# 互信息和散度

1. 互信息
2. KL-散度
3. 马尔可夫链
4. 数据处理定理
5. 费诺不等式

# 互信息和散度

1. **互信息**
2. KL-散度
3. 马尔可夫链
4. 数据处理定理
5. 费诺不等式

# 课堂小测

某城市温度 $T$ 和天气 $W$ 的联合分布为:

| $p_{\mathsf{WT}}(w,t)$ | snowy | rainy | cloudy |
|---|---|---|---|
| $0°$C | $\frac{1}{6}$ | $\frac{1}{12}$ | $\frac{1}{12}$ |
| $10°$C | $\frac{1}{12}$ | $\frac{1}{6}$ | $\frac{1}{12}$ |
| $20°$C | $\frac{1}{12}$ | $\frac{1}{12}$ | $\frac{1}{6}$ |

(a) 计算 $H(T)$ 和 $H(W)$;

(b) 计算 $H(T, W)$;

(c) 计算 $H(W|T)$;

(d) (本节内容) 计算 $I(T; W)$.

# 互信息

- $X \sim p_X$，随机变量 $X$ 包含多少信息？

  $\Rightarrow$ **熵**：$H(X) = \mathrm{E}_X\{-\log p_X\}$

- $\{X, Y\} \sim p_{X,Y}$，从 $Y$ 中能获取多少关于 $X$ 的信息？

  $\Rightarrow$ **互信息**：$I(X; Y) = \mathrm{E}_{X,Y}\{\log \frac{p_{X,Y}}{p_X p_Y}\}$

# 互信息

- $X \sim p_X$，随机变量 $X$ 包含多少信息？

  $\Rightarrow$ **熵**：$H(X) = \mathrm{E}_X\{-\log p_X\}$

- $\{X, Y\} \sim p_{X,Y}$，从 $Y$ 中能获取多少关于 $X$ 的信息？

  $\Rightarrow$ **互信息**：$I(X;Y) = \mathrm{E}_{X,Y}\{\log \frac{p_{X,Y}}{p_X p_Y}\}$

**经验 1**：若 $Y = X$：$I(X;Y) = H(X)$

# 互信息

- $X \sim p_X$，随机变量 $X$ 包含多少信息？

  $\Rightarrow$ **熵**：$H(X) = \mathrm{E}_X\{-\log p_X\}$

- $\{X, Y\} \sim p_{X,Y}$，从 $Y$ 中能获取多少关于 $X$ 的信息？

  $\Rightarrow$ **互信息**：$I(X; Y) = \mathrm{E}_{X,Y}\{\log \frac{p_{X,Y}}{p_X p_Y}\}$

**经验** 1：若 $Y = X$：$I(X; Y) = H(X)$

**经验** 2：若 $X$ 和 $Y$ 相互独立：$I(X; Y) = 0$

# 互信息

**定义**：两个联合分布的随机变量 $X, Y$ 的互信息 $I(X; Y)$ 定义为

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

# 互信息

**定义**：两个联合分布的随机变量 $X, Y$ 的互信息 $I(X;Y)$ 定义为

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

- $I(X;Y) =$ “$X$ 的不确定性” $-$ “给定 $Y$ 条件下 $X$ 的不确定性”

# 互信息

**定义**：两个联合分布的随机变量 $X, Y$ 的互信息 $I(X;Y)$ 定义为

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

- $I(X;Y) =$ "$X$ 的不确定性" $-$ "给定 $Y$ 条件下 $X$ 的不确定性"
- $I(X;Y)$ 表示：给定 $Y$ 条件下 $X$ 减少的不确定性

# 互信息

**定义**：两个联合分布的随机变量 $X, Y$ 的互信息 $I(X;Y)$ 定义为

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

- $I(X;Y) =$ "$X$ 的不确定性" − "给定 $Y$ 条件下 $X$ 的不确定性"
- $I(X;Y)$ 表示：给定 $Y$ 条件下 $X$ 减少的不确定性
- $I(X;Y)$ 表示：$Y$ 携带关于 $X$ 的信息

# 互信息

**定义**：两个联合分布的随机变量 $X, Y$ 的互信息 $I(X; Y)$ 定义为

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

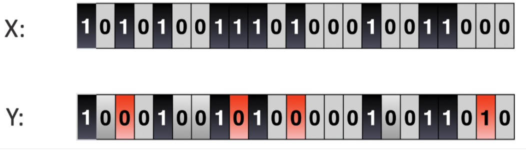
- $I(X;Y) =$ “$X$ 的不确定性” $-$ “给定 $Y$ 条件下 $X$ 的不确定性”
- $I(X;Y)$ 表示：给定 $Y$ 条件下 $X$ 减少的不确定性
- $I(X;Y)$ 表示：$Y$ 携带关于 $X$ 的信息

$$I(X;Y) = \mathbb{E}_{X,Y}\left\{\log\frac{p_{X|Y}}{p_X}\right\} = \mathbb{E}_{X,Y}\left\{\log\frac{p_{X,Y}}{p_X p_Y}\right\}.$$

