

概率破玄機，統計解迷離

嚴加安

概率論起源於中世紀的歐洲，那時盛行擲骰子賭博，提出了許多有趣的概率問題。當時法國的帕斯卡¹(Blaise Pascal)、費爾馬²(Pierre de Fermat) 和旅居巴黎的荷蘭數學家惠更斯³(Christiaan Huygens) 都對此類問題感興趣，他們用組合數學研究了許多與擲骰子有關的概率計算問題。20世紀30年代柯爾莫哥洛夫⁴(Andrey Nikolaevich Kolmogorov) 提出概率公理化，隨後概率論迅速發展成為數學領域裡一個獨立分支。

隨機現象背後是隱藏某些規律的，概率論的一項基本任務就是揭示這些規律。現在概率論已經發展成為數學領域裡一個相對充滿活力的學科，並且在工程、國防、生物、經濟和金融等領域得到了廣泛的應用。

統計學是一門具有方法論性質的應用性科學，它在概率論基礎上，發展出一系列的原理和方法，研究如何採集和整理反映事物總體資訊的數字資料，並依據這些複雜的資料（稱為樣本）對總體的特徵和現象背後隱藏的規律進行分析和推斷。

法國數學家拉普拉斯⁵(Pierre-Simon marquis de Laplace) 有句名言：“生活中最重要的問題，絕大部分其實只是概率問題”。當代國際著名的統計學家 C. R. 勞⁶(Rao) 說過：“如果世界中的事件完全不可預測的隨機發生，則我們的生活是無法忍受的。而與此相反，如果每一件事都是確定的、完全可以預測的，則我們的生活將是無趣的。”

我長期從事概率論和隨機分析研究，對概率統計學科的本質有些領悟，曾寫過下面這首“悟道詩”：

隨機非隨意，概率破玄機；

感謝嚴加安教授同意本刊登載他原發表於《中國數學會通訊》2012年第4期的文章。本文下方註解非原稿所有，為本刊所加，希望方便讀者閱讀。

¹Blaise Pascal (1623~1662)，法國十七世紀物理、數學、哲學家。數學成就如射影幾何的 Pascal 定理，與 Pierre de Fermat 是機率論的奠基者。

²Pierre de Fermat (1601~1665)，法國十七世紀律師與業餘數學家，在解析幾何，機率及光學有顯著的貢獻，更以數論的工作著名。有名的費瑪最後定理是他在 Diophantus' Arithmetica 上的眉批。

³Christiaan Huygens (1629~1695)，十七世紀荷蘭數學、物理和天文學家。

⁴Andrey Nikolaevich Kolmogorov (1903~1983)，二十世紀俄國數學家。其研究領域包括實變函數論、數學基礎論等多個分支。

⁵Pierre-Simon marquis de Laplace (1749~1827)，法國天文和數學學家，提出拉普拉斯定理。

⁶C. R. Rao (1920~)，美國統計與數學家。出生於印度，著有《統計與真理：怎樣運用偶然性》。

無序隱有序，統計解迷離。

本文試圖通過若干日常生活中的一些例子來向大家展示概率是如何破玄機和統計是如何解迷離的。

1. 什麼是隨機和隨意？

在社會和自然界中，我們經常遇到一些事件，因為有很多不確定的偶然因素很難判斷它會發生或不發生，這樣的事件就是所謂的隨機事件或偶然事件。

概率則是對隨機事件發生的可能性大小的一個度量。必然要發生的事件的概率規定為1，不可能發生的事件的概率規定為0，其他隨機事件發生的概率介乎0與1之間。

例如，拋一枚勻質的硬幣，出現正面或反面的概率均為二分之一；擲一個勻質的骰子，每個面出現朝上的概率均為六分之一。在這兩個例子中，每個簡單事件（或“場景”）都是等可能發生的。一個複合事件（如擲骰子出現的點數是偶數）發生的概率就等於使得該複合事件發生的場景數目與可能場景總數之比。

什麼是隨意？隨意就是帶有主觀意識的一種隨機。

比方說，我們知道擲一枚勻質硬幣出現正面或反面的概率都是 $1/2$ 。如果讓某人臆想一個相繼擲50次硬幣的可能結果，並用1和0分別表示出現“正面”和“反面”，在一張紙上寫下來，由於他考慮到接連多次出現正面或反面的可能性較小，在他寫1和0時，可能有意識避免連寫三個或四個以上的1或0，這樣產生的0-1序列就是“隨意的”。它看似隨機，但與真實做一次擲50次硬幣記錄下的結果在統計特性上是有區別的。

