信息論基础

張宏基編
陈太一审

（修訂版）

北京科学教育編輯室
1964年11月

信息論基础

*

出版者：北京科学教育編輯室
印刷者：中国人民解放軍535工厂
开本787×1092毫米 1/16 · 印張6 7/16 · 字数170,000
1964年11月第一版（北京）
1964年11月第一次印刷
定价：0.95元

目　　录

緒　言

通信的最主要目的，就是把消息多、快、好、省地傳輸到对方。如何描述消息，是研究通信的基本問題。描述的方法不同，研究的方法也就不同。一般說来，消息有下面两种不同的描述方法：

第一种，把消息看成是一种能够展开成为富氏級数(富氏积分)的时間函数，如果富氏級数不是无穷多項的話，也可以根据柯切尼可夫定理变为时間上离散的序列，这种方法称为經典方法。

多路通信中的頻率划分和間时划分是两个典型的例子。

第二种，把消息看成是随机过程，这种方法称为現代方法，或者統計方法。

这种方法研究通信問題，有两个主要的学派：(*a*)着重研究随机过程的关联，即前后間的关系；(*b*)着重研究随机过程的不肯定程度，后者就是信息論。

随机过程的关联和不肯定程度有着密切的关系。我們不是說信息論不考虑随机过程的关联；而是說，在信息論中，关联主要反映在不肯定程度上。

本书希望討論信息論研究通信問題的基本方法，所以只把随机过程看作离散的随机序列。

現在，簡单地介紹一下信息論的基本內容，图 1 是一个通信系統的一种模型。信源发出的随机序列称为消息。終端在通信之前，对信源发出什么消息是不肯定的，通信的目的就是希望終端能够肯定信源发出的是什么消息，为此，应該通过通信系統告訴終端足够的信息，用来解除終端原有的不肯定性，这就是信息論研究通信的方法。

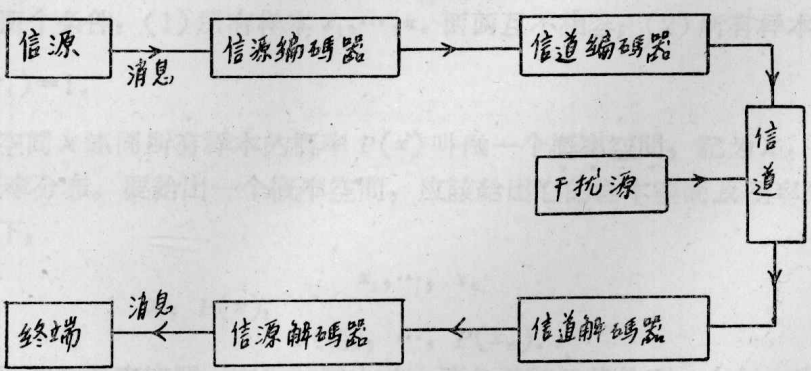

图 1　通信系統的一种模型

換句話說，信息論认为通信的作用是傳送信息，通信的目的是要解除終端对信源的不肯定性。根据这些概念，自然会产生下面两个重要問題：(1) 如何用最經济的办法解除終端的不肯定性；(2)如何使終端的不肯定性解除得最彻底。

本书从我們对不肯定性的直观概念出发，推导出反映一个概率空間的不肯定程度的函

数。(第一章)

要解除不肯定性，就得收到足够的信息量。因此，我們接着討論了信息量的定义。解除的不肯定程度与获得的信息量之間的关系。怎样才能最有效的获得信息量。(第二章)

不是所有的信源都能够用簡单的概率空間来描述，必須进一步討論N維空間及其不肯定程度，然后才能够比較滿意地描述信源和計算消息的不肯定程度。(第三、四章)

消息要經过信源編碼。信源編碼的目的是把消息編成实际上能够使用的，而且是最好的电碼。怎样的电碼才能实际上使用？最好的电碼能好到什么程度？如何編出最好的电碼？这些都是信源編碼中的主要問題。(第五章)

信息是由信道傳輸到对方的。可是由于信道中存在干扰，干扰使傳輸发生錯誤，有了錯誤，接收端就不容易判断发送的是什么消息。这样一来，干扰的影响就反映在降低了接收的信息量。在給定的干扰下，信道能够傳輸的最大信息量叫做信道容量C。(第六章)

虽然我們总可以在傳輸率等于C的情况下使用一条信道，但并不能保証不发生錯誤。可是，实际通信往往不希望有錯誤。在不减少給定的干扰的情况下，能否使得錯誤不发生？如果可以，这时信道傳輸的信息量是多少？理論上的研究告訴我們：不能使錯誤不发生，但可以使錯誤发生得任意少，少到什么程度都可以；而且这时信道傳輸的信息量可以几乎等于C，但不能大于C。(第七章)

信道編碼的目的就是要找出一些可行的办法，使得傳輸接近定理給出的理想情况。(第八章)

最后，我們分析了現行的中文电报系統，作为信息論的应用的例子。(第九章)

第一章　不肯定程度

本章从我們对不肯定程度的直观概念出发，給出不肯定程度的数学定义。根据定义，推导反映概率空間不肯定程度的函数$H(\ \ )$，最后說明$H(\ \ )$的最大值。

§1-1　概率空間的不肯定程度 H(　)

