隨機與密碼

黃文璋

一. 密碼

我們處在一個密碼的時代。年輕人以手機傳簡訊, 520表我愛你, 5120184表我要愛你一輩子, 08376表你別生氣了, 7456表氣死我了, 789表去打球 (用台語發音)。這些乍看之下不明所以, 但對習於傳簡訊者, 看這些數字大約就與看中文一樣。

在布魯斯威利 (Bruce Willis) 主演的終極密碼戰 (Mercury Rising) 裡, 一個患自閉症的男孩, 竟然破解軍方一極機密的水星密碼。美麗境界 (A Beautiful Mind) 裡的數學家納許 (John Nash, 羅素克洛 (Russell Crowe) 主演), 拿到博士學位後, 到國防部工作, 也參與解碼的任務。這類解碼, 大抵與聖經密碼 (The Bible Code, Michael Drosnin原著) 一書所描述者類似。即從一大片文字或符號中, 看出中間藏有某項訊息。例如, 從聖經的第一個字母開始, 依序每次跳過100個字母, 即將第$1, 101, 201, 301, \cdots$, 等字母連起來, 看組合出什麼字句。不過這種密碼容易淪於各說各話, 穿鑿附會, 是不是眞的被破解, 有時不得而知。

有些縱橫填字遊戲 (cross-word puzzle) 或數字謎, 也可視爲密碼。如

$$\begin{array}{ccccc} & S & E & N & D \\ + & M & O & R & E \\ \hline M & O & N & E & Y \end{array}$$

式中每一英文字母代表一不同的阿拉伯數字, 假設此爲一正確算式, 而解出其中的英文字母。這種密碼的破解, 主要是依據邏輯的合理性。另外, 人類對遺傳的奧秘一直深感興趣, 近年來遂有熱門的遺傳密碼。

大家可看出, 我們所稱的密碼, 是很廣泛的。只要是我們不知的, 常可稱其中有密碼; 而對訊息想隱藏不讓人知道, 就可說是在編密碼, 或稱編碼; 要解出隱藏的訊息, 便稱爲解碼。

比較有系統的編碼, 是把文字對應到數字 (或符號, 如以旗號或閃燈造出密碼, 通行的手語也屬於此類)。電報 (在辭彙那部字典裡, 對密碼的解釋即爲 "收發電報的秘密號碼") 現在已不流行了。我們當年出國讀書 (那已經是很久以前了), 要到外交部辦手續。那時政府對留學生設想周到, 我們每人要將自己名字的每個字, 從一本對照表上, 各找出一個四位數字的碼, 填寫交

上, 以備若在國外發生什麼意外, 駐外單位可發電報通知台灣。軍隊傳遞情報, 也常用這類方式。這種編碼雖很有效率, 但只要那本對照表被敵方取得, 情報便暴露了。

在第二次世界大戰時, 德國憑藉其優異的密碼通訊能力, 潛艦神出鬼沒, 盟軍膽顫心驚。在獵殺U-571(U-571) 裡, 便是演盟軍如何奪取北大西洋中, 德軍編號U-571潛艦上密碼解碼機的驚險過程。獵風行動 (Windtalkers) 也是以第二次世界大戰爲背景。美國海軍利用納瓦荷族的母語創造出一種密碼。尼可拉斯凱吉 (Nicolas Cage) 飾演的美國軍官, 在太平洋塞班島與日軍的浴血戰中, 其職責乃是不讓那群納瓦荷族的通信兵落入日軍手中, 以免密碼被破。所以除了保護外, 必要時得槍殺他們。

上述這種編碼方式, 顯然存在一罩門, 不要說對照表、解碼機, 或納瓦荷族人落入敵方手中是很難防止的, 只要時間夠久, 經過比對, 便沒有破解不了的。

金融機構提款卡的密碼、開鎖的密碼、航空公司訂機位的電腦代號, 及電腦開機的密碼 (password) 等, 又是一類密碼。甚至樂透彩每期的頭獎號碼, 亦可視爲此類密碼。是一種個人式或偶發性的需求, 但也都是儘量不想讓人猜中。這類密碼爲了記憶方便, 往往不致於過長。例如, 提款卡的密碼通常只有四個數字。開機密碼大抵是英文字母或數字, 七個就很多了, 太長自己都記不住。理論上這種密碼也是只要時間允許, 以及沒有限制錯誤次數 (如提款機通常按錯三次, 卡片便出不來了), 都可以經由一個個試而破解。

既然沒有破解不了的密碼, 那有沒有比較難破的編碼方式呢? 也就是破解要花很長的時間, 此時間超過保密的有效期限, 則便可充分達到保密的效果了。答案是肯定的。數學與統計在這裡倒是有好的角色可扮演。

利用巨大整數難以分解的特性, 美國麻省理工學院的三位數學家 Rivest, Shamir 及 Adleman, 於西元1977年提出一個至目前仍被認爲極安全的密碼技術, 論文並於1978年刊登, 所謂公開鑰匙密碼法 (Public-Key Cryptography)。取他們三人姓的第一個字母, 又稱 RSA 法。這方面的討論可參考楊重駿、楊照崑 (1983, 1986), 及楊淑芬 (1991)。RSA法適用於金融或軍事等對保密工作很需要的單位, 爲一種有系統的編碼法。欲操作此法, 通常要有很強的計算設備。