隨機現象背後是隱藏某些規律的，概率論的一項基本任務就是揭示這些規律。

2. 靠直覺做判斷常常會出錯

下面是一個靠直覺做判斷容易出錯的例子。某人新來鄰居是一對夫婦，只知道這對夫婦有兩個非雙胞胎孩子。某天，看到爸爸領著一男孩出門了，問這對夫婦的另一孩子也是男孩的概率是多大？許多人可能給出的答案是 $1/2$ ，因為生男生女的概率都是 $1/2$ 。但實際上正確答案應該是 $1/3$ ，因為在已知該家至少有一男孩的前提下，他家兩個小孩可能的場景是三個（按孩子出生先後次序）：“男男”，“男女”，“女男”。只有“男男”才符合“另一孩子也是男孩”這一場景。如果突然從這家傳出嬰兒的啼哭聲，“另一孩子也是男孩”的概率就變成 $1/2$ 了，因為這時可以斷定出了門的那個男孩是老大，可能的場景就變成兩個了（按出生先後次序）：男男，男女。

從這兩個簡單初等概率問題可以悟出一個道理：靠直覺做判斷常常會出錯。計算一個隨機事件發生的概率，重要的是要對此事件得以發生的所有可能場景有正確的判斷。

3. 改變生育政策會引起性別比例失衡嗎？

下面這個例子更加說明單靠直覺做判斷容易出錯。假定有一項新的生育政策，規定頭胎生男孩的不可以生第二胎，但允許頭胎生女孩的生第二胎（當然也可以選擇不再生），試問：這一政策會引起性別比例失衡嗎？從直覺上看，似乎這一政策有利於生男孩，但這一擔心是多餘的，因為生男生女的概率都是 $1/2$ ，第一胎的小孩性別比例不會失衡，第二胎的小孩性別比例也不會失衡，總體來說，生育政策不會造成性別比例失衡。由於女人生小孩的胎數有一致的上界（比如說不超過 20 胎），用概率分析可以斷言：即使有政策，允許婦女直到首次生出男孩才終止生育，理論上講也不會引起性別比例失衡。

如何解釋現實中存在的性別比例失衡呢？一方面，由於過去曾允許懷孕期間做性別檢驗，有的婦女發現懷的是女孩就墮胎了；另一方面，一些人重男輕女思想造成幼年女孩夭折的比例高於男孩。

4. “生日悖論”

n 個人中至少有兩人生日相同的概率是多少？這是有名的“生日問題”。令人難以置信的是：隨機選取的 23 人中至少兩人生日相同的概率居然超過 50%，50 人中至少兩人生日相同的概率居然達到 97%！例如，假定一個中學有二十個班，每個班平均有 50 個學生，你可以調查一下，大概會有十幾個班都有至少兩個生日相同的學生。這和人們的直覺是抵觸的。因此這一結果被稱為“生日悖論”。

其實有關概率的計算很簡單，首先計算 50 個人生日都不相同的概率。第一個人的生日有 365 個可能性，第二個人如果生日與第一個人不同，他的生日有 364 個可能性，依次類推，直到第 50 個人的生日有 316 個可能性，所以 50 人生日都不同的可能組合方式就是 365 乘 364 乘 363 一直乘到 316，但由於每個人的生日是獨立的，總的可能組合是 365 的 50 次方，這樣一來，50 個人生日都不相同的概率就等於兩個組合數之比，這個概率非常小，只有 3%，至少兩個人生日相同的概率等於 1 減去 3%，得到 97%，這樣概率就計算出來了。

注意：如果預先選定一個生日，隨機選取 125 人、250 人、500 人、1000 人，出現某人生日正好是選定生日的概率分別大約只有 30%、50%、75%、94%，比想像的小得多。

5. “三枚銀幣”騙局

某人在街頭設一賭局。他向觀眾出示了放在帽子裡的三枚銀幣（記為甲、乙、丙），銀幣甲的兩面塗了黑色，銀幣丙的兩面塗了紅色，銀幣乙一面塗了黑色，另一面塗了紅色。

遊戲規則是：他讓一個觀眾從帽子裡任意取出一枚銀幣放到桌面上（這裡不用“投擲銀幣”是為避免暴露銀幣兩面的顏色），然後由設局人猜銀幣另一面的顏色，如果猜中了，該參

與者付給他1元錢, 如果猜錯了, 他付給該參與者1元錢。

試問：這一賭局是公平的嗎？從直覺上看，無論取出的銀幣所展示的一面是黑色或紅色，另一面是紅色或黑色的概率都是 $1/2$ ，這一賭局似乎是公平的。但實際上不公平，設局者只要每次“猜”背面和正面是同一顏色，他的勝算概率是 $2/3$ ，因為從這三張牌隨機選取一枚銀幣，其兩面塗相同顏色的概率就是 $2/3$ 。如果有許多人參與賭局，大概有 $1/3$ 的人會贏錢， $2/3$ 的人會輸錢。