首先，我們举一个例子来說明几个有关的名詞。

投擲一顆骰子，如果我們研究的是它落下后，朝上一面的点数，那么，每做一次这样的試驗，其結果必然屬于1点、2点、3点、4点、5点、6点中的一个。这6个我們希望加以研究的結果，无論哪一个，都称为基本事件或者样本。所有基本事件的全体称为基本事件集或者样本空間。

同样是投擲一顆骰子，如果我們希望研究的不是朝上一面的点数，而是这些点数屬于奇数还是偶数，因为希望加以研究的結果不同，所以現在的样本空間就变成了只有两个样本——奇数、偶数。

样本虽然可以任意选擇，但它們必須符合一个条件：每做一次試驗，其結果只能屬于样本中的某一个，仍以投擲骰子为例，如果已經选用奇数作为样本，就不能同时再选用1点、3点、5点作为样本。因此，样本空間应該是这样的一个集合：每做一次試驗，其結果必定而且只能屬于該集合的某一元素。

拿数学的語言来說，一个样本空間

$$X=\{x_1,\cdots,x_n\}$$

必須滿足下列两个条件：(1)所有样本$x_1,\cdots,x_n$两两互不相容；(2)所有样本的概率之和为1，即$\sum_{i=1}^{n}P(x_i)=1$.

一个样本空間X連同所有样本的概率$P(x)$叫做一个概率空間，記为X，$P(x)$。通常称$P(x)$为X的概率分布。要給出一个概率空間，应該給出它的样本空間及概率分布。概率空間的詳細表示如下：

$$X,\ P(x):\ \begin{matrix} x_1,\cdots,\ x_n \\ P(x_1),\ \cdots,\ P(x_n). \end{matrix}$$

对于一个已知的概率空間，在做試驗之前，要知道結果将是哪一个样本出現，往往具有不肯定性。不肯定的程度，不是所有的概率空間都是一样的。虽然我們还不清楚不肯定程度的明确含义，但凭着直观的想法，我們可以把不肯定程度解釋为做試驗之前猜測試驗結果的困难程度。当然，这个解釋不能作为不肯定程度的数学定义，因为猜測的困难程度也只是一个相当含糊的概念。至于不肯定程度的数学定义，将在下一节給出。这里只希望根据直观的概念，說明不同的概率空間，它們的不肯定性的确有程度上的不同。

設有两个概率空間：

$$X, P(x): \begin{matrix} x_1, & x_2 \\ 0.5, & 0.5; \end{matrix}$$

$$Y, P(y): \begin{matrix} y_1, & y_2, & y_3, & y_4 \\ 0.25, & 0.25, & 0.25, & 0.25 \end{matrix}$$

显然，$X, P(x)$的不肯定程度比 $Y, P(y)$ 的不肯定程度小，因为前者的困难程度只是在 两 者中猜其一，但后者是从 4 个中猜一个。

再看下面一个概率空間

$$Z, P(z): \begin{matrix} z_1, & z_2 \\ 0.99, & 0.01. \end{matrix}$$

事先猜测 z_1，z_2 哪一个出現，虽然仍具有不肯定性，但可以說，大概总是 z_1 出現。因此，它的不肯定程度比 $X, P(x)$的小，比之 $Y, P(y)$的不肯定程度，那就更小了。

概率空間的不肯定程度是信息論研究的重要問題之一，以后我們用 $H(\ \ )$代表概率空間的不肯定程度，用 $H(X)$ 代表概率空間X，$P(x)$的不肯定程度。

§1-2 H( )的 定 义

現在，我們根据 $H(\ \ )$的直观概念，給出 $H(\ \ )$的数学定义，为下一节做好准备。

在一个布袋中，放入两个对人手感覚完全一样的球。一个紅色，一个白色。如果我們随意拿一个出来，并对拿出来的球的顏色进行事先猜测，其困难程度等于概率空間

$$X, P(x): \begin{matrix} \text{紅}, & \text{白} \\ 0.5, & 0.5 \end{matrix}$$

的不肯定程度。

投擲一个质地均匀的硬币，事先猜测它落下后是正面还是反面朝上，其困难程度等于概率空間

$$Y, P(y): \begin{matrix} \text{正}, & \text{反} \\ 0.5, & 0.5 \end{matrix}$$

的不肯定程度。

不难想象，上述两种猜测的困难程度应該相等，即 $H(X)=H(Y)$。因为不論我們在紅、白中选其一，或者在正、反中选其一，其困难程度都是在两个等可能性的样本中选一个，这就意味着 $H(X)$与 x 到底代表什么无关，它只与 $P(x)$发生关系，用数学公式来表示这句話，我們有

$$H(X)=H(P_1, \cdots, P_n).$$

就是說：不肯定程度只是各 P 的函数。

下面是我們对 $H(P_1, \cdots, P_n)$的一些直观想法：

(1) $H(P_1, 1-P)$是 P 的連續函数，当 $0 \leqslant P \leqslant 1$。

(2) $H(P_1, \cdots, P_n)$与各 P 的次序无关。

(3) $H(1, 0)=0$。

(4) $H(P_1, \cdots, P_n)=H(P_1, \cdots, P_n, 0)$。

(5) 如果 $P_1=\cdots=P_n=\frac{1}{n}$，则当 n 增加时，$H\left(\frac{1}{n}, \cdots, \frac{1}{n}\right)$也随之而增加。