# 理解互信息

**互信息 $I(X;Y)$ 表示：$Y$ 携带关于 $X$ 的信息**

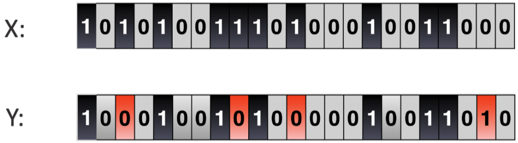
- $X$ 和 $Y$ 相似——互信息较大



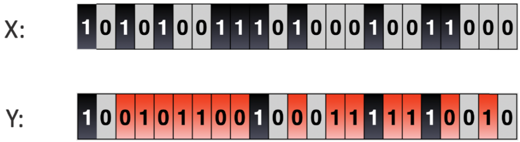
# 理解互信息

**互信息** $I(X;Y)$ **表示：** $Y$ **携带关于** $X$ **的信息**

- $X$ 和 $Y$ 相似——互信息较大




- $X$ 和 $Y$ 独立——互信息为 0



# 理解互信息

## 互信息的性质

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

# 理解互信息

## 互信息的性质

- $I(X;Y) = I(Y;X)$
- $I(X;Y) \geq 0$

# 理解互信息

## 互信息的性质

- $I(X;Y) = I(Y;X)$
- $I(X;Y) \geq 0$
- $I(X;Y) = H(Y) - H(Y|X)$

# 理解互信息

## 互信息的性质

- $I(X;Y) = I(Y;X)$
- $I(X;Y) \geq 0$
- $I(X;Y) = H(Y) - H(Y|X)$
- $I(X;Y) = H(X) + H(Y) - H(X,Y)$

# 理解互信息

## 互信息的性质

- $I(X;Y) = I(Y;X)$
- $I(X;Y) \geq 0$
- $I(X;Y) = H(Y) - H(Y|X)$
- $I(X;Y) = H(X) + H(Y) - H(X,Y)$
- $I(X;Y) \leq H(X)$

# 理解互信息

## 互信息的性质

- $I(X;Y) = I(Y;X)$
- $I(X;Y) \geq 0$
- $I(X;Y) = H(Y) - H(Y|X)$
- $I(X;Y) = H(X) + H(Y) - H(X,Y)$
- $I(X;Y) \leq H(X)$
- $I(X;Y) \leq H(Y)$

# 理解互信息

## 互信息的性质

- $I(X;Y) = I(Y;X)$
- $I(X;Y) \geq 0$
- $I(X;Y) = H(Y) - H(Y|X)$
- $I(X;Y) = H(X) + H(Y) - H(X,Y)$
- $I(X;Y) \leq H(X)$
- $I(X;Y) \leq H(Y)$
- $I(X;X) = H(X)$（自信息）

# 理解互信息

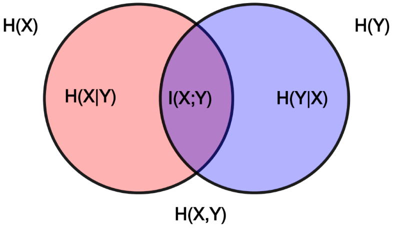
## 互信息的性质

- $I(X;Y) = I(Y;X)$
- $I(X;Y) \geq 0$
- $I(X;Y) = H(Y) - H(Y|X)$
- $I(X;Y) = H(X) + H(Y) - H(X,Y)$
- $I(X;Y) \leq H(X)$
- $I(X;Y) \leq H(Y)$
- $I(X;X) = H(X)$（自信息）
- $X$ 与 $Y$ 相互独立：$I(X;Y) = 0$

# 理解互信息

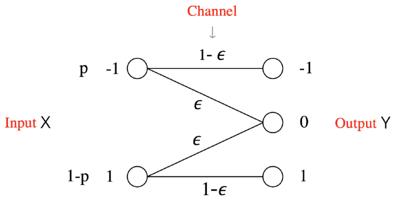
## 互信息的性质

- $I(X;Y) = I(Y;X)$
- $I(X;Y) \geq 0$
- $I(X;Y) = H(Y) - H(Y|X)$
- $I(X;Y) = H(X) + H(Y) - H(X,Y)$
- $I(X;Y) \leq H(X)$
- $I(X;Y) \leq H(Y)$
- $I(X;X) = H(X)$（自信息）
- $X$ 与 $Y$ 相互独立：$I(X;Y) = 0$

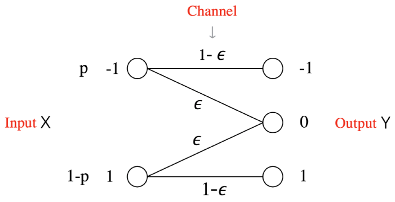



Venn 图：理解互信息的定义和性质

# 二元擦除信道（BEC）的互信息

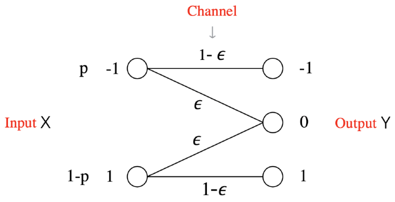


# 二元擦除信道（BEC）的互信息




- 易求：$p_Y = \left\{ p(1-\epsilon), \quad \epsilon, \quad (1-p)(1-\epsilon) \right\}$

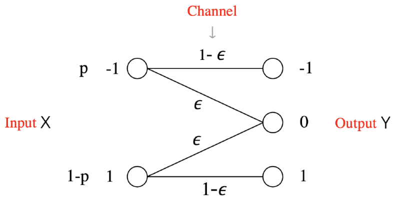
# 二元擦除信道（BEC）的互信息




- 易求：$p_Y = \Big\{ p(1-\epsilon), \quad \epsilon, \quad (1-p)(1-\epsilon) \Big\}$

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

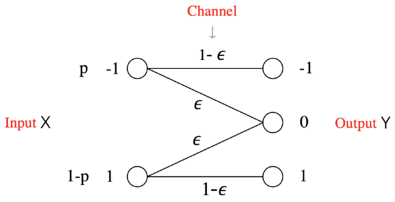
# 二元擦除信道（BEC）的互信息




- 易求：$p_Y = \Big\{ p(1-\epsilon), \quad \epsilon, \quad (1-p)(1-\epsilon) \Big\}$

