熵(Entropy)(上)

李天岩

在我們日常生活中，似乎經常存在着「不確定性」的問題。比方說，天氣預報員常說「明天下雨的可能性是 70 % 」。這是我們習以爲常的「不確定性」問題的一個例子。一般不確定性問題所包涵「不確定」(uncertainty)的程度可以用數學來定量地描述嗎？在多數的情況下是可以的。本世紀 40 年代末，由於信息理論 (Information Theory) 的需要而首次出現的 Shannon 熵，50 年代末以解決遍歷理論 (Ergodic Theory) 經典問題而嶄露頭角的 Kolmogorov 熵，以及 60 年代中期，爲研究拓樸動力系統 (Topological Dynamical System) 而產生的拓樸熵 (Topological Entropy) 等概念，都是關於不確定性的數學度量。它們在現代動力系統和遍歷理論中，扮演着十分重要的角色。在自然科學和社會科學中的應用也日趨廣泛。本文的主旨在於引導盡量多的讀者在這一引人入勝的領域中尋幽訪勝，而不必在艱深的數學語言中躑躅不前。物理、化學家們也許對他們早已熟悉的熱力學熵更覺親切。我們在最後一節也將給古典的 Boltzmann 熵作一番數學的描述。

§1. Shannon 熵

設想我們有兩枚五分硬幣，一枚硬幣表面光滑，材料均勻，而另一枚硬幣則表面粗糙，奇形怪狀。我們把硬幣上有人頭的那面叫正面，另一面稱反面。然後在一個光滑的桌面上旋轉硬幣，等它停下來後，看是正面或是反面。這是一個不確定性的問題：可能是正面，可能是反面。第一枚硬幣，由於正面和反面的對稱性，正面或反面朝上的機率各爲一半。但對第二枚硬幣來說，由於材料磨損，正面和反面不再對稱。可能正面朝上的機率爲 70 % ，反面朝上的機率爲 30 % 。對「究竟會是正面？或會是反面？」這一不確定性問題來說，第一枚硬幣「不確定」的程度顯然比第二枚硬幣要大了許多。若要下賭注的話，我想還是下第二枚硬幣的正面朝上，較爲保險，不是嗎？現在假設鑄幣局的先生們別出心裁，把硬幣設計成圖 1-1 所示的形狀，其上爲正，其下爲反，則無論我們怎樣旋轉它，最終總是正面朝上。它「不確定」的度量應該爲零——其結果在未旋轉

圖 1-1

前都已確定，那來什麼「不」確定度呢？

有了這些直接的觀察，我們可以在數學上做文章了。假設樣本空間（Sample space）X有n個基本事件（events），其基本事件w_i的概率爲p_i，$i=1,2,\cdots\cdots,n$。我們記之爲$(X;p_1,\cdots\cdots,p_n)$。當然，我們有基本關係式$\sum_{i=1}^{n} p_i=1$，$p_i\geqslant 0$，$i=1,2,\cdots\cdots,n$。我們要定義一個函數H它的定義域是所有的樣本空間，它在樣本空間$(X;p_1,\cdots\cdots,p_n)$的值，我們用$H(p_1,\cdots\cdots,p_n)$來表示（X省略掉）我們要拿這個數來刻劃具有概率分別爲$p_1,p_2,\cdots\cdots,p_n$的事件$w_1,w_2,\cdots\cdots,w_n$的樣本空間的「不確定度」。$H(p_1,\cdots\cdots,p_n)$若要精確地反映試驗結果的不確定度，似乎必須滿足下列三個基本條件：

(i)**對固定n來說，H是$(p_1,\cdots\cdots,p_n)$的連續函數：**（這是數學上很基本的要求）

代替硬幣，讓我們來擲骰子。這骰子是個材料均勻各面光滑的正六邊體。當我們將它擲到桌面上時，每個面朝上的機率都是$\frac{1}{6}$。究竟是那面朝上的不確定度，顯然比旋轉光滑對稱硬幣那面朝上的不確定度要大許多。這個事實若用H來表達，應當是$H(\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6})>H(\frac{1}{2},\frac{1}{2})$。一般來說，$H$應當滿足。

(ii)**若$p_i=\frac{1}{n}$，$i=1,\cdots\cdots,n$，則對應的$H(\frac{1}{n},\cdots\cdots,\frac{1}{n})$應當是$n$的單調遞增函數。**

現在有一筆研究經費要分配給工程系的一名教授或數學系的兩名教授之一。假設工程系教授A獲得這筆經費的可能性是$\frac{1}{2}$，數學系教授B獲此經費的可能性爲$\frac{1}{3}$，而數學系教授C獲此經費的可能性爲$\frac{1}{6}$。事實上，這筆經費現在在教務長那裡，他認爲爲了公平起見，工程系獲此資助的可能性爲$\frac{1}{2}$，而數學系獲此資助的可能性亦爲$\frac{1}{2}$。工程系若獲此資助，系主任只會給教授A，沒有其他的候選人。但在數學系教授獲資助的前提下，教授B獲資助的可能性爲$\frac{2}{3}$，而教授C獲資助的可能性爲$\frac{1}{3}$（見圖1-2），這兩種「絕對不確定」和「相對不確定」分析應給出同樣的結果，也就是說，教授A,B,C，獲此研究費的不確定度，$H(\frac{1}{2},\frac{1}{3},\frac{1}{6})$應當等於教務長將它分給工程系或數學系的不確定度，$H(\frac{1}{2},\frac{1}{2})$加上**若是**分到數學系，教授$B$或教授$C$得此資助的不確定度

