# 第1章 导 论

## 1.1 基本概念

无论是远距离的卫星图像传输，还是近距离的计算机运算，只要将程序和数字从一台机器传输到另一台机器，都不免有大量的数据流在信道中流动。数据流在公共信道中传输时，不可避免地要受到随机噪声的干扰，使之产生错误。所以数字信号处理中一个重要的方面便是如何纠正由此发生的错误。纠错码便是应通信的需要而产生的一门学科。

编码理论就在于讨论如何发现错误，并进一步将它纠正过来。当然，为了避免信息在公共信道中传输时被非法窃取或篡改，而产生了密码学，密码学也是一种编码。但通常讲到编码理论时，一般都指的是纠错码。其实这些功能迥异的两种密码都是保证通信安全所必需的。本书最后一章讨论加密-纠错级联码的构造。

### 1.1.1 二元对称信道

首先说明一下，后面的讨论，如不是特别声明，都是在伽罗华域 $GF(2)$ 上进行。信息用二进制储存在计算机系统里，并以二进制数形式在信道上传输。但信道上难免有随机的噪声干扰，使之产生错误。比如 0 传输出错为 1，1 传输出错为 0。如若 0 出错或 1 出错的概率都一样，即出错的概率都为 $p$，如图 1-1 所示。这样的信道称为二元对称信道。这是最简单、也是最重要的一种信道。

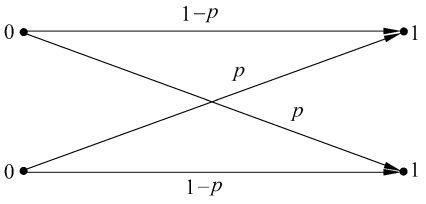

图 1-1

对于二元对称信道，最简单的纠错办法是重复码，即传输一个字符时，可重复三次或五次，然后采取“少数服从多数”的原则处理。比如传输一个字符 0 时，采用重复三次的方法，输出 000，接收的若为 000，则按传输正确处理，如若出一个错，有可能收到的是 001，010，100，因 0 的个数 2 大于 1 的个数 1，把它纠正为 000。若传输出两个错，接收到的为 011，101，110；或出三个错，接收到的为 111，则纠错将出错。不过出两个或三个错的概率相对要少得多。

现在定量地来看一看，假如对称信道出错的概率 $p=0.001$，即平均传输 1 000 个字符出一个错。信道传输正确的概率等于 $1-0.001=0.999$，若采用三重码，传输出错的概率将是怎么样？

出两个错的概率是

$$C(3,2)\times 0.001^2\times 0.999 = 3\times(1/10)^6\times 0.999$$
$$= 2.997\times(1/10)^6 < 0.000\,003$$

出三个错的概率是

$$(0.001)^3 = (1/10)^9$$

因此利用三重码传输一个字符出错，而得不到纠正的概率从 0.001 降到不到 0.000 003。

上面计算 $C(3,2)\times(0.001)^2\times 0.999$，是指出两个错的状态有 $C(3,2)=3$ 种，传输三个字符时有确定的两位出错，而其余一位不出错的概率是 $(0.001)^2\times 0.999$，其中 $0.999=1-0.001$。

但三重码的传输效率只有原来的 1/3。即传输同一个信息，传输量增大了三倍，这是不可能被接受的。编码理论是要找出效率高的编码技术。

### 1.1.2 Hamming 距离

$n$ 维 0,1 向量 $X=(x_1,x_2,\cdots,x_n)$，$x_i=0,1;i=1,2,\cdots,n$ 其中非零元素个数用 $w(X)$ 表示它，称为向量 $X$ 的权。例 $X=0011001$，则 $w(X)=3$。

