

決策的誤差

黃文璋

1. 決策

一艘俄國的核子潛艦向美國釋出投誠的訊息。在尾隨其後的美國一艘潛艦上，美方一分析家相信他們真的要投誠，但艦長半信半疑，為先發制人，隨時準備發射魚雷。在深海中狹窄的水道中航行，到一出口，分析家為了讓艦長相信他對俄國潛艦的判斷準確，說“你看他們要右轉。”果然是右轉，艦長立刻下令攻擊取消的命令。事後好奇地問分析家怎知他們要右轉？分析家說“我其實不知道，但二分之一的機會。”這是電影獵殺紅色十月 (The Hunt for Red October) 中的情節，史恩康納萊 (Sean Connery) 飾演那位俄國艦長，亞歷鮑德溫 (Alec Baldwin) 則是那位擅用機率的分析家。

我們常在做各種決策。例如，病人生什麼病？這種藥有沒有效？這位被起訴者是否真有罪？銅板出現正面的機率是否為0.6？公司下一年之獲利是多少？A 電池用較久還是 B 電池？買樂透彩該簽什麼號碼？颱風是否會登陸？

對一件事做某種決定，皆可稱為決策。諸如估計、推論、檢定、預測、選擇、判決、診斷，皆屬於決策。而自古以來，人們是如何做決策呢？求神、卜卦、扶乩、擲爻、命相、八字、測字、星座、風水、姓名學、水晶球、陰陽五行、紫微斗數、問靈媒、問智者、抽籤、丟銅板等，皆是一些常用的方法。例如，甲骨文字，是殷人卜事之辭，刻於龜甲獸骨上，清光緒25年，於河南省安陽縣出土。古代的占卜之術，是先拿利刃在龜甲上刻劃，火燒後由割痕烤出的裂痕，以對未來做決定。在屈原的卜居一文裡寫著，由於“心煩慮亂，不知所從，乃往見太卜鄭詹尹曰‘余有所疑，願因先生決之。’詹尹乃端策拂龜曰‘君將何以教之？’”最後“詹尹乃釋策而謝曰‘...，用君之心，行君之意，龜策誠不能知此事。’”原來龜策也有束手無策的時候。其他還有相傳周文王善演“天數”，能依據神農伏羲演成的八卦，以定人事之吉凶休咎。封神演義中說他隨取金錢，便能占演凶吉。書裡那些神仙更只要屈指一算，便能知吉凶。又在三國演義裡，司馬懿仰觀天文，見將星失位，就知孔明必然有病，不久將死。

有些決策方式，也許不是那麼有科學依據，純靠運氣，如擲爻。爻通常是木製的，近人為了方便也常用兩個銅板取代，甚至也有將檳榔剖成兩半來投擲。先膜拜再擲爻請示神明，要一陰一

陽（一正一反）才表神明同意。慎重時還要得到三次一陰一陽才行。以擲銅板為例，假設是公正的銅板，大家知道要得到一正一反之機率為二分之一，並非太困難。就算沒得到，多試幾次也就成了。以這種方式來做決策，只能參考用，不值得多討論。

有些決策是可以較有科學依據的。例如，產品壽命的估計、品質的檢驗、醫學上的診斷、經濟的預測、民意調查等，都會用到現代的工具：先收集資料(data)，再整理資料，然後分析資料，最後給出推論並做出決策。

對於隨機現象所做的決策，不論是純用猜的，或依據某種科學方法，皆難免有失準的時候，但總希望雖不中亦不遠矣，因此誤差的大小，是做決策時所須顧慮的。這其中又有一些令人迷惑的問題。例如，偶有氣象局宣佈降雨機率為 0.9 卻沒有下雨，降雨機率為 0.1 反而傾盆大雨。有人會以專家的口吻告訴你，不能只由一天、兩天的結果，就下結論說準或不準，要看很多天。就像二分之一的機率，表銅板投擲 1 百萬次，就約有 50 萬次是正面，50 萬次是反面。真的是這樣嗎？

2. 誤差

民國 94 年 3 月 26 日，為了表達台灣人民抗議大陸制定反分裂法，民主和平護台灣大聯盟在台北舉行遊行。遊行落幕後，台北市長馬英九表示遊行人數達 27 萬 5 千人，民進黨則認為遊行人數超過百萬，遂引發一場口水戰。如果連呈現在眼前的一群人之數目，估計值都可有如此大的差距，對於那不易見到的，如台灣雲豹總數之估計，又如何可信呢？另外，報上也屢可見到未染愛滋 (AIDS)，卻檢驗出愛滋，受檢者明明未染病，檢驗出來的結果偏偏是陽性。負責檢驗的醫院一再強調數據沒錯，那到底是誰的錯？

對非隨機現象，所做的決策可以沒有誤差。例如，欲決定 $3 + 2$ 之值，則一定要說是 5，寫成 4.99 或 5.0001 都不行。數學上的計算，精確是很重要的。我們從數學裡學到細心、一絲不苟。我們深怕失之毫釐，謬以千里。數學家有時會引用蝴蝶效應(butterfly effect)，來說明有些表面上看起來非常微小似乎毫無關聯的變化，卻可能產生關鍵性的影響，而引起巨大的改變。這是美國麻省理工學院的數學家及氣象學家，專研混沌理論(chaos theory) 的勞倫茲(Edward N. Lorenz)，在一篇發表於 1963 年的論文中首先提出。起先是以海鷗 (seagull) 為例，於 1972 年他的一場演講中，海鷗被蝴蝶取代。當時的演講題目為：

Predictability: Does the Flap of a Butterfly Wings in Brazil set off a Tornado in Texas?