(6) 当 $P_1=\cdots=P_n=\frac{1}{n}$时，$H(P_1, \cdots\cdots, P_n)$达到最大值。

(7) 如果 $P_n=q_1+q_2>0$，则

$$H(P_1, \cdots, P_{n-1}, q_1, q_2)=H(P_1, \cdots, P_n)+P_nH\left(\frac{q_1}{P_n}, \frac{q_2}{P_n}\right)。$$

关于最后一点，我們准备解釋一下。仍然用从布袋中取一个球为例。設袋中放有紅、黄、白三个球，事先猜測拿出的球的顏色，有两种猜法：

第一、直接猜它是紅、黄、白三种顏色中的哪一个，这时的困难程度等于从三个等可能性的事件中猜其一，故应为$H\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$。

第二、先猜它是不是某一种顏色（比如說紅色）。因为不是紅色的概率为 $P(\overline{紅})=1-\frac{1}{3}=\frac{2}{3}$，故其困难程度就是概率空間

$$\begin{matrix} 紅, & \overline{紅} \\ \frac{1}{3}, & \frac{2}{3} \end{matrix}$$

的不肯定程度，即$H\left(\frac{1}{3}, \frac{2}{3}\right)$。如果是紅，猜測就可以終止。如果不是紅，猜測的困难程度将比$H\left(\frac{1}{3}, \frac{2}{3}\right)$要大些，因为还得猜它是黄还是白。也就是說，这时的困难程度应为

$$H\left(\frac{1}{3}, \frac{2}{3}\right)十再猜是黄还是白的困难程度。$$

既然已知不是紅色，则黄色发生的概率为 $P\left(\frac{黄}{\overline{紅}}\right)=\frac{P(\overline{紅}黄)}{P(\overline{紅})}=\frac{\frac{1}{3}}{\frac{2}{3}}$，同理，$P\left(\frac{白}{\overline{紅}}\right)=\frac{\frac{1}{3}}{\frac{2}{3}}$。

因此，再猜是黄还是白的困难程度为$H\left(\frac{\frac{1}{3}}{\frac{2}{3}}, \frac{\frac{1}{3}}{\frac{2}{3}}\right)$。

这种猜法的困难程度，等于$H\left(\frac{1}{3}, \frac{2}{3}\right)$的概率为$\frac{1}{3}$，等于$H\left(\frac{1}{3}, \frac{2}{3}\right)+H\left(\frac{\frac{1}{3}}{\frac{2}{3}}, \frac{\frac{1}{3}}{\frac{2}{3}}\right)$的概率为$\frac{2}{3}$，平均困难程度为

$$\frac{1}{3}H\left(\frac{1}{3}, \frac{2}{3}\right)+\frac{2}{3}\left[H\left(\frac{1}{3}, \frac{2}{3}\right)+H\left(\frac{\frac{1}{3}}{\frac{2}{3}}, \frac{\frac{1}{3}}{\frac{2}{3}}\right)\right]$$

$$=H\left(\frac{1}{3}, \frac{2}{3}\right)+\frac{2}{3}H\left(\frac{\frac{1}{3}}{\frac{2}{3}}, \frac{\frac{1}{3}}{\frac{2}{3}}\right)。$$

显然，这两种猜法都是希望知道同一个概率空間中哪一个顏色出現，所以它們的困难程度应該相等，即

$$H\left(\frac{1}{3},\ \frac{1}{3},\ \frac{1}{3}\right)=H\left(\frac{1}{3},\ \frac{2}{3}\right)+\frac{2}{3}H\left(\frac{\frac{1}{3}}{\frac{2}{3}},\ \frac{\frac{1}{3}}{\frac{2}{3}}\right)。$$

这就是 $H(P_1,\ \cdots,\ P_{n-1},\ q_1,\ q_2)=H(P_1,\cdots,P_n)+P_nH\left(\frac{q_1}{P_n},\ \frac{q_2}{P_n}\right)$ 的直观想法。当然，还应該規定 $P_n>0$，否則，我們可能有 $H\left(\frac{0}{0},\ \frac{0}{0}\right)$，这种情况不能从直观概念加以解釋。

上述的7种直观想法，曾被不同的作者利用来决定函数 $H(P_1,\ \cdots,\ P_n)$。根据法捷耶夫(Фадеев. Д.К.)1956年发表的杰出研究，証明了只要 $H(P_1,\ \cdots,\ P_n)$ 滿足(1)，(2)，(7)三个想法，其他(3)、(4)、(5)、(6)几个想法必然会被滿足。所以我們可以說(1)、(2)、(7)就是不肯定程度的定义。換句話說，如果 $H(P_1,\ \cdots,\ P_n)$ 是概率空間 X，$P(x)$ 的不肯定程度，它应該滿足(1)、(2)、(7)三个条件。