**定义 1-1** 两个 $n$ 维 0,1 向量 $X=(x_1,x_2,\cdots,x_n)$，$Y=(y_1,y_2,\cdots,y_n)$，定义

$$d(X,Y) = w(X\oplus Y)$$

称 $d(X,Y)$ 为 $X$ 和 $Y$ 的 Hamming 距离。其中$\oplus$是对 $X$ 和 $Y$ 按位作异或运算：

$$0\oplus 0=0,\quad 1\oplus 1=0\quad 1\oplus 0=0\oplus 1=1$$

例如：$X=(010100)$，$Y=(110010)$

$$X = 0\ 1\ 0\ 1\ 0\ 0$$

$$Y = 1\ 1\ 0\ 0\ 1\ 0$$

$$d(X,Y) = w(100110) = 3$$

实际上 $d(X,Y)$ 就是指 $X$ 和 $Y$ 不相同位的个数，即

$$d(X,Y) = w(X\oplus Y) = w(Y\oplus X)$$

$X$ 和 $Y$ 的 Hamming 距离有以下几个性质：

① $d(X,Y)=0$，若 $X=Y$。

② $d(X,Y)\neq 0$，若 $X\neq Y$。

③ $d(X,Y)=d(Y,X)$。

④ $d(X,Y)+d(Y,Z)\geqslant d(X,Z)$

性质④就是所谓距离的三角不等式，两边之和大于(或等于)第三边。在欧几里德几何里是一条公理，但现在的距离是 Hamming 距离，还必须证明如下：

设 $X=(x_1,x_2,\cdots,x_n)$，$Y=(y_1,y_2,\cdots,y_n)$，$Z=(z_1,z_2,\cdots,z_n)$。因 $x_i,y_i,z_i$ 都属于 (0,1)，故有：

若 $x_i\neq y_i$，$y_i\neq z_i$，则 $x_i\neq z_i$。

明白了这个关系，三角不等式的性质④便不难理解了。

⑤ $0\leqslant w(X)\leqslant n$。

⑥ $w(X)=0$，当且仅当 $X=0$。

### 1.1.3 码字

编码的传输过程如图 1-2 所示。

其中 $m_1m_2\cdots m_k$ 为 $k$ 个字符的信息字，$c_1c_2\cdots c_n$，为 $n$ 个字符的码字，$e_1e_2\cdots e_n$ 为干扰，$e_i=0$ 表示第 $i$ 位无干扰，$e_j=1$ 表示第 $j$ 位有一干扰引起的错误。

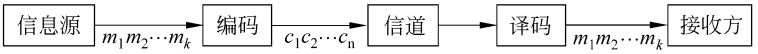


图 1-2

**例 1-1** 一组 3 位信息字对应一 6 位码字编码 $\boldsymbol{C}$，如表 1-1 所示。

**表 1-1**

| 信息字 | 码字 |
|---|---|
| 000 | 000000 |
| 001 | 001110 |
| 010 | 010101 |
| 011 | 011011 |
| 100 | 100011 |
| 101 | 101101 |
| 110 | 110110 |
| 111 | 111000 |

如 000 正确传输，则收到的码字为 000000，若出现错误，比如收到的是 001000 不是码字，8 个码字中没有它。以后就称之为字，字不一定是码字。

一组编码 $\boldsymbol{C}$，含有若干个码字，码字中两两间都有一个距离 $d$，其中最小的距离有称之为 $\boldsymbol{C}$ 的最短距离，用 $d_{\min}$ 表示它。即

$$d_{\min}=\min_{c_i,c_j\in C}\{d(C_i,C_j)\}$$

**定理 1-1** 编码 $\boldsymbol{C}$ 当它的 $d_{\min}$ 至少是 $d$ 时能检出 $d-1$ 位或小于 $d-1$ 位错误。

所谓能检出 $d-1$ 位错误，是指码字传输过程出现了 $d-1$ 位错误能被发现提出来，之所以能被发现，因为出错后的字不是码字。如果出错后变成了另一个码字，则识别不了。检查出错，但不能断定正确的应是什么，只好请求重发。计算机常用的奇偶校验码便是一例。

例 1-1 中计算两两 Hamming 距离如表 1-2 所示，可知 $d_{\min}=3$。

**表 1-2**

| $X$ \ $Y$ | 000000 | 001110 | 010101 | 011011 | 100011 | 101101 | 110110 | 111000 |
|---|---|---|---|---|---|---|---|---|
| 000000 | — | 3 | 3 | 4 | 3 | 4 | 4 | 3 |
| 001110 | 3 | — | 4 | 3 | 4 | 3 | 3 | 4 |
| 010101 | 3 | 4 | — | 3 | 4 | 3 | 3 | 4 |
| 011011 | 4 | 3 | 3 | — | 3 | 4 | 4 | 4 |
| 100011 | 3 | 4 | 4 | 3 | — | 3 | 3 | 4 |
| 101101 | 4 | 3 | 3 | 4 | 3 | — | 4 | 3 |
| 110110 | 4 | 3 | 3 | 4 | 4 | 4 | — | 3 |
| 111000 | 3 | 4 | 4 | 4 | 4 | 3 | 3 | — |