下面進一步用“場景分析”來戳穿“三枚銀幣”騙局。假定參與者取出並放到桌面上的銀幣展示面是黑色，則這枚銀幣只可能是銀幣甲或乙。“銀幣展示面是黑色”這一隨機事件有三種可能場景：銀幣甲的“某一面”和“另一面”，或銀幣乙的“塗黑一面”。因此，這枚銀幣是銀幣甲的概率是 $2/3$ 。展示面是紅色情形完全類似。因此，每次“猜”另一面和展示面是同一顏色的勝算概率是 $2/3$ 。

下面這個例子是從“三枚銀幣”騙局衍生出來的。假設在你面前放置三個盒子，盒子裡分別放了金幣兩枚、銀幣兩枚、金銀幣各一枚。你隨機選取一個盒子並從中摸出一枚錢幣，發現是一枚金幣。試問：該盒子裝有兩枚金幣的概率有多大？請你給出答案。

6. 在猜獎遊戲中改猜是否增大中獎概率？

這一問題出自美國的一個電視遊戲節目，問題的名字來自該節目的主持人蒙提·霍爾，上世紀90年代曾在美國引起廣泛和熱烈的討論。假定在臺上有三扇關閉的門，其中一扇門後面有一輛汽車，另外兩扇門後面各有一隻山羊。主持人是知道哪扇門後面有汽車的。當競猜者選定了一扇門但尚未開啟它的時候，節目主持人去開啟剩下兩扇門中的一扇，露出的是山羊。主持人會問參賽者要不要改猜另一扇未開啟的門。問題是：改猜另一扇未開啟的門是否比不改猜贏得汽車的概率要大？正確的答案是：改猜能增大贏得汽車的概率，從原來的 $1/3$ 增大為 $2/3$ 。這是因為競猜者選定的一扇門後面有汽車的概率是 $1/3$ ，在未選定的兩扇門後面有汽車的概率是 $2/3$ ，主持人開啟其中一扇門把這門後面有汽車給排除了，所以另一扇未開啟的門後面有汽車的概率是 $2/3$ 。

也許有人對此答案提出質疑，認為在剩下未開啟的兩扇門後有汽車的概率都是 $1/2$ ，因此不需要改猜。為消除這一質疑，不妨假定有10扇門的情形，其中一扇門後面有一輛汽車，另外9扇門後面各有一隻山羊。當競猜者猜了一扇門但尚未開啟時，主持人去開啟剩下9扇門中的8扇，露出的全是山羊。顯然：原先猜的那扇門後面有一輛汽車的概率只是 $1/10$ ，這時改猜另一扇未開啟的門贏得汽車的概率是 $9/10$ 。

7. 條件概率和全概率公式

在上面好幾個例子中, 都涉及“條件概率問題”。設 A, B 是兩個事件, 如果已知 A 和 B 各自發生的概率為 $P(A)$ 和 $P(B)$, 又知道 A 和 B 同時都發生的概率為 $P(AB)$, 則在事件 A 發生的條件下事件 B 發生的概率 (稱為事件 B 關於事件 A 的條件概率, 記為 $P(B | A)$) 顯然為 $P(B | A) = P(AB)/P(A)$ 。這裡所謂的“條件事件 A ”和“事件 B ”的發生, 在時間上沒有先後次序。在實際問題中, 通常“條件事件 A ”表示結果, 即知道事件 A 已經發生了, 而“事件 B ”則表示導致這一結果的可能原因。關於這一點, 千萬不要把“條件”二字誤導為“原因”了。

現在設 A_1, \dots, A_n 是 n 個事件, 假定其中之一會發生, 但其中任意兩個事件不會同時發生, 已知這些事件發生的概率分別為 $P(A_1), \dots, P(A_n)$ 。另外, 假定另有某事件 B , 已知條件概率 $P(B | A_1), \dots, P(B | A_n)$ 。試問: 事件 B 發生的概率 $P(B)$ 是多大? 顯然, $P(B)$ 可以通過以下的“全概率公式”給出:

$$P(B) = \sum_{j=1}^n P(B | A_j)P(A_j).$$

8. “競賽規則”藏玄機

假定有甲、乙兩個乒乓球運動員參加比賽, 已知甲的實力強於乙。現有兩個備選的競賽規則: “3局2勝制”, 或“5局3勝制”。試問: 哪一種競賽規則對甲有利?

