信息论基础

李　莹
liying2009@ecust.edu.cn

第二章：信息的度量

一、自信息和互信息

1. 自信息
2. 互信息

二、平均自信息

三、平均互信息

1. 自信息（量）

感受下如下信息的信息量:

- 明天太阳从东边升起
- 今晚的新闻联播时长40分钟
- 今晚有流星雨
- 明天预计有雪
- 明天停课放假一天

1. 自信息（量）

公理性条件：

(1) 如果$p(x_1) < p(x_2)$，则$I(x_1) > I(x_2)$，$I(x_i)$是 $p(x_i)$ 的单调递减函数；

(2) 如果$p(x_i)=0$，则$I(x_i) \to \infty$；如果$p(x_i)=1$，则$I(x_i)=0$；

(3)由两个相对独立的事件所提供的信息量，应等于它们分别提供的信息量之和：$I(x_i y_j)=I(x_i)+I(y_j)$

1. 自信息（量）

随机事件 x_i 的自信息定义为该事件发生概 率的对数的负值：

$$I(x_i) \triangleq -\log p(x_i)$$

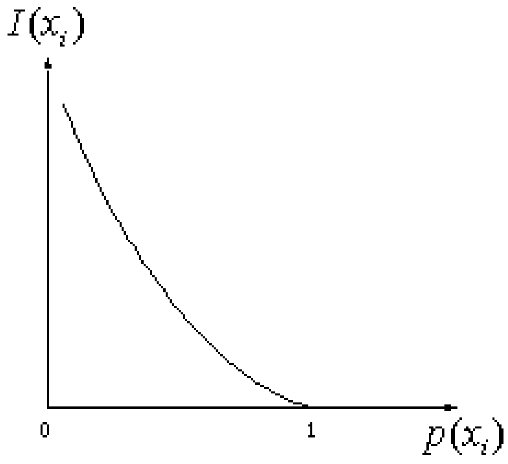


图 2.1 自信息量

关于对数底的选取：

- **以2为底，单位为比特(bit)**
- **以e为底，单位为奈特(nat)**
- **以10为底，单位为哈特莱 (Hartley)**

一般都采用以2为底的对数，为了书写简洁，有时把底数2略去不写。

1. 自信息（量）

单位之间的换算关系：

- **1奈特= $\log_2 e$ 比特=1.443比特**
- **1哈特莱= $\log_2 10$ 比特=3.322比特**
- **1 r进制单位 = $\log_2 r$ 比特**

自信息可以从两个方面来理解：

- 自信息是事件发生前，事件发生的不确定性。
- 自信息表示事件发生后，事件所包含的信息量。

1. 自信息（量）

例1: 设在甲袋中放入n个不同阻值的电阻，随意取出一个，求当被告知“取出的电阻阻值为 i ”时所获得的信息量。

解: 由于是随意取出一个电阻,所以取出任意阻值的电阻的概率相等:

$$p(x_i)=\frac{1}{n} \qquad i=1，2，\cdots，n$$

$$I(x_i)=-\log p(x_i)=\log_2 n$$ 比特

1. 自信息（量）

例2：在乙袋中放入 $\frac{1}{2}n(n+1)$ 个电阻，其中阻值为$1\,\Omega$ 的1个，$2\,\Omega$的2个，…，$n\,\Omega$的 n个，随意取出一个，求被告知“取出的电阻阻值为$1\,\Omega$”和“取出的电阻阻值为$n\,\Omega$”时分别获得的信息量。

解：

$$p(x_1)=\frac{1}{\frac{1}{2}n(n+1)} \quad \cdots \quad p(x_n)=\frac{n}{\frac{1}{2}n(n+1)}$$

$$I(x_1)=\log\left[\frac{1}{2}n(n+1)\right] \quad \cdots \quad I(x_n)=\log\left[\frac{1}{2}(n+1)\right]$$