巴西一隻蝴蝶揮舞翅膀，輕微改變周圍的氣壓，最後居然引起數千里外美國德州的一場颶風。看起來頗不可思議，但有時就是有這種現象發生。

對於隨機現象，例如投擲一有兩個面的銅板，是無法每次都說對出現那一面。其實對於非隨機現象，決策也常會有誤差。例如，前述遊行人數，雖為一定值，不過除非對所有參加遊行者皆點名，否則只要是估計，總難免有誤差。但參加遊行者，來自四面八方，加入時間不一致，離去時間也不一致，更何況人數眾多，很難做到將所有參加者都數一遍而不出錯。這類例子很多。例如，對一特定的山及一特定的時刻，山上的雲豹數並非隨機，只是往往很難做到把它們都抓來數一遍而不出差錯。

人們一向知道誤差的重要，也知道必須要能掌握誤差的大小。以圓周率為例，這是長久以來，一直引人入勝的一個數。人們早就知道圓周長比直徑為一定值，並稱之為圓周率。圓周率的符號 π 則是大數學家歐拉(Leonard Euler) 最先引進的。在舊約聖經列王紀上的第7章第23節：

他又鑄一個銅海，樣式是圓的，高五肘，徑十肘，圍三十肘。

其中“他”指所羅門王。當時的人認為 $\pi = 3$ 。在西元一世紀寫成的周髀算經中，亦有“圓徑一而周三”的說法。那時中國人亦認為 $\pi = 3$ 。古希臘時代，阿基米德 (Archimedes) 則求出

$$\frac{223}{71} < \pi < \frac{22}{7}.$$

即 $3.1408 < \pi < 3.1429$ ，至小數第二位正確。南北朝時的祖沖之(429-500)，給出

$$3.1415926 < \pi < 3.1415927,$$

至小數第七位正確。將 π 表示為介於此兩數之間，顯示誤差不超過 10^{-6} 。附帶一提，2002年9月，日本東京大學的安正金田(Yasumasa Kanada) 博士的研究小組，求出 π 至小數 1.2411 兆位。此數若以 A4 紙印出，假設每頁可印 4,000 位，則須 $3.10275 \cdot 10^8$ 頁。如果每 1,000 頁裝訂成一冊，將超過 31 萬冊，真是天文數字。

祖沖之所算的圓周率，是採用三國魏晉時代，劉徽(生平不詳，約西元三世紀) 所提出的割圓術。先做一半徑為 1 之圓，然後作內接正 6 邊形，其周長為 6。由此逐步算出內接正 $6 \cdot 2^n$ 多邊形的周長， $n \geq 1$ 。割圓術見之於劉徽所著的九章算術注。書中提到：

割之彌細，所失彌少；割之又割，以至於不可割，則與圓周合體，而無所失矣。

割圓術已有近代積分的概念。在微積分裡，經由分割，以小長方形的面積和，做為曲線所圍出的面積之近似值。這當然會有誤差，若誤差值會隨著分割愈細而愈小，且趨近至 0，便可以小長方形面積和之極限，定義為曲線所圍出的面積。這種要嘛能很精確(如 $3 + 2$)，要嘛能掌握誤差，且讓誤差趨近至 0，通常只有非隨機現象才辦得到。在隨機現象裡，做決策時，誤差常難以避免，誤差不會很小的情況屢屢發生。底下給一例。

例 1: 機率裡有一著名的秘書問題(secretary problem)。假設你要面試以挑選一位秘書。每面試完一位, 在面試下一位前, 便要做個要或不要的決定, 而且不能回頭。又假設你對已面試過的人選皆能做個排序。則如何挑選才能挑中最佳者呢? 直觀上, 由於不能回頭, 故此問題不會有保證選中最佳者之策略。但若改為如何挑選, 會使挑中最佳者的機率最大? 則此問題便有解了。其解為一開始面試的幾位都放棄, 而自約 $1/e$ ($\doteq 0.368$) 比例起的面試者, 若有比前面都更佳的人, 便挑選他, 其中 e 為自然對數的底。又可證明, 當面試的總人數 n 趨近至 ∞ , 會挑中最佳者之機率為 $1/e$ 。以面試 30 位為例, 由於 30×0.368 約為 11, 所以放棄前 10 位; 自第 11 位起, 若有比前 10 位都好的, 便挑選他。這是不能回頭之情況下, 會挑中最佳者之機率最大的策略。即使如此, 採用此策略, 並不保證能挑中最佳者, 挑中最佳者之機率才約 0.368。換句話說, 約有 0.632 的機率, 你得不到最想要的。

在上例中, 雖找到最佳者之機率比找不到最佳者之機率低, 但這仍是最佳策略。換句話說, 在相同條件下, 若採其他策略, 也許有時運氣好, 一挑就挑中最佳者, 但若比挑中最佳者之機率, 均不如例 1 中所提供之方法。再看一例。