在“3局2勝制”規則下, 只有“甲甲”、“甲乙甲”和“乙甲甲”這三種可能場景導致甲最終獲勝。因此, 設在單局中甲勝的概率為 p , 則甲最終獲勝的概率為這三種場景的概率之和, 等於 $f(p)[1 + 2(1 + p)]P^2$ 。同理, 在“5局3勝制”規則下, 進行三局甲獲勝只有“甲甲甲”這一場景; 進行四局甲獲勝有“甲乙甲甲”、“乙甲甲甲”、“甲甲乙甲”三種可能場景; 進行五局甲獲勝有六種可能場景 (具體描述留給讀者)。因此甲最終獲勝的概率為這十種場景的概率之和, 等於 $g(p) = [1 + 3(1 - p) + 6(1 - p)^2]p^3$ 。當 $p > 1/2$ 時, 容易證明 $g(p) > f(p)$ 。因此, “5局3勝制”規則對甲有利。

9. 為什麼在多人博弈中弱者有時反倒有利?

假定甲、乙、丙三個商家為搶佔市場而進行競爭。在競爭中, 甲淘汰乙和丙的概率分別是 0.6 和 0.8, 乙淘汰甲和丙的概率分別是 0.4 和 0.7, 丙淘汰甲和乙的概率分別為 0.2 和 0.3。競爭的結局可以是兩敗俱傷, 或者兩個依舊倖存, 也可能是一個倖存、另一個被淘汰。競爭的規則: 每個商家只能選中一名對手來競爭, 未被淘汰的進入下一輪競爭。問: 到第二輪結束時, 各個商家倖存下來的概率有多大?

首先，從非合作博弈角度分析，第一輪的最優策略是：甲和乙競爭，丙與乙結成“暫時聯盟”，共同對付三者中最強的甲。第一輪結束時，丙肯定倖存下來；甲和乙都被淘汰的概率是 0.312，它等於甲淘汰乙的概率 (0.6) 乘以甲被乙和 (或) 丙淘汰的概率 $(1 - 0.6 \times 0.8)$ ；甲和乙都倖存的概率是 0.192，它等於甲未被乙和丙淘汰的概率 (0.6×0.8) 乘以乙未被甲淘汰的概率 (0.4)；甲倖存、乙被淘汰的概率是 0.288，它等於甲未被乙和丙淘汰的概率 (0.6×0.8) 乘以乙被甲淘汰的概率 (0.6)；乙倖存、甲被淘汰的概率是 0.208，它等於乙未被甲淘汰的概率 (0.4) 乘以甲被乙和 (或) 丙淘汰的概率 $(1 - 0.6 \times 0.8)$ 。如果甲和乙都被淘汰，丙單獨倖存，競爭結束。否則進入第二輪競爭，這時分兩種情形：(1) 如果第一輪結束時甲和乙都倖存，則與第一輪情形相同；(2) 如果第一輪結束時甲 (或乙) 倖存，這時甲 (或乙) 和丙競爭，甲 (或乙) 繼續倖存的概率是 0.8 (或 0.7)，丙繼續倖存的概率是 0.2 (或 0.3)。利用全概率公式計算，第二輪結束時，甲最終倖存下來的概率是 0.322，它等於 $0.192 \times 0.192 + 0.192 \times 0.288 + 0.288 \times 0.8$ ；類似可以推出乙和丙最終倖存下來的概率分別是 0.222 和 0.624。由此可見，丙倖存下來的概率最大。

當然，這一模型是理想化的數學模型，但它給我們很好的啟示。弱者在競爭的夾縫中倖存下來的例子在商界層出不窮。

10. 計算條件概率的貝葉斯公式

設 A, B 是兩個事件，如果已知 A 和 B 各自發生的概率為 $P(A)$ 和 $P(B)$ ，又知道事件關於事件的條件概率 $P(B | A)$ ，如何求事件 A 關於事件 B 的條件概率？由於 A 和 B 同時發生的概率為 $P(AB) = P(B | A)P(A)$ ，所以有

$$P(A | B) = P(AB)/P(B) = P(B | A)P(A)/P(B).$$

這就是 18 世紀中葉英國學者貝葉斯 (Bayes) 提出的“由結果推測原因”的概率公式，即著名的“貝葉斯公式”。

下面考慮多個場景情形。設 A_1, \dots, A_n 是 n 個事件，假定其中之一會發生，但其中任意兩個事件不會同時發生，已知這些事件發生的概率分別為 $P(A_1), \dots, P(A_n)$ 。另外，假定另有某事件 B ，已知條件概率 $P(B | A_1), \dots, P(B | A_n)$ 。這裡每個事件 A_j 通常代表導致事件 B 發生的可能場景。試問：在已知事件 B 發生的條件下，某個事件 A_j 發生的概率是多大？由貝葉斯公式，我們有