**证明** 因码 $\boldsymbol{C}$ 的最短距离是 $d$，任何一个码字受到小于 $d$ 位的干扰，不可能使之错误到被识别为另一个码字，所以能被检查出来。如果对于某个码字，错误的位达到 $d$ 位，有可能错误地被识别为别的码字，错误无法被查出来。

**定理 1-2** 已知编码 $\boldsymbol{C}$ 的 $d_{\min}=2t+1$，则能纠正不超过 $t$ 位的错误。

**证明** 设 $Z$ 是被传输的码字，接收到的是 $R$，$d(Z,R)\leqslant t$。则不存在码字 $Y$，使得 $d(Y,R)\leqslant t$，否则

$$d(Z,Y)\leqslant d(Z,R)+d(R,Y)\leqslant t+t=2t<2t+1$$

与定理的假设相矛盾。

$d_{\min}=2t+1$ 不仅是纠 $t$ 个错的充分条件，有时也是必要条件。

还是以本节的例 1-1 来说明。

明文 3 比特，直接传输，信道出错概率若为 1/10，不出错的概率为 $1-1/10=9/10$，3 位都不出错的概率为 $(9/10)^3=0.729$。

编码 $\boldsymbol{C}$ 为 6 比特，码 $\boldsymbol{C}$ 的最短距离 $d_{\min}=3$，根据定理能纠一个错。故译码不出错的概率为

$$(9/10)^6+\boldsymbol{C}(6,1)(9/10)^5(1/10)=0.5314+0.3543=0.9357$$

说明编码 $\boldsymbol{C}$ 传输正确的概率有很大的提高。而传输的长度扩大了一倍。其中 $(9/10)^6$ 是码字 6 位都不出错的概率。$\boldsymbol{C}(6,1)(9/10)^5(1/10)$ 是 6 位中 5 位不出错，只有 1 位出错的概率。

若采用三重码，传输中最多 1 位出错的概率是

$$(9/10)^3+C(3,1)(9/10)^2(1/10)=0.729+0.243=0.972$$

而传输的长度为原来信息位的 3 倍。

### 1.1.4 熵的概念

1. 熵的概念

随机事件的特点在于它的不确定性，它出现的概率是它的不确定性的一种度量，概率越小，其不确定性越大。概率等于 1 为确定性事件，即必然事件，概率为零表示不可能的事件。

信息论首先要讨论对信息量的度量，直觉上对不确定性愈大的信息越重视，因为表示它的信息量越大，确定性事件的消息被看作毫无信息量，或信息量等于零。比如听到一个人在说："太阳从东方升起"，你会感到这句话等于没有说，因为它没给出任何新的信息。若听到消息："一个不明的飞行物出现在天空"，将无疑会引起你的极大的好奇而争相一睹，因为那是难得出现的吸引人的事件。

从直观上看，$k$ 个等概率事件的不确定性将随着 $k$ 的增大而增大。$k=1$ 时不确定性为零，不确定性度量结果是 $k$ 的函数，设为 $f(k)$，或归纳为：

(1) $f(1)=0$。

(2) 若 $k_1<k_2$，则 $f(k_1)<f(k_2)$。

若事件 $A$ 有 $h$ 个等概率事件，事件 $B$ 有 $k$ 个等概率事件，则事件 $A$ 与 $B$，$A\cap B$ 有 $hk$ 个等概率事件 3，它的不确定性将超过 $A$ 或 $B$，即 $f(hk)>f(h)$，$f(hk)>f(k)$，令

$$f(hk) = f(h) + f(k),$$

不妨因此令

$$f(k) = \log k = -\log \frac{1}{k}$$

这里“位”指比特位，以后不特别注明的对数均取 2 为底，即以两个等概率事件的不确定性的信息量作为一个单位：比特(bit)。一位有(0,1)两种状态。比如掷一银币，出正反两面的概率各$\frac{1}{2}$，$-\log \frac{1}{2}=\log 2=1$ 比特。