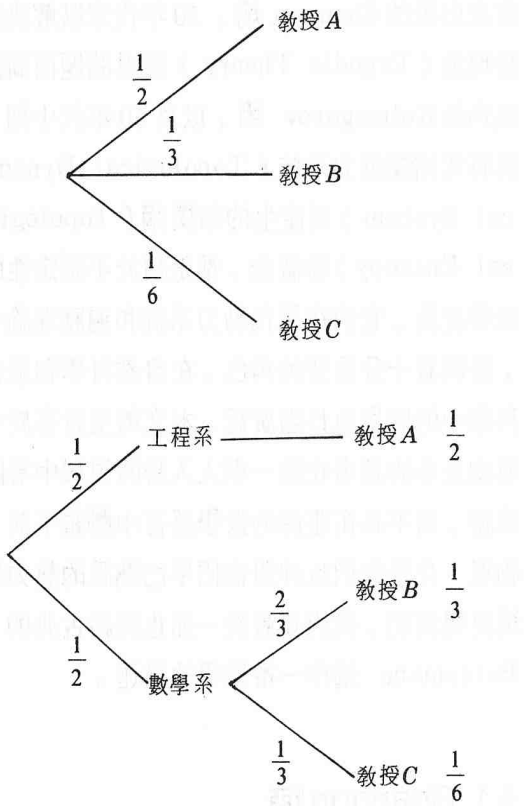

圖 1-2

$H(\frac{2}{3},\frac{1}{3})$，但這個不確定度是在此經費分到數學系的前提下。這種可能只有$\frac{1}{2}$，因此，

$$H(\frac{1}{2},\frac{1}{3},\frac{1}{6})$$

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

將此分析一般化，我們有下列的條件：

(iii)**若某一試驗分解成多個相繼的試驗，則原先的H值應爲相應的各個H值之加權和（weighted sum）。**

下面我們來證明一個重要結論：

定理1-1：滿足條件(i)、(ii)和(iii)的函數H恰好具有形式

$$H(p_1,\cdots,p_n)=-K\sum_{i=1}^{n}p_i\log p_i \quad (*)$$

其中K爲某個固定正常數。

證明：我們分三步來證明此定理。

第一步：

記$A(n)=H(\frac{1}{n},\frac{1}{n},\cdots\cdots,\frac{1}{n})$，$n$爲正整數。

斷言：

$A(s^m)=mA(s)$，其中s和m均爲正整數。

我們先對$s=2$，$m=3$用下列圖1-3所示來證明此斷言。即我們要證明$H(\frac{1}{8},\cdots,\frac{1}{8})=3H(\frac{1}{2},\frac{1}{2})$。由條件(iii)得

$$H(\frac{1}{8},\cdots\cdots,\frac{1}{8})$$

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

$$+\left[\frac{1}{4}H(\frac{1}{2},\frac{1}{2})+\frac{1}{4}H(\frac{1}{2},\frac{1}{2})\right]$$

$$+\frac{1}{4}H(\frac{1}{2},\frac{1}{2})+\frac{1}{4}H(\frac{1}{2},\frac{1}{2})\Big]$$

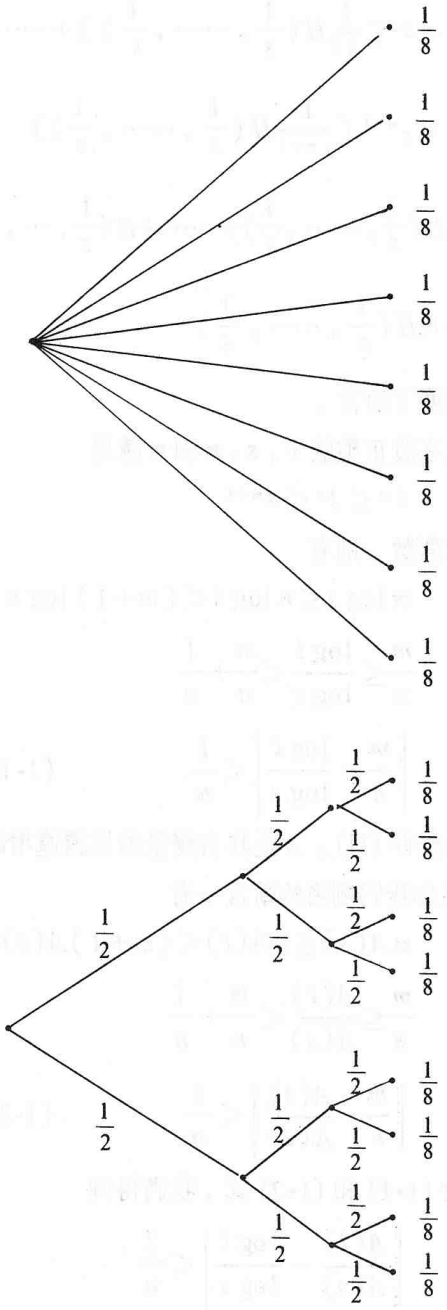

圖1-3

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

$$=3H(\frac{1}{2},\frac{1}{2})$$

由歸納法易知，一般地有

$$H(\frac{1}{s^m},\cdots\cdots,\frac{1}{s^m})$$

$$=H(\frac{1}{s},\cdots,\frac{1}{s})+s\left[\frac{1}{s}H(\frac{1}{s},\cdots,\frac{1}{s})\right]$$

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

$$+s^{m-1}\left(\frac{1}{s^{m-1}}H\left(\frac{1}{s},\cdots\cdots,\frac{1}{s}\right)\right]$$

$$=H\left(\frac{1}{s},\cdots\cdots,\frac{1}{s}\right)+\cdots\cdots+H\left(\frac{1}{s},\cdots,\frac{1}{s}\right)$$

$$=mH\left(\frac{1}{s},\cdots\cdots,\frac{1}{s}\right)$$

這就證明了斷言。

現在設正整數 t，s，n 和 m 滿足

$$s^m\leq t^n<s^{m+1}$$

兩邊取對數，則有

$$m\log s\leq n\log t<(m+1)\log s,$$

即
$$\frac{m}{n}\leq\frac{\log t}{\log s}<\frac{m}{n}+\frac{1}{n}$$

故有
$$\left|\frac{m}{n}-\frac{\log t}{\log s}\right|<\frac{1}{n}\qquad(1\text{-}1)$$

