決策與誤差

黃文璋
國立高雄大學應用數學系

1 前言

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眞的要投誠, 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? 這位被起訴者是否眞有罪? 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? 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。眞的是這樣嗎? 本文將從統計的角度, 對決策的形成, 及誤差的掌握, 做初步的介紹。並藉此釐清有關隨機的一些基本概念。

2 決策與誤差簡介

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。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, 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, 是先拿利刀在龜甲上刻劃, 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, 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 ", 能依據神農伏義演成的八卦, 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, 為了方便也常用兩個銅板取代, 近人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。愼重時還要得到三次一陰一陽才行。以銅板為例, 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。這中間每一環節, 皆可能產生誤差。失之毫釐, 差之千里, 對同一現象, 得到不同的決策並不足為奇。

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, 舉行"三二六民主和平護台灣大遊行"。遊行落幕後, 台北市長馬英九表示三二六遊行人數達27萬5千人。民進黨則認為遊行人數超過百萬, 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?

雖然做決策誤差難以避免, 但如何以較好的設計, 來減小誤差, 就要仰賴能善用統計了。

3 資料品質

資料品質不佳(包含所取得之原始資料不佳, 及處理過程不佳), 及未能正確掌握機率的內涵, 是造成誤差的兩大主因。後者我們下一節再討論。對於隨機現象, 不論多麼謹慎, 皆難免有因運氣而得到品質不佳資料的時候。我們只能以較好的方法, 使有較大的機會得到品質較佳的資料。

收集資料是統計分析的一重要步驟, 如果收集到的是偏差或未具代表性的樣本, 據此所做的推論, 誤差大便不足為奇。假設要估計袋子中白球所佔全部球的比例。一個簡單的方法是取出若干個(每次取出後不放回), 數看看其中白球所佔比例, 以此比例做為全袋中白球所佔比例之估計值。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。處理球還算容易, 如果是處理動物或人, 就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 。此過程看起來只是前述估計白球比例問題的一應用, 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, 更何況如果是處理有關人的問題 。我們來看底下取材自Freedman et al.(1991)著名的選舉實例 。

例1. 西元1936年, 美國羅斯福(Franklin Delano Roosevelt, 1882-1945)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, 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 。

文學文摘的錯誤是令人驚訝的 。要知那幾乎可說是有史以來最大一次的民調, 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户 、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(平均每四户才有一具) 。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。1千萬份問卷中, 只有比例不高(24%)的240萬份回覆, 這240萬份回覆者的意見, 不見得能代表1千萬位被挑選出之選民。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2. 某製藥公司宣稱發明一種對延長壽命有效的藥, 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能延壽? 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, 造成壽命有延長的傾向。

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。可參考羅夢娜(1987)一文。

收集到的資料若未妥善的處理, 有時會得到很荒謬的結論。底下為一例。

例3. 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(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。

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。這種例子很多。如曾有人指出, 當可樂銷售量較大時, 到醫院腸胃科就診人數常亦增加。喝可樂會引起腸胃的毛病? 這可是相當重要的發現。其實很可能是夏天天氣炎熱, 可樂需求量

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%	

因而增大。而天氣炎熱時病蟲較易滋長, 因此腸胃有問題的人也變多。至於喝可樂與腸胃的毛病, 倒不見得有因果關係。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, 較願意接受對身體有利的建議, 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。大家看過金庸的倚天屠龍記嗎?書中說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, 得到不滿65歲的婦女中, 抽煙者之死亡率較高的推論。

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

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%	

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(如量取全世界每個人的身高), 否則依據取樣所做的推論, 與眞實的值有差異, 並不足為奇, 此點我們下一節會討論 。

4 一些機率的概念

最佳決策為何, 依可供選擇的範圍及評比準則而定 。就像選美, 有依服裝 、泳裝 、才藝 、機智問答等, 服裝尚可分自選服裝 、晚宴禮服等來評審, 各項百分比也可不同 。因應網路時代, 有時尚有一項網路人氣票選 。而所選出的第一名, 也只是參賽者中的第一 。因此一般而言, 依評比準則的不同, 及可選擇範圍的不同, 最佳決策往往也就不同 。而且最佳決策可能不存在, 或不唯一 。