本文則針對前述第二類密碼, 說明如何利用隨機性來編碼使較難破解。此法雖然卑之無甚高論, 且並不需太多的設備, 卻常爲一般人所忽略。又鑑於一般人對隨機性的概念往往未能充分掌握, 本文也將對此概念加以闡釋。

二. 隨機抽樣

賭博老手到賭場可能不會立刻就賭, 而是先觀察一番, 看看莊家出牌有沒有什麼規律, 看看骰子是否那一面出現的頻率較高。一般大賭場, 如美國拉斯維加 (Las Vegas) 及大西洋城

(Atlantic City) 等地的賭場, 每天進出的客人很多, 賭場大抵不會詐財, 而是從玩法的設計, 使得對賭徒而言, 不是一公正的賭局。

有些事件彼此間有關係。如父親與兒子的身高; 兩次考試的成績; 明天的氣溫與今天的氣溫; 要拿報告給上司看之前, 先打聽上司今天的心情, 因認爲上司心情的好壞, 會影響看你報告的評語。警察辦案會研究犯罪者的行爲, 因認爲犯罪者有一些行爲模式, 由作案手法類似的案件, 猜測嫌犯可能爲同一人, 再由過去案件發生地點, 找出地緣關係, 推測其下一作案處。

我們常會根據過去的資料, 以對未來做預測。有時則儘量想防止被別人預測中 (如賭場之出牌, 或老師之命題)。究竟什麼樣的情況會較難預測?

假設要猜下次月考班上誰會考第一。班上雖四十多人, 依過去成績可做一些推斷。其中有少數幾人考第一名的機會較大, 大部分人則機會很小。因此通常不會太難猜, 會考第一的總不出那兩三個人。但若要猜這次年終摸彩誰會中頭獎, 就不容易了。因如果每人一張彩券, 則每人會被抽中的機會皆相同, 即使知道過去若干年誰中頭獎, 顯然也沒有幫助。雖有時我們說某人一向運氣好、手氣佳, 但多半是事後諸葛, 事前我們倒不見得眞認爲有誰被抽中的機會較大。

在統計上抽樣的方法很多, 如系統抽樣 (systematic sampling), 分層抽樣 (stratified sampling), 及叢聚抽樣 (cluster sampling) 等, 都是常用的方法。大部分的民調、抽獎等, 包含各國風行的樂透彩, 常是採用以不重複的簡單隨機抽樣 (simple random sampling without replacement, 底下只稱簡單隨機抽樣) 的方式, 產生所要的號碼。以北銀樂透彩爲例, 由42個數字, 每次產生不重複且順序不計的6碼, 總共

$$\binom{42}{6} = 5{,}245{,}786$$

組號碼, 每組產生的機率都相同。如果是n取r, 則共有$\binom{n}{r}$種組合。如果是以n個相異數字來編長度爲r之密碼 (可重複且計順序, 這是通常的情況), 則共有n^r組, 數目顯然增加很多。假設每組產生之機率皆相同, 我們便稱此爲隨機密碼 (stochastic code)。簡單隨機抽樣, 與隨機密碼, 其中皆含有兩個概念: 獨立及均勻。

獨立表每次抽出的號碼與以前的不相干, 均勻表每組號碼被抽中的機會都一樣。在資訊理論 (information theory) 裡, 一項試驗若有$A_1, \cdots, A_k$等k種可能的結果, 發生的機率分別爲$p_1, \cdots, p_k$, 其中$p_i \geq 0$, $\sum_{i=1}^k p_i = 1$, 則可以

$$H(p_1, \cdots, p_k) = -\sum_{i=1}^{k} p_i \log p_i$$

來量測此實驗中所含不確定性 (uncertainty)。其中對數可取任何一不爲1之固定正數爲底, 而若某$p_i = 0$, 則定義$p_i \log p_i = 0$。在物理上$H(p_1, \cdots, p_k)$稱爲此實驗之熵 (entropy), 此處不擬多談。

當$p_1,\cdots,p_k$之值爲何, 會使$H(p_1,\cdots.p_k)$最大? 也就是這種試驗何時會有最大的不確定性? 對固定的k, 當每一結果之可能性皆相同, 即$p_i=1/k$, $i=1,\cdots,k$, $H(p_1,\cdots,p_k)$達到最大値, 即此時會有最大的不確定性。可利用下述不等式 (此爲 Jensen's inequality 之一特例) 來證明:

若$\phi(x)$爲一凸函數 (convex function), 則對任意正數$a_1,\cdots,a_k$,

$$\phi(\frac{1}{k}\sum_{i=1}^{k}a_i)\le\frac{1}{k}\sum_{i=1}^{k}\phi(a_i).$$

現取$\phi(x)=x\log x$, $x>0$, $\phi(0)=0$, 爲一定義於$[0,\infty)$之凸函數, 取$a_i=p_i$, 利用$\sum_{i=1}^{k}p_i=1$, 可得

$$\begin{aligned}-\frac{1}{k}\log k=\frac{1}{k}\log\frac{1}{k}&=\phi(\frac{1}{k})=\phi(\frac{1}{k}\sum_{i=1}^{k}p_i)\\&\le\frac{1}{k}\sum_{i=1}^{k}\phi(p_i)=\frac{1}{k}\sum_{i=1}^{k}p_i\log p_i=-\frac{1}{k}H(p_1,\cdots,p_k),\end{aligned}$$