§1-3 H( ) 的 推 导

这一节的內容是証明下面一个定理：

〔**定理**〕 除了一个常数 K 外，下列三个条件决定概率空間 X，$P(x)$ 的不肯定程度

$$H(P_1,\ \cdots,\ P_n)=-K\sum_{i=1}^{n}P_i\log P_i,$$

这三个条件是：

(1) $H(P_1,\ 1-P)$ 为 P 的連續函数，当 $0\leqslant P\leqslant 1$；

(2) $H(P_1,\ \cdots,\ P_n)$ 与各 P 的次序无关；

(3) 如果 $P_n=q_1+q_2>0$，則

$$H(P_1,\ \cdots,\ P_{n-1},\ q_1,\ q_2)=H(P_1,\ \cdots,\ P_n)+P_nH\left(\frac{q_1}{P_n},\ \frac{q_2}{P_n}\right).$$

我們将通过下列一些引理来証明定理。

〔**引理 1**〕

$$H(1,\ 0)=0$$

証：利用条件(3)，我們有

$$H\left(\frac{1}{2},\ \frac{1}{2},0\right)=H\left(\frac{1}{2},\ \frac{1}{2}\right)+\frac{1}{2}H(1,\ 0);$$

$$H\left(0,\ \frac{1}{2},\ \frac{1}{2}\right)=H(0,\ 1)+H\left(\frac{1}{2},\ \frac{1}{2}\right).$$

根据条件(2)，得

$$\frac{1}{2}H(1,\ 0)=H(1,\ 0),$$

所以

$$H(1,\ 0)=0.$$

〔**引理 2**〕 如果 $P_n>0$，則

$$H(P_1,\ \cdots,\ P_n,\ 0)=H(P_1,\ \cdots,\ P_n)$$

証：利用条件(3)及引理1，立即可以得到証明。

〔**引理 3**〕 如果 $P_n=q_1+\cdots+q_m>0$，則

$$H(P_1,\ \cdots,\ P_{n-1},\ q_1,\ \cdots,q_m)=H(P_1,\ \cdots,P_n)+P_nH\left(\frac{q_1}{P_n},\ \cdots,\ \frac{q_m}{P_n}\right)。$$

証：如果 $q_1,\ \cdots,\ q_m$ 中，有一些 q 等于零，根据引理 2，可以把它們除去，所以我們可以假設 $q_1,\ \cdots,\ q_m$ 均不等于零。

首先，我們假設 m 个 q 引理是成立的，然后証明 $m+1$ 个 q 引理也能成立。

$$H(P_1,\ \cdots,\ P_{n-1},\ q_1,\ \cdots,\ q_{m+1})$$

$$=H(P_1,\ \cdots,\ P_{n-1},q_1,\ Q)+QH\left(\frac{q_2}{Q},\ \cdots,\ \frac{q_{m+1}}{Q}\right)$$

$$=H(P_1,\ \cdots,\ P_n)+P_nH\left(\frac{q_1}{P_n},\ \frac{Q}{P_n}\right)+QH\left(\frac{q_2}{Q},\ \cdots,\ \frac{q_{m+1}}{Q}\right)$$

其中 $Q=q_2+\cdots+q_{m+1}>0$。因为

$$H\left(\frac{q_1}{P_n},\ \cdots,\ \frac{q_{m+1}}{P_n}\right)=H\left(\frac{q_1}{P_n},\ \frac{Q}{P_n}\right)+\frac{Q}{P_n}H\left(\frac{q_2}{Q},\ \cdots,\ \frac{q_{m+1}}{Q}\right),$$

于是証明了

$$H(P_1,\ \cdots,\ P_{n-1},\ q_1,\cdots,q_{m+1})=H(P_1,\ \cdots,\ P_n)+P_nH\left(\frac{q_1}{P_n},\ \cdots,\ \frac{q_{m+1}}{P_n}\right)。$$

其中 $P_n=q_1+\cdots+q_{m+1}>0$。由于 $m=2$ 时，引理变成了条件(3)，所以 m 等于任意自然数时，引理均能成立。

〔引理 4〕

$$H\left(\frac{1}{mn},\cdots,\frac{1}{mn}\right)=H\left(\frac{1}{m},\cdots,\frac{1}{m}\right)+H\left(\frac{1}{n},\cdots,\frac{1}{n}\right)$$

証：利用引理 3

$$H\Big(\overbrace{\frac{1}{mn},\cdots,\frac{1}{mn}}^{n\text{个}},\cdots,\overbrace{\frac{1}{mn},\cdots,\frac{1}{mn}}^{n\text{个}}\Big)=H\Big(\overbrace{\frac{1}{mn},\cdots,\frac{1}{mn}}^{n\text{个}},\cdots,\overbrace{\frac{1}{mn},\cdots,\frac{1}{mn}}^{n\text{个}},\ \frac{1}{m}\Big)+\frac{1}{m}H\left(\frac{1}{n},\cdots,\frac{1}{n}\right)$$