$$P(A_j | B) = P(B | A_j)P(A_j)/P(B),$$

其中 $P(B)$ 由全概率公式給出：

$$P(B) = \sum_{j=1}^n P(B | A_j)P(A_j).$$

合併這兩個公式, 我們得到“貝葉斯公式”的最一般形式。

下面舉一個簡單的例子。假定有甲、乙兩個容器, 容器甲裡有 7 個紅球和 3 個白球, 容器乙裡有 1 個紅球和 9 個白球, 隨機從這兩個容器中抽出一個球, 發現是紅球, 問這個紅球是來自容器甲的概率是多大?

設“球是從容器甲抽出”為事件 A , “抽出的球是紅球”為事件 B , 則有:

$$P(A) = 1/2, \quad P(B | A) = 7/10,$$

$$P(B) = 1/2 \times 1/10 + 1/2 \times 7/10 = 2/5,$$

按照貝葉斯公式, 抽出的紅球來自容器甲的概率是

$$P(A | B) = (7/10) \times 1/2 / (2/5) = 7/8.$$

下面我們還將繼續給出應用貝葉斯公式計算概率的兩個典型例子。

11. 如何評估疾病診斷的確診率?

假想有一種通過檢驗胃液來診斷胃癌的方法, 胃癌患者檢驗結果為陽性的概率為 99.9%, 非胃癌患者檢驗結果為陽性 (“假陽性”) 的概率為 0.1%。問題是:

- (1) 檢驗結果為陽性者確實患胃癌的概率 (即確診率) 是多大?
- (2) 如果“假陽性”的概率降為 0.01%、0.001% 和 0, 確診率分別上升多少?
- (3) 用重複檢驗方法能提高確診率嗎?

我們用 “+” 表示“檢驗結果為陽性”, 用 H 表示被檢者為“胃癌患者”, 則由貝葉斯公式, 確診率為: $P(H | +) = P(+ | H)P(H)/P(+)$ 。從這一公式看出, 我們要預先知道被檢者所在地區胃癌患病率 $P(H)$ 。假定該地區胃癌患病率為 0.01%。問題 (1) 的答案是: 確診率為 1/11; 問題 (2) 的答案是: 如果“假陽性”的概率降為 0.01%、0.001% 和 0, 確診率分別上升為 50%、90.9% 和 100%; 問題 (3) 的答案是: 有一定的提高, 但大幅度提高的可能性很小。原因是“假陽性”主要是檢驗技術本身問題造成的, 重複檢驗的結果相關性很大, 不能按獨立事件對待。

12. 如何設計對敏感性問題的社會調查?

設想要對研究生論文抄襲現象進行社會調查。如果直接就此問題進行問卷調查, 就是說要你直說你是否抄襲, 即使這樣的調查是無記名的, 也會使被調查者感到尷尬。設計如下方案可使被調查者願意做出真實的回答: 在一個箱子裡放進 1 個紅球和 1 個白球。被調查者在摸到球後記住顏色並立刻將球放回, 然後根據球的顏色是紅和白分別回答如下問題: 你的生日是否在

7月1日以前？你做論文時是否有過抄襲行為？回答時只要在一張預備好的白紙上打 \checkmark 或打 \times ，分別表示是或否。假定被調查者有 150 人，統計出共有 60 個 \checkmark 。問題是：有抄襲行為的比率大概是多？已知： $P(\text{紅}) = 0.5$ ， $P(\checkmark|\text{紅}) = 0.5$ ， $P(\checkmark) = 0.4$ ，求條件概率 $P(\checkmark|\text{白})$ ，用貝葉斯公式算出的答案是 30%。

13. 如何理解社會和大自然中出現的奇蹟？

對單個彩民和單次抽獎來說，中樂透頭獎的概率大概是 2250 萬分之一。到 2008 年，在“紐約樂透”史上發生過 3 次有一人中過兩次頭獎的事件。例如，2007 年 8 月 30 日美國紐約的安傑洛夫婦喜中“紐約樂透”頭獎，獲得 500 萬美元獎金。他們 1996 年與另外 3 人共分了 1000 萬美元頭獎。這堪稱一個奇蹟。“紐約樂透”每週三和六晚間各開獎一次，每年開獎 104 次，40 年間經歷約 4100 次開獎。假定以前中過“紐約樂透”頭獎的人還經常買“紐約樂透”彩票，而且每次下的注數都比較大，那麼在 40 年間他們之中有三人兩次中頭獎的概率就不是非常小了。