例 2: 持續投擲一公正銅板 10,000 次, 每次出現正面贏得 1 元, 出現反面則失去 1 元。令

$$S_n = X_1 + \cdots + X_n, \quad n = 1, \dots, 10,000,$$

表投擲 n 次後之淨所得, 其中 X_n 表第 n 次之所得 (1 或 -1)。 $\{S_n, 1 \leq n \leq 10,000\}$ 形成一隨機過程(stochastic process), 這種特別的隨機過程又稱隨機漫步(random walk)。依假設 $\{X_n, n \geq 1\}$ 為一數列之獨立且有共同分佈的隨機變數, 且

$$P(X_1 = 1) = P(X_1 = -1) = \frac{1}{2}.$$

圖 1 給出 6 個模擬圖, 橫軸為 n , 縱軸為 S_n , 其中 $S_i > 0$ 表至第 i 次投擲正面數領先, $S_i < 0$ 則表反面數領先。可看出 6 個圖形差異很大。事實上即使總投擲數趨近至 ∞ , 圖形仍有很大變異; $n \rightarrow \infty$ 時, S_n 之極限也不存在, 雖然 S_n 之期望值 $E(S_n) = 0, \forall n \geq 1$ 。這點與前述以分割求面積, 逐漸地小長方形面積和會趨近至一定值是完全不一樣的。

由於變異的存在, 對於隨機現象所做的決策, 如何能儘量減小誤差, 就要仰賴能善用統計了。雖然會有誤差, 但還是可以依據誤差的大小來做決策。

例 3: 對於是否有親子關係的家事糾紛, 常雙方各執一詞。今日科技發達, DNA 的鑑定遂成為一可靠的依據。假設有一案例, 法官囑託 A 醫院及 B 醫院鑑定血緣關係。結果 A 醫院及 B 醫院所提供的親子關係確定率分別為 99.9993%, 及 99.988%。醫院認為均“不能排除”兩人有親子關係。由於誤差均夠小, 法官遂判定兩人有親子關係。

雖然誤差存在，但兩家醫院的鑑定結果相當一致，法官做此判定，當事人即使不服，有時考慮再上訴要翻案並不易，也只好就接受了。

這種利用發生機率的大小來做決策，就發展出假設檢定(testing hypothesis)，是統計裡一種重要的決策依據，其想法可參考黃文璋 (2005) 一文。

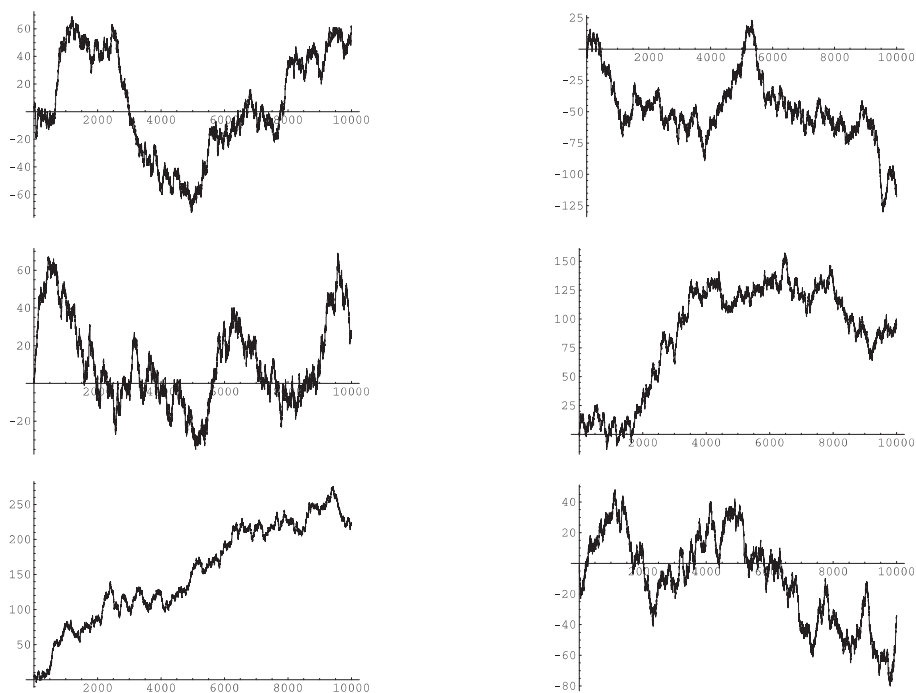


圖1. 6個 $\{S_n, 1 \leq n \leq 10,000\}$ 的模擬圖。

3. 數據說話有誤

收集資料是統計分析的一重要步驟。我們有時批評人以偏概全，其實統計裡常在做以偏概全的工作：以所收集到之樣本，對母體做推估。雖是以偏概全，我們可不想取樣有偏。也就是希望所收集到的樣本能代表母體。如此對母體之推估，誤差才不至於太大。如果收集到的是偏差而未具代表性的樣本，據此所做的推論，誤差大便不足為奇。假設一袋子中有若干個形狀大小完全一樣的白色及紅色球。現要估計白球佔袋中全部球的比例。一個簡單的方法是取出若干個(每次取出後不放入)，數看看其中白球所佔比例，以此比例來估計全袋中白球所佔比例。這樣做之一先決條件是，袋中的球要混合的很均勻，即每次取球，袋中每一球皆要有相同的機會被取中，也就是要隨機取樣(random sampling)，否則估計便不會太準。

把球混合得很均勻感覺上似乎較容易。底下一個看起來類似的情況就沒那麼容易了。如何藉由統計估計圓周率 π ？劃一正方形及其內切圓，圓的半徑設為 r 。“隨機地”灑下一些芝麻，