还可以将上面的概念进一步推广，设事件 $A=\{A_1, A_2, \cdots, A_n\}$，其中 $A_i$ 出现的概率为 $p_i, i=1,2,\cdots,n$。

$$p_1 + p_2, + \cdots + p_n = 1$$

**定义 1-2** $I(A_i)=-\log p_i, i=1,2,\cdots,n$。作为对 $A_i$ 的信息量的度量，令

$$H(A) = -p_1 \log p_1 - p_2 \log p_2 - \cdots - p_n \log p_n \tag{1-1}$$

作为事件 $A$ 的熵，用它来度量事件 $A$ 的不确定性。$H(A)$越大表示 $A$ 的不确定性也越大。

从 $H(A)$的定义可见，$H(A)$实际上是 $A$ 的各事件 $A_i$ 的信息量 $H(A_i)$的平均值，或$\{H(A_i)\}$的数学期望。

**例 1-2** 有一箱里装有 20 个球，其中白球 10 个，红球 5 个，黑球 5 个。从箱中任取一个球作为事件 $A$：取出的球为白球、红球或黑球的概率依次为

$$p_1 = \frac{1}{2}, p_2 = \frac{1}{4}, p_3 = \frac{1}{4}$$

则 $H(A)=-\frac{1}{2}\log \frac{1}{2}-\frac{1}{4}\log \frac{1}{4}-\frac{1}{4}\log \frac{1}{4}=\frac{1}{2}\log 2+\frac{1}{2}\log 4=\frac{1}{2}+1=1.5$(比特)

掷一个骰子只有 6 种状态 $A$，而且机会均等，则

$$H(A) = \log 6 = 2.585\,0$$

掷两个骰子有 36 种状态 $A$；和为 2 的有 1 种，和为 3 的有 2 种，和为 4 的有 3 种，和为 5 的有 4 种，和为 6 的有 5 种，和为 7 的有 6 种，和为 8 的有 5 种，和为 9 的有 4 种，和为 10 的有 3 种，和为 11 的有 2 种，和为 12 的有 1 种，则

$$H(A) = -\frac{2}{36}\log \frac{1}{36} - \frac{4}{36}\log \frac{2}{30} - \frac{6}{36}\log \frac{3}{36}$$

$$-\frac{8}{36}\log \frac{4}{36} = \frac{10}{36}\log \frac{5}{36} - \frac{12}{36}\log \frac{6}{36} = 3.274\,4,$$

由此可见，掷两个骰子较掷一个骰子的不确定性大。

**例 1-3** 对某地进行了 15 年的观察，发现 6 月份下雨的概率为 0.4，晴天的概率为 0.6；10 月份下雨的概率为 0.65，下雪的概率为 0.15，晴天的概率为 0.2，分别求它们的熵。

设 $A$ 为 6 月份的天气状况，$B$ 为 10 月份的天气状况。则

$$H(A) = -0.4\log 0.4 - 0.6\log 0.6 \approx 0.969\,4$$

$$H(B) = -0.65\log 0.65 - 0.15\log 0.15 - 0.2\log 0.2 \approx 1.277\,2$$

若只考虑是否晴天，则

$$H(B) = -0.8\log 0.8 - 0.2\log 0.2 = 0.720\,4$$

则

$$H(B) < H(A)$$

2. 熵的性质

① $\lim\limits_{p\to 0}(-p\log p)=0$

$$\lim_{p\to 0} p\ln p = \lim_{p\to 0}\frac{\ln p}{\frac{1}{p}}$$

根据罗比塔法则，有

$$\lim_{p\to 0} p\ln p = \lim_{p\to 0}\frac{\frac{1}{p}}{-\frac{1}{p}} = \lim_{p\to 0}(-p) = 0$$

$$y = -x\ln x$$

求 $y=-x\ln x$ 的极值：

$$y = -x\ln x,\quad y' = -1-\ln x = 0,\quad \ln x = -1,\quad x = \frac{1}{\mathrm{e}}$$