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &= (1-\epsilon)h(p) + h(\epsilon) - h(\epsilon) \end{aligned}$$

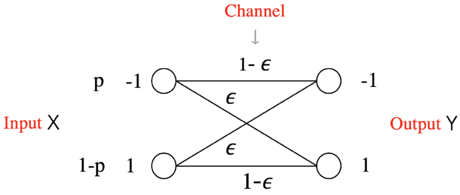
# 二元擦除信道（BEC）的互信息



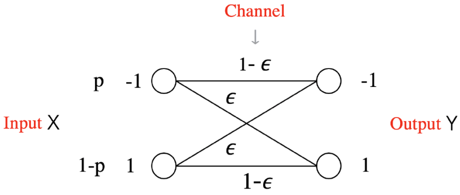

- 易求：$p_Y = \left\{ p(1-\epsilon), \quad \epsilon, \quad (1-p)(1-\epsilon) \right\}$

$$
\begin{aligned}
I(X;Y) &= H(Y) - H(Y|X) \\
&= (1-\epsilon)h(p) + h(\epsilon) - h(\epsilon) \\
&= (1-\epsilon)h(p)
\end{aligned}
$$

# 二元对称信道（BSC）的互信息

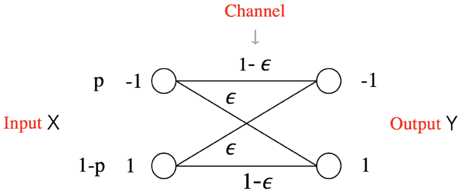


# 二元对称信道（BSC）的互信息



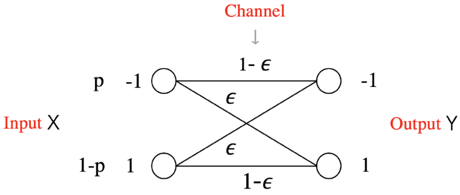

$$p_Y = \{p(1-\epsilon) + (1-p)\epsilon, p\epsilon + (1-p)(1-\epsilon)\}$$

# 二元对称信道（BSC）的互信息




$$
\begin{aligned}
p_Y &= \{p(1-\epsilon) + (1-p)\epsilon, p\epsilon + (1-p)(1-\epsilon)\} \\
&= \{p + \epsilon - 2p\epsilon, 1 - p - \epsilon + 2p\epsilon\}
\end{aligned}
$$

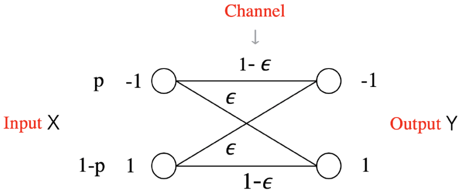
# 二元对称信道（BSC）的互信息




$$p_Y = \{p(1-\epsilon) + (1-p)\epsilon, p\epsilon + (1-p)(1-\epsilon)\}$$
$$= \{p + \epsilon - 2p\epsilon, 1 - p - \epsilon + 2p\epsilon\}$$

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

# 二元对称信道（BSC）的互信息




$$
\begin{aligned}
p_Y &= \{p(1-\epsilon) + (1-p)\epsilon, p\epsilon + (1-p)(1-\epsilon)\} \\
&= \{p + \epsilon - 2p\epsilon, 1 - p - \epsilon + 2p\epsilon\}
\end{aligned}
$$

$$
\begin{aligned}
I(X;Y) &= H(Y) - H(Y|X) \\
&= h\big(p + \epsilon - 2p\epsilon\big) - h(\epsilon)
\end{aligned}
$$

# 条件互信息和链式法则

- **条件互信息**：给定 $Z$ 的条件下，$X$ 和 $Y$ 互信息定义为

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

# 条件互信息和链式法则

- **条件互信息**：给定 $Z$ 的条件下，$X$ 和 $Y$ 互信息定义为

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

- **链式法则**：

$$I(X;Y,Z) = I(X;Z) + I(X;Y|Z)$$

# 条件互信息和链式法则

- **条件互信息**：给定 $Z$ 的条件下，$X$ 和 $Y$ 互信息定义为

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

- **链式法则**：

$$I(X;Y,Z) = I(X;Z) + I(X;Y|Z)$$

**证明**：

$$I(X;Y,Z) = \quad H(X) - H(X|Y,Z)$$

# 条件互信息和链式法则

- **条件互信息**：给定 $Z$ 的条件下，$X$ 和 $Y$ 互信息定义为

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

- **链式法则**：

$$I(X;Y,Z) = I(X;Z) + I(X;Y|Z)$$

**证明**：

$$\begin{aligned} I(X;Y,Z) &= H(X) - H(X|Y,Z) \\ &= \Big[H(X) - H(X|Z)\Big] + \Big[H(X|Z) - H(X|Y,Z)\Big] \end{aligned}$$

# 条件互信息和链式法则

- **条件互信息**：给定 $Z$ 的条件下，$X$ 和 $Y$ 互信息定义为

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

- **链式法则**：

$$I(X;Y,Z) = I(X;Z) + I(X;Y|Z)$$

**证明**：

$$\begin{aligned} I(X;Y,Z) &= H(X) - H(X|Y,Z) \\ &= \Big[H(X) - H(X|Z)\Big] + \Big[H(X|Z) - H(X|Y,Z)\Big] \\ &= I(X;Z) + I(X;Y|Z) \end{aligned}$$