（m 組；$m-1$ 組）

把上式右边第一項（ ）中的$\frac{1}{m}$移到（ ）的最左边，重复上述步驟，引理就可被証明。

〔引理 5〕

$$\frac{H\left(\frac{1}{n},\cdots,\frac{1}{n}\right)}{\log n}=K(\text{常数})。$$

証：見附录 1。

$H\left(\frac{1}{n},\cdots,\frac{1}{n}\right)$其实是 $H(P_1,\ \cdots,\ P_n)$在 $P_1=\cdots=P_n=\frac{1}{n}$这个条件限制下的特殊情况，这时的 $H(\ \ )$只是 n 的函数，所以我們可以写成

$$h(n)=H\left(\frac{1}{n},\ \cdots,\ \frac{1}{n}\right)。$$

根据引理 4，$h(n)$具有下列两个性质：

$$h(mn)=h(m)+h(n),$$

$$h(m^n)=nh(m).$$

显然，

$$\log(mn)=\log m+\log n$$

$$\log m^n=n\log m$$

这虽不能作为引理 5 的証明，但的确給了我們一个强有力的暗示：$h(n)$可能等于 $K\log n$。

有了上述 5 个引理，現在我們可以証明本节开始所提出的定理，令 S 及 r 为任意正整数，且 $S>r$，从引理 3

$$H\Big(\overbrace{\frac{1}{S},\cdots,\frac{1}{S}}^{r\text{ 个}},\ \overbrace{\frac{1}{S},\cdots,\frac{1}{S}}^{(S-r)\text{个}}\Big)$$

$$=H\Big(\frac{r}{S},\ \frac{S-r}{S}\Big)+\frac{r}{S}H\Big(\frac{1}{r},\cdots,\frac{1}{r}\Big)+\frac{S-r}{S}H\Big(\frac{1}{S-r},\cdots,\frac{1}{S-r}\Big)$$

如果 $P=\frac{r}{S}$，即 P 为有理数，则

$$\begin{aligned}H(P,\ 1-P)&=H\Big(\frac{1}{S},\cdots,\frac{1}{S}\Big)-PH\Big(\frac{1}{r},\cdots,\frac{1}{r}\Big)-(1-P)H\Big(\frac{1}{S-r},\cdots,\frac{1}{S-r}\Big)\\&=K\log S-KP\log S+KP\log S-KP\log r-K(1-P)\log(S-r)\\&=K\Big[P\log\frac{S}{r}+(1-P)\log\frac{S}{S-r}\Big]\\&=-K[P\log P-(1-P)\log(1-P)].\end{aligned}$$

由于条件 (1) 規定 $H(P,\ 1-P)$是 P 的連續函数，故上述結果馬上可以推广到 P 为无理数的情况。

进一步

$$\begin{aligned}H(P_1,\ P_2,\ P_3)&=H(P_1,\ P_2+P_3)+(P_2+P_3)H\Big(\frac{P_2}{P_2+P_3},\ \frac{P_3}{P_2+P_3}\Big)\\&=-KP_1\log P_1-K(P_2+P_3)\log(P_2+P_3)\\&\quad-KP_2\log\frac{P_2}{P_2+P_3}-KP_3\log\frac{P_3}{P_2+P_3}\\&=-K[P_1\log P_1+P_2\log P_2+P_3\log P_3]\end{aligned}$$

继續下去，显然可得

$$H(P_1,\ \cdots,\ P_n)=-K\sum_{i=1}^{n}P_i\log P_i,$$

于是定理得到証明。

为了方便，我們往往把上式写成

$$H(X)=-K\sum_{X}P(x)\log P(x)$$

定理留下常数 K 尚未解决，下一节我們将討論这个問題。

§1-4 常数 K 的决定

本节主要討論 $H(P_1,\ \cdots,\ P_n)=-K\sum_{i=1}^{n}P_i\log P_i$ 中常数 K 的决定問題。

对于給定的 $P_1,\ \cdots,\ P_n$，这个 K 影响 $H(P_1,\ \cdots,\ P_n)$的大小。經驗告訴我們，

$$X, P(x): \begin{matrix} x_1, & x_2 \\ \frac{1}{2}, & \frac{1}{2} \end{matrix}$$

是一个常遇到的概率空間。因此，我們希望这个空間的不肯定程度

$$H\left(\frac{1}{2}, \frac{1}{2}\right)=1.$$

为了达到上述目的，我們可以令

$$H\left(\frac{1}{2}, \frac{1}{2}\right)=-K\left[\frac{1}{2}\log\frac{1}{2}+\frac{1}{2}\log\frac{1}{2}\right]=K\log 2$$

中的 $K=1$，$\log_2$ 代替 log。这样得到的 $H(\ \ )$值，我們給它一个特殊名称，叫做 bit（二进单位）。

如果令 $K=1$，$\log_e$ 代替 log，那么，得到的 $H(\ \ )$值叫做 nit（自然单位）。

当然，我們还可以有各种各样的假設。除非特别注明，本书以后的 $H(P_1, \cdots, P_n)$都是按 bit 計算的，而且为了簡单起見，把 bit 也略去，不写出来，因此得到

$$H(P_1, \cdots, P_n)=-\sum_{i=1}^{n} P_i \log P_i$$

一个只具有两个样本的概率空間，如果这两个样本的概率相等，則它的不肯定程度为 1 bit。換句話說，事先猜測两个等可能性事件哪一个出現，其困难程度为 1。也就是說，我們約定，这种猜測具有单位困难程度，猜測的困难程度只是和单位比較的相对数值。下面我們举一个例子来帮助說明这一段話。例如：我們要事先猜測概率空間。

$$X, P(x): \begin{matrix} x_1, & x_2, & x_3, & x_4 \\ \frac{1}{4}, & \frac{1}{4}, & \frac{1}{4}, & \frac{1}{4} \end{matrix}$$

