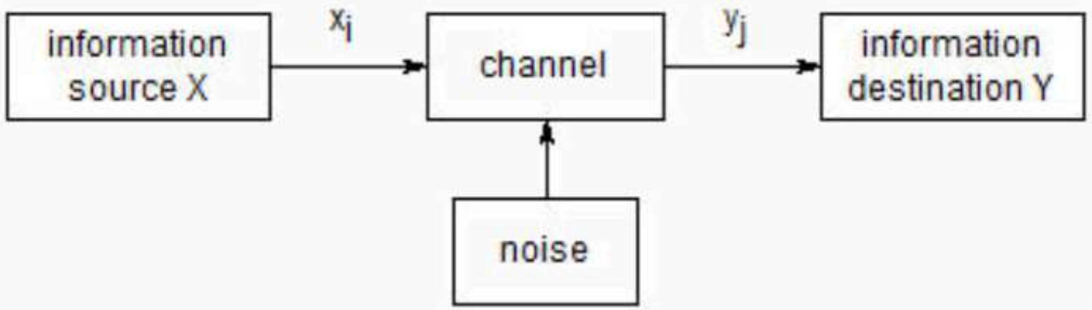


- Mutual information of **realization at the micro-level**
  - $I(x_i; y_j) = \log\left[\frac{p(x_i, y_j)}{p(x_i)p(y_j)}\right] = \log\left[\frac{p(x_i|y_j)}{p(x_i)}\right] = \log\left[\frac{1}{p(x_i)}\right] - \log\left[\frac{1}{p(x_i|y_j)}\right]$
  - At destination: $I(x_i; y_j) = I(x_i) - I(x_i|y_j)$
  - At source: $I(y_j; x_i) = I(y_j) - I(y_j|x_i)$
  - From system: $I(x_i; y_j) = I(x_i) + I(y_j) - I(x_i, y_j)$

- Mutual information at the macro-level

$$I(X;Y) = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log\left[\frac{p(x,y)}{p(x)p(y)}\right] = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x)p(y|x)\log\left[\frac{p(y|x)}{p(y)}\right]$$

Mutual information: micro-level vs. macro-level



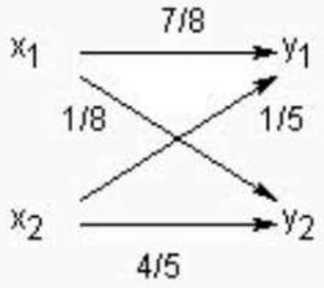
$$I(x_i; y_j) = \log\left[\frac{p(x_i, y_j)}{p(x_i)p(y_j)}\right]$$

$$\begin{aligned} I(X;Y) &= D\left[p(x,y) || p(x)p(y)\right] \\ &= \sum_{x_i \in \mathcal{X}} \sum_{y_j \in \mathcal{Y}} p(x_i, y_j) \log\left[\frac{p(x_i, y_j)}{p(x_i)p(y_j)}\right] \\ &= \sum_{x_i \in \mathcal{X}} \sum_{y_j \in \mathcal{Y}} p(x_i, y_j) I(x_i; y_j) \\ &= E_{X,Y}\left[I(x;y)\right] \end{aligned}$$

Non-negativity: ***I(X;Y) ≥ 0***

An example communication system

Given a discrete source of $\begin{bmatrix} X \\ p(X) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \\ 0.2 & 0.8 \end{bmatrix}$, the output messages pass through a noise channel; then, the received messages are modeled using $Y = [y_1, y_2]$.



- self-information in event x_1: $I(x_1) = \log \frac{1}{p(x_1)} = 2.322$ bits

$$p(y_1) = \sum_{x_i} p(x_i) p(y_1|x_i) = 0.335$$

$$I(x_1; y_1) = \log_2 \left(\frac{p(y_1|x_1)}{p(y_1)} \right) = \log_2 \left(\frac{7/8}{0.335} \right) = 1.39 \text{ bits}$$

$$I(x_1; y_2) = \log_2 \left(\frac{p(y_2|x_1)}{p(y_2)} \right) = -2.42 \text{ bits}$$

What does it mean?

Outline

- Model of communication systems
- How to characterize the information source?
- How much information a message contains?
- What is entropy?
- Joint and conditional entropy
- Relative entropy and mutual information
- Entropies in communications
- Chain Rules
- Jensen's Inequality and Log Sum Inequality
- Entropy rate: from single-outcome to sequence-outcome
- What is a Markov source?
- Differential Entropy: from discrete to continuous

本节学习目标

1. 写出定义与表达式，进行计算
   - ❑Joint entropy
   - ❑Conditional entropy
   - ❑Relative entropy
   - ❑Mutual information
2. 说出≥2条互信息的性质
3. 根据Venn Diagram，说出≥4个数学关系
4. 说出熵等相关概念在通信系统中的物理意义

重难点：

- ➢ 概念及其表达式
- ➢ 概念之间的关系
- ➢ 计算
- ➢ 物理意义

My Homepage

Thank you!

Yayu Gao

School of Electronic Information and Communications

Huazhong University of Science and Technology

Email: yayugao@hust.edu.cn