假設有 n 粒芝麻落進正方形內，其中又有 a 粒落進圓中。令 a 比 n 等於圓面積與正方形面積之比。即

$$\frac{a}{n} = \frac{\pi r^2}{4r^2}.$$

解出

$$\pi = \frac{4a}{n},$$

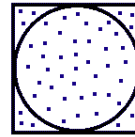


圖2. 如何估計 π 。

以此當做 π 的估計值。問題是隨機地灑芝麻容易辦到嗎？如果有往中間灑的傾向，則說不一定會有

$$\frac{a}{n} > \frac{\pi r^2}{4r^2}.$$

因此以這種方法來估計 π ，誤差有時可能不小。

處理球或灑芝麻其實還算容易，如果是處理動物或人，就更不易做到隨機取樣了。例如，要估計魚池中有多少尾魚 (N)。取出若干條 (k)，做上記號後一一放回；再“隨機”取若干條 (n)，算出其中有做記號的魚數 (l)；令第二次取出的魚中，有做記號的比例 (l/n)，等於魚池中有做記號的魚之比例 (k/N)，解出 $N = nk/l$ ，做為池中魚數之估計值。此問題看起來與前述估計白球比例或估計圓周率的問題似乎是同一類，但誤差可能大很多。如果某些魚有較易被取中的傾向，或被取出的魚，因受驚嚇而躲藏不易再被取中，則 k/N 與 l/n 可能便有很大差異了。隨機取樣可說是收集資料的一基本步驟，但連對魚都不易做到，更何況是對處理有關人的問題。甚至大部分的人對“隨機”的意義常有誤解，造成隨機取樣實際上不易辦到，可參考黃文璋 (2004) 一文。

我們再看底下一實例。

2002年11月，英國廣播公司 (BBC)，對英國境內一百萬聽眾與觀眾，票選有史以來最偉大的十位英國人。結果曾任首相的邱吉爾排名第一，他獲得約 44.7% 的票，絕對領先。數學家及物理學家牛頓 (Issac Newton) 則排名第六。隔年8月對全球聽眾所做的票選，牛頓變成排名第一，獲 21.6% 的票。邱吉爾的票則減少，只得 16%，成為第二。即使樣本數多達百萬，同一廣播公司的聽眾與觀眾，在英國與在全球，其看法差異很大。對人的調查可說一向相當困難，不要說不同地區的人，即使對同一群人，隨著情況的改變，或某些事件的發生，都可能影響一些人的想法。甚至不需要有任何原因，對同一個人做兩次調查，其對同一件事的看法都可能有異。

底下給一取材自 Freedman et al.(1991) 著名的選舉實例，以再說明對人的調查之不易精準。

例 4: 西元 1936 年，美國羅斯福 (Franklin D. Roosevelt) 總統準備競選連任，對手是代表共和黨的候選人，堪薩斯州的州長藍頓 (Alfred Landon) 先生。此時美國正從經濟大蕭條 (Great Depression) 中復甦，雖全國仍有九百萬失業人口，人民的實際收入比 1929-1933 那段

時期少了約三分之一，但情況正開始好轉。藍頓提出政府經濟計畫的政見，而羅斯福則為其財務赤字而辯護。

大部分的觀察者均預測羅斯福可輕易地連任，但文學文摘 (Literary Digest) 雜誌卻預測藍頓會以 57% 比 43% 大勝羅斯福。他們是依據高達約 240 萬份回答的問卷所做之預測。文學文摘在總統選舉的預測素負聲望，因他們自 1916 年開始做預測以來，從未錯過，只是這回栽了大跟斗。選舉結果羅斯福以 62.5% 比 37.5% 獲得壓倒性的勝利。選舉完不久，文學文摘也就破產了。

文學文摘會有這麼離譜的錯誤是很令人驚訝的。要知那幾乎可說是有史以來最大的一次民調，回收的問卷數也很多。那時蓋洛普 (Gallup) 公司才剛成立，僅使用了 50,000 個樣本，便正確地預測羅斯福會贏，雖然得票率方面有些誤差 (預測羅斯福會得 56% 的票)。

文學文摘何以會犯這麼大的錯？我們先來看他們是如何挑選樣本。如前所述，在抽樣調查裡，樣本的挑選必須很公正，才能獲得有效的資訊。若在選樣過程中，有排除 (或多取) 某一類樣本的傾向，便稱選擇偏差 (selection bias)。文學文摘寄出 1 千萬份問卷，姓名及地址的來源是他們的訂戶、電話簿及一些俱樂部的會員。訂閱他們雜誌者，顯然是一群特定的人。又在 1936 年，電話尚非那麼普及 (平均每四戶才有一具)。此外，沒有參加任何俱樂部者也被排除了。換句話說，這種抽樣過程，有排除窮人的傾向。在 1936 年以前，這種選擇偏差對於預測還沒有很大的影響，因當時富人與窮人的投票行為差異並不太大。但在 1936 年，因經濟的因素，造成選民政治傾向有很大的分野：窮人較多選擇羅斯福，而富人則傾向支持藍頓。選擇偏差是造成文學文摘犯這麼大錯誤的主因之一。