即在 $x=\frac{1}{\mathrm{e}}$ 时，$y=-x\ln x$ 取极大值(见图 1-3)。

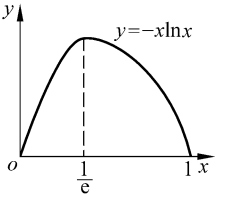

图 1-3

故当事件 $A$ 有 $k$ 个结果，设为 $A_1, A_2, \cdots, A_k$，其概率分别是 $p_1, p_2, \cdots, p_k$。当且仅当其中之一，设 $p_i=1$，其余全为 0 时，有

$$H(A) = 0$$

若要求在约束条件 $p_1+p_2+\cdots+p_k=1$ 下，求

$$H(A) = p_1\log\frac{1}{p_1} + p_2\log\frac{1}{p_2} + \cdots + p_k\log\frac{1}{p_k}$$

的极值，根据拉格朗日乘数法，求得

$$H(A) + \lambda\sum_{i=1}^{k} p_i = p_1\log\frac{1}{p_1} + p_2\log\frac{1}{p_2} + \cdots + p_k\log\frac{1}{p_k} + \lambda(p_1+p_2+\cdots+p_k)$$

的无条件极值。

$$\frac{\partial H}{\partial p_i} + \lambda\frac{\partial}{\partial p_i}\left(\sum_{j=1}^{k}\lambda_j\right) = 0$$

或

$$-(\log p_j + 1) + \lambda = 0,\quad j = 1,2,\cdots,k$$

所以

$$p_1 = p_2 = \cdots = p_k$$

但

$$p_1 + p_2 + \cdots + p_k = 1$$

所以有

$$p_1 = p_2 = \cdots = p_k = \frac{1}{k}$$

② $H(A)$作为 $p_1,p_2,\cdots,p_k$ 的函数 $H(p_1,p_2,\cdots,p_k)$，不因 $p_1,p_2,\cdots,p_k$ 的顺序改变而改变，而且

$$H(p_1,p_2,\cdots,p_k)=H(p_1+p_2,p_3,p_4,\cdots,p_k)+(p_1+p_2)H\left(\frac{p_1}{p_1+p_2},\frac{p_2}{p_1+p_2}\right)$$

**证明**

$$\begin{aligned}&H(p_1+p_2,p_3,p_4,\cdots,p_k)+(p_1+p_2)H\left(\frac{p_1}{p_1+p_2},\frac{p_2}{p_1+p_2}\right)\\&=-(p_1+p_2)\log(p_1+p_2)-p_3\log p_3-\cdots-p_k\log p_k-\\&\quad(p_1+p_2)\left[\frac{p_1}{p_1+p_2}\log\frac{p_1}{p_1+p_2}+\frac{p_2}{p_1+p_2}\log\frac{p_2}{p_1+p_2}\right]\\&=-(p_1+p_2)\log(p_1+p_2)-p_3\log p_3-\cdots-p_k\log p_k+\\&\quad p_1\log(p_1+p_2)+p_2\log(p_1+p_2)-p_1\log p_1-p_2\log p_2\\&=-p_1\log p_1-p_2\log p_2-\cdots-p_k\log p_k=H(p_1,p_2,\cdots,p_k)\end{aligned}$$

还可以证明更一般的结果：

$$\begin{aligned}&H(p_1,p_2,\cdots,p_{i_1-1},p_{i_1},p_{i_1+1},\cdots,p_{i_2-1},p_{i_2},p_{i_2+1},\cdots,p_{i_s-1},p_{i_s},p_{i_s+1},\cdots,p_k)\\&=H(p_1+p_2+\cdots+p_{i_1},p_{i_1+1}+\cdots+p_{i_2},\cdots,p_{i_s+1}+\cdots+p_{i_k})+\\&(p_1+p_2+\cdots+p_{i_1})H\left(\frac{p_1}{p_1+p_2+\cdots+p_{i_1}},\cdots,\frac{p_{i_1}}{p_1+p_2+\cdots+p_{i_1}}\right)+\\&(p_{i_1+1}+p_{i_1+2}+\cdots+p_{i_2})H\left(\frac{p_{i_1+1}}{p_{i_1+1}+p_{i_1+2}+\cdots+p_{i_2}},\cdots,\frac{p_{i_2}}{p_{i_1+1}+p_{i_1+2}+\cdots+p_{i_2}}\right)\\&+\cdots+(p_{i_s+1}+\cdots+p_{i_k})H\left(\frac{p_{i_s+1}}{p_{i_s+1}+p_{i_s+2}+\cdots+p_k},\cdots,\frac{p_k}{p_{i_s+1}+p_{i_s+2}+\cdots+p_k}\right)\end{aligned}$$