由條件 (ii)，A 是其自變量的單調遞增函數，且由我們剛證的斷言，有

$$mA(s)\leq nA(t)<(m+1)A(s)$$

即
$$\frac{m}{n}\leq\frac{A(t)}{A(s)}<\frac{m}{n}+\frac{1}{n}$$

故有
$$\left|\frac{m}{n}-\frac{A(t)}{A(s)}\right|<\frac{1}{n}\qquad(1\text{-}2)$$

由 (1-1) 和 (1-2) 式，我們得到

$$\left|\frac{A(t)}{A(s)}-\frac{\log t}{\log s}\right|<\frac{2}{n}$$

因爲 n 可以取任意自然數，而上式左邊與 n 無關，故有

$$\frac{A(t)}{A(s)}=\frac{\log t}{\log s}$$

或
$$\frac{A(t)}{\log t}=\frac{A(s)}{\log s}\equiv K$$

其中 K 爲一固定正常數，這樣我們有

$$A(t)=K\log t$$

由此，

$$H\left(\frac{1}{n},\cdots\cdots,\frac{1}{n}\right)=K\log n=-K\sum_{i=1}^{n}\frac{1}{n}\log\frac{1}{n}$$

即，本定理對特殊情形 $p_i=\frac{1}{n}$，$i=1$，……，n 成立。

第二步：

現在對 p_i 取一般的非負有理數來證明此定理，我們對 $p_1=\frac{1}{2}$，$p_2=\frac{1}{3}$，$p_3=\frac{1}{6}$ 來描述證明的思想，作出下列圖 1-4 。

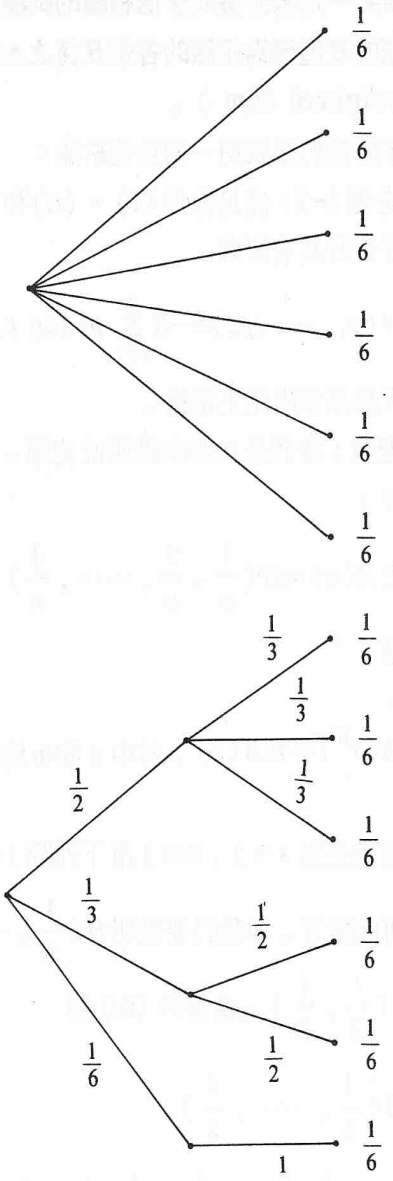

圖 1-4

根據條件 (iii)

$$H\left(\frac{1}{6},\cdots\cdots,\frac{1}{6}\right)$$

$$=H(\frac{1}{2},\frac{1}{3},\frac{1}{6})+\frac{1}{2}H(\frac{1}{2},\frac{1}{3},\frac{1}{3})$$

$$+\frac{1}{3}H(\frac{1}{2},\frac{1}{2})+\frac{1}{6}H(1)$$

故有

$$H(\frac{1}{2},\frac{1}{3},\frac{1}{6})$$

$$=H(\frac{1}{6},\cdots\cdots,\frac{1}{6})-\frac{1}{2}H(\frac{1}{3},\frac{1}{3},\frac{1}{3})$$

$$-\frac{1}{3}H(\frac{1}{2},\frac{1}{2})-\frac{1}{6}H(1)$$

這樣分解的目的在於我們可用第一步證明的結果來證明第二步。

令 $n_1=3$，$n_2=2$，$n_3=1$，則

$$p_1=\frac{1}{2}=\frac{n_1}{n_1+n_2+n_3}$$

$$p_2=\frac{1}{3}=\frac{n_2}{n_1+n_2+n_3}$$

$$p_3=\frac{1}{6}=\frac{n_3}{n_1+n_2+n_3}$$

將上面的結果抽象化，我們就有，

$$H(p_1,p_2,p_3)$$

$$=A(\sum_{i=1}^{3}n_i)-\sum_{i=1}^{3}p_i\,A(n_i)\ 。$$

對一般情形，我們可依同法處理。設 $p_1,\cdots\cdots,p_r$ 爲非負有理數，滿足 $\sum_{i=1}^{r}p_i=1$，則存在自然數 $n_1,\cdots\cdots,n_r$，使得

$$p_i=\frac{n_i}{\sum_{j=1}^{r}n_j},\quad i=1,\cdots\cdots,r$$

利用條件 (iii)，我們得到如下等式

$$H(p_1,p_2,\cdots\cdots,p_r)$$

$$=A(\sum_{i=1}^{r}n_i)-\sum_{i=1}^{r}p_i\,A(n_i)$$

由第一步證明之結果，$A(n)=K\log n$ 代入上式有

$$H(p_1,\cdots\cdots,p_r)$$

$$=K\log(\sum_{i=1}^{r}n_i)-\sum_{i=1}^{r}p_i(K\log n_i)$$

$$=K[\sum_{i=1}^{r}p_i\log(\sum_{j=1}^{r}n_j)]-K\sum_{i=1}^{r}p_i\log n_i$$

$$=-K\sum_{i=1}^{r}p_i\log\frac{n_i}{\sum_{j=1}^{r}n_j}$$

$$=-K\sum_{i=1}^{r}p_i\log p_i$$

故我們證明了 (*) 式對任何滿足 $\sum_{i=1}^{r}p_i=1$ 的非負有理數 $p_1,\cdots\cdots,p_r$ 成立。

第三步：

設 $p_1,\cdots\cdots,p_r$ 爲任意非負實數，$\sum_{i=1}^{r}p_i=1$。由條件 (i)，H 爲 $p_1,\cdots\cdots,p_r$ 的連續函數，而任何實數均可由有理數列來任意逼近，故第二步證明結果隱含了 (*) 式在實數情形之正確性。定理證畢。

由定理中 (*) 式可知，若對某一個 i 有 $p_i=1$，則 $H(p_1,\cdots\cdots,p_n)=0$，這正好和我們的願望相符：$p_i=1$ 意味着對應的事件總是發生的，因而不確定度爲零。

