基础信息论

互信息量的物理解释

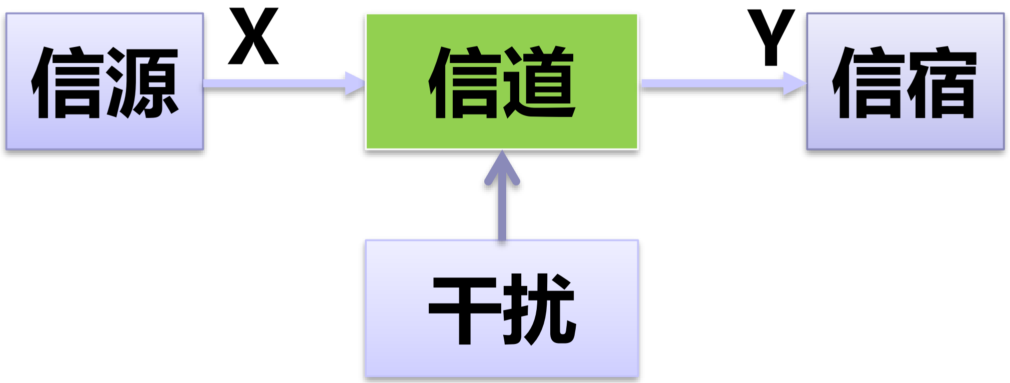
华中科技大学电信学院

学习目标

- 分析互信息量的物理解释

A. 互信息量

简化的通信系统模型



受噪声影响，信源发出的消息在信道传输过程中可能会出现错误。

- 信源

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ p(x_1) & p(x_2) & \cdots & p(x_n) \end{bmatrix}$$

- 信宿

$$\begin{bmatrix} Y \\ P(Y) \end{bmatrix} = \begin{bmatrix} y_1 & y_2 & \cdots & y_m \\ p(y_1) & p(y_2) & \cdots & p(y_m) \end{bmatrix}$$

观察通信过程

一般情况下，信源发出的是消息 x_i，但受噪声影响，接收端收到的消息是 y_j 。

在通信之前，信宿端猜测发出的是 x_i的不确定度：

先验不确定度

$$\underbrace{I(x_i)} = -\log p(x_i) = \log \frac{1}{\underbrace{p(x_i)}}$$

先验概率

在通信之后，信宿端收到 y_j之后，再来猜测信源发出的是x_i的不确定度：

$$\underbrace{I(x_i / y_j)} = -\log p(x_i / y_j) = \log \frac{1}{\underbrace{p(x_i / y_j)}}$$

后验概率

后验不确定度

互信息量定义

- 定义：
  - 后验不确定度，相对于先验不确定度的减少量，为本次通信过程中从收到的 y_j 中获得的关于 x_i 的互信息量。

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

$$=-\log p(x_i)-\left[-\log p(x_i/y_j)\right]$$

$$=\log p(x_i/y_j)-\log p(x_i)=\log\frac{p(x_i/y_j)}{p(x_i)} \quad *$$

注意： $I(x_iy_j)$ $I(x_i/y_j)$ $I(x_i;y_j)$ **容易混淆**

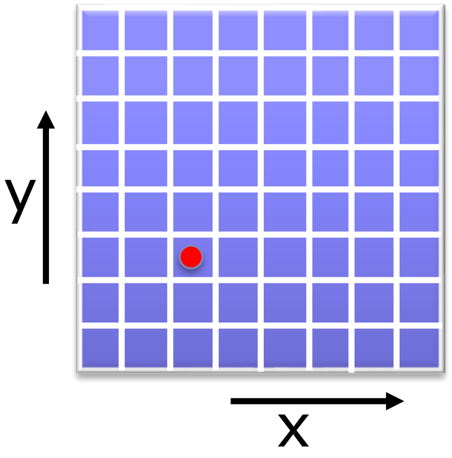
例题

- 甲在一个8×8的方格棋盘上随意放入一个棋子，棋子所放位置是等概率的。
- (1) 若甲告知乙，棋子落入方格的行号，这时乙得到了多少信息量?
- (2) 若甲将棋子落入方格的行号和列号都告知乙，这时乙得到了多少信息量?