# 条件互信息和链式法则

- **条件互信息**：给定 $Z$ 的条件下，$X$ 和 $Y$ 互信息定义为

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

- **链式法则**：

$$I(X;Y,Z) = I(X;Z) + I(X;Y|Z)$$

**证明**：

$$\begin{aligned} I(X;Y,Z) = & \quad H(X) - H(X|Y,Z) \\ = & \quad \Big[H(X) - H(X|Z)\Big] + \Big[H(X|Z) - H(X|Y,Z)\Big] \\ = & \quad I(X;Z) + I(X;Y|Z) \end{aligned}$$

- $I(X_1, X_2, \cdots, X_n; Y) = I(X_1; Y) + \sum_{i=1}^{n} I(X_i; Y|X_{i-1}, \cdots, X_1)$

# 互信息的凹凸性

**命题**：令 $(X, Y) \sim p_{X,Y} = p_X p_{Y|X}$，

1. 给定 $p_{Y|X}$，$I(X;Y)$ 是 $\{p_x\}$ 的上凸函数；
2. 给定 $p_X$，$I(X;Y)$ 是 $\{p_{y|x}\}$ 的下凸函数。

# 互信息的凹凸性

**命题**：令 $(X, Y) \sim p_{X,Y} = p_X p_{Y|X}$，

1. 给定 $p_{Y|X}$，$I(X; Y)$ 是$\{p_x\}$的上凸函数；
2. 给定 $p_X$，$I(X; Y)$ 是$\{p_{y|x}\}$的下凸函数。

**证明**：

1. $I(X; Y) = H(Y) - H(Y|X) = H(Y) - \sum_x p_x H(Y|x)$

   其中 $H(Y)$关于 $\{p_y\}$ 上凸，给定 $\{p_{y|x}\}$ 时 $\{p_y\}$ 为$\{p_x\}$ 的线性函数

# 互信息的凹凸性

**命题**：令 $(X, Y) \sim p_{X,Y} = p_X p_{Y|X}$，

1. 给定 $p_{Y|X}$，$I(X;Y)$ 是$\{p_x\}$的上凸函数；
2. 给定 $p_X$，$I(X;Y)$ 是$\{p_{y|x}\}$的下凸函数。

**证明**：

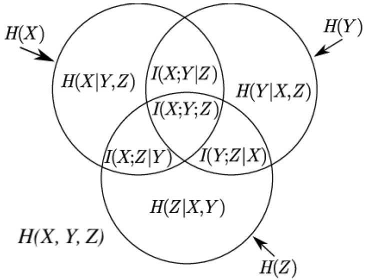
1. $I(X;Y) = H(Y) - H(Y|X) = H(Y) - \sum_x p_x H(Y|x)$

   其中 $H(Y)$关于 $\{p_y\}$ 上凸，给定 $\{p_{y|x}\}$ 时 $\{p_y\}$ 为$\{p_x\}$ 的线性函数

2. $I(X;Y) = \sum_x p_x D_x(p_{y|x} \| p_y)$

   其中 $D_x(p_{y|x} \| p_y)$关于 $\{p_{y|x}, p_y\}$ 下凸（见下节），给定 $p_X$ 时 $\{p_y\}$ 是关于 $\{p_{y|x}\}$ 的线性函数

# 理解条件互信息及其性质




$X, Y, Z$ 的 Venn 图

利用 Venn 图理解

- 熵，条件熵，联合熵
- 互信息，条件互信息

# 互信息和散度

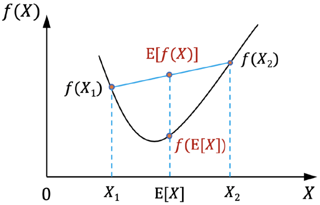
# 相对熵/KL 散度

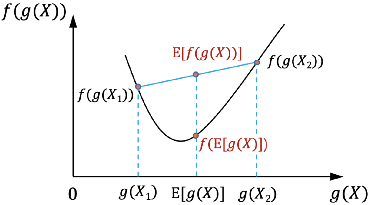
**定义**：概率分布 $p_X$ 和 $q_X$ 的相对熵或 KL 散度 $D(p_x\|q_x)$ 或 $(D(p\|q))$ 为

$$D(p\|q) = \sum_{x\in\mathcal{X}} p_x \log\frac{p_x}{q_x} = \mathrm{E}_X\left\{\log\frac{p_x}{q_x}\right\}$$

注：$D(p\|q)$ 表示**真实分布**$p_X$ 和**近似分布**$q_X$的**距离**。

# Jensen 不等式

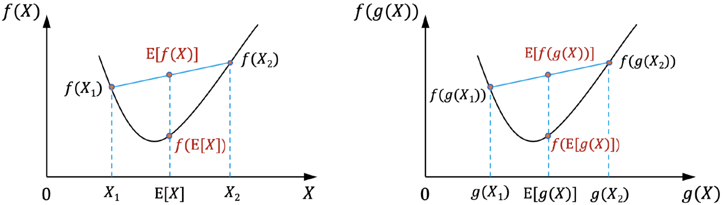



1. Jensen **不等式**：令 $f$ 为一个下凸函数，$X$ 为一个随机变量，那么

$$\mathrm{E}[f(X)] \geq f(\mathrm{E}[X])$$

# Jensen 不等式




1. Jensen **不等式**：令 $f$ 为一个下凸函数，$X$ 为一个随机变量，那么

$$\mathrm{E}[f(X)] \geq f(\mathrm{E}[X])$$

2. **广义** Jensen **不等式**：令 $f$ 为一个下凸函数，$X$ 为一个随机变量，那么

$$\mathrm{E}[f(g(X))] \geq f\big(\mathrm{E}[g(X)]\big)$$

# KL 散度的非负性

- 广义 Jensen 不等式：令 $f$ 为一个下凸函数，$X$ 为一个随机变量，那么

$$\mathrm{E}[f(g(X))] \geq f\big(\mathrm{E}[g(X)]\big)$$

# KL 散度的非负性

- 广义 Jensen 不等式：令 $f$ 为一个下凸函数，$X$ 为一个随机变量，那么

$$\mathrm{E}[f(g(X))] \geq f\big(\mathrm{E}[g(X)]\big)$$

- 非负性：$D(p\|q) \geq 0$

# KL 散度的非负性

- 广义 Jensen 不等式：令 $f$ 为一个下凸函数，$X$ 为一个随机变量，那么

$$\mathrm{E}[f(g(X))] \geq f\big(\mathrm{E}[g(X)]\big)$$

- 非负性：$D(p\|q) \geq 0$

**证明**：

$$D(p\|q) = -\sum_{x \in \mathcal{X}} p_x \log \frac{q_x}{p_x} = \mathrm{E}_X[f(g(X))]$$