故有

$$H(p_1,\cdots,p_k)\le\log k=H(\frac{1}{k},\cdots,\frac{1}{k}).$$

利用上述結果, 欲自n個數字中, 產生r個不重複且順序不計的號碼, 以簡單隨機抽樣產生, 會有最大的不確定性。而以n個數字來編長度爲r之密碼, 隨機密碼會有最大的不確定性。換句話說, 這種情況是最難猜中的, 印證我們之前的想法。

在簡單隨機抽樣裡, 知道過去的抽樣結果, 對未來之預測毫無幫助。如果是n取r, 每組號碼會出現的機率永遠是$1/\binom{n}{r}$。號碼的出現如果只是獨立, 而不均匀 (例如有些號碼球較重, 因此較易出現), 則當然要猜那些較易出現的號碼。對於樂透彩, 經過一段時間的觀察, 可統計出各號碼出現的頻率。只是有人對出現頻率較低的號碼, 會認爲應快出現了, 有人則對出現頻率較高的號碼, 認爲該號碼“氣 ”較旺, 應較易再出現。到底那一種看法才正確呢? 這就牽涉到對隨機的概念是否能正確掌握。

三. 你了解隨機嗎?

民國92年1月1日起, 環保署實施第二階段的塑膠袋限用政策, 塑膠業者與民衆均感到困擾。中國時報92年1月1日15版有一則公視記者馬台興的投書, 其中有底下的一些句子:

昨天筆者支援採訪此則新聞, 經“隨機採樣” 受訪者, $\cdots$。而「平口, 無提把」塑膠袋可用的細節幾乎都能 “隨機” 答出。$\cdots$。於是筆者又鍥而不捨的“隨機” 多問了許多間店家, $\cdots$。

短短的文章裡, 用了三次 “隨機” 的字眼。但作者是否眞了解隨機的意義呢? 隨機與隨便的意思一樣嗎? 就算該文作者了解隨機的意義, 但個人容不容易做到隨機採樣呢?

在機率裡, 隨機的意義本來是很一般的。只要是一事先不能預知結果的試驗, 便稱隨機試驗。假設有一銅板, 出現正面的機率爲 0.2, 反面的機率爲 0.8, 連續投擲 10 次, 看得到幾個正面, 這便是一隨機試驗。若以 X 表所得正面數, 則由排列組合裡的結果知

$$P(X=k)=\binom{10}{k}0.2^k0.8^{10-k},\quad k=0,1,\cdots,10.$$

X 便稱有二項分佈 $\mathcal{B}(10,0.2)$。一般的二項分佈則以 $\mathcal{B}(n,p)$ 表之。但在簡單隨機抽樣裡, 或是說將 10 個球 “隨機地” 放進 10 個箱子中, 如前所述, 此處之 “隨機” 便含有獨立及均勻的意思。有時我們會說有一均勻的骰子, 或說將撲克牌洗得很均勻。

由於含義爲 “均勻 ”, 一般人會將之視爲與水泥塗抹得很 “均勻”, 儀隊隊員的身高很 “均勻” 的意義相同。也就是將均勻與相等視爲同義, 而忽略了此爲隨機現象。要知在隨機現象裡, 均勻乃表出現之機率相等, 而非出現之頻率相等。

先看底下的例子。

例 1. 假設有 2 個箱子, 將 2 個球分別隨機地放進箱中。即每一個球皆有 $1/2$ 的機率放進任何一箱中。而每箱中各恰有一球的機率爲

$$\frac{2!}{2^2}=\frac{2}{4}=\frac{1}{2}.$$

如果是 3 個球隨機地放進 3 個箱子中, 則每箱中各恰有一球的機率爲

$$\frac{3!}{3^3}=\frac{6}{27}=\frac{2}{9}.$$

如果是 10 個球隨機地放進 10 個箱子中, 則每箱中各恰有一球的機率爲

$$\frac{10!}{10^{10}}=\frac{3{,}628{,}800}{10^{10}}=0.00036288.$$

n 個球隨機地放進 n 個箱子中, 則每箱中各恰有一球的機率爲 $n!/n^n$, 此值隨著 n 之增大而漸減。

上述現象, 可能違反一般人的直觀。將 10 個球隨機地放進 10 個箱子中, 每個箱子中各恰有一個球, 應是最均勻的, 結果卻是極不容易發生。反而是不均勻的情況, 即至少有一箱子中有兩個以上的球很容易發生, 機率爲

$$1-0.00036288=0.99963712\doteq 1.$$

而一箱中有 2 球, 8 箱中各有一球, 另一空箱的機率爲

$$\frac{\binom{10}{1}\binom{9}{8}\binom{10}{2}\cdot 8!}{10^{10}}=45\cdot\frac{10!}{10^{10}},$$

是每箱中各恰有一球的機率之45倍。

令每箱中各恰有一球的機率爲a。再給一些例子如下:

(a) 1箱中有3球, 7箱中各有一球, 另2空箱, 其機率爲$60a$。

(b) 2箱中各有3球, 2箱中各有2球, 另6空箱, 其機率爲$(35/4)a$。

(c) 2箱中各有4球, 2箱中各有1球, 另6空箱, 其機率爲$(35/16)a$。

(d) 1箱中有1球, 1箱中有2球, 1箱中有3球, 1箱中有4球, 另6空箱, 其機率爲$(35/2)a$。

(e) 1箱中有6球, 1箱中有2球, 2箱中各有1球, 另6空箱, 其機率爲$(7/4)a$。

(f) 5箱中各有2球, 另5空箱, 其機率爲$(63/8)a$。

有些看起來很偏頗的事件, 其發生的機率卻比很均勻的事件之機率大。我們給出恰有i個空箱的機率如下:

(i) 0空箱之機率爲a。

(ii) 1空箱之機率爲$45a$。

(iii) 2空箱之機率爲$375a$。

(iv) 3空箱之機率爲$980a$。

(v) 4空箱之機率爲$(7609/8)a$。

(vi) 5空箱之機率爲$(2835/8)a$。

(vii) 6空箱之機率爲$(6821/144)a$。

(viii) 7空箱之機率爲$(311/168)a$。

(ix) 8空箱之機率爲$(511/40320)a$。

(x) 9空箱之機率爲$a/9!$。

可看出有3個空箱之機率最大, 有4個空箱之機率次大。

現在很多高中教室, 置有一竹籤筒, 以方便老師上課時隨機地點學生上台。一學期下來, 就是有幾位學生多次被點中, 有幾位學生卻從未被點過。這也可以解釋爲何即使上天對每一個人可能無意有差別待遇, 但結果是抱怨禍不單行者不少, 慶幸好事連連者也不少。

在紅樓夢的第八回, 賈寶玉去探望薛寶釵, 正在閒聊。一語未了, 忽聽外面的人說:『林姑娘來了。』話猶未完, 黛玉已搖搖擺擺的進來, 一見寶玉, 便笑道:『哎喲! 我來的不巧了!』寶玉等忙起身讓坐。寶釵笑道:『這是怎麼說?』黛玉道:『早知他來, 我就不來了。』寶釵道:『這是什麼意思?』黛玉道:『什麼意思呢? 來呢, 一齊來, 不來, 一個也不來。今兒他來, 明兒我來, 間錯開了來, 豈不天天有人來呢? 也不至太冷落, 也不至太熱鬧。姐姐有什麼不解的呢?』

再看一例。

例2. 設一箱中有20個有編號的球, 自其中隨機地依序取兩個, 每次取出後放回。則兩球皆相異之機率爲

$$\frac{20}{20}\cdot\frac{19}{20}=0.95,$$

會有重複之機率爲

$$1-0.95=0.05.$$

從$10n$個有編號的球中, 依序隨機地取n個, 每次取出後放回。則會有重複之機率爲

$$1-\frac{10n\cdot(10n-1)\cdots(10n-n+1)}{(10n)^n}.$$

易見此值隨著n之增大而漸增。$n=30$時, 此值約0.098, 也就是自30個球中取3個還不太會重覆; 但$n=300$時, 此值就已約0.777。若$n\to\infty$, 則此值趨近至1。

球數愈多, 愈容易有取樣重複的現象。一個類似的問題是, 如果做芝麻餅, 希望芝麻很均勻地散佈, 可否隨機地撒呢? 你現在知道了, 不可以, 否則芝麻必是有些地方很多, 有些地方很稀疏。隨機下的後果, 往往是不均勻。

在沈默的羔羊 (The Silence of the Lambs) 那部電影裡, 有底下一句話:

Doesn't this random scattering site seem desperately random, like an elaboration of bad liar.

這些隨機散佈的地點, 不是極度地隨機嗎? 就像差勁的騙子精心設計的謊言。

看起來很 "隨機", 反而會像精心設計的謊言! 上課時老師點名, 如果是隨機地點, 是很難每次都點不同的人。統計樂透彩過去的頭獎號碼, 如果每個號碼累積出現的次數都一樣, 或連續幾期開出的號碼都不一樣, 反而才該懷疑其隨機性。

有時我們會懷疑事件之隨機性, 此因看到過多的巧合。以樂透彩爲例, 從每期開出的6個頭獎號碼, 要找到一些特殊的組合, 往往並非太困難的事。

例3. 在42取6的樂透彩裡, 偶數共有21個。故6碼全爲偶數之機率爲

$$\frac{\binom{21}{6}}{\binom{42}{6}}=\frac{54{,}264}{5{,}245{,}786}\doteq 0.0103,$$

同理, 6碼全爲奇數, 6碼全在1至21, 6碼全在22至42, 機率均約爲0.0103。所以每期頭獎號碼全爲偶數, 或全爲奇數, 或全在1至21, 或全在22至42, 其機率約爲

$$4\cdot 0.0103=0.0412.$$

甚至6碼全爲3的倍數, 全不爲3的倍數, $\cdots$, 認眞地找, 總可從每期開出的6碼中, 找到一些有趣的現象。當期數夠多後, 更易從其間找到一些有趣的現象 (如北銀樂透彩39號曾連續5期出現)。這並不奇怪, 除非經過統計檢定, 否則不要輕易判定號碼並非隨機地出現。

另外, 有些我們以爲不容易發生的事件, 其發生的機率其實並沒有想像中的小。見底下的兩個例子。

例4. 在n取r的樂透彩中, 頭獎號碼會有連號的機率爲

$$1-\frac{\binom{n-r+1}{r}}{\binom{n}{r}}.$$

若是42取6, 則此機率爲

$$1-\frac{\binom{37}{6}}{\binom{42}{6}}\doteq 0.5568,$$

超過二分之一。因此看到連號不用太驚訝。但是否因此簽注時該簽連號, 使中頭獎的機率較大呢? 此點留給讀者自己回答。

例5. 對北銀發行的樂透彩, 假設每期簽5注, 連續50年, 又假設北銀樂透彩的發行方式一直未改變。則至少會中一次的機率爲何?