解：(1) 互信息量 = 先验不确定度—后验不确定度

$$= -\log\frac{1}{64} - [-\log\frac{1}{8}] = 3 \text{ 比特}$$

(2) 互信息量 $= -\log\frac{1}{64} - 0 = 6$ **比特**

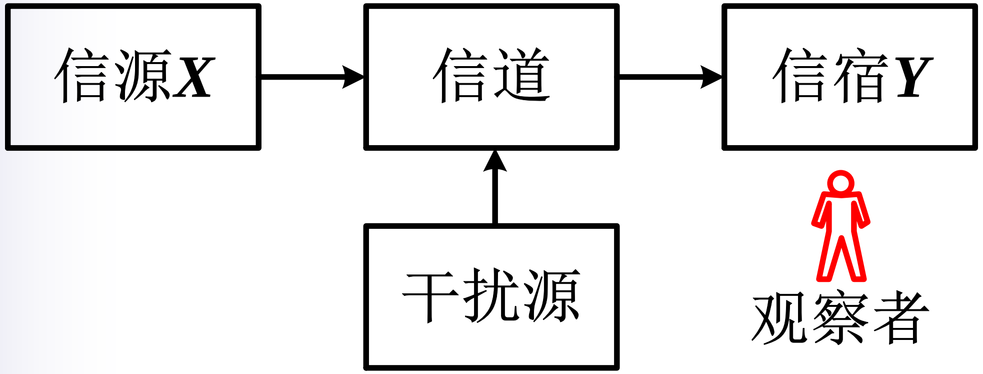


互信息的三种形式1

1. 第一种形式

$$I(x_i;y_j)=I(x_i)-I(x_i/y_j)=\log\frac{p(x_i/y_j)}{p(x_i)}$$ *

物理意义：



观察者站在信宿端。通信后，从y_j获得的关于x_i的信息量。

通信前问：这次信源发送的会是x_i吗?

回答：不确定度 $I(x_i)$

通信后问：（收到的是y_j）这次信源发送的会是x_i吗?

回答：不确定度 $I(x_i/y_j)$

互信息的三种形式2

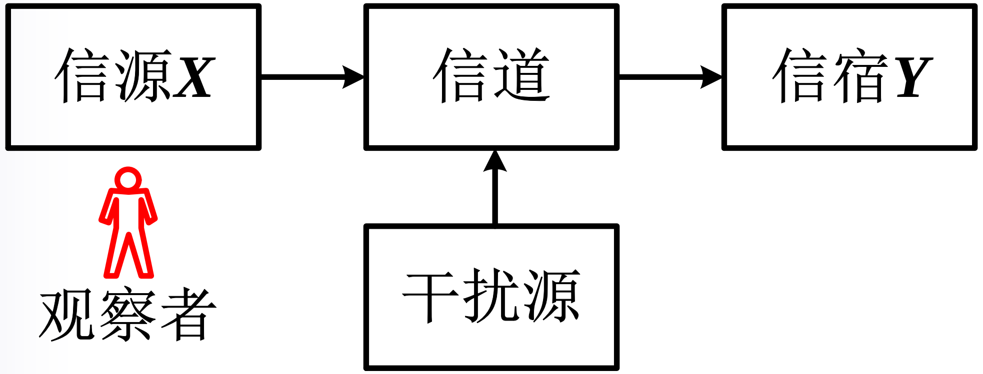
2. 第二种形式

$$I(y_j;x_i)=I(y_j)-I(y_j/x_i)=\log\frac{p(y_j/x_i)}{p(y_j)}$$ *

第一种形式：从y_j获得的关于x_i的信息量。

第二种形式：从x_i获得的关于y_j的信息量。

物理意义：



观察者站在信源端。通信后，从x_i获得的关于y_j的信息量。

通信前问：这次信宿收到的会是y_j吗?