當有選擇偏差時，樣本數雖多便不見得有用。

文學文摘還犯了另一嚴重的錯誤：一旦決定了受訪名單，就要盡力去獲得他們的意見，這部分工作可說是高難度。當取出的樣本中，有過多沒有回覆或拒絕受訪，將可能造成一嚴重的扭曲，我們稱之不回答的偏差 (non-response bias)。有時不回答者與回答者的意見可能有很大的差異。舉例而言，文學文摘在芝加哥所發出的問卷數，大約是芝加哥選民的三分之一，不可謂不多，但其中回覆者才約 20%。回收問卷中，支持藍頓的超過半數。但選舉結果，羅斯福獲得芝加哥約三分之二的選票。文學文摘對全國所發出的 1 千萬份問卷中，只有比例不高的 24% 的回覆。這 24% 回覆者的意見，不見得能代表全部被挑選出之選民。所以文學文摘既犯了選擇偏差，又犯了不回答的偏差兩種錯，調查結果會準確才是奇怪。

一般而言，低收入與高收入者，不回答問卷之比例較高。也就是回收問卷中，中收入者超過該有之比率。而中收入者的意見與高收入或低收入者，不見得相同。由於有這種回答的偏差，現代民意調查機構，對於重大議題，在時間及經濟因素不成問題下，傾向採用面訪，而非郵寄問卷。面訪成功率通常可達 65% 以上，而郵寄問卷之回收率常不到 25%。不過即使採用面訪，不回

答的偏差之問題仍然存在。那些面訪時不在家者，可能與面訪時在家者，習性有很大差異：工作類別、家庭狀況、社會背景等，想法可能也就不太相同。拒絕受訪者的情況也類似。好的抽樣調查設計，會正視不回答偏差的問題，而採用較巧妙的方法以設法克服。

在醫學上，進行一實驗以收集數據，往往要先對實驗過程做一些必要的設計。我們給一例如下。

例 5：民國 94 年有一項調查指出，約有四分之一的大學生有嚴重的憂鬱情緒。看來得憂鬱症的人不少，已是一嚴重的社會問題。某製藥公司宣稱發明一種治憂鬱症的藥，該如何評估其效果？找一些人來做實驗，你可能會想到。沒錯，但通常應將接受實驗者分為兩組，一組服用此新藥，稱為處理組(treatment group)，一組則不服用此新藥，稱為控制組(control group)。接受實驗者該屬何組，採隨機的方式決定，以消除外來的偏差。如果沒有控制組，如何判定服用此藥確實有效？又實驗常採“二重隱瞞隨機化控制設計”(doubly blind randomized control experiment, 簡稱 DBRC 設計)。即接受實驗者，與負責檢查診斷者，均不知接受實驗者中，何者為處理組，何者為控制組。所以被分在控制組的人，也接受一種看起來像新藥的安慰劑(placebo)。採 DBRC 設計，可避免因主觀的期望及偏見，而造成數據的偏差。想想若檢查者知道誰屬於處理組，可能不自覺地對他多照顧些；而受測者若知道自己是在處理組，心情可能愉快些，造成憂鬱症有康復的傾向。

實驗設計的好壞，影響到所收集到的資料之品質。如果對藥效的推論，所依據的實驗不是基於 DBRC 設計，其結果往往不會太可靠。即使找再多的人來接受實驗，都可能得到誤差很大的結論。這方面的討論，可參考羅夢娜 (1987) 一文。

收集到的資料若未先經妥善的處理，便進行分析，有時會得到很荒謬的結論。底下為一例。

例 6：西元 1972-1974 年，英國進行一有關甲狀腺疾病與心臟病的研究。二十年後做後續的追蹤研究。Appleton et al.(1996) 一文對其中婦女抽煙與死亡的數據做一些分析，部分數據列在表 1，不抽煙者之死亡率明顯地較高。經統計檢定，得到不抽煙者之死亡率高於抽煙者之死亡率的推論。

表 1. 婦女抽煙與死亡的數據

| | 抽煙 | 不抽煙 | 總數 |
|-----|--------|--------|-------|
| 死亡 | 139 | 230 | 369 |
| 存活 | 443 | 502 | 945 |
| 總數 | 582 | 732 | 1,314 |
| 死亡率 | 23.88% | 31.42% | |

這樣的結果自然很令人驚訝，違反一般人的認知。不過統計上的推論，只是“證實”抽煙與不抽煙兩群人的死亡率不同，且不抽煙那群人之死亡率較高，並未證實抽煙是造成死亡率較低的原因，此點務必要留意，可參考 Simpson(1951) 對此方面的探討。

由於辛普生 (Simpson) 首先注意到此問題，後來對於兩個變數的關係，因另一變數的介入而反過來，便統稱辛普生詭論 (Simpson's paradox)。

如果仔細分析數據，發現有一重要的變數不能忽略：年齡。初次調查時，高齡者中較少是抽煙的，原因很可能是年紀大的人，較願意接受對健康有利的建議，抽煙者因而減少。我們對年齡分群，而得表2，其中以“+”表抽煙，“-”表不抽煙。

表2. 對年齡分群之抽煙與死亡的數據

| 年齡分群 | 18-24 | | 25-34 | | 35-44 | | 45-54 | | 55-64 | | 65-74 | | ≥ 75 | |
|------|-------|----|-------|-----|-------|-----|-------|----|-------|----|-------|-----|------|----|
| | + | - | + | - | + | - | + | - | + | - | + | - | + | - |
| 死亡人數 | 2 | 1 | 3 | 5 | 14 | 7 | 27 | 12 | 51 | 40 | 29 | 101 | 13 | 64 |
| 存活人數 | 53 | 61 | 121 | 152 | 95 | 114 | 103 | 66 | 64 | 81 | 7 | 28 | 0 | 0 |