由於每週發行兩期, 1年104期, 50年共5,200期。則50年間至少中一次頭獎之機率爲

$$1-(1-\frac{5}{5{,}245{,}786})^{5{,}200}\doteq 0.004944\doteq\frac{1}{202}.$$

約爲兩百分之一。對中頭獎而言, 這是一不算小的機率。不過仔細一想, 也做了不少投資。50年間共簽了$5\cdot 5,200=26,000$(注), 佔全部注數

$$\frac{26{,}000}{5{,}245{,}786}\doteq 0.004956\doteq\frac{1}{202}.$$

利息不計, 共花了一百三十萬元 (每注50元)。

由上例可得到一些啓示: 在一個人的一生中, 自己或認識的人裡, 有中頭獎 (或發生很特殊的事件) 者, 是不太稀奇的。民國90年12月, 台北市新開幕的京華城購物中心, 爲了促銷, 推出一百名休旅車抽獎活動, 每天抽10部, 購物每滿2,000元就可兌換一張抽獎券。一對夫婦合計抽中7部車, 造成不小的轟動。這對夫婦共花了三百多萬元, 換來1,500餘張抽獎券。抽獎活動期間, 共投進約十四、五萬張彩券, 每天箱內究竟有多少張彩券並不確定, 要算他們中7部車的機率並不容易。不過利用波松近似 (Poisson approximation, 見例6之後的註1), 估計此機率約萬分之一左右, 當然是很小。但從新聞的觀點, 只要有一這類幸運發生皆會引起注意 (也不一

定要7部車, 只要5部以上大約就有新聞價值了), 並不限京華城, 任何一家百貨公司, 任何一種抽獎活動, 或任何一特殊事件皆行, 當然也不一定要發生在台北市。如此一來發生的機率便更高了。

一件事若發生在每個人身上的機率爲百萬分之一, 則台灣兩千三百萬人, 每天發生二十餘件是毫不稀奇的。樂透彩中獎機率雖很低, 但若每一期賣出上千萬張, 則有幾個人中頭獎, 是很合理的。這個道理應不難弄明白。不用因此常揣測那些人是如何中頭獎的。

例6. 美國紐約時報曾在第一版 (1986年2月14日) 報導一位名叫 Adams 的女士第二度贏得紐澤西 (New Jersey) 州的樂透彩頭獎。1985年10月24日, 她第一次得三百九十萬美元, 第二次則得一百五十萬美元。這是紐澤西州第一次有人得到兩次百萬美元以上獎金的樂透彩。

第一次中的樂透彩是39取6, 中頭獎之機率爲

$$\frac{1}{\binom{39}{6}}=\frac{1}{3{,}262{,}623}.$$

第二次中的樂透彩是42取6, 中頭獎之機率爲

$$\frac{1}{\binom{42}{6}}=\frac{1}{5{,}245{,}786}.$$

樂透彩主辦單位說, 一個人一生中中兩次頭獎之機率爲

$$\frac{1}{3{,}262{,}623}\cdot\frac{1}{5{,}245{,}786}\doteq\frac{1}{1.7115\cdot 10^{13}},$$

約十七兆分之一。

這樣算對嗎?

上述計算是假設 Adams 兩種彩券各買一張。事實上 Adams 每週買好幾張且買了好幾年。而且在第一次中頭獎後, 便增加每週買的張數。若在39取6的玩法裡, 每週買3張, 在42取6的玩法裡, 每週買5張, 則每週有大於百萬分之一的機率中頭獎:

$$1-(1-\frac{3}{3{,}262{,}623})(1-\frac{5}{5{,}245{,}786})\doteq 1.87265\cdot 10^{-6}.$$

就用百萬分之一計好了, 在4年 (約200期, 每週一期) 裡, 一次頭獎皆未中的機率約爲

$$(1-\frac{1}{1{,}000{,}000})^{200}\doteq e^{-\frac{200}{1{,}000{,}000}}=e^{-\frac{1}{5{,}000}}.$$

利用波松近似, 四年裡恰好中一次頭獎的機率約爲

$$\frac{1}{5{,}000}e^{-\frac{1}{5{,}000}}\doteq\frac{1}{5{,}000},$$

恰好中兩次頭獎的機率約爲

$$\frac{1}{2}(\frac{1}{5{,}000})^2e^{-\frac{1}{5{,}000}} \doteq \frac{1}{50{,}000{,}000},$$

約五千萬分之一。至於一個人終身 (以30年, 1,500期計) 恰好中兩次頭獎之機率則約爲

$$\frac{1}{2}(\frac{1{,}500}{1{,}000{,}000})^2e^{-\frac{1{,}500}{1{,}000{,}000}} \doteq 1.125\cdot 10^{-6}.$$

