如何找出劣幣?

—簡介訊息與熵的概念

蔡聰明

機率、訊息、熵、混沌與碎形是現代資訊革命所產生之新的科學典範 (paradigm)。作爲將要邁入21世紀的現代人, 很有必要去掌握這些概念。

本文我們僅輕輕觸及「機率、訊息、熵」這一面, 作最初步的淺介。

一、一個益智問題

> 在益智遊戲裡, 有這樣一個問題:
>
> 給27個外表完全相同的硬幣, 其中恰有一個的重量跟其它的不同, 叫做劣幣, 要利用天平把它秤出來, 問應如何秤法? 最少要秤幾次?

這是施展分析法與實驗試誤法的好問題, 求得解答並不困難。表面上看起來, 這個問題只是作爲 「茶餘飯後」 的益智談論, 但是如果認眞追根究柢, 它的背後卻涉及到「熵」(entropy) 的概念、訊息理論 (information theory)、遍歷理論 (ergodic theory) 等等, 這些就非常重要了。

因此, 這是一個好例子, 足以引動好奇心, 產生追尋過程, 進入一個美麗的天地。換言之, 這個「特例」具有推展到「一般理論」的妙趣。

二、已知劣幣是輕或重

不妨假設劣幣較輕 (較重的情形同理可求解)。首先將27枚硬幣均分成 A, B, C 三堆, 每一堆都含有9枚, 再按下面三個步驟就可以找到劣幣:

第一步: 任選兩堆, 例如 A 與 B, 分別置於天平的兩邊, 得到三種可能的結果:

(1) A ——Δ—— B

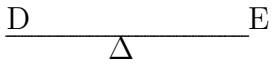
(2) A —Δ— B

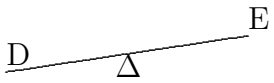
(3) A —Δ— B

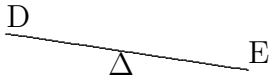
我們由 (1),(2),(3) 可知, 劣幣分別在 C, B 或 A 這一堆。

第二步: 將含劣幣這一堆, 再均分成 D, E, F 三堆, 每一堆都含有3枚, 再仿第一

步的辦法, 任取兩堆, 例如 D 與 E, 分別置於天平的兩邊, 得到三種可能結果:

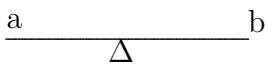
(4) 

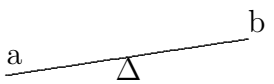
(5) 

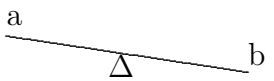
(6) 

對應於 (4),(5),(6) 三種情形, 可知劣幣分別在 F, E 或 D 這一堆, 令其三枚硬幣爲 a, b, c。

第三步: 從 a, b, c 中任取兩枚, 例如 a 與 b, 分別置於天平兩邊, 得到三種可能結果:

(7) 

(8) 

(9) 

由此立知, 對應於 (7),(8),(9) 三種情形, 劣幣分別爲 c 或 b 或 a。

結論: 秤三次必可找出劣幣, 並且秤法如上所述。

三、不知劣幣是輕或重

這種情形必須多秤一次, 即總共要秤四次, 如下:

若第一次秤的結果是 (2), 則知 C 堆皆是良幣。取 A 與 C 秤第二次:

(i) 如果平衡, 則知劣幣在 B 堆之中且較輕;

(ii) 如果 A 重 C 輕, 則知劣幣在 A 堆之中且較重。

注意, A 輕 C 重的情形不可能發生。

接著不妨考慮 (i) 的情形。將 B 堆等分成 D, E, F 三堆, 每堆皆含三枚硬幣。任取 D 與 E 兩堆秤第三次: 如果平衡, 則劣幣在 F 堆且較輕; 如果 D 重 E 輕, 則劣幣在 E 堆且較輕; 如果 D 輕 E 重, 則劣幣在 D 堆且較輕。