因此，我們可以給出如下關於**熵**的定義，這個定義的**熵**（entropy），又稱爲 shannon 熵。

定義 1-2：由式 $H(p_1,\cdots\cdots,p_n)=-\sum_{i=1}^{n}p_i\log p_i$ 定義的數，$H(p_1,\cdots\cdots,p_n)$ 稱爲對應於樣本空間 $(X,p_1,\cdots\cdots,p_n)$ 的**熵**。

在本節之初，我們已知旋轉光滑硬幣時，正面朝上有 $\frac{1}{2}$ 的機率，反面朝上也有 $\frac{1}{2}$ 的機率

，它的不確定度爲最大。既然熵是關於不確定度的一種數學度量，這就自然地要求當 $p_1=p_2=\cdots\cdots=p_n=\frac{1}{n}$ 時，H給出最大值。要注意的是，我們在推導H表達式的三個基本條件中，並無強加此項要求。現在我們要證明：這個直觀的要求，事實上可由上述三個基本條件推出結論。

命題 1-3：

設 $H(p_1,\cdots,p_n)=-\sum_{i=1}^{n} p_i\log p_i$ 則

$$H(\frac{1}{n},\cdots\cdots,\frac{1}{n})=\log n$$

$$=\max\{H(p_1,\cdots,p_n):p_i\geqslant 0,\ \sum_{i=1}^{n} p_i=1\}。$$

證明：由初等微積分知函數 $\log\mu$ 是 μ 的嚴格凹函數。任給 $p_1,\cdots,p_n>0$，$\sum_{i=1}^{n} p_i=1$，

$$H(p_1,\cdots\cdots,p_n)$$

$$=-\sum_{i=1}^{n} p_i\log p_i$$

$$=\sum_{i=1}^{n} p_i\log\frac{1}{p_i}\leqslant\log(\sum_{i=1}^{n}\frac{p_i}{p_i})$$

$$=\log n=H(\frac{1}{n},\cdots\cdots,\frac{1}{n})。$$

當某一 p_i 爲零時，比如說 $p_i=0$。這就好像一個只有 $n-1$ 個基本事件的樣本空間。由上面的推論

$$H(0,p_2,\cdots,p_n)\leqslant H(\frac{1}{n-1},\cdots,\frac{1}{n-1})$$

$$<H(\frac{1}{n},\cdots\cdots,\frac{1}{n})$$

第二個不等號是由於條件 (ii)。

這節定義的熵起源於信息理論的研究。是 C. Shannon 在 1948 年引進的。在此基礎上，蘇聯數學家 A. N. Kolmogorov 在 1958 年給出了動力系統熵的概念。從而揭開了現代遍歷理論研究的新篇章。

§2. Kolmogorov熵

我們再來做旋轉光滑硬幣的遊戲。爲了方便起見，我們稱硬幣的正面爲 1，反面爲 0。讓我們考察連續旋轉 n 次，其每次正反面出現的各種可能性。旋轉一次，有兩個可能性，或正面朝上，或反面朝上，即 1，0；旋轉兩次有 $4=2^2$ 種可能性，即 11, 10, 01, 00; 一般來說，旋轉 n 次則有 2^n 種可能性。把連續旋轉 n 次的任一可能結果看成一個「基本事件」，我們則得到一個具有 2^n 個基本事件的樣本空間，其每一基本事件有同樣的概率 2^{-n}。上節中所談 Shannon 熵給出了這個樣本空間的不確定度——$n\log 2$。現在我們要進一步問的是：如果我們已知旋轉硬幣第一次，第二次，……，第 $n-1$ 的結果，那麼第 n 次會是正面或會是反面的不確定度該是多少？

我們希望能用數學上的語言來描述這個問題。首先讓我們來考慮定義在〔0，1〕上的函數 $f(x)=2x\ (\mathrm{mod}\ 1)$，也就是

$$f(x)=\begin{cases}2x & 0\leq x<\frac{1}{2}\\ 2x-1 & \frac{1}{2}\leqslant x\leqslant 1\end{cases}。$$

（見圖2-1），取Lebesgue 測度m做爲〔0，1〕上的測度，令 $\bar{A}=\{〔0,\frac{1}{2}〕,〔\frac{1}{2},1〕\}$爲〔0，1〕上的一個劃分（partition），則 $f^{-1}(\bar{A})=\{f^{-1}(〔0,\frac{1}{2}〕),f^{-1}(〔\frac{1}{2},1〕)\}=\{〔0,\frac{1}{4}〕\cup〔\frac{1}{2},\frac{3}{4}〕,〔\frac{1}{4},\frac{1}{2}〕\cup〔\frac{3}{4},1〕\}$ 也是〔0，1〕上的一個劃分。任給兩個劃分 $\bar{A}$ 和 $\bar{B}$，令 $\bar{A}\vee\bar{B}$ 爲由

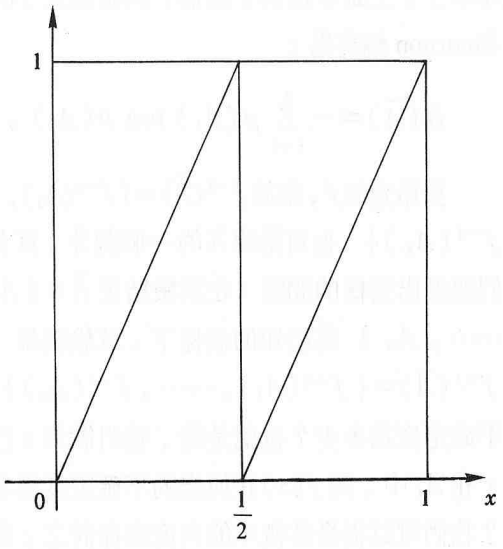

圖 2-1　$f(x)=2x\ (\mathrm{mod}\ 1)$

下式定義的劃分

$$\overline{A}\vee\overline{B}=\{A\cap B: A\in\overline{A},\ B\in\overline{B}\}。$$

由此，我們則有 $f^{-1}(\overline{A})\vee\overline{A}=\{[0,\frac{1}{4}),[\frac{1}{4},\frac{1}{2}),[\frac{1}{2},\frac{3}{4}),[\frac{3}{4},1)\}$。