回答：不确定度 $I(y_j)$

通信后问：（发送的是x_i）这次信宿收到的会是y_j吗?

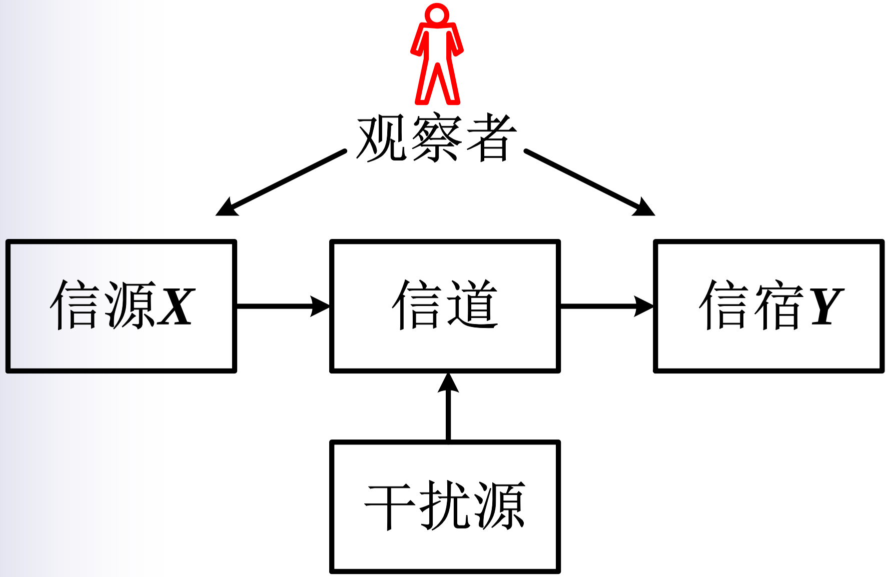
回答：不确定度 $I(y_j/x_i)$

互信息的三种形式3

3. 第三种形式

第一/二种形式：观察者站在信宿/信源端。

第三种形式：观察者站在系统整体（宏观角度观察）。



通信前：x_i和y_j相互独立

通信后：x_i和y_j相互关联

通信前问：这次信源发送的会是x_i吗？信宿收到的会是y_j吗？

回答：不确定度 $I(x_i)+I(y_j)$

通信后问：这次信源发送的会是x_i吗？信宿收到的会是y_j吗？

回答：不确定度 $I(x_i y_j)$

互信息的三种形式3(续)

第三种形式的物理意义：

观察者站在系统整体进行宏观观察。
通信前的整体不确定度： $I(x_i)+I(y_j)$
通信后的整体不确定度： $I(x_i y_j)$
通信后获得的信息量为整体不确定度的减少。

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

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

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

*

B. 互信息的性质1

1. 对称性（互易性）

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

证:
$$I(x_i;y_j)=\log\frac{p(x_i/y_j)\cdot p(y_j)}{p(x_i)\cdot p(y_j)}=\log\frac{p(x_iy_j)}{p(x_i)\cdot p(y_j)}$$

第一种形式：
信宿端

$$=\log\frac{p(y_j/x_i)}{p(y_j)}=I(y_j;x_i)$$

第二种形式：
信源端

物理意义：

信源 信宿 “你中有我，我中有你”。
事件y_j 提供的有关于事件x_i 的信息量等于由事件x_i 提供的关于事件y_j 信息量

互信息的性质2

- **2. 互信息量可为0**

当X、Y独立时，

$$I(x_i;y_j)=\log\frac{p(x_iy_j)}{p(x_i)\cdot p(y_j)}=\log\frac{p(x_i)\cdot p(y_j)}{p(x_i)\cdot p(y_j)}=0$$