在河北省著名旅遊景點野三坡的螞蟻嶺左側，斷崖邊緣有一塊直徑十米、高四米的“風動石”，此石著地面積不足覆蓋面積的 $1/20$ ，尤其基部接觸處只有兩個支點。這也算是一個奇蹟。

從概率論觀點看，上述兩個奇蹟的發生並不奇怪，因為即使是極小概率事件，如果重複很多次，會有很大概率發生。假設一事件發生概率為 p ，重複 n 次還不發生的概率為 $(1-p)^n$ ，當 n 足夠大，這一概率就很小，從而該事件發生的概率為 $1 - (1-p)^n$ 就變得很大了。大自然中的奇蹟是地殼在億萬年的變遷中偶然發生的，但這種奇蹟在歷史的長河中最終出現是一種必然現象。

14. 從概率學家眼光看“華南虎照事件”

2007 年 10 月 12 日，陝西省林業廳宣佈陝西發現華南虎，並公佈據稱為陝西安康市鎮坪縣城關鎮文采村村民周正龍 2007 年 10 月 3 日拍攝到的華南虎照片。但這一轟動性的消息隨即引來廣大網友質疑，指可能是紙老虎造假。11 月 16 日，一網友稱“華南虎”的原型實為自家牆上年畫。同時，義烏年畫廠證實確曾生產過老虎年畫。2007 年 12 月 3 日，來自六個方面的鑒定報告和專家意見匯總認為虎照為假。當我從網上看到虎照和年畫對照圖片後立刻做出“華南虎照”是假的判斷，理由是虎照和年畫相似率達到百分之九十九以上的概率幾乎是零。

2007 年 12 月以來我在多次科普報告中都公開了我的這一觀點。在網上可以搜索到我 2008 年 4 月在華中科技大學做報告的如下新聞報導：題目為：“四院士齊聚華科，同台共述概率之美”。報導中寫道：“嚴加安院士分別揭示了‘生日悖論’等日常生活中人們常見問題背後隱藏的概率論原理。他還以一個概率學家的身份判斷華南虎照片肯定是假的，理由是虎照和年畫相

似率達到百分之九十九以上的概率是零。”

2008年6月29日, 陝西省人民政府向新聞媒體宣佈: 2007年10月5日, 陝西省鎮坪縣農民周正龍拍攝的野生華南虎照片為造假。2008年11月17日8時30分, 陝西省安康市中級人民法院宣判周正龍有期徒刑兩年半緩期3年, 並處罰金2000元。

15. “辛普森悖論”

分組對比中占優勢總體上一定占優勢嗎? 答案是: 不一定! 下面是一個例子。假定有兩種藥 (A 和 B), 要通過分組臨床試驗對比其療效。以下是試驗結果的統計表: 從甲乙兩組試驗結果看, 藥物 A 的療效都優於藥物 B, 但總體來看, 藥物 B 的療效反而優於藥物 A。

組別	病人數 (用藥 A)	治癒數 (比例)	病人數 (用藥 B)	治癒數 (比例)
甲	50	20 (40%)	30	10 (33%)
乙	40	30 (75%)	70	50 (71%)
總計	90	50 (56%)	100	60 (60%)

早在20世紀初, 當人們為探究兩種因數是否具有某種相關性而進行分組研究時就發現了這種現象: 在分組比較中都佔優勢的一方, 在總評中反而居劣勢。直到1951年英國統計學家辛普森在他發表的論文中才正式對這一現象給與理論解釋。後人就把這一現象稱為“辛普森悖論”。

16. “統計平均”的陷阱 (例1)

下面這個例子在現實生活中更加典型, 它是“辛普森悖論”的一種表現形式。假定有一公司現有員工100人, 另有一研究所, 職工150人。在一次普查體檢中, 發現公司有糖尿病患者16人, 研究所有糖尿病患者36人。從糖尿病患者的患病率來看, 研究所的情況比公司嚴重, 其患病率分別是24%和16%。但實際情況恰恰相反, 這怎麼可能呢?