假設劣幣在 $F = \{a, b, c\}$ 這一堆, 任取 a 與 b 兩枚硬幣秤第四次: 如果平衡, 則劣幣爲 c 且較輕; 如果 a 重 b 輕, 則劣幣爲 b 且較輕; 如果 a 較 b 重, 則劣幣爲 a 且較輕。

結論: 秤四次必可找出劣幣, 並且知道劣幣是較重或較輕。

四、更多的疑惑

數學的美妙之一就是, 當我們解決一個問題時, 往往會生出更多的問題, 好像是神話故事中的怪蛇, 每砍斷一個頭, 立刻又長出三個頭。數學家的主要興趣在於, 由特殊問題引出一大類問題, 然後一舉解決整類問題。在解決問題的過程中, 特殊問題通常扮演著引導與照亮暗路的燈光。

在上述問題中, 27 枚硬幣, 秤 3 次或 4 次, 這些數字是偶然的嗎? 它們有什麼關連? 如果改爲 12 枚或 30 枚又如何? 一般而言, N 枚硬幣需秤幾次? 如果允許秤 5 次, 最少與最多可鑑定幾枚硬幣? 秤 n 次呢?

對於這一整類問題, 欲求得一個全盤的解決辦法, 這就是訊息理論所要研究的一個論題。事實上, 這只是訊息理論的一個簡單應用。

在 N 枚硬幣中, 含有一枚劣幣, 但不知是哪一枚。這表示我們對此系統有某種程度的「無知」(ignorance), N 越大無知的程度也越大。當我們得知劣幣是哪一枚時, 就獲得了「訊息」(information), 從「無知」變成「知」。

我們用天平每秤一次, 就得到一些「訊息」, 把「無知」的程度減少一點。逐步秤就逐次累積「訊息」, 終至把無知完全消除。

對於「訊息」每個人都有或多或少的直觀了解 (正如對美醜有所了解一樣), 但是我們要的是「定量」(quantitative) 的掌握。從直觀的「定性」(qualitative) 描述走到精確的「定量」刻劃, 乃是科學探索的發端, 然後定量與定性迴旋交互運用。因此, 我們的問題是:

如何度量一個事件的「訊息」?

由此可展開一個驚心動魄的「觀念探險」(adventures of an idea) 之旅。

五、事件的局部訊息

解讀「自然之書」(the Book of Nature) 是科學求知活動最重要且最有趣的主題。大自然像古羅馬 Janus 神一樣有許多面, 甚至是無窮多面。機率論 (probability theory) 就是解讀大自然的反覆無常 (caprice)、說不準 (uncertainty) 這一面而發展出來的數學理論。

自然或人文現象大致可分成定命的 (deterministic) 與隨機的兩個範疇 (categories)。前者說得準, 即在給定條件或原因之下, 鐵定會得到某個果 (具有一定的因果關係), 所謂「種瓜得瓜, 種豆得豆」; 而後者在給定條件下, 可能得到這個果, 也可能得到那個果, 即有種種可能的結果 (outcomes)。

對一個隨機現象作實驗觀測, 就叫做「隨機實驗」(random experiment)。我們對「實驗」採取廣義的解釋, 它可以是天然發生的或人爲設計的, 例如丟一個骰子, 觀測生男或生女, 預測明天的天氣, 一個袋子含有27枚硬幣從中任取出一枚, 這些都隨機實驗。

令 Ω 表示一個隨機實驗所有可能出現的結果, 叫做樣本空間 (sample space)。例如丟一枚硬幣時,$\Omega = \{H, T\}$; 丟兩個骰子時,

$$\Omega = \{(1,1), \cdots, (1,6), \cdots, (6,6)\}。$$

樣本空間的一個子集 $A \subset \Omega$ 就叫做一個事件 (an event)。我們用 $P(A)$ 表示事件 A 發生的機率 (probability), 它是介於0與1之間的一個數。