# KL 散度的非负性

- 广义 Jensen 不等式：令 $f$ 为一个下凸函数，$X$ 为一个随机变量，那么

$$\mathrm{E}[f(g(X))] \geq f(\mathrm{E}[g(X)])$$

- 非负性：$D(p\|q) \geq 0$

**证明**：

$$D(p\|q) = -\sum_{x\in\mathcal{X}} p_x \log \frac{q_x}{p_x} = \mathrm{E}_X[f(g(X))]$$

其中 $g(x) = \frac{q_x}{p_x}$，且 $f(\cdot) = -\log(\cdot)$ 是下凸函数，根据广义 Jensen 不等式，

$$\mathrm{E}_X[f(g(X))] \geq f(\mathrm{E}_X[g(X)]) = \log 1 = 0$$

# KL 散度的性质

- 一般 $D(p\|q) \neq D(q\|p)$，因此 KL 散度不是真正的距离。

  **例**：$p_X = \{r, 1-r\}, q(X) = \{s, 1-s\}$.

# KL 散度的性质

- 一般 $D(p\|q) \neq D(q\|p)$，因此 KL 散度不是真正的距离。

  **例**：$p_X = \{r, 1-r\}, q(X) = \{s, 1-s\}$.

- 与互信息的关系

$$I(X;Y) = D\big(p_{x,y}\|p_x p_y\big)$$

# KL 散度的性质

- 一般 $D(p\|q) \neq D(q\|p)$，因此 KL 散度不是真正的距离。

  **例**：$p_X = \{r, 1-r\}, q(X) = \{s, 1-s\}$.

- 与互信息的关系

$$I(X;Y) = D\big(p_{x,y}\|p_x p_y\big)$$

- 条件 KL 散度

$$D\left(p_{y|x}\|q_{y|x}\right) = \sum_{x\in\mathcal{X}} p_x \sum_{y\in\mathcal{Y}} p_{y|x} \log \frac{p_{y|x}}{q_{y|x}}$$

# KL 散度的性质

- 一般 $D(p\|q) \neq D(q\|p)$，因此 KL 散度不是真正的距离。

  **例**：$p_X = \{r, 1-r\}, q(X) = \{s, 1-s\}$.

- 与互信息的关系

$$I(X;Y) = D\big(p_{x,y} \| p_x p_y\big)$$

- 条件 KL 散度

$$D\left(p_{y|x} \| q_{y|x}\right) = \sum_{x \in \mathcal{X}} p_x \sum_{y \in \mathcal{Y}} p_{y|x} \log \frac{p_{y|x}}{q_{y|x}}$$

- KL 散度的链式法则

$$D\big(p_{x,y} \| q_{x,y}\big) = D\big(p_x \| q_x\big) + D\big(p_{y|x} \| q_{y|x}\big)$$

# KL 散度的性质

- 一般 $D(p\|q) \neq D(q\|p)$，因此 KL 散度不是真正的距离。

  **例**：$p_X = \{r, 1-r\}, q(X) = \{s, 1-s\}$.

- 与互信息的关系

$$I(X;Y) = D\big(p_{x,y} \| p_x p_y\big)$$

- 条件 KL 散度

$$D\left(p_{y|x} \| q_{y|x}\right) = \sum_{x \in \mathcal{X}} p_x \sum_{y \in \mathcal{Y}} p_{y|x} \log \frac{p_{y|x}}{q_{y|x}}$$

- KL 散度的链式法则

$$D\big(p_{x,y} \| q_{x,y}\big) = D\big(p_x \| q_x\big) + D\big(p_{y|x} \| q_{y|x}\big)$$

- $D(p\|q)$ 关于 $(p,q)$ 是上凸函数，即，对于任意 $0 \leq \lambda \leq 1$,

$$D\big(\lambda p_1 + (1-\lambda) p_2 \| \lambda q_1 + (1-\lambda) q_2\big) \leq \lambda D\big(p_1 \| q_1\big) + (1-\lambda) D\big(p_2 \| q_2\big).$$

# 互信息和散度

1. 互信息
2. KL-散度
3. 马尔可夫链
4. 数据处理定理
5. 费诺不等式

# 马尔可夫链

**定义**: $X \to Y \to Z$ 为一个马尔可夫链，当且仅当

$$p_{x,y,z} = p_x p_{y|x} q_{z|y}$$

# 马尔可夫链

**定义**： $X \to Y \to Z$ 为一个马尔可夫链，当且仅当

$$p_{x,y,z} = p_x p_{y|x} q_{z|y}$$

- $X \to Y \to Z \iff$ 给定 $Y$ 时，$X$ 和 $Z$ 条件独立，即

$$q_{x,z|y} = q_{x|y} q_{z|y}$$

# 马尔可夫链

**定义**： $X \to Y \to Z$ 为一个马尔可夫链，当且仅当

$$p_{x,y,z} = p_x p_{y|x} q_{z|y}$$

- $X \to Y \to Z \iff$ 给定 $Y$ 时，$X$ 和 $Z$ 条件独立，即

$$q_{x,z|y} = q_{x|y} q_{z|y}$$

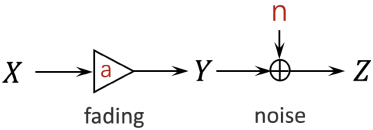
- $q_{z|x,y} = q_{z|y} \iff H(Z|X,Y) = H(Z|Y)$