1. 自信息（量）

例3：设在A袋放入n个不同阻值的电阻，随意取出一个，求当被告知“取出的电阻阻值为i”时所获得的信息量。

在B袋中放入m种不同功率的电阻，任意取出一个，求被告知“取出的电阻功率为j”时获得的信息量。

在C袋中放入n种不同阻值，而每种阻值又有m种不同功率的电阻，即共有nm个电阻，随意选取一个，被告知“取出的电阻阻值为i，功率为j”时获得的信息量。

1. 自信息（量）

解：对应A,B,C三袋，随意取出一个电阻事件的概率分别为：

$$p(x_i)=\frac{1}{n} \qquad p(y_j)=\frac{1}{m} \qquad p(x_i y_j)=\frac{1}{nm}$$

因此

$I(x_i)=-\log p(x_i)=\log n$ 比特

$I(y_j)=-\log p(y_j)=\log m$ 比特

$I(x_i y_j)=-\log p(x_i y_j)=\log(n\,m)=I(x_i)+I(y_j)$比特

1. 自信息（量）

例4：设在一正方形棋盘上共有64个方格，如果甲将一粒棋子随意的放在棋盘中的某方格且让乙猜测棋子所在位置。

（1） 将方格按顺序编号，令乙猜测棋子所在的顺序号。问猜测的难易程度。

（2）将方格按行和列编号，甲将棋子所在方格的列编号告诉乙之后，再令乙猜测棋子所在行的位置。问猜测的难易程度。

1. 自信息（量）

解： $p(x_i y_j)=1/64$ $i=1,2,\ldots,8; j=1,2,\ldots,8$

（1） $I(x_i y_j)=-\log p(x_i y_j)=6$ 比特

（2） $I(x_i | y_j)=-\log p(x_i | y_j)=-\log[p(x_i y_j)/p(y_j)]=3$ 比特

$I(x_i)=-\log p(x_i)=3$ 比特

$I(y_j)=3$ 比特

1. 自信息——联合自信息

例 箱中有90个红球，10个白球。现从箱中随机地取出一个球。求：

（1）事件“取出一个红球”的不确定性；

（2）事件“取出一个白球”所提供的信息量；

随机取出两个球，求：

（3）事件“两个球都是红球”所提供的信息量；

1. 自信息——联合自信息

解（1）设a_1表示事件"取出一个红球".

$$p(a_1)=0.9,\quad I(a_1)=-log0.9=0.152\ bit$$

（2）设a_2表示事件"取出一个白球".

$$p(a_2)=0.1,\quad I(a_2)=-log0.1=3.323\ bit$$

（3）事件"两个球都是红球"所提供的信息量；

$$p(a_1a_1)=\frac{C_{90}^2}{C_{100}^2}=\frac{89}{110},$$

$$I(a_1a_1)=-log\frac{89}{110}=0.306\ bit$$

1. 自信息——条件自信息

例 箱中有90个红球，10个白球。现从箱中随机地取出 两个个球。求：

（1）事件“第一个球是红球的条件下，第二个球是白球”的不确定性；

（2）事件“第一个球是红球的条件下，第二个球是红球”所提供的信息量；

1. 自信息——条件自信息

解　设a_1表示事件“取出一个红球”，a_2表示事件“取出一个白球”。

那么，$a_2|a_1$表示事件“第一个球是红球的条件下，第二个球是白球”