略超過百萬分之一。

不論是五千萬分之一, 或百萬分之一的機率當然都很小。但紐澤西州人口超過八百萬, 若其中有一百萬 $(=10^6)$ 人, 一生中每期皆以上述方式買彩券 (兩種各買3張及5張), 則該州會有人一生中至少中兩次頭獎之機率便很大了:

$$1-(1-1.125\cdot 10^{-6})^{10^6} \doteq 1-e^{-1.125} \doteq 0.6753.$$

若全美有五千萬 $(=5\cdot 10^7)$ 人, 每期皆以上述方式買彩券, 則即使只在4年裡, 至少有一人中兩次頭獎的機率便已不算小了:

$$1-(1-\frac{1}{5\cdot 10^7})^{5\cdot 10^7} \doteq 1-e^{-1} \doteq 0.6322.$$

1998年, Humphries 第二度贏得賓州樂透彩頭獎, 兩次合計有六百八十萬美元的獎金。千萬不要小看大數的威力。

有關巧合事件之討論, 可參考黃文璋 (1999b) 一文。

註1. 若$n\to\infty$時, $a_n\to 0$, 且$a_nb_n\to c$, 其中$|c|<\infty$, 則$n\to\infty$時, $(1+a_n)^{b_n}\to e^c$。又若隨機變數X_n有參數爲n及p_n之二項分佈, $n\geq 1$, 且滿足$\lim_{n\to\infty} np_n=\lambda$, $0<\lambda<\infty$, 則

$$\lim_{n\to\infty} P(X_n=k)=\frac{e^{-\lambda}\lambda^k}{k!}, \quad k=0,1,\cdots.$$

這就是所謂波松近似, 爲機率中一重要的結果。

人的天性很可能是不具有隨機性的。Boland and Pawitan(1999) 一文曾做底下的實驗: 他們在所開設的初等統計學課程中, 以愛爾蘭國家樂透彩的玩法 (亦爲42取6), 要學生每人隨機地寫出一組頭獎號碼, 如此得到234組號碼。結果這234組號碼通不過隨機性的檢定。

隨機性的檢定是什麼呢? 我們以下例來說明。

例7. 你拿到一個銅板, 想看它是否爲公正。也就是想知道銅板正、反面出現之機率是否均爲1/2。假設此銅板爲公正, 隨機投擲10次, 令X表所得正面數。下表給出 $X\leq c$ 之機率 $P(X\leq c)$, $c=0,1,\cdots,10$。

c	0	1	2	3	4	5	6	7	8	9	10
$P(X \le c)$	.001	.011	.055	.172	.377	.623	.828	.945	.989	.999	1

我們不會要求得到5個正面才相信此銅板爲公正, 因

$$P(X = 5) \doteq 0.623 - 0.377 = 0.246,$$

機率小於四分之一, 並非那麼大。但若得到8個正面, 可能就會懷疑此銅板出現正面之機率可能大於1/2, 此因至少得到8個正面之機率

$$P(X \ge 8) = 1 - P(X \le 7) \doteq 1 - 0.945 = 0.055,$$

並不太大。若得到9個正面懷疑便更強烈:

$$P(X \ge 9) = 1 - P(X \le 8) \doteq 0.011,$$

此值更小。若得到10個正面 (機率爲$1/1{,}024 \doteq 0.000977$) 懷疑心當然更強烈了。至於機率多小才該懷疑, 乃視不同情況而定。一般來說0.1就算小, 0.05可說夠小, 0.01則是很小了。

上例說明統計裡假設檢定的基本想法, 與刑事訴訟法上的無罪推定原則 (被告未經審判證明有罪確定前, 推定其爲無罪) 類似。在隨機性的檢定裡, 便是先相信各號碼出現的機率相同, 然後看會出現如此異常的機率是否夠小, 以判定該不該推翻出現的機率相同之假設。

由於在北銀42取6的樂透彩裡, 共有五百多萬種不同的組合, 而一年也僅開出104期, 每一組號碼, 平均要五萬多年, 才會出現一次, 所以目前無法以各組號碼的出現頻率是否符合該有的頻率, 來做檢定。因此須以其他方式檢定。通過檢定倒不一定表示號碼爲隨機產生, 只是說尚無不合; 但若不通過, 大約便不相信號碼爲隨機產生。

假設有42注樂透彩號碼:

$$\begin{gathered} 1,2,3,4,5,6; 7,8,9,10,11,12; \cdots; 37,38,39,40,41,42; \\ \vdots \\ 1,2,3,4,5,6; 7,8,9,10,11,12; \cdots; 37,38,39,40,41,42. \end{gathered}$$

即依序從1開始每次寫6個數字, 共6循環。這42注號碼隨機嗎? 雖然1至42每個號碼出現的次數一樣多, 都是6次。但卻無諸如$(1,7)$, $(21,42)$這種 “對” 出現, 即每注中號碼之差異沒有大於5者。因此這42組有規律的號碼, 是通不過檢定的。

我們也可對偶數個數W, 最小間距MG, 最大間距L, 數字和S, 總間距數D, 及連號等做檢定。以總間距數爲例, 在簽注時, 等差數列爲許多人所愛好, 等差數列之總間距數爲1, 但會

出現等差數列之機率其實很低。表1至表6給出隨機變數W等之機率分佈。很多證據顯示, 一般人“ 隨意寫” 的號碼是不易符合隨機性的。讀者可試著寫50個1至42的數字, 許多人認為奇數較隨機, 因此隨意寫的數字常以奇數居多, 看你的結果如何? 這方面的討論可參考黃文璋、洪宛蘋及羅夢娜 (2002) 一文。大家再回想本節一開始所提的那位記者, 自行 “隨機採樣” 很可能不是眞正隨機, 而只是隨意罷了。