3. 条件熵

本节考虑 $U$ 和 $V$ 是两个互相独立的随机事件，设 $U$ 有 $k$ 个可能的结果：$A_1,A_2,\cdots,A_k$，对应的概率依次为 $p(A_1),p(A_2),\cdots,p(A_k)$；$V$ 有 $l$ 个可能的结果：$B_1,B_2,\cdots,B_l$，对应的概率依次为 $p(B_1),p(B_2),\cdots,p(B_l)$。

$$p(A_1)+p(A_2)+\cdots+p(A_k)=1$$
$$p(B_1)+p(B_2)+\cdots+p(B_l)=1$$

$UV$ 作为事件可能有以下可能的结果：

$$\begin{aligned}&A_1B_1,A_1B_2,\cdots,A_1B_l\\&A_2B_1,A_2B_2,\cdots,A_2B_l\\&\vdots\\&A_kB_1,A_kB_2,\cdots,A_kB_l\end{aligned}$$

**定理 1-3** 若 $U$ 和 $V$ 是互相独立的事件，则 $H(U,V)=H(U)+H(V)$

**证明** $H(U,V)=-p(A_1B_1)\log(A_1B_1)-p(A_1B_2)\log(A_1B_2)-\cdots-p(A_1B_l)\log(A_1B_l)-p(A_2B_1)\log(A_2B_1)-p(A_2B_2)\log(A_2B_2)-\cdots-p(A_2B_l)\log(A_2B_l)-\cdots-p(A_kB_1)\log(A_kB_1)-p(A_kB_2)\log(A_kB_2)-\cdots-p(A_kB_l)\log(A_kB_l)$

由于 $U$ 和 $V$ 相互独立，故

$$p(A_iB_j)=p(A_i)p(B_j),\qquad i=1,2,\cdots,k,\quad j=1,2,\cdots,l$$

而且

$$p(A_1)+p(A_2)+\cdots+p(A_k)=1$$
$$p(B_1)+p(B_2)+\cdots+p(B_l)=1$$

因此

$$\begin{aligned}&-p(A_iB_1)\log p(A_iB_1)-p(A_iB_2)\log p(A_iB_2)-\cdots-p(A_iB_l)\log p(A_iB_l)\\=&-p(A_i)p(B_1)[\log p(A_i)+\log p(B_1)]-p(A_i)p(B_2)[\log p(A_i)+\log p(B_2)]-\cdots\\&-p(A_i)p(B_l)[\log p(A_i)+\log p(B_l)]\\=&-p(A_i)\log p(A_i)[p(B_1)+p(B_2)+\cdots+p(B_l)]-p(A_i)[p(B_1)\log p(B_1)+\\&p(B_2)\log p(B_2)+\cdots+p(B_l)\log p(B_l)]\\=&-p(A_i)\log p(A_i)+p(A_i)H(V)\end{aligned}$$

故

$$\begin{aligned}H(U,V)=&-p(A_1)\log p(A_1)+p(A_1)H(V)-p(A_2)\log p(A_2)+p(A_2)H(V)-\\&\cdots-p(A_k)\log p(A_k)+p(A_k)H(V)\\=&[p(A_1)+p(A_2)+\cdots+p(A_k)]H(V)+[-p(A_1)\log p(A_1)-\\&p(A_2)\log p(A_2)-\cdots-p(A_k)\log p(A_k)]=H(U)+H(V)\end{aligned}$$