如此這般繼續下去，我們會有

$$\bigvee_{i=0}^{n-1} f^{-i}(\overline{A})$$

$$=\{[\frac{i-1}{2^n},\frac{i}{2^n}): i=1,\cdots\cdots,2^n\}$$

的劃分，這個劃分裡的每個區間 $[\frac{i-1}{2^n},\frac{i}{2^n})$ 都有 2^{-n} 的Lebesgue 概率測度。事實上，它和旋轉硬幣 n 次那個樣本空間裡的 2^n 個基本事件是一一對應的。

拿 $n=3$ 其中的一個簡單情況來看。把 $[\frac{3}{8},\frac{4}{8})$ 這個區間左端的 $\frac{3}{8}$ 寫成

$$\frac{3}{8}=\frac{0}{2}+\frac{1}{2^2}+\frac{1}{2^3}$$

然後將 $[\frac{3}{8},\frac{4}{8})$ 這個區間和 011（第一次反面，第二次正面，第三次反面）對應。一般來說，我們可以把 $[\frac{i-1}{2^n},\frac{i}{2^n})$ 這個區間左端的 $\frac{i-1}{2^n}$ 寫成

$$\frac{i-1}{2^n}=\frac{a_1}{2}+\frac{a_2}{2^2}+\cdots\cdots+\frac{a_n}{2^n}$$

其中 $a_k=0$ 或 1，$k=1,\cdots\cdots,n$。這個區間對應的是旋轉硬幣 n 次，出現 $a_1a_2\cdots\cdots a_n$ 的基本事件。總的來說，旋轉硬幣 n 次，2^n 個基本事件，大家的機率都是 2^{-n} 的樣本空間，拿 $f(x)=2x\ (\mathrm{mod}\ 1)$ 和劃分 $\overline{A}=\{[0,\frac{1}{2}),[\frac{1}{2},1)\}$ 來描述，則是：拿劃分 $\bigvee_{i=0}^{n-1} f^{-1}(\overline{A})$ 裡的 2^n 個元素 $[\frac{i-1}{2^n},\frac{i}{2^n})$ 做基本事件，大家的Lebesgue 概率測度都是 2^{-n} 的樣本空間。

「已知旋轉硬幣第一次，第二次，……，第 $n-1$ 次的結果，那麼第 n 次會是正面或反面的不確定度是多少？」的這一問題，拿 $f(x)=2x\ (\mathrm{mod}\ 1)$ 和劃分 $\overline{A}=\{[0,\frac{1}{2}),[\frac{1}{2},1)\}$ 來描述，事實上是在問：已知 x，……，$f^{n-1}(x)$ 在劃分 $\overline{A}$ 裡的位置，那麼 $f^n(x)$ 會在 $[0,\frac{1}{2})$ 裡或在 $[\frac{1}{2},1)$ 裡的不確度是多少？

讓我們來看 $n=4$ 這個特殊情形。比如說我們已知前三次的結果，它們是 101（第一次正面，第二次反面，第三次正面），這在 $\bigvee_{i=0}^{2} f^{-i}(\overline{A})$ 中所對應的區間是 $[\frac{5}{2^3},\frac{6}{2^3})$，因爲

$$\frac{5}{2^3}=\frac{1}{2}+\frac{0}{2^2}+\frac{1}{2^3}。$$

仔細的看，這個間區事實上是，$[\frac{1}{2},1)$ 和 $f^{-1}([0,\frac{1}{2}))=[0,\frac{1}{4})\cup[\frac{1}{2},\frac{3}{4})$ 以及

$f^{-2}([\frac{1}{2},1])=[\frac{1}{8},\frac{1}{4}]\cup[\frac{3}{8},\frac{1}{2}]\cup[\frac{5}{8},\frac{3}{4}]\cup[\frac{7}{8},1]$ 的交集，也就是說

$$[\frac{5}{2^3},\frac{6}{2^3}]$$

$$=[\frac{1}{2},1]\cap f^{-1}([0,\frac{1}{2}])\cap f^{-2}([\frac{1}{2},1])$$

元素 x 在這交集所代表的意義是：$x\in[\frac{1}{2},1]$，$f(x)\in[0,\frac{1}{2}]$ 和 $f^2(x)\in[\frac{1}{2},1]$。一般來說，已知前三次旋轉硬幣的結果相當於已知 x，$f(x)$，$f^2(x)$ 在劃分 $\overline{A}=\{[0,\frac{1}{2}],[\frac{1}{2},1]\}$ 中的位置。問第四次是正面還是反面的不確定度，相當於問 $f^3(x)$ 究竟是在 $[0,\frac{1}{2}]$ 中還是在 $[\frac{1}{2},1]$ 中的不確定度。

已知 x，$f(x)$，……，$f^{n-1}(x)$ 在那裡，問 $f^n(x)$ 在那裡的不確定度，當 n 趨近於無窮大時的變化就是我們在這一節要談的Kolmogorov 熵。

我們將把我們的着眼點放在一般的**概率測度空間**（Probability measure space）和定義在它上面的**可測變換**（measurable function）。設 (X,Σ,μ) 爲一概率測度空間。即 X 爲一集合，Σ 爲 X 上的一些子集合所構成的一個 σ-代數，μ 爲 Σ 上的概率測度，也就是說 $\mu(X)=1$。假設 $f:X\to X$ 爲一個**可測變換**。這是指，Σ 中每個元素 A 的逆像 $f^{-1}(A)$ 仍在 Σ 中。我們任取 X 的一個**有限劃分**（finite partition）$\overline{A}=\{A_1,\cdots\cdots,A_m\}$ 即 $\overline{A}$ 中每個集合 A_i 屬於 Σ，它們之間互不相交（交集的測度爲 0）且聯集恰爲 X。這樣 $\overline{A}$ 可看成具有「基本事件」$A_1,A_2,\cdots\cdots,A_m$ 且有概率分布 $\mu(A_1),\cdots\cdots,\mu(A_m)$ 的一個有限樣本空間。這個樣本空間經常被稱爲「**試驗結果**」。上節中談到，這個「試驗結果」的 Shannon 熵應爲：