的哪一个样本出現，其困难程度为

$$H(X)=-\sum_{i=1}^{4} P_i \log P_i = \log 4=2$$

2 表示这种猜測的困难程度比从两个等可能性事件中猜其一要困难 2 倍。我們可以这样想象，把 X 分成两組，第一組包含 x_1 与 x_2，第二組包含 x_3 与 x_4。由于这两組的出現概率各为 $\frac{1}{2}$，所以猜測是哪一組出現的困难程度为 1。另外，猜測 x 的注脚是奇数还是偶数，其困难程度也是 1。知道了 x 屬于哪一組，也知道了 x 的注脚是奇数还是偶数，当然也就知道了 x 是哪一个。比方第二組，注脚为奇数，那么一定是 x_3。因此，这个 2 是和单位困难程度比較的相对数值。

特别指出，不能說凡是从两个事件中猜其一，它的困难程度都是 1，只有当这两个事件的概率相等时，它的困难程度才是 1，例如在布袋中放进 1000 个球，其中 107 个为紅球，893 个为白球，随意拿一个出来。猜測这个球是什么顏色的困难程度为 $H(0.107, 0.893)=0.4908$。

$-\log_2 P$，$-P\log_2 P$，$H(P, 1-P)$都可以从本书末的附表中直接查到。

§1-5 H( ) 的 最 大 值

在前面的 §1-2 节中，我們曾談到过，即使从直观概念出发，也会想到当 $P_1=\cdots=P_n$ 时，$H(P_1, \cdots, P_n)$应該达到最大值。但是，在推导 $H(\ \ )$的过程中，我們却沒有利用这一点。本节主要証明 §1-3 节得到的 $H(\ \ )$的确符合上述直观想法。

〔引理〕 如果

$$\sum_{i=1}^{n} P_i = 1, \qquad \sum_{i=1}^{n} q_i = 1$$

則
$$-\sum_{i=1}^{n} P_i \log P_i \leqslant -\sum_{i=1}^{n} P_i \log q_i$$

且只当 $P_i = q_i (i = 1, \cdots, n)$时，上式才取等号。

証　对 P_1 个$\frac{q_1}{P_1}$，P_2 个$\frac{q_2}{P_2}$，…，P_n 个$\frac{q_n}{P_n}$取它們的几何平均及算术平均，由于几何平均值小于或者等于算术平均值，所以

$$\left(\frac{q_1}{P_1}\right)^{P_1} \cdots \left(\frac{q_n}{P_n}\right)^{P_n} \leqslant \frac{P_1\left(\frac{q_1}{P_1}\right) + \cdots + P_n\left(\frac{q_n}{P_n}\right)}{P_1 + \cdots + P_n} = 1.$$

等式成立的必要和充分条件为 $P_i = q_i (i = 1, \cdots, n)$。因此

$$\sum_{i=1}^{n} P_i \log \frac{q_i}{P_i} = \sum_{i=1}^{n} P_i \log q_i - \sum_{i=1}^{n} P_i \log P_i \leqslant \log 1 = 0$$

移项后，引理就得到証明。

〔定理〕

$$H(P_1, \cdots, P_n) \leqslant \log n$$

当而且只当 $P_1 = \cdots = P_n = \frac{1}{n}$时，上式才取等号。

証：利用引理，令 $q_1 = \cdots = q_n = \frac{1}{n}$，馬上得到

$$-\sum_{i=1}^{n} P_i \log P_i \leqslant -\sum_{i=1}^{n} P_i \log \frac{1}{n} = \log n.$$

这就是定理的証明。

定理說明一个概率空間，当而且只当所有的样本具有相同的概率时，它的不肯定程度最大。

习　題

(1) 如果說：一个概率空間包含的样本愈多，它的不肯定程度愈大，对嗎？

(2) 如果說：一个概率空間，不論它包含几个样本，只要有一个样本的概率为 1，則它的不肯定程度为 0，对嗎？

(3) 如果 $P_1 = \cdots = P_n = \frac{1}{n}$，当 n 增加时，証明 $H(P_1, \cdots, P_n)$也随之而增加。

(4) 設有概率空間

$$X, P(x): \begin{matrix} x_1, & x_2, & x_3, & x_4, & x_5 \\ \frac{1}{2}, & \frac{1}{4}, & \frac{1}{8}, & \frac{1}{16}, & \frac{1}{16} \end{matrix}$$

求这个概率空間的不肯定程度 $H(X)$。

(5) 設 $P_1 = 0.2$，$P_2 = 0.19$，$P_3 = 0.18$，$P_4 = 0.17$，$P_5 = 0.16$，$P_6 = 0.15$。

求 $-\sum_{i=1}^{6} P_i \log P_i$，并解釋为什么 $-\sum_{i=1}^{6} P_i \log P_i > \log 6$，与§1-5 节的定理不符。

(6) 用曲綫画出 $H(P, 1-P)$与 P 的关系。

第二章 信 息 量

为了减小不肯定程度，必須采用某些手段(称为檢驗)来获得情报，所得到的情报的数量叫做信息量。本章主要討論信息量与不肯定程度的关系及获得信息量所付出的代价。

§2-1 信息量的定义

先让我們举一个例子来看看如何通过檢驗来减小不肯定程度。假設有一部机器，它由8个元件組成，各个元件損坏的概率相等。再假設現在有一个元件坏了，在我們尚未对該机进行檢驗之前，要知道哪一个元件損坏的不肯定程度，显然就是下面概率空間的不肯定程度。

$$X,\ P_1(x):\ \begin{matrix} x_1, & x_2, & x_3, & x_4, & x_5, & x_6, & x_7, & x_8 \\ \frac{1}{8}, & \frac{1}{8}, & \frac{1}{8}, & \frac{1}{8}, & \frac{1}{8}, & \frac{1}{8}, & \frac{1}{8}, & \frac{1}{8} \end{matrix}$$