$$p(a_2|a_1)=\frac{10}{99}\qquad I(a_2|a_1)=-log\frac{11}{99}=3.307.$$

$$p(a_1|a_1)=\frac{89}{99}\qquad I(a_1|a_1)=-log\frac{89}{99}=0.154.$$

2. 互信息（量）

设X为信源发出的离散消息集合；Y为信宿收到的离散消息集合；

信源发出的消息，经过有噪声的信道传递到信宿；

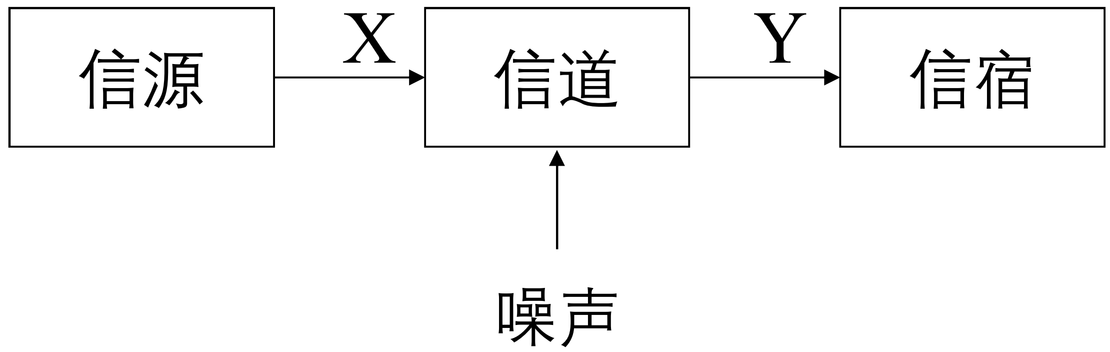
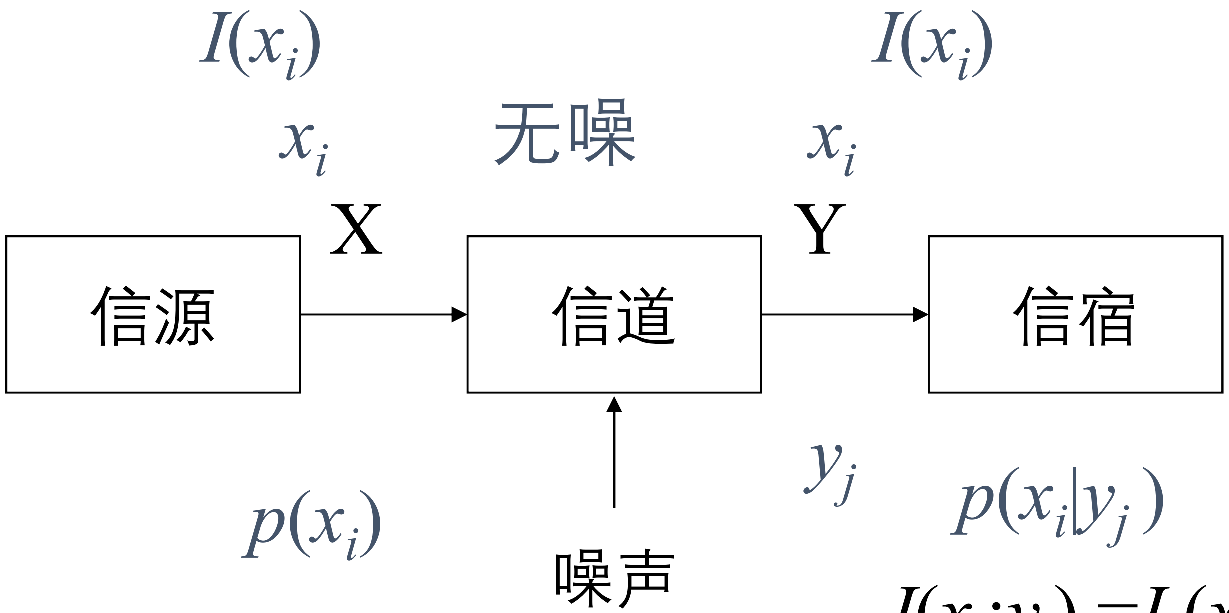


图1 通信系统的简化模型

2. 互信息（量）



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

先验概率：信源发出消息 x_i 的概率 $p(x_i)$ 。

后验概率：信宿收到消息 y_j 后推测信源发出 x_i 的概率，即条件概率 $p(x_i \mid y_j)$ 。

2. 互信息（量）

互信息定义为：

$$
\begin{aligned}
I(x_i;y_j) &= I(x_i) - I(x_i \mid y_j) \\
&= -\log p(x_i) + \log p(x_i \mid y_j) \qquad (i=1,2,\ldots,n \;\; ; \;\; j=1,2,\ldots,m) \\
&= \log \frac{p(x_i \mid y_j)}{p(x_i)}
\end{aligned}
$$