經過20年，不論是否抽煙，年紀大的總是較易死亡。只要看表2中，當初75歲以上者，20年後無一存活即可得知。大家看過金庸 (1996) 的倚天屠龍記嗎？書中說 (p. 380) 張三丰於過百歲大壽時，已成名垂七十年，當年跟他動過手的人，已死得乾乾淨淨，世上再無一人。這也是類似的情況。若將表2中，65歲以上之資料去掉，而得表3。則抽煙者之死亡率便高於不抽煙者了。再經過統計檢定，將得到不滿65歲的婦女中，抽煙者之死亡率較高的推論。

表3. 不滿65歲婦女抽煙與死亡的數據

| | 抽煙 | 不抽煙 | 總數 |
|-----|--------|--------|-------|
| 死亡 | 97 | 65 | 162 |
| 存活 | 436 | 474 | 910 |
| 總數 | 533 | 539 | 1,072 |
| 死亡率 | 18.20% | 12.06% | |

上三例顯示，資料的品質不佳，常會使決策產生不小的誤差。所謂數據會說話，但若所取得原始數據之品質便不佳，或對數據的處理過程有太大瑕疵，甚至對數據的解讀有誤，所說出的話自然不會太正確。另一方面，對於隨機現象，除非是對母體(如全世界的人口)全部取樣，否則依據取樣所做的推論，與真實的值有差異，並不足為奇。

4. 掌握誤差

在例2中我們看到獨立隨機變數 $X_1, \dots, X_n, n \geq 1$, 之和 S_n , 不見得會收斂。也就是 S_n 不會隨著 n 之增大而能掌握, S_n 可能會一直跳躍。那有何我們可以掌握到的呢? 隨機並非隨意, 也是有須依循的法則。機率裡有一重要的法則, 就是大數法則(law of large numbers), 有兩個版本, 一為強大數法則, 一為弱大數法則。本文裡我們只看“弱”的版本。

設 $X_1, \dots, X_n, n \geq 1$, 為獨立且有共同分佈之隨機變數, 且設 $E(X_1) = \mu$ 存在。令樣本平均

$$\bar{X}_n = \frac{S_n}{n} = \frac{X_1 + \dots + X_n}{n}, n \geq 1.$$

則 $n \rightarrow \infty$ 時, \bar{X}_n 會機率收斂至 μ 。即 $n \rightarrow \infty$ 時, 對 $\forall \varepsilon > 0$,

$$P(|\bar{X}_n - \mu| > \varepsilon) \rightarrow 0, \tag{1}$$

或等價地說

$$P(|\bar{X}_n - \mu| \leq \varepsilon) \rightarrow 1. \tag{2}$$

(1) 式顯示 n 很大時, \bar{X}_n 與 μ 的差異超過 ε 之“機率”會很小。注意, 並未保證 \bar{X}_n 一定很接近 μ , 而是很接近 μ 的機率很大 (即機率接近1)。

在例2中因 $E(X_1) = 0$, 便有 $n \rightarrow \infty$ 時, S_n/n 機率收斂至 0。

由大數法則, 我們理解到何以估計時通常所取的樣本數要夠大。以估計銅板出現正面的機率為例。投擲一銅板, 第 n 次若正面出現, 則令 $X_n = 1$, 否則令 $X_n = 0$ 。則 $E(X_n) = p$, 其中 p 表正面出現之機率。大數法則告訴我們, 以 \bar{X}_n 做為 p 之估計, 當 n 很大時, 此估計不會太離譜。

不過光是知道事件 $|\bar{X}_n - \mu|$ 的機率很小, 我們可能仍不滿足, 此機率究竟有多小? 可不可以估計? 機率裡另一重要的法則—中央極限定理(central limit theorem) 會告訴我們。

誤差理論是大數學家高斯(Karl F. Gauss) 對機率論的主要貢獻。統計裡常要做估計, 估計會有誤差, 誤差的大小如何掌握? 在一些條件下, 高斯導出誤差有常態分佈。德國今日已用歐元, 昔日德國10馬克, 乃以高斯為人像。紙幣上伴隨著他的, 不是高斯其他數學上的成就, 而是一常態分佈的圖形。

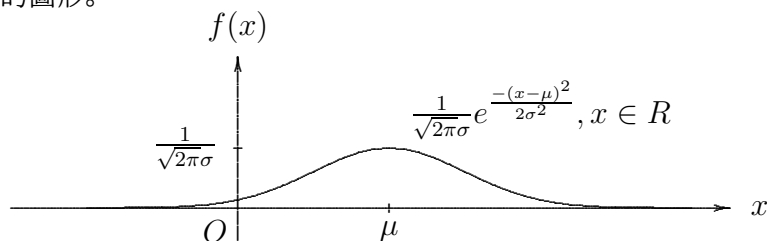


圖3. 常態分佈 $\mathcal{N}(\mu, \sigma^2)$ 機率密度函數圖形

對 $\forall n \geq 1$, 設 X_1, \dots, X_n 為獨立且有共同分佈之隨機變數, 又設 $\mu = E(X_1)$, $\sigma = \sqrt{\text{Var}(X_1)} > 0$ 皆存在。令 $\bar{X}_n = \sum_{i=1}^n X_i/n$, $n \geq 1$ 。則中央極限定理指出, n 很大時, $(\bar{X}_n - \mu)/(\sigma/\sqrt{n})$ 有近似的標準常態(以 $\mathcal{N}(0, 1)$ 表之) 分佈。即

$$\lim_{n \rightarrow \infty} P\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq z\right) = \Phi(z), \quad z \in R, \quad (3)$$