$$H(\overline{A})=-\sum_{i=1}^{n}\mu(A_i)\log\mu(A_i)。$$

對給定的 f，集族 $f^{-1}(\overline{A})=\{f^{-1}(A_1),\cdots,f^{-1}(A_m)\}$ 也可給出 X 的一個劃分。首先我們要提出這樣的問題：在試驗結果 $\overline{A}=\{A_1,\cdots\cdots,A_m\}$ 爲已知的前提下，試驗結果 $f^{-1}(\overline{A})=\{f^{-1}(A_1),\cdots\cdots,f^{-1}(A_m)\}$ 的不確定度爲多少？也就是說，我們欲知：已知 x 在 A_i 中，問 $f(x)$ 在何處的不確定度爲多少？我們可以從條件概率的角度來探討之。爲簡單起見，設 $n=3$，即 $\overline{A}=\{A_1,A_2,A_3\}$。假如，已知 x 在 A_1 中我們來看 $f(x)$ 在 A_1，A_2 或 A_3 的概率爲如何。對 $i=1,2,3$，$f(x)\in A_i$，當且僅當 $x\in f^{-1}(A_i)$，故 x 在 A_1 中且 $f(x)$ 在 A_i 中之集合爲 $A_1\cap f^{-1}(A_i)$，因而其條件概率爲 $\mu(A_1\cap f^{-1}(A_i))/\mu(A_1)$。由 Shannon 熵的定義知，在 $x\in A_1$ 的條件下，$f(x)$ 會在 A_1，或 A_2，或 A_3 的不確度應爲

$$H_1=-\sum_{i=1}^{3}\frac{\mu(A_1\cap f^{-1}(A_i))}{\mu(A_1)}\times\log\frac{\mu(A_1\cap f^{-1}(A_i))}{\mu(A_1)}。$$

類似地，在 $x\in A_2$ 或 $x\in A_3$ 的條件下，試驗結果 $f^{-1}(\overline{A})=\{f^{-1}(A_1),f^{-1}(A_2),f^{-1}(A_3)\}$ 的不確定度應分別爲

$$H_2=-\sum_{i=1}^{3}\frac{\mu(A_2\cap f^{-1}(A_i))}{\mu(A_2)}\times\log\frac{\mu(A_2\cap f^{-1}(A_i))}{\mu(A_2)}$$

和

$$H_3=-\sum_{i=1}^{3}\frac{\mu(A_3\cap f^{-1}(A_i))}{\mu(A_3)}\times\log\frac{\mu(A_3\cap f^{-1}(A_i))}{\mu(A_3)}。$$

由推導 Shannon 熵定義的條件 (ii) 易知，在試驗結果 $\overline{A}=\{A_1,A_2,A_3\}$ 爲已知的條件下，試驗結果 $f^{-1}(\overline{A})=\{f^{-1}(A_1),f^{-1}(A_2),$

$f^{-1}(A_3)\}$的不確定度$H(f^{-1}(\overline{A})\mid\overline{A})$爲$H_1$, H_2和H_3的加權和，即

$$H(f^{-1}(\overline{A})\mid\overline{A})=\sum_{i=1}^{3}\mu(A_i)H_i$$

$$=-\sum_{i=1}^{3}\sum_{j=1}^{3}\mu(A_1\cap f^{-1}(A_j))$$

$$\times\log\frac{\mu(A_i\cap f^{-1}(A_j))}{\mu(A_i)}。$$

如法炮製，對一般的有限劃分$\overline{A}=\{A_1,\cdots\cdots,A_m\}$，我們可得到所謂的「劃分$f^{-1}(\overline{A})$關於劃分$\overline{A}$的條件shannon熵」，

$$H(f^{-1}(\overline{A})\mid\overline{A})$$

$$=-\sum_{j=1}^{m}\sum_{i=1}^{m}\mu(A_i\cap f^{-1}(A_j))$$

$$\times\log\frac{\mu(A_i\cap f^{-1}(A_j))}{\mu(A_i)}。$$

下面，我們來給出上述$H(f^{-1}(\overline{A})\mid\overline{A})$的另一等價形式以便後面推廣。

命題2-1：

$$H(f^{-1}(\overline{A})\mid\overline{A})=H(\overline{A}\vee f^{-1}(\overline{A}))-H(\overline{A})$$

證明：

$$H(f^{-1}(\overline{A})\mid\overline{A})$$

$$=-\sum_{i=1}^{m}\sum_{j=1}^{m}\mu(A_i\cap f^{-1}(A_j))$$

$$\times\log\frac{\mu(A_i\cap f^{-1}(A_j))}{\mu(A_i)}$$

$$=-\sum_{i=1}^{m}\sum_{j=1}^{m}\mu(A_i\cap f^{-1}(A_j))$$

$$\times[\log\mu(A_i\cap f^{-1}(A_j))-\log\mu(A_i)]$$

$$=-\sum_{i=1}^{m}\sum_{j=1}^{m}\mu(A_i\cap f^{-1}(A_j))$$

$$\times\log\mu(A_i\cap f^{-1}(A_j))$$

$$+\sum_{i=1}^{m}\sum_{j=1}^{m}\mu(A_i\cap f^{-1}(A_j))\log\mu(A_i)$$

$$=H(\overline{A}\vee f^{-1}(\overline{A}))+\sum_{i=1}^{m}\mu(A_i)\log\mu(A_i)$$

$$=H(\overline{A})\vee f^{-1}(\overline{A}))-H(\overline{A})\quad\#$$

命題2-1在直觀上看也很顯然：試驗結果$\overline{A}\vee f^{-1}(\overline{A})$的不確定度$H(\overline{A}\vee f^{-1}(\overline{A}))$應爲試驗結果$\overline{A}$的不確定度$H(\overline{A})$和在試驗結果$\overline{A}$爲已知條件下，試驗結果$f^{-1}(\overline{A})$的不確定度$H(f^{-1}(\overline{A})\mid\overline{A})$之和。

上述已知試驗結果$\overline{A}$，問試驗結果$f^{-1}(\overline{A})$的不確定度，相當於已知x在$\overline{A}$中的位置，我們問$f(x)$在$\overline{A}$中的位置的不確定度。已知x, $f(x)$, $\cdots\cdots$, $f^{n-1}(x)$在分劃$\overline{A}$中的位置，問$f^n(x)$在$\overline{A}$中的位置的不確定度，則相當於已知試驗結果$\bigvee_{i=0}^{n-1}f^{-i}(\overline{A})=\overline{A}\vee f^{-1}(\overline{A})\vee\cdots\cdots\vee f^{-(n-1)}(\overline{A})$，問試驗結果$f^{-n}(\overline{A})$的不確定度。Kolmogorov熵基本上是在刻劃這個不確定度在當n趨近於無窮大時的漸近性質。