由於缺乏隨機性的概念, 大部分人雖欲追求明牌, 但其實所追逐的往往卻是 “名牌” 。樂透彩除了普獎外, 是由中獎人均分該獎獎金。而每組號碼中獎機率又相同, 所以該簽注熱門號碼還是冷門號碼, 道理應很容易明白。德國的樂透彩為49取6, 1993年10月16日那期共賣出6,803,090張彩券,

表7給出最熱門的20組號碼。諸位看, 如果簽中頭獎, 卻要與$4,000$人共分獎金, 頭獎獎金如果是一億元, 則每人只分到兩萬五千元。這將是件多麼令人難過的事。等差數列、過去的頭獎號碼、修改過去頭獎號碼、別國頭獎號碼、與重大事件有關的號碼等, 都是一般人喜歡簽注的, 這些其實是名牌而非明牌。由表7可看出追求名牌之不智。與其追求明牌卻追成名牌, 倒還不如聽天由命 (隨機地選, 或採電腦選號), 至少結果不會更壞。

附帶一提, 那是否電腦選號較個人選號, 有較大之中獎機率呢? 我們看底下中國時報92年3月29日14版記者蔡沛恆的一則報導。

> 昨日彩券銷售額降至五億五千七百萬元, 是去年底以來新低。北銀彩券部經理楊瑞東表示, 面對樂透彩銷售金額出現 “盤跌” 走勢, 北銀確實傷透腦筋, 甚至連 “取消電腦選號” 的方式都考慮過, 後來因為影響層面過大而暫時作罷。
>
> 採用電腦選號可適度提升中獎率, 北銀評估暫停電腦選號主要是為了增加「槓龜」機會, 頭彩可以累積, 買氣自然上升。
>
> 楊瑞東進一步指出, 目前電腦選號比重約占六成, 六億元的銷售量等於有三億六千萬元採電腦選號。換算每二億六千三百萬元的銷售額就能開出一個頭獎, 與最近每期頭獎得主一到二名的實際情況相比, 就能證明電腦選號果然保證每期都能開出頭獎, 北銀樂彩的銷售量就欲高不易。

究竟電腦選號是否可適度提升中獎率? 暫停電腦選號是否可增加槓龜機會? 電腦選號是否可保證每期都能開出頭獎? 這幾點有對的也有錯的, 也留給讀者自行思索。

表1. 42取6樂透彩 $W=i$ 之機率。

i	0	1	2	3	4	5	6
機率	0.01034	0.08146	0.23959	0.33720	0.23959	0.08146	0.01034

表2. 42取6之樂透彩 $MG=l$ 之機率。

l	1	2	3	4	5	6	7	8
$P(MG=l)$	0.5568	0.2704	0.1163	0.0422	0.01186	0.002183	0.0001748	0.000001334

表3. 42取6樂透彩 $L=k$ 之機率。

k	1	2	3	4	5	6	7	8
機率	0.000007	0.000203	0.001272	0.004276	0.010326	0.020233	0.034169	0.051322
k	9	10	11	12	13	14	15	16
機率	0.069558	0.085374	0.095205	0.097488	0.093070	0.084134	0.072984	0.061337
k	17	18	19	20	21	22	23	24
機率	0.050294	0.040461	0.032071	0.025100	0.019396	0.014778	0.011083	0.008167
k	25	26	27	28	29	30	31	32
機率	0.005898	0.004163	0.002862	0.001908	0.001227	0.000755	0.000440	0.000240
k	33	34	35	36	37			
機率	0.000120	0.000053	0.000020	0.000006	0.000001			

表4. 42取6樂透彩 S 落在各區間之機率。

區間	$[21,99)$	$[99,113)$	$[113,124)$	$[124,135)$	$[135,146)$	$[146,160)$	$[160,238)$
機率	0.14039	0.14079	0.14244	0.15276	0.14244	0.14079	0.14039

表5. 42取6樂透彩 $D=i$ 之機率。

i	1	2	3	4	5
機率	0.000030	0.005250	0.107826	0.466809	0.420086

表6. 連號情況之機率。

情況	111111	21111	2211	222	3111	321
機率	0.44317	0.41547	0.07554	0.00148	0.05036	0.00889
情況	33	411	42	51	6	
機率	0.00013	0.00444	0.00025	0.00025	0.000007	

註. 111111表無連號, 21111表恰有一組二連號, 餘類推。

四. 隨機密碼

在電影裡屢有底下這類場景: 想潛入某人之電腦, 一再試他的密碼都不對。突然看到他桌上貼著女友的照片, 你知道他女友叫 Jeniffer, 一試果然對了。再看底下91年11月25日, 中廣

新聞網的一則報導。

表7. 1993年10月16日德國樂透彩最熱門之20組號碼。

排名	組合						張數	排名	組合						張數
1	7	13	19	25	31	37	4004	11	8	14	21	25	36	39	2083
2	7	14	21	28	35	42	3817	12	6	25	27	30	34	39	1896
3	5	27	34	35	37	49	3698	13	9	17	20	21	26	41	1868
4	1	2	3	4	5	6	3249	14	2	10	18	26	34	42	1551
5	4	11	18	25	32	39	2821	15	5	10	15	20	25	30	1527
6	13	19	25	31	37	43	2335	16	44	45	46	47	48	49	1489
7	6	12	18	24	30	36	2288	17	12	24	32	36	40	42	1459
8	9	17	25	33	41	49	2227	18	1	10	20	30	40	49	1387
9	1	9	17	25	33	41	2116	19	43	44	45	46	47	48	1341
10	8	16	24	32	40	48	2097	20	1	7	22	28	43	49	1317