其中

$$\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx, \quad z \in R. \quad (4)$$

(3) 式導致

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| \leq z\sigma/\sqrt{n}) = 2\Phi(z) - 1, \quad z > 0, \quad (5)$$

且

$$\lim_{n \rightarrow \infty} P(|S_n - n\mu| \leq z\sigma\sqrt{n}) = 2\Phi(z) - 1, \quad z > 0. \quad (6)$$

其中如前 $S_n = X_1 + \dots + X_n$, $n \geq 1$ 。由 (5) 式, 當樣本數 n 愈大, 對相同的機率, \bar{X}_n 會落在一愈窄的區間。由 (6) 式, 對相同的機率, 當 n 愈大, 樣本和 S_n 會落在一愈寬的區間。雖 n 很大時, 樣本平均 \bar{X}_n 很可能會接近期望值 $E(X_1) = \mu$, 樣本和 S_n 倒不見得會接近 $E(S_n) = n\mu$ 。而且隨著 n 的增大, S_n 與 $n\mu$ 的差異有變大的趨勢。分別取 $z = 1, 2, 3$, 由標準常態分佈的機率值表, (5) 式導致當樣本數 n 很大時,

$$\begin{aligned} P(\bar{X}_n \in [\mu - \sigma/\sqrt{n}, \mu + \sigma/\sqrt{n}]) &\doteq 0.6827, \\ P(\bar{X}_n \in [\mu - 2\sigma/\sqrt{n}, \mu + 2\sigma/\sqrt{n}]) &\doteq 0.9545, \\ P(\bar{X}_n \in [\mu - 3\sigma/\sqrt{n}, \mu + 3\sigma/\sqrt{n}]) &\doteq 0.9973. \end{aligned} \quad (7)$$

上三式又分別等價於

$$\begin{aligned} P(S_n \in [n\mu - \sigma\sqrt{n}, n\mu + \sigma\sqrt{n}]) &\doteq 0.6827, \\ P(S_n \in [n\mu - 2\sigma\sqrt{n}, n\mu + 2\sigma\sqrt{n}]) &\doteq 0.9545, \\ P(S_n \in [n\mu - 3\sigma\sqrt{n}, n\mu + 3\sigma\sqrt{n}]) &\doteq 0.9973, \end{aligned} \quad (8)$$

又當 σ 已知, (7) 式即分別給給出 μ 的 68.27%, 95.45%, 99.73% 信賴區間 (confidence interval):

$$\begin{aligned} P(\mu \in [\bar{X}_n - \sigma/\sqrt{n}, \bar{X}_n + \sigma/\sqrt{n}]) &\doteq 0.6827, \\ P(\mu \in [\bar{X}_n - 2\sigma/\sqrt{n}, \bar{X}_n + 2\sigma/\sqrt{n}]) &\doteq 0.9545, \\ P(\mu \in [\bar{X}_n - 3\sigma/\sqrt{n}, \bar{X}_n + 3\sigma/\sqrt{n}]) &\doteq 0.9973. \end{aligned} \quad (9)$$

也就是可以 \bar{X}_n 來估計 μ 。但若想知道 μ 大致落在那一區間，則可以信賴區間來表示。給出信賴區間，有雖不中亦不遠矣的意思。例如，約有 0.9545 的機率， μ 會落在區間 $[\bar{X}_n - 2\sigma/\sqrt{n}, \bar{X}_n + 2\sigma/\sqrt{n}]$ 。此區間的半徑或者說對 μ 估計的誤差 $2\sigma/\sqrt{n}$ ，通常希望不要太大。在 σ 不變下，要減小估計誤差，就只好靠提高樣本數 n 了。

回到第 2 節最後所提的投擲銅板 1 百萬次之例。

令 $X_n = 1$ 表第 n 次投擲出現正面， $X_n = 0$ 表第 n 次投擲出現反面，則 $S_n = X_1 + \cdots + X_n$ ， $n \geq 1$ ，表 n 次投擲共得之正面數。假設銅板為公正，則 $\mu = E(X_1) = 1/2$ ， $\sigma^2 = \text{Var}(X_1) = 1/4$ 。對於 $n = 1,000,000$ ，由 (8) 式 (為了簡便，正面數仍以 S_n 表之)

$$\begin{aligned} P(S_n \in [499,500, 500,500]) &\doteq 0.6827, \\ P(S_n \in [499,000, 501,000]) &\doteq 0.9545, \\ P(S_n \in [498,500, 501,500]) &\doteq 0.9973. \end{aligned}$$

約有 0.3173 的機率，正面數與期望值 50 萬之差異超過 500；約有 0.0455 的機率，正面數與期望值 50 萬之差異超過 1,000；約有 0.0027 的機率，正面數與期望值 50 萬之差異超過 1,500。即使投擲數高達 1 百萬，正面數並非就一定很接近期望值 50 萬。