現在我們換一種統計方式來考察結果, 分成年輕人(24~45歲)和中、老年人(46~65歲)兩個組來計算患病率。該公司有90位年輕人, 其中患糖尿病12人(患病率13.3%), 有中、老年人10人, 其中患糖尿病4人(患病率40%); 該研究所有50位年輕人, 患糖尿病4人(患病率8%), 有中、老年人100人, 其中患糖尿病32人(患病率32%)。

後一種統計方式的結果表明, 公司的人, 無論是年輕人還是中、老年人, 患糖尿病的比例都顯著高於研究所的相應人群, 這可能和他們經常加班和中午吃盒飯有關。這一分組統計結果比總體統計結果更有說服力。

17. “統計平均”的陷阱 (例2)

下面的例子再次表明分組平均往往比總體平均更有說服力。假定某大學數學系有教授15人、副教授40人、講師和助教25人，這三類人的平均年收入分別是15萬、12萬、8萬，該單位職工平均年收入為10萬。又假定科學院某研究所有研究員60人、副研究員30人、助研30人，這三類人的平均年收入分別是14萬、11萬、7萬，但該研究所職工平均年收入為11.5萬，高出那個系職工平均年收入1.5萬。這一例子表明：由於各單位人員構成比例不同，單位職工平均年收入這一指標不能真實反映單位職工的收入狀況。

這一例子給了我們一啟示：有些新聞報導中的統計平均數字沒有實際意義。例如，2010年2月中國國家統計局公佈稱，2009年中國70個大中城市房價同比上漲1.5%，這與大城市居民的實際感受完全背離，被網友戲稱為“房價被拉低”。事實上，“70個大中城市房價的平均漲幅或跌幅”在統計學上沒有實際意義。接受這次教訓，國家統計局於2011年2月16日正式宣佈，今後將不再發佈全國70城市房價漲幅平均數，理由是“平均數在個體差異較大的情況下，往往會削峰填谷，抹平個體間的差異。”這是一個明智的決定。

18. 統計調查資料的誤讀

前幾年，曾經在報刊和網路上出現一條聳人聽聞的新聞：“中國國家有關部門公佈的一個專項調查結果表明，我國知識份子平均壽命為58歲，低於全國平均壽命10歲左右；北京中關村知識份子平均死亡年齡為53.34歲，比10年前縮短了5.18歲。”

這一消息可信嗎？大部分人肯定不相信。問題出在哪裡？可以猜測，這一資料是根據知識份子中在職期間死亡的資料統計出來的。中關村知識份子在職期間平均死亡年齡為53.34歲並不奇怪，因為男女平均退休的年齡是57.5歲。這條新聞錯誤在於把知識份子在職期間死亡的平均年齡誇大為知識份子平均壽命。其實絕大多數知識份子是在他們退休以後的若干年內才死亡的，我國知識份子平均壽命至少應該超過70歲。

19. “抽樣調查”的陷阱

在做抽樣調查時，如果資料的採集缺乏代表性，可能導致錯誤的結論，下面是一個著名的例子。在1936年美國大選中，羅斯福總統以62.5%的得票率獲勝連任，擊敗了共和黨候選人蘭登。在選舉前，1935年才由美國統計學家蓋洛普創立的美國民意研究所，只用了5萬多個調查問卷，便成功預測了羅斯福會贏得大選（儘管後來實際得票率比預測高了約7%）。與此成鮮明對照的是，老牌的著名雜誌《文學文摘》依據高達約240萬份的問卷調查結果，卻預測蘭登將以57%對43%的絕對優勢大勝羅斯福。選舉後不久，《文學文摘》由於這一重大醜聞就倒閉了。

《文學文摘》的預測為什麼會失敗？問題就出在抽樣調查樣本的代表性有嚴重偏差。首先，

該雜誌寄出了大約1千萬份問卷, 選擇的對象主要來自雜誌的訂戶和一些俱樂部的會員, 這些人大都相對比較富裕。當時美國剛從經濟大蕭條中恢復, 富人比較傾向支持蘭登, 而窮人較多傾向支持羅斯福。另外, 問卷的回收率太低, 只有 24%, 這進一步降低了樣本的代表性, 因為收入較低者回答問卷的比例通常要比收入較高者低。

該例子說明, 在做統計調查時, 要精心設計好方案。例如, 採用分層抽樣, 並隨機選擇調查物件, 這樣才能使抽樣調查的樣本具有代表性。

20. 分組混合血標本篩查檢驗

某醫院對一群人進行體檢, 其中有一項是愛滋病血清檢驗。如果對每個人的血液標本單獨檢驗, 成本將很高。採用“分組混合血標本篩查檢驗”可以節省成本和時間。假定採集到 N 個血標本, 把每個血標本平分成兩份, 一份留做備用。另一份平均分成 M 組, 將每組的血標本混合在一起進行檢驗。如果某組血液檢測為陽性, 再用該組的每個備用血標本進行逐個篩查。