"**未来**$Z$ 取决于**现在**$Y$ 而非**过去**$X$"

**注**：对于 $X \to Y \to Z$，$X$ 和 $Z$ 并不独立。

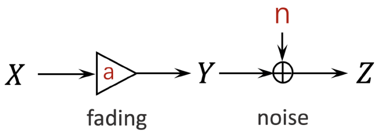
# 马尔可夫链举例

- 衰落信道：$X \to$ “$Y = aX$” $\to$ “$Z = Y + n$”

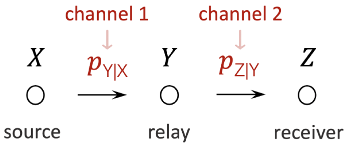


# 马尔可夫链举例

- 衰落信道：$X \to$ “$Y = aX$” $\to$ “$Z = Y + n$”




- 中继信道：“$X \to Y \to Z$”



# 马尔可夫链的性质

- $X \to Y \to Z \;\Rightarrow\; Z \to Y \to X$

  马尔可夫链也可表示为 $X \leftrightarrow Y \leftrightarrow Z$.

# 马尔可夫链的性质

- $X \to Y \to Z \ \Rightarrow \ Z \to Y \to X$
  马尔可夫链也可表示为 $X \leftrightarrow Y \leftrightarrow Z.$

- $X \to Y \to g(Y)$ 组成一个马尔可夫链

# 马尔可夫链的性质

- $X \to Y \to Z \;\Rightarrow\; Z \to Y \to X$

  马尔可夫链也可表示为 $X \leftrightarrow Y \leftrightarrow Z$.

- $X \to Y \to g(Y)$ 组成一个马尔可夫链

- $X \to Y \to Z \Rightarrow I(X;Z|Y) = 0$

# 马尔可夫链的性质

- $X \to Y \to Z \;\Rightarrow\; Z \to Y \to X$

  马尔可夫链也可表示为 $X \leftrightarrow Y \leftrightarrow Z$.

- $X \to Y \to g(Y)$ 组成一个马尔可夫链

- $X \to Y \to Z \Rightarrow I(X;Z|Y) = 0$

**提示**：从 Venn 图理解。

# 互信息和散度

1. 互信息
2. KL-散度
3. 马尔可夫链
4. **数据处理定理**
5. 费诺不等式

# 数据处理定理

**数据处理不等式**：如果 $X \to Y \to Z$, 那么

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

# 数据处理定理

**数据处理不等式：** 如果 $X \to Y \to Z$, 那么

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

**证明：** 根据链式法则，

$$I(X;Y,Z) = I(X;Z) + \underbrace{I(X;Y|Z)}_{\geq 0} = I(X;Y) + \underbrace{I(X;Z|Y)}_{=0}.$$

# 数据处理定理

**数据处理不等式**：如果 $X \to Y \to Z$, 那么

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

**证明**：根据链式法则，

$$I(X;Y,Z) = I(X;Z) + \underbrace{I(X;Y|Z)}_{\geq 0} = I(X;Y) + \underbrace{I(X;Z|Y)}_{=0}.$$

**推论**：

- $I(X;Y) \geq I(X;Y|Z)$, 因为 $I(X;Z) \geq 0$（马尔可夫链的性质）

# 数据处理定理

**数据处理不等式**：如果 $X \to Y \to Z$, 那么

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

**证明**：根据链式法则，

$$I(X;Y,Z) = I(X;Z) + \underbrace{I(X;Y|Z)}_{\geq 0} = I(X;Y) + \underbrace{I(X;Z|Y)}_{=0}.$$

**推论**：

- $I(X;Y) \geq I(X;Y|Z)$, 因为 $I(X;Z) \geq 0$（马尔可夫链的性质）
- $I(X;Y) \geq I(X;g(Y))$

# 数据处理定理

**数据处理不等式**：如果 $X \to Y \to Z$, 那么

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

**证明**：根据链式法则，

$$I(X;Y,Z) = I(X;Z) + \underbrace{I(X;Y|Z)}_{\geq 0} = I(X;Y) + \underbrace{I(X;Z|Y)}_{=0}.$$

**推论**：

- $I(X;Y) \geq I(X;Y|Z)$, 因为 $I(X;Z) \geq 0$（马尔可夫链的性质）
- $I(X;Y) \geq I(X;g(Y))$
- $H(X|Y) \leq H(X|Z)$, $H(X|Y) \leq H(X|g(Y))$

# 数据处理定理

**数据处理不等式**：如果 $X \to Y \to Z$, 那么

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

**证明**：根据链式法则，

$$I(X;Y,Z) = I(X;Z) + \underbrace{I(X;Y|Z)}_{\geq 0} = I(X;Y) + \underbrace{I(X;Z|Y)}_{=0}.$$

**推论**：

- $I(X;Y) \geq I(X;Y|Z)$, 因为 $I(X;Z) \geq 0$（马尔可夫链的性质）
- $I(X;Y) \geq I(X;g(Y))$
- $H(X|Y) \leq H(X|Z)$, $H(X|Y) \leq H(X|g(Y))$