第三种形式：全局

物理意义：

当X、Y独立时，从 y_j 中得不到关于x_i的任何信息。

互信息的性质3

3. 互信息量可负

对比：自信息量为非负值

解释：$$I(x_i;y_j)=\log\frac{p(x_i/y_j)}{p(x_i)}$$ 后验概率（分子） 先验概率（分母）

当 $p(x_i/y_j)<p(x_i)$ 时，$I(x_i;y_j)<0$

所表达的含义：$$I(x_i;y_j)=\underbrace{-\log p(x_i)}_{\text{先验不确定度}}-\underbrace{[-\log p(x_i/y_j)]}_{\text{后验不确定度}}$$

先验不确定度 < 后验不确定度

物理意义：

- 当收到 y_j 后，对 x_i 是否会发生的不确定度不仅没有减少，反而还增加了。
- 这通常是由于通信过程中出现传输错误或受到干扰所引起的。

C. 条件互信息量和联合互信息量

- 条件互信息量：在随机事件 z_k 已经发生的条件下，随机事件 y_j 发生后，所间接提供的随机事件 x_i 的信息（不确定度的消减）
- 数学表达式

$$I(x_i;y_j/z_k)=\log\frac{p(x_i/y_jz_k)}{p(x_i/z_k)}$$

- 联合互信息量：联合事件 $y_j z_k$ 发生后，所间接提供的另一个随机事件 x_i 的信息（不确定度的消减）
- 数学表达式

$$I(x_i;y_jz_k)=\log\frac{p(x_i/y_jz_k)}{p(x_i)}$$

三个互信息量的关系

- 互信息量、联合事件互信息量、条件互信息量三者都是随机变量，其值随着变量x_i， y_j , z_k的变化而变化。
- 三者关系为：

$$I(x_i;y_jz_k)=I(x_i;y_j)+I(x_i;z_k/y_j)$$

- 说明：
- 一联合事件 $y_j z_k$ 出现后所提供的有关 x_i 的信息量等于事件 y_j 出现后所提供的有关 x_i 的信息量加上在给定事件 y_j 的条件下再出现事件 z_k 所提供的有关 x_i 的信息量。

例

- 某人A预先知道他的三位朋友B、C、D中必定将有一人晚上到他家来，并且这三人来的可能性均相同
  - 其先验概率为：$p(B)=p(C)=p(D)=1/3$
- 但是上午A接到D的电话不能来了
  - 把这次电话作为事件E，那么有后验概率$p(D/E)=0,p(B/E)=p(C/E)=1/2$
- 下午A又接到C的电话，说晚上开会不能来
  - 把这次电话作为事件F，那么有后验概率