問題是: 如何根據數量 N 和患愛滋病的概率 p 來確定分組數 M , 或每組的血標本個數 k , 其中 $k = N/M$ (假定 k 為整數), 使得平均檢驗次數達到最低? 令 $q = 1 - p$, 一組血標本檢驗是陰性的概率為 q^k , 一組血標本檢驗是陽性的概率為 $1 - q^k$ 。如果某組血液檢測為陽性, 則該組共計需要進行 $k + 1$ 次檢驗。因此平均檢驗總次數為 $N/k[q^k + (k + 1)(1 - q^k)]$ 。由此通過電腦計算可以確定最佳的 k 。

21. 概率分析在不確定投資決策中並非萬能

假定有兩個投資項目。第一個項目分兩階段進行, 第一階段以 0.90 概率進入下一階段, 以 0.10 概率出局。進入第二階段後, 以 0.90 的概率取得成功, 獲得 400 萬元的利潤; 第二個項目以 0.80 的概率直接盈利 400 萬元。儘管從概率論分析來看, 第一個項目以 0.81 的概率獲得 400 萬元的利潤, 優於第二個項目, 但多數人選擇第二個項目, 原因是兩次面臨風險比一次面臨風險給人造成的心理壓力更大。

再舉一個例子。假定投資者有兩個投資項目可供選擇: 項目 A 確保盈利 400 萬元; 項目 B 以 70% 的概率盈利 500 萬元, 以 30% 的概率盈利 200 萬元。儘管項目 B 的平均盈利 410 萬元, 高於項目 A, 但大多數人為了規避風險, 寧願選擇項目 A 而不選擇 B。另外, 假定投資者有一項目分兩階段進行, 在第一階段確保盈利 300 萬元, 第二階段有兩個項目 C 和 D 可供他選擇: 項目 C 確保再盈利 100 萬元; 項目 D 以 70% 的概率盈利 200 萬元, 以 30% 的概率虧損 100 萬元。這時, 大多數人可能更傾向於選擇 D。儘管從概率分析來看, 項目 A 或 B 分別等同於盈利 300 萬元前提下的項目 C 或 D, 但在兩種不同的投資境況下, 投資者做出了不同的決策, 原因是對兩種境況下的風險認知不同。

22. 抽樣調查的結論依賴於樣本量的大小

現在有一種說法：抽煙者患老年癡呆症的比率較低。爲了檢驗這一說法是否可信，設想某醫療機構在某個城市從 65~75 歲的人群中隨機調查了 1000 人，分別統計抽煙者和非抽煙者老年癡呆症患病人數。調查結果是：1000 人中有 250 人是抽煙者，其中老年癡呆症患者 10 人；750 人非吸煙者，其中有 45 人患老年癡呆症。這兩類人患老年癡呆症的比率分別是 4% 和 6%。表面上看，差異比較顯著，但能否根據這一差異就來斷定吸煙有助於預防老年癡呆症呢？我們可以用統計中的“假設檢驗”來回答這一問題。從統計學知道，如果“吸煙不降低老年癡呆症患病率”這一假設成立，如下定義的統計量近似服從均值爲 0 方差爲 1 的正態分佈：

$$\xi = \frac{p_2 - p_1}{\sqrt{(1/n_1 + 1/n_2)(p(1-p))}},$$

其中 n_1 和 n_2 分別是吸煙者和非吸煙者的人數， p_1 和 p_2 分別爲吸煙者和非吸煙者患老年癡呆症比率， p 爲這兩類人總體患老年癡呆症比率。根據抽樣調查結果算得 $\xi = 1.2$ 。從標準正態分佈表查出， $\xi \geq 1.2$ 的概率超過 15%，因此這一結果還不足以否定“吸煙不降低老年癡呆症患病率”這一假設。但是，如果抽樣調查的人數擴大到 4000 人，假定兩類人也相應地擴大 4 倍，而且患病率仍然分別是 4% 和 6%，這時算得 $\xi = 2.4$ 。從標準常態分布表查出的概率小於 1%。這時我們可以有 99% 的把握斷定吸煙能夠降低老年癡呆症患病率了。這一例子告訴我們，抽樣調查的結論不僅要看統計資料，還要看抽樣調查的樣本量的大小。

—本文作者任教中科院數學與系統科學研究院應用數學研究所—

2014 Taipei workshop on Analysis and Geometry in Several Complex Variables

日期：2014 年 12 月 15 日 (星期一) ~ 2014 年 12 月 19 日 (星期五)

地點：台北市大安區羅斯福路四段 1 號 天文數學館

詳見中研院數學所網頁 <http://www.math.sinica.edu.tw>