借助 Venn 图理解！

# 互信息和散度

1. 互信息
2. KL-散度
3. 马尔可夫链
4. 数据处理定理
5. 费诺不等式

# 费诺不等式

**考虑估计问题**： $\hat{X} = g(Y)$，即 $X \to Y \to \hat{X} = g(Y)$，我们关心

$$p_e = \Pr\{\hat{X} \neq X\}$$

# 费诺不等式

**考虑估计问题**：$\hat{X} = g(Y)$，即 $X \to Y \to \hat{X} = g(Y)$，我们关心

$$p_e = \Pr\{\hat{X} \neq X\}$$

- 当 $H(X|\hat{X}) = 0$，即 $X = g(Y)$ 时，$p_e = 0$

# 费诺不等式

**考虑估计问题**： $\hat{X} = g(Y)$，即 $X \to Y \to \hat{X} = g(Y)$，我们关心

$$p_e = \Pr\{\hat{X} \neq X\}$$

- 当 $H(X|\hat{X}) = 0$，即 $X = g(Y)$ 时，$p_e = 0$
- 直觉上，$H(X|Y)$ 越小，$p_e$ 也应该越小

# 费诺不等式

考虑估计问题：$\hat{X} = g(Y)$，即 $X \to Y \to \hat{X} = g(Y)$，我们关心

$$p_e = \Pr\{\hat{X} \neq X\}$$

- 当 $H(X|\hat{X}) = 0$，即 $X = g(Y)$ 时，$p_e = 0$
- 直觉上，$H(X|Y)$ 越小，$p_e$ 也应该越小

问：$H(X|Y)$ 和 $p_e$ 的关系？

# 费诺不等式

- **费诺不等式**：令 $X \to Y \to \hat{X}, p_e = \Pr\{\hat{X} \neq X\}$, 那么

$$\underbrace{h(p_e)}_{\leq 1} + p_e \log|\mathcal{X}| \geq H(X|\hat{X}) \geq H(X|Y),$$

# 费诺不等式

- **费诺不等式**：令 $X \to Y \to \hat{X}, p_e = \Pr\{\hat{X} \neq X\}$, 那么

$$\underbrace{h(p_e)}_{\leq 1} + p_e \log|\mathcal{X}| \geq H(X|\hat{X}) \geq H(X|Y),$$

也可进一步弱化为

$$1 + p_e \log|\mathcal{X}| \geq H(X|Y),$$

# 费诺不等式

- **费诺不等式**：令 $X \to Y \to \hat{X}, p_e = \Pr\{\hat{X} \neq X\}$, 那么

$$\underbrace{h(p_e)}_{\leq 1} + p_e \log|\mathcal{X}| \geq H(X|\hat{X}) \geq H(X|Y),$$

也可进一步弱化为

$$1 + p_e \log|\mathcal{X}| \geq H(X|Y),$$

或者

$$p_e \geq \frac{H(X|Y) - 1}{\log|\mathcal{X}|}.$$

# 费诺不等式

- **费诺不等式**：令 $X \to Y \to \hat{X}, p_e = \Pr\{\hat{X} \neq X\}$, 那么

$$\underbrace{h(p_e)}_{\leq 1} + p_e \log|\mathcal{X}| \geq H(X|\hat{X}) \geq H(X|Y),$$

也可进一步弱化为

$$1 + p_e \log|\mathcal{X}| \geq H(X|Y),$$

或者

$$p_e \geq \frac{H(X|Y) - 1}{\log|\mathcal{X}|}.$$

- 费诺不等式给了错误概率的一个下界 (能好到什么程度)

# 费诺不等式

- **费诺不等式**：令 $X \to Y \to \hat{X}, p_e = \Pr\{\hat{X} \neq X\}$, 那么

$$\underbrace{h(p_e)}_{\leq 1} + p_e \log|\mathcal{X}| \geq H(X|\hat{X}) \geq H(X|Y),$$

也可进一步弱化为

$$1 + p_e \log|\mathcal{X}| \geq H(X|Y),$$

或者

$$p_e \geq \frac{H(X|Y) - 1}{\log|\mathcal{X}|}.$$

- 费诺不等式给了错误概率的一个下界 (能好到什么程度)
- $p_e = 0 \Rightarrow H(X|Y) = 0$

# 费诺不等式

- **费诺不等式**：令 $X \to Y \to \hat{X}, p_e = \Pr\{\hat{X} \neq X\}$, 那么

$$\underbrace{h(p_e)}_{\leq 1} + p_e \log|\mathcal{X}| \geq H(X|\hat{X}) \geq H(X|Y),$$

也可进一步弱化为

$$1 + p_e \log|\mathcal{X}| \geq H(X|Y),$$

或者

$$p_e \geq \frac{H(X|Y) - 1}{\log|\mathcal{X}|}.$$

- 费诺不等式给了错误概率的一个下界 (能好到什么程度)
- $p_e = 0 \Rightarrow H(X|Y) = 0$
- 一般 $\hat{\mathcal{X}} \neq \mathcal{X}$，将 $\log|\mathcal{X}|$ 替换为 $\log(|\mathcal{X}| - 1)$ 可做到更好

# 费诺不等式

- **费诺不等式**：令 $X \to Y \to \hat{X}, p_e = \Pr\{\hat{X} \neq X\}$, 那么

$$\underbrace{h(p_e)}_{\leq 1} + p_e \log|\mathcal{X}| \geq H(X|\hat{X}) \geq H(X|Y),$$