其中 $x_1,\cdots,x_8$ 分别代表8个元件，$P_1(x)$代表这时的概率分布，經过一次檢驗后，确知其中4个元件(比如說 x_1, x_2, x_3, x_4)沒有損坏，但剩下的4个元件，仍然不知哪一个損坏，故經过檢驗后，保留的不肯定程度等于下面概率空間的 $H(\ \ )$。

$$X,\ P_2(x):\ \begin{matrix} x_1, & x_2, & x_3, & x_4, & x_5, & x_6, & x_7, & x_8 \\ 0, & 0, & 0, & 0, & \frac{1}{4}, & \frac{1}{4}, & \frac{1}{4}, & \frac{1}{4} \end{matrix}$$

其中 $P_2(x)$ 表示經过檢驗后的概率分布，可以认为檢驗在这里所起的作用是使 $P_1(x)$ 变成 $P_2(x)$。由于 $H\left(\frac{1}{8},\ \cdots,\ \frac{1}{8}\right)=3$ 大于 $H\left(\frac{1}{4},\cdots,\frac{1}{4}\right)=2$，所以檢驗减小了不肯定程度。

值得注意的是檢驗减小了多少不肯定程度，决定于檢驗方法。例如我們換一种方法，它一次檢驗并不能决定4个元素有沒有損坏，而只决定2个元素(比如說 x_1, x_2)有沒有損坏。这种檢驗方法，經过檢驗后保留的不肯定程度可以分为下列两种情况：

(a) 如果眞的是 x_1 或者 x_2 損坏，那末，

$$X,\ P_2(x):\ \begin{matrix} x_1, & x_2, & x_3, & x_4, & x_5, & x_6, & x_7, & x_8 \\ \frac{1}{2}, & \frac{1}{2}, & 0, & 0, & 0, & 0, & 0, & 0 \end{matrix}$$

(b) 如果不是 x_1 或者 x_2 損坏，那末，

$$X,\ P_3(x):\ \begin{matrix} x_1, & x_2, & x_3, & x_4, & x_5, & x_6, & x_7, & x_8 \\ 0, & 0, & \frac{1}{6}, & \frac{1}{6}, & \frac{1}{6}, & \frac{1}{6}, & \frac{1}{6}, & \frac{1}{6} \end{matrix}$$

(a) 这种情况出現的概率为 $\frac{2}{8}$，保留的不肯定程度为 $H\left(\frac{1}{2},\ \frac{1}{2}\right)=1$；

(b) 这种情况出現的概率为 $\frac{6}{8}$，保留的不肯定程度为 $H\left(\frac{1}{6},\cdots,\frac{1}{6}\right)=2.58496$，因此，平均保留的不肯定程度为

$$\frac{2}{8}H\left(\frac{1}{2},\ \frac{1}{2}\right)+\frac{6}{8}H\left(\frac{1}{6},\cdots,\frac{1}{6}\right)=2.18872.$$

从这个例子可以看出經过一次檢驗后，保留的不肯定程度决定于：

(1) 檢驗方法。

(2) 檢驗結果是哪一种情况。

現在，我們可以給出信息量的定义：

信息量＝不肯定程度减小的量。

不肯定程度减少的量等于原有的不肯定程度减去保留的不肯定程度，既然保留的不肯定程度决定于檢驗方法及檢驗結果是哪一种情况，信息量也同样如此。

假設原来的概率分布为 $P_1(x)$，經过一次檢驗后，概率分布可能变成 $P_1(x),\cdots,P_m(x)$中的任意一种。根据信息量的定义，我們有

$$I[P_1(x):P_j(x)]=H[P_1(x)]-H[P_j(x)]$$

其中 $I[P_1(x):P_j(x)]$表示使概率分布从 $P_1(x)$变为 $P_j(x)$，$1\leqslant j\leqslant m$ 的这一次檢驗所供給的信息量；$H[P_1(x)]$ 表示原有的不肯定程度；$H[P_j(x)]$ 表示經过檢驗后保留的不肯定程度。

在上述例子的第二种檢驗方法中，如果是 x_1 或者 x_2 損坏，則該次檢驗供給的信息量为

$$I[P_1(x):P_2(x)]=H[P_1(x)]-H[P_2(x)]$$

$$=H\left(\frac{1}{8},\cdots,\frac{1}{8}\right)-H\left(\frac{1}{2},\ \frac{1}{2}\right)=2.$$

如果不是 x_1 或者 x_2 損坏，則該次檢驗供給的信息量为

$$I[P_1(x):P_3(x)]=H[P_1(x)]-H[P_3(x)]$$

$$=H\left(\frac{1}{8},\cdots,\frac{1}{8}\right)-H\left(\frac{1}{6},\cdots,\frac{1}{6}\right)=0.41504$$