    $p(C/EF)=p(D/EF)=0,p(B/EF)=1$

续例

$p(B)=p(C)=p(D)=1/3$

$p(D/E)=0, p(B/E)=p(C/E)=1/2$ $p(C/EF)=p(D/EF)=0, p(B/EF)=1$

- 事件E（上午的电话）发生后，A获得关于B,C,D的互信息为：

$$I(B;E)=\log\frac{p(B/E)}{p(B)}=\log\frac{1/2}{1/3}=0.585\ bit$$

$$I(C;E)=I(B;E)=0.585\ bit$$

因为$p(D/E)=0$，即在事件E发生的条件下不会出现D事件，

所以，无须考虑D事件与E事件之间的互信息量。

- 事件EF（两次电话）发生后， A获得关于B,C,D的互信息为：

$$I(B;EF)=\log\frac{p(B/EF)}{p(B)}=\log\frac{1}{1/3}=1.585\ bit$$

因为其它两个条件概率$p(C/EF), p(D/EF)$均为零，

所以，不必考虑C,D事件与EF事件之间的互信息量。

- 由此例可以看出，由于I(B;EF)=1.585bit,I(B;E)=0.585bit，因此事件EF的出现有助于肯定事件B的出现。

续例

$p(B)=p(C)=p(D)=1/3$

$p(D/E)=0, p(B/E)=p(C/E)=1/2$

$p(C/EF)=p(D/EF)=0, p(B/EF)=1$

- 在事件E（上午的电话）发生的条件下，计算条件互信息量

$$I(B;F/E)=\log\frac{p(B/EF)}{p(B/E)}=\log\frac{1}{1/2}=1\,bit$$

前面已算出 $I(B;EF)=1.585\,bit;\ \ I(B;E)=0.585\,bit$

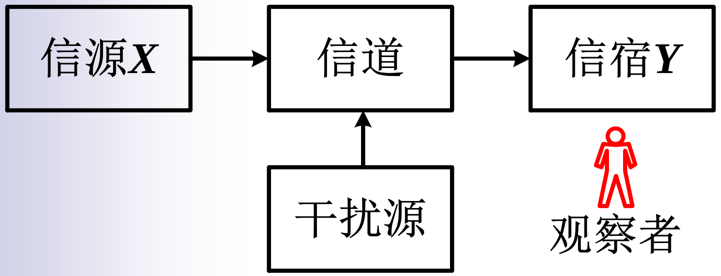
可见 $I(B;EF)=I(B;E)+I(B;F/E)$

- 事件EF出现后所提供的有关B的信息量I(B;EF)，等于事件E出现后所提供的有关B的信息量I(B;E)加上在给定事件E的条件下，再出现事件F所提供的有关B的信息量。

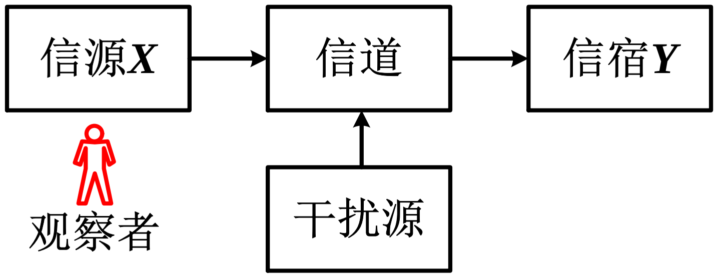
总结

- 从微观不同角度分析互信息量的物理解释，结果自洽

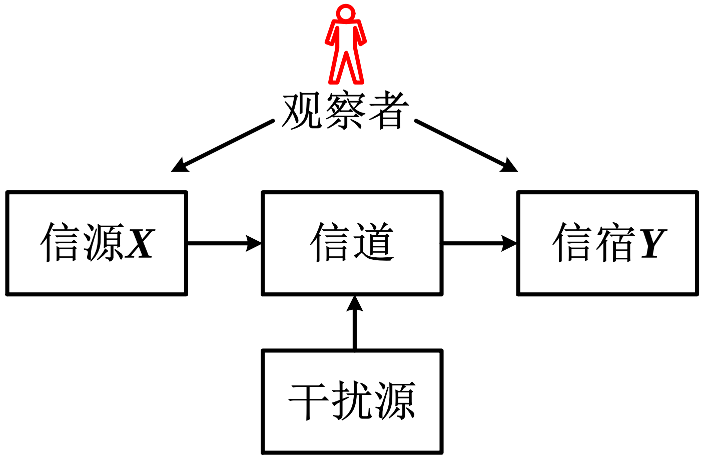
$$I(x_i;y_j)=I(x_i)-I(x_i/y_j)=\log\frac{p(x_i/y_j)}{p(x_i)}$$



$$I(y_j;x_i)=I(y_j)-I(y_j/x_i)=\log\frac{p(y_j/x_i)}{p(y_j)}$$



$$I(x_i;y_j)=\log\frac{p(x_iy_j)}{p(x_i)\cdot p(y_j)}$$



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

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

谢谢！

黑晓军
华中科技大学
电子信息与通信学院
Email: heixj@hust.edu.cn
网址： http://eic.hust.edu.cn/aprofessor/heixiaojun

参考资料

- 陈运，信息论与编码（第三版）第4章，电子工业出版社出版，2015