也可进一步弱化为

$$1 + p_e \log|\mathcal{X}| \geq H(X|Y),$$

或者

$$p_e \geq \frac{H(X|Y) - 1}{\log|\mathcal{X}|}.$$

- 费诺不等式给了错误概率的一个下界 (能好到什么程度)
- $p_e = 0 \Rightarrow H(X|Y) = 0$
- 一般 $\hat{\mathcal{X}} \neq \mathcal{X}$，将 $\log|\mathcal{X}|$ 替换为 $\log(|\mathcal{X}| - 1)$ 可做到更好
- 费诺不等式对**信道容量逆定理**非常重要

# 费诺不等式的证明

**定义随机变量**

$$E = \begin{cases} 0, & \text{如果} \hat{X} \neq X \\ 1, & \text{如果} \hat{X} = X \end{cases}$$

# 费诺不等式的证明

**定义随机变量**

$$E = \begin{cases} 0, & \text{如果} \hat{X} \neq X \\ 1, & \text{如果} \hat{X} = X \end{cases}$$

- 根据链式法则,

$$H(E, X|\hat{X}) = \quad H(X|\hat{X}) + \underbrace{H(E|\hat{X}, X)}_{=0}$$

# 费诺不等式的证明

**定义随机变量**

$$E = \begin{cases} 0, & \text{如果}\ \hat{X} \neq X \\ 1, & \text{如果}\ \hat{X} = X \end{cases}$$

- 根据链式法则,

$$\begin{aligned} H(E, X|\hat{X}) &= H(X|\hat{X}) + \underbrace{H(E|\hat{X}, X)}_{=0} \\ &= \underbrace{H(E|\hat{X})}_{\leq h(p_e)} + \underbrace{H(X|\hat{X}, E)}_{\leq p_e H(X) \leq p_e \log|\mathcal{X}|} \end{aligned}$$

# 费诺不等式的证明

**定义随机变量**

$$E = \begin{cases} 0, & \text{如果}\ \hat{X} \neq X \\ 1, & \text{如果}\ \hat{X} = X \end{cases}$$

- 根据链式法则，

$$\begin{aligned} H(E, X|\hat{X}) = & \quad H(X|\hat{X}) + \underbrace{H(E|\hat{X}, X)}_{=0} \\ = & \quad \underbrace{H(E|\hat{X})}_{\leq h(p_e)} + \underbrace{H(X|\hat{X}, E)}_{\leq p_e H(X) \leq p_e \log|\mathcal{X}|} \end{aligned}$$

因此，

$$H(X|\hat{X}) \leq h(p_e) + p_e \log|\mathcal{X}|$$

# 费诺不等式的证明

**定义随机变量**

$$E = \begin{cases} 0, & \text{如果 } \hat{X} \neq X \\ 1, & \text{如果 } \hat{X} = X \end{cases}$$

- 根据链式法则，

$$\begin{aligned} H(E, X|\hat{X}) = & \quad H(X|\hat{X}) + \underbrace{H(E|\hat{X}, X)}_{=0} \\ = & \quad \underbrace{H(E|\hat{X})}_{\leq h(p_e)} + \underbrace{H(X|\hat{X}, E)}_{\leq p_e H(X) \leq p_e \log|\mathcal{X}|} \end{aligned}$$

因此，

$$H(X|\hat{X}) \leq h(p_e) + p_e \log|\mathcal{X}|$$

- 根据马尔可夫链 $X \to Y \to \hat{X}$,

$$H(X|Y) \leq H(X|\hat{X})$$

# 费诺不等式的性质

- $H(p) + p\log|\mathcal{X}| \geq H(X|Y)$，其中 $p = Pr(Y \neq X)$。(令 $\hat{X} = Y$)

# 费诺不等式的性质

- $H(p) + p\log|\mathcal{X}| \geq H(X|Y)$，其中 $p = Pr(Y \neq X)$。(令 $\hat{X} = Y$)
- 令 $\hat{X}: \mathcal{Y} \to \mathcal{X}$，那么

$$H(X|\hat{X}, E) = p_e H(X|\hat{X}, \hat{X} \neq X) \leq p_e \log(|\mathcal{X}| - 1)$$

因此，$h(p_e) + p_e \log(|\mathcal{X}| - 1) \geq H(X|Y)$。

# 费诺不等式的性质

- $H(p) + p\log|\mathcal{X}| \geq H(X|Y)$，其中 $p = Pr(Y \neq X)$。(令 $\hat{X} = Y$)
- 令 $\hat{X}: \mathcal{Y} \to \mathcal{X}$，那么

  $$H(X|\hat{X}, E) = p_e H(X|\hat{X}, \hat{X} \neq X) \leq p_e \log(|\mathcal{X}| - 1)$$

  因此，$h(p_e) + p_e \log(|\mathcal{X}| - 1) \geq H(X|Y)$。
- 在特定情况下费诺不等式的等号可以成立。例如，令 $Y$ 与 $X$ 独立，$X \in \mathcal{X} = \{0, 1\}$ 且 $p_0 \geq p_1 = 1 - p_0$，则最优估计为 $\hat{X} = 0$。那么，$p_e = 1 - p_0, |\mathcal{X}| = 2$，且

  $$h(p_e) + p_e \log(|\mathcal{X}| - 1) = h(p_0) = H(X) = H(X|Y)$$

# 总结

- **互信息**：一个变量包含的另一个变量的信息
- KL **距离**：两个概率分布的相似程度
- **马尔可夫链**：未来由现在决定，而不是过去
- **数据处理定理**：处理不会增加信息量
- **费诺不等式**：错误概率 $p_e$ 的下界
- **许多重要的性质**
- **理解各种概念的物理意义，借助 Venn 图！**

# 作业

- 复习授课内容
- 预习单信源编码
- 独立完成习题
  - 2.9
  - 2.10
  - 2.12
  - 2.13
  - 2.17