互信息有两方面的含义：

- 表示事件 y_j 出现前后关于事件 x_i 的不确定性减少的量；
- 事件 y_j 出现以后信宿获得的关于事件 x_i 的信息量。

2. 互信息（量）

讨论: $I(x_i;y_j)=\log p(x_i|y_j)-\log p(x_i)=I(x_i)-I(x_i|y_j)$

(1) 统计独立 $p(x_i|y_j)=p(x_i)$, $I(x_i|y_j)=I(x_i)$

$I(x_i;y_j)=0$

(2) 若 $p(x_i)<p(x_i|y_j)$ 则 $I(x_i;y_j)>0$

若 $p(x_i)>p(x_i|y_j)$ 则 $I(x_i;y_j)<0$

(3) $p(x_i|y_j)=1$

$$I(x_i;y_j)=\log\frac{p(x_i\mid y_j)}{p(x_i)}=\log\frac{1}{p(x_i)}=I(x_i)$$

2. 互信息（量）

例5：某地二月份天气构成的信源为

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} x_1\,(\text{晴}) & x_2\,(\text{阴}) & x_3\,(\text{雨}) & x_4\,(\text{雪}) \\ 1/2 & 1/4 & 1/8 & 1/8 \end{bmatrix}$$

某一天有人告诉你："今天不是晴天"，把这句话作为收到的消息y_1,求当收到y_1后, y_1与各种天气的互信息量。

解：

2. 互信息（量）

$$I(x_1;y_1)=\log\frac{p(x_1\mid y_1)}{p(x_1)} \qquad p(x_1\mid y_1)=\frac{p(x_1y_1)}{p(y_1)}=0 \qquad I(x_1;y_1)=\infty$$

$$p(x_2\mid y_1)=\frac{p(x_2y_1)}{p(y_1)}=\frac{p(x_2)p(y_1\mid x_2)}{p(y_1)}=\frac{\frac{1}{4}.1}{\frac{1}{4}+\frac{1}{8}+\frac{1}{8}}=\frac{1}{2}$$

$$I(x_2;y_1)=\log\frac{p(x_2\mid y_1)}{p(x_2)}=\log\frac{1/2}{1/4}=1$$

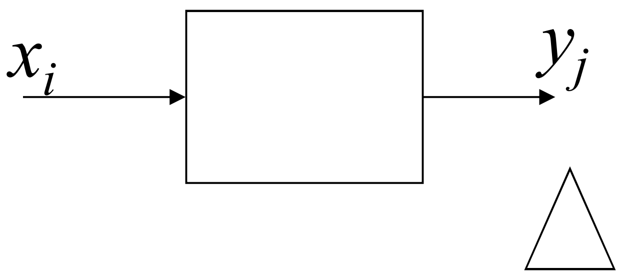
$$p(x_3\mid y_1)=\frac{1}{4}$$

$$I(x_3;y_1)=\log\frac{p(x_3\mid y_1)}{p(x_3)}=\log\frac{1/4}{1/8}=1$$

$$p(x_4\mid y_1)=\frac{1}{4}$$

$$I(x_4;y_1)=\log\frac{p(x_4\mid y_1)}{p(x_4)}=\log\frac{1/4}{1/8}=1$$

2. 互信息（量）



❖ 观察者站在输出端

$$I(x_i;y_j)=\log p(x_i|y_j)-\log p(x_i)=I(x_i)-I(x_i|y_j)$$

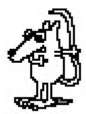
- $I(x_i)$：对 y_j 一无所知的情况下 x_i 存在的不确定度；
- $I(x_i|y_j)$：收到 y_j 后 x_i 仍然存在的不确定度；
- 互信息：收到 y_j 前和收到 y_j 后不确定度被消除的部分。