信息量也可以用平均值来表示。即

平均信息量＝不肯定程度平均减小的量。

当然，这里所謂平均，指的是概率平均。平均信息量，指的是一次檢驗所能供給的平均信息量。上面例子中的第一种檢驗方法，它的平均信息量与任意一种檢驗結果所供給的信息量是相等的。所以

$$\text{平均信息量}=H\left(\frac{1}{8},\cdots,\frac{1}{8}\right)-H\left(\frac{1}{4},\cdots,\frac{1}{4}\right)=1.$$

至于第二种檢驗方法，它的

$$\text{平均信息量}=H\left(\frac{1}{8},\cdots,\frac{1}{8}\right)-\left[\frac{2}{8}H\left(\frac{1}{2},\ \frac{1}{2}\right)+\frac{6}{8}H\left(\frac{1}{6},\cdots,\frac{1}{6}\right)\right]$$

$$=0.81128.$$

§2-2 二态檢驗定理

一种檢驗，如果它的結果只能告訴我們两种状态(比如說是、否)，那就叫做二态檢驗。前节例子中的檢驗是一种二态檢驗，因为檢驗結果只能告訴我們某些元素是否損坏。用电压表檢驗电源电压，則不是二态檢驗；但如果只檢驗电源电压是否超过 220 伏，那就是二态檢驗了。

〔定理〕 任意一种二态檢驗，一次檢驗所能供給的平均信息量小于或者等于 1 bit。

証：設有一个概率空間 X，$P_1(x)$

$$X,\ P_1(x):\ \begin{matrix} x_1, & \cdots, & x_s, & x_{s+1}, & \cdots, & x_n \\ P_1, & \cdots, & P_s, & P_{s+1}, & \cdots, & P_n \end{matrix}$$

如果檢驗結果落在 $x_1, \cdots, x_s\ (1 \leqslant S \leqslant n)$中的任一个，我們把这种情况称为“是”。令 P(是)表示“是”发生的概率，H(是)表示是发生后保留的不肯定程度。結果落在 $x_{s+1}, \cdots, x_n$ 中任意一个的情况称为“否”。用 P(否)表示“否”发生的概率，H(否)表示否发生后保留的不肯定程度。根据信息量的定义，一次檢驗所供給的平均信息量，可以用下式計算。

$$\text{平均信息量} = H(P_1, \cdots, P_n) - [P(\text{是})H(\text{是}) + P(\text{否})H(\text{否})]$$

$$= -\sum_{i=1}^{n} P_i \log P_i + P(\text{是})\left[\frac{P_1}{P(\text{是})}\log\frac{P_1}{P(\text{是})} + \cdots + \frac{P_s}{P(\text{是})}\log\frac{P_s}{P(\text{是})}\right] +$$

$$+ P(\text{否})\left[\frac{P_{s+1}}{P(\text{否})}\log\frac{P_{s+1}}{P(\text{否})} + \cdots + \frac{P_n}{P(\text{否})}\log\frac{P_n}{P(\text{否})}\right]$$

$$= -P(\text{是})\log P(\text{是}) - P(\text{否})\log P(\text{否})。 \qquad (2\text{-}2\text{-}1)$$

利用 § 1-5 节的定理，我們有

$$\text{平均信息量} \leqslant \log 2 = 1,$$

定理得到証明。

定理指出一个很重要的事实，那就是当而且只当 P(是)$=P$(否)$=\frac{1}{2}$时，二态檢驗的平均信息量才等于 1 bit。檢驗的目的就是要得到信息量，所以我們决定檢驗方法的时候，应該尽可能使 P(是)等于 P(否)。下面我們举两个例子来加以說明。

例 1：事先猜測概率空間

$$\begin{matrix} x_1, & x_2, & x_3, & x_4, & x_5, & x_6, & x_7, & x_8 \\ \frac{1}{8}, & \frac{1}{8}, & \frac{1}{8}, & \frac{1}{8}, & \frac{1}{8}, & \frac{1}{8}, & \frac{1}{8}, & \frac{1}{8} \end{matrix}$$

哪一个样本出現，具有 $H\left(\frac{1}{8}, \cdots, \frac{1}{8}\right) = 3$ 的不肯定程度，所以我們至少要用三次二态檢驗才能把不肯定性全部取消。当然，全部取消了不肯定性就等于我們能够肯定哪一个样本出現了，現在我們分三次进行檢驗。

第一次，檢驗出現的样本是否落在 x_1, x_2, x_3, x_4 中。

第二次，檢驗出現的样本是否落在 x_1, x_2, x_5, x_6 中。

第三次，檢驗出現的样本是否落在 x_1, x_3, x_5, x_7 中。

由于每次檢驗的 P(是)$=P$(否)$=\frac{1}{2}$，所以三次就能把所有的不肯定性消除。比如，第一次为是，第二次为否，第三次为否，則一定是 x_4 出現。

要是我們另外換一种檢驗程序，第一次檢驗是否 x_1 出現，第二次檢驗是否 x_2 出現，……直到我們能肯定为止，那末，平均說来，三次檢驗是不够的，原因是一次这样的檢驗并沒有供給我們 1 bit 的信息量。

例 2：有一部机器，它包含三个电子管 x_1, x_2, x_3。x_1 損坏的概率为$\frac{1}{2}$，x_2 損坏的概率为