當我們觀測到或聽到一個事件發生時, 就得到了一些「訊息」。例如考慮下面兩個事件:

「太陽從東邊出來」與「人咬狗」。

第一句話幾乎是必然事件, 機率爲1, 我們聽到它發生, 一點都不覺得驚訝 (surprise), 沒有得到「訊息」。第二句話是稀有事件, 機率很小, 我們聽到它發生, 覺得很驚訝, 得到很多「訊息」。因此, 訊息的多寡跟機率很有關係, 一個事件的機率越小, 但發生了, 我們就

得到越多的訊息。換言之, 一個事件的訊息是其機率的函數。

一般而言, 作一個隨機實驗, 令事件 A 的機率$P(A) = p$, 今觀測到 A 發生, 我們得到訊息

$$I = f(p)$$

問題: f是什麼函數呢?

通常要追尋一個函數並不容易, 好在「大自然」會透露一些線索, 提供我們有關 f 的一些訊息, 由此常可唯一決定出 f。例如, 當 A 爲一個必然事件 (sure event), 即$P(A) = 1$ 時, A 發生我們沒有得到訊息。換言之,

(i) $f(1) = 0$

其次, 當 p 越小時, $f(p)$ 越大, 即

(ii) f 爲一個遞減函數

進一步, 因爲當 p 微小改變時, 只導致訊息的微小改變, 故我們可以假設

(iii) f 爲一個連續函數

這是欲作數學討論的一個起碼要求。

這三個條件能唯一決定 f 嗎? 不能! 因爲

$$f(p) = \alpha(1-p)^n,\ \alpha > 0,\ n \in \mathbb{N}$$

$$f(p) = -k\log p,\ k > 0$$

$$f(p) = 1 - e^{p-1}$$

顯然都滿足上述三個條件。事實上, 有無窮多個函數滿足 (i) 至 (iii) 的條件。要唯一決定 f 的形式, 還需要其它條件。

一個隨機實驗通常並不是單純的如丟一個銅板或一個骰子, 而是由許多小實驗合成的, 例如丟 n 次銅板的隨機實驗就是由丟一個銅板的小實驗獨立地作 n 次整個合起來的。這當然也可以看成是幾個相同的銅板丟一次的隨機實驗。

今考慮一個複合的隨機實驗。設 A 與 B 爲兩個事件, 則 $A \cap B$ 表示 A 與 B 都發生的事件。如果

$$P(A \cap B) = P(A)P(B)$$

則稱 A 與 B 是獨立的 (independent)。獨立性的概念在機率論與統計學中佔有核心的地位, 在重覆獨立的實驗中, 才浮現出「機運」的規律性。

假設 A, B 是兩個獨立事件, 今 $p = P(A)$, $q = P(B)$, 則 $P(A \cap B) = pq$。A 發生並且 B 又發生, 我們得到 $f(pq)$ 的訊息。那麼 $f(pq)$ 與 $f(p), f(q)$ 是什麼關係呢?

A 發生, 我們得到訊息 $f(p)$。由於 A, B 是獨立的, 故 A 發生後 B 再發生, 我們又得到訊息 $f(q)$, 完全跟 p 無關, 兩者合起來就得到 $f(pq)$。換言之,

(iv) $f(pq) = f(p) + f(q)$

這個條件大大地局限了 f 的可能性。事實上, (i)-(iv) 唯一決定 f 的形式, 這就是下面的結果:

定理 1: 如果函數 $f : (0, 1] \to [0, \infty)$ 滿足上述 (i)-(iv) 的條件, 則 f 必形如

$$f(p) = -K\log_2 p,\ K > 0.$$

證明: 由 (iv) 知

$$f(p^2) = f(p) + f(p) = 2f(p).$$

由數學歸納法可知

$$f(p^m) = mf(p), \quad \forall m \in \mathbb{N} \qquad (1)$$

從而, 對任意自然數n, 我們有

$$f(p) = f(p^{1/n} \dots p^{1/n}) = nf(p^{1/n})$$

於是

$$f(p^{1/n}) = \frac{1}{n} f(p) \qquad (2)$$

由 (1) 與 (2) 兩式得知, 對任意非負有理數r

$$f(p^r) = rf(p) \qquad (3)$$

再由 f 的連續性知, 對任意非負實數 x, 恆有

$$f(p^x) = xf(p) \qquad (4)$$

今對任意實數p, $0 < p \leq 1$, 令$x = -\log_2 p$, 則$p = (\frac{1}{2})^x$。由 (4) 式知

$$\begin{aligned} f(p) &= f\Big((\frac{1}{2})^x\Big) = xf(\frac{1}{2}) \\ &= -K \log_2 p \end{aligned} \qquad (5)$$

其中$K = f(\frac{1}{2})$ 並且由 (i) 與 (ii) 知

$$K = f(\frac{1}{2}) > f(1) = 0$$

至此證明完畢。

通常取 $K = f(\frac{1}{2}) = 1$, 這意指取定一個度量單位, 此時度量出的訊息單位叫做 bit(位元, binary digit)。從而公式 (5) 變成

$$f(p) = -\log_2 p \quad \text{bits} \qquad (6)$$

假設事件 A 的機率爲 p, 那麼得知 A 發生所得到的訊息爲 $-\log_2 p$ bits, 這叫做事件 A 的局部訊息量 (local information)。丟一個公正銅板, 正面與反面出現的機率各 $\frac{1}{2}$, 當我們得知出現正面時, 就得到 1 bit 之局部訊息量。

六、一個機率分佈的熵

我們先考慮下面三個例子:

甲. 在賽馬場上, 兩匹勢均力敵的馬, 不易預測其結果; 但是兩匹相差懸殊的馬, 幾可預知其結果。

乙. 丟一個骰子比丟一個銅板的不確定性更大。

丙. 丟一個公正銅板比丟一個有偏銅板 (如出現正、反面的機率分別爲0.8與0.2) 更讓人捉摸不定。

這些大家都「直觀地」承認, 但是我們的問題是:

如何衡量一個隨機實驗的不確定性程度?

機率空間 $(\Omega, \mathcal{F}, P)$ 是隨機實驗所含資訊的精簡與理想化, 是一切機率演算的基礎。但是, 這並不夠, 我們還需要隨機變數 (random variable) 的概念

$$X : \Omega \to \mathbb{R}$$

使我們更自由靈活地探討機率問題。隨機變數可以解釋成賭徒的輸贏金額, 對事件的重新改訂, 一種出象機制, 一種觀測度量,...等等。

一般而言, 觀察一個隨機實驗, 其結果是一個隨機變數X。假設 X 只取有限多個值

$x_1, x_2, \ldots, x_n$, 令事件 $A_k = \{X = x_k\}$ 的機率爲 p_k。顯然

$$p_k \geq 0,\ k = 1, 2, \ldots, n \text{ 並且 } \sum_{k=1}^{n} p_k = 1$$

我們稱 $(p_1, p_2, \ldots, p_n)$ 爲隨機變數X的機率分佈 (probability distribution)。於是我們的問題變成:

如何衡量隨變數 X 或機率分佈 $(p_1, \ldots, p_n)$ 的不確定性程度?