若 $U$ 和 $V$ 不独立时，则

$$H(UV)=-\sum_{h=1}^{k}\sum_{j=1}^{l}p(A_hB_j)\log p(A_hB_j)$$

但

$$p(A_hB_j)=p(A_h)p(B_j\mid A_h)$$
$$\log p(A_hB_j)=\log p(A_h)+\log p(B_j\mid A_h)$$

所以

$$\begin{aligned}H(U,V)=&-p(A_1)p(B_1\mid A_1)[\log p(A_1)+\log p(B_1\mid A_1)]-p(A_1)(p(B_2\mid A_1)\\&[\log p(A_1)+\log p(B_2\mid A_1)]-\cdots\\&-p(A_1)p(B_l\mid A_1)[\log p(A_1)+\log p(B_l\mid A_1)]-\cdots\\&-p(A_k)p(B_1\mid A_k)[\log p(A_k)+\log p(B_1\mid A_k)]-\cdots\\&-p(A_k)p(B_l\mid A_k)[\log p(A_k)+\log p(B_l\mid A_k)]\end{aligned}$$

另一方面

$$p(B_1\mid A_i)+p(B_2\mid A_i)+\cdots+p(B_l\mid A_i)$$
$$=\frac{1}{p(A_i)}[p(A_iB_1)+p(A_iB_2)+\cdots+p(A_iB_l)]=1,i=1,2,\cdots,k$$

所以

$$\begin{aligned}H(U,V)=&-\sum_{i=1}^{k}\{p(A_i)[p(B_1\mid A_i)+p(B_2\mid A_i)+\cdots+p(B_l\mid A_i)]\cdot\log p(A_i)\\&+p(A_i)[p(A_i\mid B_1)\log p(A_i\mid B_1)+p(A_i\mid B_2)\log p(A_i\mid B_2)+\cdots\\&+p(A_i\mid B_l)\log p(A_i\mid B_l)]\}\\=&-p(A_1)\log p(A_1)+p(A_1)H(V\mid A_1)-p(A_2)\log p(A_2)+p(A_2)H(V\mid A_2)\end{aligned}$$

$$-\cdots-p(A_k)\log p(A_k)+p(A_k)H(V\mid A_k)$$

$$H(V\mid A_i)=-p(B_1\mid A_i)\log p(B_1\mid A_i)-p(B_2\mid A_i)\log p(B_2\mid A_i)-\cdots$$
$$-p(B_l\mid A_i)\log p(B_l\mid A_i),i=1,2,\cdots,k$$

令

$$H(V\mid U)=-\sum_{i=1}^{k}\sum_{j=1}^{l}p(A_iB_j)\log p(B_j\mid A_i)$$

所以

$$H(U,V)=H(U)+H(V\mid U) \tag{1-2}$$

称 $H(V|A_i)$为 $A_i$ 已知条件下 $V$ 的条件熵。

令

$$H(V\mid U)=p(A_1)H(V\mid A_1)+p(A_2)H(V\mid A_2)+\cdots+p(A_k)H(V\mid A_k)$$

若 $U$ 和 $V$ 不互相独立，则有

$$H(U,V)=H(U)+H(V\mid U) \tag{1-3}$$

$H(V|U)$为 $U$ 已知条件下事件 $V$ 的条件熵，或简称为条件熵。用以度量 $U$ 给定后留给 $V$ 的不确定性，也称为 $U$ 给定后 $V$ 的暧昧度。

**例 1-4** 有一种病的发病率为 2%，即每 100 个人得此病的人有两人。为了诊断出此病，要做一种试验，有病的人对这种试验必有阳性反应，而健康的人阳性和阴性反应各半。设事件 $V$ 有两种结果：$B_1$——健康，$B_2$——病人。而事件 $U$ 的结果为 $A_1$——阳性反应，$A_2$——阴性反应。求 $H(V)$和 $H(V|U)$。