2. 互信息（量）

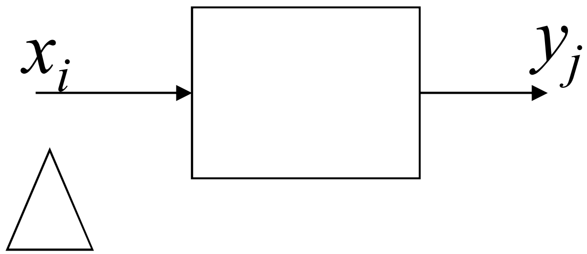
- 观察者站在输入端

$$I(y_j;x_i)=\log p(y_j\mid x_i)-\log p(y_j)=I(y_j)-I(y_j\mid x_i)$$

观察者得知输入端发出 x_i 前、后对输出端出现 y_j 的不确定度的差。



$I(y_j;x_i)=I(x_i;y_j)$ ？



2. 互信息（量）

❖ 观察者站在通信系统总体立场上

➢ 互信息等于通信前后不确定度的差值

$$I(x_i;y_j)=\log\frac{1}{p(x_i)p(y_j)}-\log\frac{1}{p(x_iy_j)}$$
$$=I^{'}(x_iy_j)-I(x_iy_j)$$
$$=I(x_i)+I(y_j)-I(x_iy_j)$$

➢ **通信前：** X和Y之间没有任何关系，即X、Y统计独立， $p(x_iy_j)=p(x_i)p(y_j)$，

先验不确定度为 $I^{'}(x_iy_j)=I(x_i)+I(y_j)$

➢ **通信后：** $p(x_iy_j)=p(x_i)p(y_j|x_i)=p(y_j)p(x_i|y_j)$，后验不确定度 $I(x_iy_j)$

2. 互信息（量）

互信息量的性质

1) 互信息的对称性

2) 互信息可为正值、负值，或为**0**

3) 任何两个事件之间的互信息不可能大于其中任一事件的自信息

2. 互信息（量）

1) 互信息的对称性

用公式表示为： $I(x_i;y_j)=I(y_j;x_i)$

互信息的对称性表明：

- 从$\boldsymbol{y_j}$得到的关于$\boldsymbol{x_i}$的信息量$I(x_i;y_j)$与从x_i得到的关于y_j的信息量$I(y_j;x_i)$是一样的，只是观察的角度不同而已。

2. 互信息（量）

2) 互信息可正可负，可为零

当后验概率大于先验概率时，互信息为正。

说明事件y_j的出现有助于消除事件x_i的不确定度。

当后验概率小于先验概率时，互信息为负。

说明收信者未收到 y_j 以前，对消息x_i是否出现的猜测难度较小，但接收到消息 y_j 后对 x_i 是否出现的猜测的难度增加了，也就是收信者接收到消息 y_j 后对 x_i出现的不确定性反而增加，所以获得的信息量为负值。

当后验概率与先验概率相等时，互信息为零。

这就是两个随机事件相互独立的情况。表明x_i和y_j之间不存在统计约束关系，从y_j得不到关于x_i的任何信息，反之亦然。

2. 互信息（量）

3) 任何两个事件之间的互信息不可能大于其中任一事件的自信息

$$I(x_i;y_j)=\log\frac{p(x_i\mid y_j)}{p(x_i)}\le\log\frac{1}{p(x_i)}=I(x_i)$$

$$I(y_j;x_i)=\log\frac{p(y_j\mid x_i)}{p(y_j)}\le\log\frac{1}{p(y_j)}=I(y_j)$$

2. 互信息（量）

例6：居住某地区的女孩中有25%是大学生，在女大学生中有75%是身高1.6m以上的，而女孩中身高1.6m以上的占总数一半。假如我们得知“身高1.6m以上的某女孩是大学生”的消息，问获得多少信息量？