機率論對初學者構成困擾的原因之一, 就是沒有分清楚「已然」與「未然」、「現實」與「可能」。在隨機實驗未作之前有各種可能結果, 但是作了實驗之後, 出現一個結果, 就叫做一個「現實」(a realization)。只有在未作實驗之前的「未然的」與「可能的」世界, 才有機率可言。

今作一個隨機實驗, 報告說事件 A_k 發生了, 我們得到的驚訝量或局部訊息爲 $-\log_2 p_k$。可是在未然的世界, 我們只知道可能發生 A_1, 也可能發生 $A_2, \ldots$, 等等, 這是說不準的。由於發生 A_k 的機率是 p_k, 我們自然應該考慮期望值, 將 $-\log_2 p_k$ 乘以 p_k, 再對 k 求和得到

$$S(p_1, \ldots, p_n) = S(X) = -\sum_{k=1}^{n} p_k \log_2 p_k \tag{7}$$

其中若出現$0 \log_2 0$就定義爲$\lim_{p\to 0} p \log_2 p = 0$。$S(p_1, \ldots, p_n)$ 或 $S(X)$ 就是局部訊息量的期望值 (expectation) 或平均值, 我們又稱之爲 X 或機率分佈 $(p_1, \ldots, p_n)$ 的熵 (entropy), 這是夏農 (Shannon) 在1948年引進的一個重要的概念。

熵是對整個隨機實驗, 表現爲機率分佈 $(p_1, \ldots, p_n)$, 所呈現出的不確定性或混亂程度之度量。由長期觀點來看, 每作一次隨機實驗, 我們從「無知」到「知」, 平均就得到 $S(p_1, \ldots, p_n)$ 的訊息量。

注意到, 在訊息論中, 隨機變數不重要, 它的取值也不重要, 重要的是它的機率分佈。事實上, 熵只跟機率分佈 $(p_1, \ldots, p_n)$ 的組合有關, 而跟其排列順序無關。作爲度量「不確定性」, 熵比機率分佈更模糊, 因爲可以有各種不同的機率分佈對應相同的熵。

如果機率分佈是均勻分佈, 即

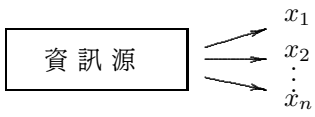
$$p_k = \frac{1}{n}, \quad k = 1, 2, \ldots, n$$

則它的熵爲

$$H(\frac{1}{n}, \ldots, \frac{1}{n}) = \log_2 n \tag{8}$$

此式叫做 Hartley 公式, 因爲 Hartley 在1928年首度建議採用對數函數作爲度量訊息的工具。

我們可以想像有一個資訊源 (an information source), 可能出現的結果是 $x_1, x_2, \ldots, x_n$ 並且機率均等:



那麼它的熵就是 $\log_2 n$, 這恰好也等於觀察到任何一個值 x_k 所得的局部訊息, 兩者合一。

例1: 一張身份證的號碼含有多少訊息?

目前台灣約有21,000,000人, 每個人一個號碼, 機率均等, 故訊息量爲

$$\log_2(21000000) = 24.3238 \text{ bits}$$

例2: 觀看電視的一個畫面, 得到多少訊息?

一個畫面由許多小光點組成, 例如假設畫面分成500行與600列, 一共有 $500 \times 600 = 300,000$ 個光點, 再假設每一個光點有10種可分辨的顏色選擇, 於是總共有 $10^{300,000}$ 種可能的畫面, 機率都相等。因此, 觀看一個畫面的訊息量爲

$$300,000 \log_2 10 \doteq 10^6 \text{bits}$$

我們不妨作一個比較: 從收音機收聽到1000個字的廣播, 得到多少訊息? 假設常用字有10,000個, 從中任意選取出1000個字, 總共有 $(10000)^{1000}$ 種取法, 並且機率均等, 所以收聽到1000個字所得的訊息量爲

$$1000 \log_2 10000 \doteq 1.3 \times 10^4 \text{bits}$$

這正好應驗一句諺語: 「一個圖勝過千言萬語」(One picture is worth a thousand words.)

問題: 考試卷上, 一個是非題的訊息是多少? 一個「五選一」的選擇題, 其訊息是多少? 一個有五個選項的多重選擇題, 其訊息是多少?