**解** 从已知条件有

$$p(B_1)=0.98,p(B_2)=0.02$$

$$H(V)=-0.98\log 0.98-0.02\log 0.02=0.141\,1(\text{比特})$$

因占 2%的病人对试验有阳性反应，而占 98%的健康人对试验有阳性和阴性反应各占一半。因此

$$p(A_1)=0.02+0.49=0.51,\quad p(A_2)=0.49$$

若试验有阳性反应，即当事件 $U$ 有 $A_1$ 结果时，有

$$p(B_1\mid A_1)=49/51,p(B_2\mid A_1)=2/51$$

$$H(V\mid A_1)=-(49/51)\log(49/51)-(2/51)\log(2/51)=0.235\,4(\text{比特})$$

若试验有阴性反应，即当事件 $U$ 有 $A_2$ 结果时，结论只能是 $B_1$，即为健康人。故

$$p(B_1\mid A_2)=0,p(B_2\mid A_2)=0,\ H(V\mid A_2)=0$$

$$H(V\mid U)=0.51H(V\mid A_1)+0.49H(V\mid A_2)=0.51\times 0.235\,4=0.120\,1(\text{比特})$$

即试验的结果使事件 $V$ 的不确定性下降为 0.141 1−0.120 1=0.021 0。

**例 1-5** 若一盒子里放了 $n$ 个球，其中 $m$ 个是红颜色的，其余 $n-m$ 个为白色的球。事件 $U$ 是从中随机取出第一个球，事件 $V$ 是从中取出第二个球。求 $H(U)$，$H(V)$，$H(V|U)$，$H(U|V)$。

**解** 事件 $U$ 有两种结果：$u_1$——取得红球，$u_2$——取得白球。

事件 $V$ 有两种结果：$v_1$——取得红球，$v_2$——取得白球。

事件 $V$ 是取第 2 个球，取第 1 个球可能是白的，也可能是红的，故有

$$p(u_1)=\frac{m}{n},p(u_2)=\frac{n-m}{n}$$

$$p(v_1)=\frac{m}{n}\cdot\frac{m-1}{n-1}+\frac{n-m}{n}\cdot\frac{m}{n-1}=\frac{m}{n}$$

$$p(v_2)=\frac{m}{n}\cdot\frac{n-m}{n}+\frac{n-m}{n}\cdot\frac{n-m-1}{n-1}=\frac{n-m}{n}$$

因此

$$H(U)=H(V)=-\frac{m}{n}\log\frac{m}{n}-\frac{n-m}{n}\log\frac{n-m}{n}$$

若已知 $U$ 或 $V$ 的结果，例如已知第一个球为白球，取第二个球也为白球的概率 $p(v_1|u_1)$ 为

$$p(v_1\mid u_1)=\frac{m-1}{n-1}$$

同理

$$p(v_2\mid u_1)=\frac{n-m}{n-1},\quad p(v_1\mid u_2)=\frac{m}{n-1},\quad p(v_2\mid u_2)=\frac{n-m-1}{n-1}$$

从而

$$H(V\mid u_1)=-\frac{m-1}{n-1}\log\frac{m-1}{n-1}-\frac{n-m}{n-1}\log\frac{n-m}{n-1}$$

$$H(V\mid u_2)=-\frac{m}{n-1}\log\frac{m}{n-1}-\frac{n-m-1}{n-1}\log\frac{n-m-1}{n-1}$$

可以证明若 $m<n-m$，则

$$H(V\mid u_1)<H(V)$$

而

$$H(V\mid u_2)>H(V)$$

即白球数目少于红球数目，已知第一个球取的是白球的条件下，事件 $V$ 的不确定性减少了。极而言之，只有一个白球，取走了，$V$ 便是确定性的事件了。

$$\begin{aligned}H(V\mid U)&=p(u_1)H(V\mid u_1)+p(u_2)H(V\mid u_2)\\&=\frac{m}{n}\left(\frac{m-1}{n-1}\log\frac{m-1}{n-1}-\frac{n-m}{n-1}\log\frac{n-m}{n-1}\right)\end{aligned}$$

**定理 1-4** $H(U)-H(U|V)=H(V)-H(V|U)$

**证明** 根据公式

$$H(U,V)=H(U)+H(V\mid U)$$

以及

$$H(U,V)=H(V,U)=H(V)+H(U\mid V)$$

可得

$$H(U)+H(V\mid U)=H(V)+H(U\mid V)$$

或

$$H(U)-H(U\mid V)=H(V)-H(V\mid U)$$

**定理 1-5** 设事件 $U$ 可能出现的结果为 $A_1,A_2,\cdots,A_k$，$V$ 可能出现的结果为