2. 互信息（量）

解： x **=某女孩是大学生；** y **=某女孩身高1米6以上。则有**

$$p(x)=0.25 \qquad p(y)=0.5 \qquad p(y|x)=0.75$$

"身高1米6以上的某女孩是女大学生"为事件 $x|y$

$$\therefore I(x|y)=-\log p(x|y)=-\log\frac{p(xy)}{p(y)}$$

$$=-\log\frac{p(y|x)\cdot p(x)}{p(y)}$$

$$=-\log\frac{0.75\times 0.25}{0.5}$$

$$=-\log 0.375=1.415 \qquad bit$$

2. 互信息（量）

例7：已知信源发出 a_1 和 a_2 两种消息，且 $p(a_1)=p(a_2)=\frac{1}{2}$ 。此消息在二进制对称信道上传输，信道传输特性为

$$p(b_1\mid a_1)=p(b_2\mid a_2)=1-\varepsilon,\quad p(b_1\mid a_2)=p(b_2\mid a_1)=\varepsilon$$

求互信息量 $I(a_1;b_1)$ 和 $I(a_1;b_2)$ 。

2. 互信息（量）

解：由已知 $p(a_1)=p(a_2)=\frac{1}{2}$ $\quad p(b_1|a_1)=p(b_2|a_2)=1-\varepsilon$

$p(b_1|a_2)=p(b_2|a_1)=\varepsilon$

可得

$$p(a_1b_1)=p(b_1|a_1)\cdot p(a_1)=\frac{1}{2}(1-\varepsilon)$$

$$p(a_1b_2)=p(b_2|a_1)\cdot p(a_1)=\frac{1}{2}\varepsilon$$

$$p(a_2b_1)=p(b_1|a_2)\cdot p(a_2)=\frac{1}{2}\varepsilon$$

$$p(a_2b_2)=p(b_2|a_2)\cdot p(a_2)=\frac{1}{2}(1-\varepsilon)$$

$\therefore$ $$p(b_1)=\sum_{i=1}^{2}p(a_ib_1)=\frac{1}{2}\quad p(b_2)=\sum_{i=1}^{2}p(a_ib_2)=\frac{1}{2}$$

2. 互信息（量）

$$I(a_1;b_1) = \log\frac{p(a_1|b_1)}{p(a_1)} = \log\frac{p(a_1b_1)}{p(a_1)p(b_1)}$$

$$= \log\frac{\frac{1}{2}(1-\varepsilon)}{\frac{1}{2}\cdot\frac{1}{2}} = 1 + \log_2(1-\varepsilon) \qquad bit$$

$$I(a_1;b_2) = \log\frac{p(a_1|b_2)}{p(a_1)} = \log\frac{p(a_1b_2)}{p(a_1)p(b_2)}$$

$$= \log\frac{\frac{1}{2}\varepsilon}{\frac{1}{2}\cdot\frac{1}{2}} = 1 + \log_2\varepsilon \qquad bit$$

练习：

设e表示事件“降雨”，f表示事件“空中有乌云”，且$P(e)=0.125$, $P(e\,|\,f)=0.8$, 求：

（1）事件“降雨”的自信息；
（2）在“空中有乌云”条件下“降雨”的自信息；
（3）事件“无雨”的自信息；
（4）在“空中有乌云”条件下“无雨”的自信息；
（5）“降雨”与“空中有乌云”的互信息；
（6）“无雨”与“空中有乌云”的互信息；

解：

$(1)I(e)= -\log 0.125 = 3\ bit$；

$(2)I(e \mid f)= -\log 0.8 = 0.322\ bit$

$(3)I(\bar{e})= -\log 0.875 = 0.193\ bit$

$(4)I(\bar{e} \mid f)= -\log 0.2 = 2.322\ bit$

$(5)I(e, f)= 3-0.322 = 2.678\ bit$

$(6)I(\bar{e}, f)= 0.193-2.322 = -2.129\ bit$