我們要強調, 訊息論不作價值判斷, 要談「一句話的訊息量」決不能談這句話是好消息或壞消息。例如, 一位男士問他的女朋友: 你喜歡吃牛肉麵嗎? 你嫁給我好嗎? 假設回答 Yes 或 No 的機率皆爲0.5, 那麼不論是得到什麼答案, 這兩句話的訊息量都是1 bit。但是, 若考慮價值判斷, 後一句話的答案顯然重要得多。

這也警告我們, 數學的訊息並不是唯一重要的考量。我們常聽說: 寫文章要「言之有物」, 不要「言之無物」; 好的文章每一句都恰到好處, 而又有難以預期的驚訝; 詩比散文含有更多的訊息。這些都很重要, 但不是訊息論所能完全掌握。

總之, 有不確定性或渾沌不明就有機率分佈, 有機率分佈就有訊息或熵。

七、熵的性質與刻劃

根據 (7) 式的定義, 我們容易驗得熵具有下列性質:

(A_1) $S(p_1, \ldots, p_n) \geq 0$ 並且等號成立 $\Leftrightarrow$ 存在某個 $p_k = 1$。

這表示不確定性恆爲非負數並且當隨機性不存在時, 等於0。

(A_2) $S(p_1, \ldots, p_n, 0) = S(p_1, \ldots, p_n)$

(A_3) $S(\frac{1}{n}, \ldots, \frac{1}{n}) \leq S(\frac{1}{n+1}, \ldots, \frac{1}{n+1})$

(A_4) $S(p_1, \ldots, p_n)$ 爲一個連續的且對稱的函數。

(A_5) $S(p_1, \ldots, p_n) \leq H(\frac{1}{n}, \ldots, \frac{1}{n})$

這表示機率是均勻分佈時, 有最大熵。

證明: Jensen 不等式告訴我們: 如果 f 爲定義在 $[a, b]$ 上的一個凸函數 (convex function), 則對於任意 $x_1, \ldots, x_n \in [a, b]$ 以及 $\lambda_k \geq 0$, $\sum_{k=1}^n \lambda_k = 1$, 恆有

$$f(\sum_{k=1}^n \lambda_k x_k) \leq \sum_{k=1}^n \lambda_k f(x_k) \qquad (9)$$

今因 $f(x) = -x \log_2 x$ 爲一個凹函數, 取

$$a = 0, b = 1, x_k = p_k, \lambda_k = \frac{1}{n}$$

則得

$$-\sum_{k=1}^{n}\frac{1}{n}p_k\log_2 p_k$$
$$\leq -(\sum_{k=1}^{n}\frac{1}{n}p_k)\log_2(\sum_{k=1}^{n}\frac{1}{n}p_k)$$

從而

$$S(p_1,\ldots,p_n)\leq \log_2 n = S(\frac{1}{n},\ldots,\frac{1}{n})$$

證畢。

其次, 我們考慮兩個隨機實驗複合成一個大實驗之訊息量。假設兩個實驗之觀測值分別是隨機變數 X 與 Y, 那麼複合實驗之觀測值是隨機變數 $Z=(X,Y)$, 令其機率分佈爲 (r_{ij})。再令

$$p_i=\sum_j r_{ij},\ q_j=\sum_i r_{ij}$$
$$i=1,\ldots,n, j=1,\ldots,m,$$

則 (p_i) 與 (q_j) 分別是 X 與 Y 的機率分佈。考慮 (r_{ij}) 的訊息量 $S(Z)=S(r_{ij})$, 這又可以看成是兩個步驟之和: 先作實驗 X, 得到訊息量 $S(X)=S(p_i)$; 然後如果知道 $X=x_i$, 則 $Y=y_j$ 的條件機率爲 ${}_i\pi_j=r_{ij}/p_i$, 故已知 $X=x_i$ 時, Y 的「條件訊息」爲 $S({}_i\pi_1,{}_i\pi_2,\ldots,{}_i\pi_m)$, 整個合起來, 已知 X 時, Y 的「條件訊息」爲