矇對別人的提款卡密碼123456, 盜領十多萬港幣。

> 香港一名三十四歲的江姓男子, 今年四月初在一部自動櫃員機上, 撿到一張先前客户未取走的提款卡。他隨便亂按六個號碼, 竟給他矇對了, 於是他先後盜領十餘萬港幣。
>
> 這名江姓男子撿到提款卡後, 隨意按下 [123456]這六個號碼, 沒想到竟然成功進入銀行系統, 於是他立即盜領兩萬港幣。在這張卡失效前, 他前前後後一共盜領十二萬六千多塊港幣。

在針鋒相對 (Insomnia) 那部電影裡, 飾演警探的艾爾帕西諾 (Al Pacino), 受到嫌犯羅賓威廉斯 (Robin Williams) 的要脅。帕西諾將一把槍藏在空調的排氣口, 以爲應很隱密, 奈何人同此心, 被威廉斯找到而拿走。

如果要藏東西該如何藏呢? 將適合藏的地點編號, 隨機地挑一個, 應是最難被找到的。假設要以英文字母或數字造密碼, 隨機挑選應是最難被破解的。如果你不夠 “隨機” (如前指出, 一般人缺乏隨機性的, 有人一寫就是123456, 以爲別人必想不到), 可用抽籤或藉助隨機數表。當然這樣做也是要付出代價的。由於那串密碼可能毫無意義, 不容易記憶, 自己可能會忘記。

我們再引一民國92年3月29日, 中廣新聞網記者韓啓賢的一則報導。

> 最近新出現的 “網路芳鄰” 電腦病毒, 主要是入侵“密碼可以輕易被破解” 的電腦伺服主機。因此, 防毒公司呼籲網路使用者, 取個特殊的密碼, 並經常更換, 才是避免被電腦病毒入侵破壞的最好辦法。$\cdots$。

> 防毒軟體公司對此表示, 破解密碼已成爲駭客快速入侵企業網路的模式。而且, 駭客通常是使用 “字典攻擊法” 進行攻擊。這種“字典攻擊法”, 就是以特定程式將所有字典上的單字逐一嘗試, 破解密碼。而這隻 “ 網路芳鄰” 電腦病毒, 就是採用“字典攻擊法” 滲透上萬部電腦系統。
>
> 因此, 防毒軟體公司呼籲網路使用者, $\cdots$, 取個沒有邏輯可循的密碼, 避免使用個人或親朋好友的生日或電話號碼, 英文字或是純數字組合, $\cdots$。

我們來簡單算一下好了。一般英文字典可能有爲數十萬左右的單字。但若以英文單字爲密碼, 很可能是採常見字。這種常用字, 總不超過兩萬個。若以26個英文字母加上$0, 1, \cdots, 9$等10個阿拉伯數字混和編碼, 共36個字母或數字。則長度爲6的字母數字串 (這是國內航空公司訂位代號的編碼方式), 共有

$$36^6 = 2,176,782,336$$

種組合, 爲一般人會想到的英文字 (如前以20,000個計) 的10萬倍以上。如果原先的密碼, 字典攻擊法平均一天可破解, 採隨機編碼, 平均而言, 便要十萬天 (約273年) 才能破解, 安全性當然大幅度地提高。駭客大約就一籌莫展了。假若還不放心, 用長度爲7的字母數字串, 那就更保險了。

在世說新語雅量篇, 周顗指著州官顧和的心問他 “此中何所有?” 顧和答以 “此中最是難測地”。後來周顗去見丞相王導, 對他說“卿州吏中, 有一令僕才!”

令僕乃指尚書令及僕射, 都是官名, 在唐、宋爲宰相之職。一語道出心最難測, 便被後來也當到左僕射的周顗, 驚爲有蓋世之才。但我們已指出, 並非每個人的心皆很難測, 除非心像隨機密碼一般。如果你明白此一道理, 說不定便有令僕才了。

參考文獻

1. 黃文璋 (1999a), 數學欣賞, 華泰文化事業股份有限公司, 台北市。
2. 黃文璋 (1999b), 純屬巧合, 數學傳播季刊, 第23卷第4期, 6-21。
3. 黃文璋、洪宛蘋、羅夢娜 (2002), 樂透彩開出號碼隨機性之檢定, 中國統計學報, 第40卷第3期, 249-273。
4. 楊淑芬 (1991), 踏著歷史的足跡學數學—數學在數論教學上之應用, 科學月刊, 第22卷第1期, 64-71。
5. 楊重駿、楊照崑 (1983), 數論在密碼上的應用 (上)、(下), 數學傳播季刊, 第7卷第2期, 16-22, 第3期, 2-7。
6. 楊重駿、楊照崑 (1986), 數字密碼的一些新研究, 數學傳播季刊, 第10卷第3期, 29-34。
7. Boland, P. J. and Pawitan, Y. (1999), Trying to be random in selecting random numbers for lotto, Journal of Statistics Education 7, 1-9.

—本文作者任教於國立高雄大學應用數學系—