任給自然數n，$\bigvee_{i=0}^{n-1}f^{-i}(\overline{A})$和$f^{-n}(\overline{A})$都是$X$的有限劃分。在已知試驗結果，$\bigvee_{i=0}^{n-1}f^{-i}(\overline{A})$的條件下，試驗結果$f^{-n}(\overline{A})$的不確定度，實際上是劃分$f^{-n}(\overline{A})$關於劃分$\bigvee_{i=0}^{n-1}f^{-i}(\overline{A})$的條件Shannon熵，它是

$$H(f^{-n}(\overline{A})\Big|\bigvee_{i=0}^{n-1}f^{-i}(\overline{A}))$$

$$=H(\bigvee_{i=0}^{n}f^{-i}(\overline{A}))-H(\bigvee_{i=0}^{n-1}f^{-i}(\overline{A}))。$$

定義2-2：設$\overline{A}=\{A_1,\cdots\cdots,A_m\}$爲$X$的有限劃分，則可測變換$f:X\to X$關於$\overline{A}$的熵定義爲

$$h_\mu(f,\overline{A})$$

$$=\lim_{n\to\infty}\sup H(f^{-n}(\overline{A})\Big|\bigvee_{i=0}^{n-1}f^{-i}(\overline{A}))$$

定義2-3：設(X,Σ,μ)爲一概率空間，$f:X\to X$爲一可測變換，則f的Kolmogorov熵定義爲

$$h_\mu(f)=\sup\{h_\mu(f,\overline{A}):\overline{A}\text{ 爲}X\text{的有限劃分}\}。$$

對一般的可測變換 $f:X\to X$，上述定義2-2中的上極限符號不能改爲極限符號。但對遍歷理論中所研究的一類重要可測變換——保測變換（measure preserving transformation）我們可以證明極限 $\lim_{n\to\infty} H(f^{-n}(\overline{A})\mid\bigvee_{i=1}^{n-1} f^{-i}(\overline{A}))$ 確實存在並有另一等價定義。該定義顯然不及前者直觀易懂，但它却給出了計算上的許多方便。所謂保測變換是指 $f:X\to X$，任給 $A\in\Sigma$，$f^{-1}(A)\in\Sigma$ 且有

$$\mu(f^{-1}(A))=\mu(A)。$$

定義 2-2：設 $\overline{A}=\{A_1,\cdots\cdots,A_m\}$ 爲 X 的有限劃分，則保測變換 $f:X\to X$ 關於 $\overline{A}$ 的熵定義爲

$$h_\mu(f,\overline{A})=\lim_{n\to\infty}\frac{1}{n}H(\bigvee_{i=0}^{n-1} f^{-i}(\overline{A}))。$$

在證明此定義合理，且與定義2-2等價之前，我們首先注意到如下事實：若 f 爲保測變換，則 $H(f^{-1}(\overline{A}))=H(\overline{A})$ 這由條件 $\mu(f^{-1}(A))=\mu(A)$ 易見。若 $\overline{C}$ 和 $\overline{D}$ 爲 X 的兩個有限劃分，我們記 $\overline{C}\leqslant\overline{D}$，若 $\overline{C}$ 的每一元素是 $\overline{D}$ 中某些元素之聯（Union）（即 $\overline{D}$ 是 $\overline{C}$ 的一個細分（refinement））我們需要下列引理，其證明稍後給出。

引理 2-4：若 $\overline{C}\leqslant\overline{D}$，則 $H(\overline{A}\mid\overline{C})\geqslant H(\overline{A}\mid\overline{D})$。

現在可以敍述並證明我們的等價定理了。

定理 2-5：若 $f:X\to X$ 爲保測變換，則對 X 的任一有限劃分 $\overline{A}$

$$\lim_{n\to\infty} H(f^{-n}(\overline{A})\mid\bigvee_{i=0}^{n-1} f^{-i}(\overline{A}))$$

$$=\lim_{n\to\infty}\frac{1}{n}H(\bigvee_{i=0}^{n-1} f^{-i}(\overline{A}))$$

證明：$n=1$，則

$$H(f^{-1}(\overline{A})\mid\overline{A})$$
$$=H(f^{-1}(\overline{A})\vee\overline{A})-H(\overline{A})$$
$$=H(f^{-1}(\overline{A})\vee\overline{A})-H(f^{-1}(\overline{A}))$$
$$=H(\overline{A}\mid f^{-1}(\overline{A}))$$

當 $n=2$ 時

$$H(f^{-2}(\overline{A})\mid\overline{A}\vee f^{-1}(\overline{A}))$$
$$=H(f^{-2}(\overline{A})\vee f^{-1}(\overline{A})\vee\overline{A})$$
$$\quad-H(f^{-1}(\overline{A})\vee\overline{A})$$
$$=H(f^{-2}(\overline{A})\vee f^{-1}(\overline{A})\vee\overline{A})$$
$$\quad-H(f^{-2}(\overline{A})\vee f^{-1}(\overline{A}))$$
$$=H(\overline{A}\mid f^{-2}(\overline{A})\vee f^{-1}(\overline{A}))$$

用歸納法易證，一般地有

$$H(f^{-n}(\overline{A})\mid\bigvee_{i=0}^{n-1} f^{-i}(\overline{A}))$$

$$=H(\overline{A}\mid\bigvee_{i=0}^{n} f^{-i}(\overline{A}))。\quad\#$$

由上述引理2-4，$H(\overline{A}\mid\bigvee_{i=0}^{n} f^{-i}(\overline{A}))$ 是 n 的單調遞減函數，故極限存在。從而，定義2-2中的極限實際上存在。另一方面，對 $i=1,2,\cdots\cdots,n-1$，由

$$H(f^{-i}(\overline{A})\mid\bigvee_{j=0}^{i-1} f^{-j}(\overline{A}))$$

$$=H(\bigvee_{j=0}^{i} f^{-j}(\overline{A}))-H(\bigvee_{j=0}^{i-1} f^{-j}(\overline{A}))$$