$$S(Y|X)=\sum_{i=1}^{n}p_iS({}_i\pi_1,\ldots,{}_i\pi_m)\quad(10)$$

從而, 我們有

(A_6) $S(Z)=S(X)+S(Y|X)$

特別地, 當 X 與 Y 獨立時, 即 $r_{ij}=p_i\cdot q_j$, 有 $S(Y|X)=S(Y)$ 並且

$$S(Z)=S(X)+S(Y)。$$

由Jensen 不等式也可證得

$$S(Y|X)\leq S(Y)$$

反過來, 一個函數 H 若滿足 (A_1) 至 (A_6) 的條件, 則 H 必形如

$$H(p_1,\ldots,p_n)=-\sum_{k=1}^{n}p_k\log_2 p_k\ \text{bits}$$

其中 1 bit$=H(\frac{1}{2},\frac{1}{2})$。我們省略掉證明。

因此, 我們可以採用公理化的手法來定義熵。所謂一個機率分佈的熵是指滿足 (A_1) 至 (A_6) 的函數。

八、熵與劣幣的找尋

找尋劣幣的問題, 有了熵的概念, 使我們看得更清楚, 心中有數。

天平秤一次 (即作一次隨機實驗) 得到多少訊息?

因爲天平兩端必須置同數的硬幣, 故只能是下列兩種情形:

(i) 將硬幣等分成兩堆: 但這限於總硬幣數爲偶數, 並且秤一次只可能是「左輕右重」或「左重右輕」, 故最多可得1 bit 之訊息量。

(ii) 將硬幣分成三堆, 其中至少有兩堆的硬幣數相等, 秤一次的可能結果是「平衡」、「左輕右重」 或 「左重右輕」, 故最多可得 $\log_2 3$ bits 之訊息量。

對照起來, 分成三堆的情形, 可得較多的訊息。

現在回到27枚硬幣的問題。將硬幣分成 A,B,C三堆, 各含有 n, n 及 $27-2n$ 枚硬

幣, 取 A 與 B 兩堆來秤。由於每一枚的機率均等, 故

$$\text{平衡的機率爲 } \frac{27-2n}{27}$$
$$\text{左輕右重的機率爲 } \frac{n}{27}$$
$$\text{左重右輕的機率爲 } \frac{n}{27}$$

欲得最大訊息 $\log_2 3$, 則必須三者的機率相等

$$\frac{n}{27}=\frac{n}{27}=\frac{27-2n}{27}=\frac{1}{3}$$

從而, $n=9$。因此, 最好的秤硬幣的方式是: 等分成三堆, 然後任取兩堆在天平上作比較。

其次考慮總共需秤幾次的問題。27枚硬幣的訊息量是多少? 我們仍然分成兩種情形討論:

(a) 已知劣幣是輕或重的情形。

因爲27枚硬幣的機率均等, 故總訊息量(或熵) 爲 $\log_2 27$ bits。又知秤 n 次最多可得 $n\log_2 3$ 之訊息量。若欲

$$n\log_2 3 \geq \log_2 27$$

則必 $n\geq 3$。因此, 最少需秤3次。

(b) 不知劣幣是輕或重的情形

因爲每一枚硬幣可能是輕或重, 故總共有 $2\times 27=54$ 種可能, 且機率均等, 從而總訊息量爲 $\log_2 54$。若欲

$$n\log_2 3 \geq \log_2 54$$

則必 $n\geq 4$。因此, 最少需秤4次。

事實上, 我們實際秤量的結果, 前者秤3次, 後者秤4次, 就可以找到劣幣, 且知較輕或較重。

最後考慮秤 n 次最多可鑑定幾枚硬幣的問題。假設我們不知道所含的劣幣是較輕或較重。

根據上述的討論, 如果最多可鑑定 N_n 枚, 則

$$n\log_2 3 \sim \log_2 2N_n$$

亦即

$$N_n \sim \frac{3^n}{2} \qquad (11)$$

這只是初步的估計。

爲了求得精確公式。我們採用歸納法。首先觀察特例, 容易驗知:

(i) 當 $n=2$ 時, $N_2=3=3^1$

(ii) 當 $n=3$ 時, $N_3=12=3^1+3^2$

(iii) 當 $n=4$ 時, $N_4=39=3^1+3^2+3^3$

由此我們猜測, 對於一般 n,

$$N_n=3^1+3^2+\cdots+3^{n-1}=\frac{3^n}{2}-\frac{3}{2} \qquad (12)$$

如何證明呢? 物理學家戴森 (F. J. Dynson) 利用三進位法證明了 (12) 式, 參見 [6]。

至此有關找尋劣幣的問題完全解決。例如, 由 (12) 式知, 秤5次可鑑定硬幣數的範圍是40到120枚。

九、熵的歷史演化

熵的概念最初發源於熱力學, 這是 Clausius 在1864年引入的一個重要概念。Boltzmann 在1896年強調熱力學的熵與機率具有密切關係。1928年 Hartley 將 Boltzmann 的想法引進訊息論中, 得到 Hartley 公式 [即 (8) 式]。1948年 Shannon 再推廣成 Shannon 公式 [即 (7) 式]。Kolmogorov

在1958年將 Shannon 熵的公式引入動力系統, 用來研究兩個動力系統的同構問題。今日熵的概念已遍及科學、數學, 甚至人文、藝術的領域, 變成日常生活的用語。

古希臘哲學家柏拉圖 (Plato, 427-347 B.C.) 非常重視幾何學, 他說:

(i) 不懂幾何學的人不得進入此門 (柏拉圖學院的門)。

(ii) 不知道正方形的邊與對角線是不可共度 (即 $\sqrt{2}$ 不是有理數) 者愧生爲人。

隨著文明的進展, 到了近代, 邏輯哲學家羅素 (B. Russell, 1872-1970) 說:

> 不懂牛頓如何從刻卜勒 (Kepler) 定律推導出萬有引力定律的人是沒有受過教育的。

現代物理學家惠勒 (Wheeler) 改述爲:

> 在過去, 只有當一個人了解熵的概念後, 才能說是受過科學教育。在將來, 如果一個人不懂碎形(fractal), 則會被認爲是科學文盲。

不確定性與混沌, 表現爲機率分佈, 一方面讓我們不可預期, 甚至困惑, 另一方面也讓我們得到驚訝與啓示。只有機運能夠對我們說話, 從中我們解讀出豐富的含義。然而, 機運仍然是一個謎。莎士比亞 (Shakespeare, 1564-1616) 問得好:

> 如果你能洞穿時間的種子, 並且知道哪一粒會發芽, 哪一粒不會, 那麼請告訴我吧! (If you can look into the seeds of time and say which grain will grow and which will not. Speak then to me!)

參考資料

1. G. Raisbeck, Information Theory, The M.I.T. Press, 1965.
2. G. Caglioti, The Dynamics of Ambiguity, Springer-Verlag, 1992.
3. A. Renyi, A Diary on Information Theory, John Wiley and Sons, 1984.
4. L. Brillouin, Science and Information Theories, Academic Press, 1956.
5. J. D. Beasley, The Mathematics of Games, Oxford Univ. Press, 1990.
6. F. J. Dyson, The problem of pennies, Mathematical Gazette, 231-234, 1946.
7. J. D. Fast, Entropy, the significance of the concept of entropy and its applications in science and technology, Macmillan, 1982。
8. P. Walters, An Introduction to Ergodic Theory, Springer-Verlag, 1982.
9. R. K. Guy and R. J. Nowakowski, Coin-weighing problems, American Math. Monthly, 164-167, 1995.

—本文作者任教於台灣大學數學系—