最後要強調的是，誤差到底算大還是小，常與樣本數 n 有關，而非只看誤差絕對值的大小。見下例。

例 7：投擲一銅板若干次，正面數出現比率為 50.114%，僅比 50% 略多一些。試問是否不足以推翻此銅板為公正？

解：結論為何與投擲數 n 有關。我們先以下述模型來描述問題。設 X_1, \dots, X_n 為獨立且有共同分佈之隨機變數， $n \geq 1$ ，且設 $P(X_1 = 1) = p = 1 - P(X_1 = 0)$ 。則 $\mu = E(X_1) = p$ ， $\sigma^2 = \text{Var}(X_1) = p(1 - p)$ 。我們想判斷究竟 $p = 0.5$ ，還是 $p > 0.5$ 。直觀上 \bar{X}_n 較大，譬如說 $\bar{X}_n > c$ ，其中 c 為一常數，就較不相信 $p = 0.5$ ，而接受 $p > 0.5$ 。由中央極限定理， n 很大時，

$$\frac{\sqrt{n}(\bar{X}_n - p)}{\sqrt{p(1 - p)}} \text{ 有近似的 } \mathcal{N}(0, 1) \text{ 分佈。}$$

現依題意觀測到 $\bar{X}_n = 0.50114$ 。

(1) 設 $n = 13,000,000$ 。如果 $p = 0.5$ ，則

$$P(\bar{X}_n \geq 0.50114) \doteq P(Z \geq \frac{\sqrt{13,000,000}(0.50114 - 0.5)}{\sqrt{0.5 \cdot 0.5}})$$

$$\begin{aligned} &\doteq P(Z \geq 3605.55 \cdot 0.00228) \\ &\doteq P(Z \geq 8.22) \\ &\doteq 1.03 \cdot 10^{-16}, \end{aligned}$$

其中 Z 表有一 $\mathcal{N}(0, 1)$ 分佈之隨機變數。 $P(Z \geq 8.22)$ 並無法由查表獲得, 要由 x 較大時, $1 - \Phi(x)$ 之近似公式求出近似值, 在此略過。當 $p = 0.5$, $\bar{X}_n \geq 0.50114$ 的機率可說微乎其微。換句話說, 在樣本數高達 13,000,000 之下, 若 $p = 0.5$, 此時會觀測到 $\bar{X}_n = 0.50114$ 是件很不合理的事。即應判定銅板並非公正。

(2) 設 $n = 1,000,000$ 。如果 $p = 0.5$, 則

$$\begin{aligned} P(\bar{X}_n \geq 0.50114) &\doteq P\left(Z \geq \frac{\sqrt{1,000,000}(0.50114 - 0.5)}{\sqrt{0.5 \cdot 0.5}}\right) \\ &= P(Z \geq 1,000 \cdot 0.00228) \\ &= P(Z \geq 2.28) \\ &\doteq 0.0113. \end{aligned}$$

此機率雖不大, 不過約百分之一的機率也不是不會發生。這時的判斷為何就要依個人而定。

(3) 再看 $n = 10,000$ 會如何 (實際 n 不可能為 10,000, 因 \bar{X}_n 算至小數第 5 位)? 如果 $p = 0.5$, 則

$$\begin{aligned} P(\bar{X}_n \geq 0.50114) &\doteq P\left(Z \geq \frac{\sqrt{10,000}(0.50114 - 0.5)}{\sqrt{0.5 \cdot 0.5}}\right) \\ &= P(Z \geq 100 \cdot 0.00228) \\ &= P(Z \geq 0.228) \\ &\doteq 0.40978. \end{aligned}$$

這麼大的機率, 可說很容易發生, 算是很平常。對大部分的人而言, 很可能會接受 $p = 0.5$ 。

舉以上三情況為例, 已足以說明 50.114% 與 50% 之差異是否夠大, 乃與投擲數 n 有關。即 n 愈大, 此差異就可能夠大, n 較小時, 此差異便可能不夠大。如果換種問法: 正面數比反面數多 30,000 是否夠多? 此問題留給讀者自行回答。

結語

統計給的決策, 不見得百分之百可靠, 甚至常不可靠, 此與數學不同。在數學裡每一數字都力求精準, 誤差的存在常是難以忍受的。在統計分析裡, 由於誤差的存在, 有些初學者常會懷疑

是否該相信統計。我們必須正確了解誤差的內涵，也接受合理的誤差。在隨機的世界，明智地依據統計給出的結果來做判斷。

參考文獻

1. 金庸 (1996). 倚天屠龍記, 第三版. 遠流出版社, 台北。
2. 黃文璋 (2005). 統計顯著性. 數學傳播季刊, 29(4), 29-38。
3. 黃文璋 (2004). 隨機與密碼. 數學傳播季刊, 28(2), 3-17。
4. 羅夢娜 (1987). 百農嘗一草—實驗設計之應用. 科學月刊, 第18卷第5期, 338-340。
5. Appleton, D.R., French, J.M. and Vanderpump, M.P.J.(1996). Ignoring a covariate: an example of Simpson's paradox. *The American Statistician*, 50, 340-341.
6. Freedman, D., Pisani, R., Purves, R. and Adhikari, A.(1991). *Statistics*, 2nd ed. W.W. Norton & Company, New York.
7. Simpson, E.H.(1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society Series B*, 13, 238-241.

—本文作者任教於國立高雄大學應用數學系—