各式相加，我們有

$$H(\bigvee_{i=0}^{n-1} f^{-i}(\overline{A}))$$

$$=H(\overline{A})+\sum_{i=1}^{n-1} H(f^{-i}(\overline{A})\mid\bigvee_{j=0}^{i-1} f^{-j}(\overline{A}))$$

$$=\sum_{i=0}^{n-1} H(f^{-i}(\overline{A})\mid\bigvee_{j=0}^{i-1} f^{-j}(\overline{A}))$$

$$=\sum_{i=0}^{n-1} H(\overline{A}\mid\bigvee_{j=0}^{i} f^{-j}(\overline{A}))$$

故有

$$\frac{1}{n}H(\bigvee_{i=0}^{n-1} f^{-i}(\overline{A}))$$

$$=\frac{1}{n}\sum_{i=0}^{n-1} H(\bar{A}\,|\,\bigvee_{j=0}^{i} f^{-j}(\bar{A}))$$

借用初等微積分的已知結果：$\lim_{n\to\infty} a_n = L \Rightarrow$

$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1} a_i = L$，我們得到

$$\lim_{n\to\infty}\frac{1}{n}H(\bigvee_{i=0}^{n-1} f^{-i}(\bar{A}))$$

$$=\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1} H(\bar{A}\,|\,\bigvee_{j=0}^{i} f^{-j}(\bar{A}))$$

$$=\lim_{n\to\infty} H(\bar{A}\,|\,\bigvee_{i=0}^{n} f^{-i}(\bar{A}))$$

$$=\lim_{n\to\infty} H(f^{-n}(\bar{A})\,|\,\bigvee_{i=0}^{n-1} f^{-i}(\bar{A}))\ 。$$

現在我們來證明引理 2-4 。

設$\bar{A}=\{A_i\}$，$\bar{C}=\{C_j\}$，$\bar{D}=\{D_k\}$，我們要證

$$-\sum_j\sum_i \mu(C_j)\frac{\mu(A_i\cap C_j)}{\mu(C_j)}\log\frac{\mu(A_i\cap C_j)}{\mu(C_j)}$$

$$\geqslant -\sum_k\sum_i \mu(D_k)\frac{\mu(A_i\cap D_k)}{\mu(D_k)}\log\frac{\mu(A_i\cap D_k)}{\mu(D_k)}$$

$$=-\sum_k\sum_i\sum_j \mu(C_j\cap D_k)\frac{\mu(A_i\cap D_k)}{\mu(D_k)}$$

$$\times\log\frac{\mu(A_i\cap D_k)}{\mu(D_k)}$$

只須證明對每一 i 和 j

$$\mu(C_j)\frac{\mu(A_i\cap C_j)}{\mu(C_j)}\log\frac{\mu(A_i\cap C_j)}{\mu(C_j)}$$

$$\leqslant\sum_k \mu(C_j\cap D_k)\frac{\mu(A_i\cap D_k)}{\mu(D_k)}\log\frac{\mu(A_i\cap D_k)}{\mu(D_k)}$$

令$\phi(x)=x\log x$　$\phi(0)=0$　則上式爲

$$\phi(\frac{\mu(A_i\cap C_j)}{\mu(C_j)})$$

$$\leqslant\sum_k\frac{\mu(C_j\cap D_k)}{\mu(C_j)}\,\phi\,(\frac{\mu(A_i\cap D_k)}{\mu(D_k)})。$$

由於ϕ是凸函數（這由$\phi''(x)=\frac{1}{x}>0$可知）和假設$\bar{C}\leqslant\bar{D}$，易知

$$\sum_k\frac{\mu(C_j\cap D_k)}{\mu(C_j)}\,\phi(\frac{\mu(A_i\cap D_k)}{\mu(D_k)})$$

$$\geqslant\phi(\sum_k\frac{\mu(C_j\cap D_k)}{\mu(C_j)}\,\frac{\mu(A_i\cap D_k)}{\mu(D_k)})$$

$$=\phi(\frac{\mu(A_i\cap C_j)}{\mu(C_j)})$$

即我們證明了$H(\bar{A}\,|\,\bar{C})\geqslant H(\bar{A}\,|\,\bar{D})$。　#

歷史上，引進 Kolmogorov 熵概念的主要動力是關於概率空間保測變換之間共軛關係的不變量的研究。設(X_1,Σ_1,μ_1)和(X_2,Σ_2,μ_2)爲二個概率空間，$T_1: X_1\to X_1$ 和 $T_2: X_2\to X_2$ 爲保測變換。我們說 T_1 和 T_2 共軛（conjugate）是指存在一個保測同構 $\phi:(X_2,\Sigma_2,\mu_2)\to(X_1,\Sigma_1,\mu_1)$ 使得 $\phi\circ T_2^{-1}=T_1^{-1}\circ\phi$。我們稱一個數量爲共軛保測變換的「不變量（invariance）」是指二個保測變換若是共軛，這個數量一定一樣。這個數量若不一樣，這兩個保測變換一定不共軛。共軛的保測變換具有同樣的遍歷性質。我們若能找到關於共軛保測變換的不變量，我們就可從本質上刻劃不同共軛類保測變換的特徵：Kolmogorov 熵就是這樣的一個重要的不變量。

早在 1943 年，人們就知道 Bernoulli 的 $(\frac{1}{2},\frac{1}{2})$－雙邊移位算子（two side shift）和 $(\frac{1}{3},\frac{1}{3},\frac{1}{3})$－雙邊移位算子都具有可數個 Lebesgue 譜點，因而是譜同構的，但不知道它們是否共軛。直到 1958 年才由 Kolmogorov 證明了它們分別具有 log 2 和 log 3 的 Kolmogorov 熵，故非共軛。從而消除了遍歷理論這個重大懸念，並開創了一個嶄新的研究領域。我們這裡介紹的 Kolmogorov 熵的概念是由 Kolmogorov 的學生 Sinai 在 1959 年改進的，和 Kolmogorov 1958 年給出的原始定義稍有不同。

(本文作者任教於美國密西根州立大學數學系)