依統計理論找出的某一最佳決策，當然不會永遠比別的決策好。此有如選美比賽裡的第一名, 不見得就是才藝最好者 。為了對決策與誤差有較正確的了解, 本節我們介紹一些機率的基本概念 。

4.1 機率的意義

假設有兩個銅板, A銅板出現正面機率為0.6, B銅板為0.5 。投擲10次, 得到較多正面者獲勝。那應該挑選那一個呢? 對此簡單的問題, 直觀上最佳決策自然是選A銅板 。但我們不乏下述經驗: 投擲之下, A銅板得到4正面, B銅板得到6正面, B銅板贏了 。這種不按牌裡出牌的例子很多 。樂透彩每期開出6個頭獎號碼及1特別號 。小樂透(42取6)的39號, 曾連續5期開

出(含特別號, 平均6期開出一次), 大樂透(49取6)的29號, 曾連續56期連特別號都未開出(平均7期開出一次), 創下所謂29魔咒。由於樂透彩的發行, 引發大眾對機率的興趣, 每次開獎, 都是一次檢驗機率值的機會。由於號碼常不按彩迷們預期的出現(於連續56期未出現後, 讓彩迷覺得29號"很冷", 下一期29號卻連莊), 常引發彩迷們懷疑開獎號碼是否眞的隨機產生。欲釐清此點, 底下我們先略微解釋機率的意義。

日常生活裡, 大家常提到機率, 到底機率的意義是什麼? 大學生常覺得微積分很難學, 但許多大學教師認為, 機率論其實是較微積分更難講清楚的。微積分有物理的意義, 諸如極限、連續、導數及積分, 都可藉物理上的現象來描述。反觀機率論中, 從機率一詞, 至期望值、變異數, 都不易藉物理來說明。機率有幾種不同的解釋, 我們先以**頻率來解釋**, 針對的是可以重複做實驗的事件, 如投擲銅板, 樂透彩開獎等。重複做一實驗n次, 假設其中事件A出現n_A次, 則以出現次數之相對頻率n_A/n的極限$\lim_{n\to\infty} n_A/n$表A之機率$P(A)$。由於$0 \le n_A \le n$, 故如此定義的機率, 其值介於0與1之間。這其中的一個困難是, 在有限的時間內, 並無法做無限多次實驗, 所以只能以有限次的實驗結果來估計A之機率$P(A)$。另一困難是, 在微積分裡大家學過, 一數列$\{a_n, n \ge 1\}$, 若$\lim_{n\to\infty} a_n = a$, 則$n$很大時, a_n與a差異的絕對值$|a_n - a|$要很小。但對隨機數列$\{n_A/n, n \ge 1\}$, 不論n多大, n_A/n仍有可能不接近$P(A)$。例如, 假設有一公正的銅板, 公正的意思是正、反面出現機率各為$1/2$。以A表正面出現的事件, 則$P(A) = 1/2$。投擲1,000(n)次, 會得到0(n_A)次正面(此時$n_A/n = 0$, 且$|n_A/n - P(A)| = 1/2$)之機率為$(1/2)^{1,000}$, 雖然很小, 但仍是正的。亦即不論n多大, $|n_A/n - P(A)|$都有可能不會很小。

但機率理論告訴我們, n很大時, $|n_A/n - P(A)|$會很小的機率很大。機率很大表可任意接近1。即$\forall \varepsilon > 0, \delta > 0$, 存在一$n_0 \ge 1$, 使得$n \ge n_0$時,

$$P(|\frac{n_A}{n} - P(A)| < \varepsilon) > 1 - \delta \text{。} \tag{1}$$

也可寫成$\forall \varepsilon > 0$,

$$\lim_{n\to\infty} P(|\frac{n_A}{n} - P(A)| < \varepsilon) = 1 \text{。} \tag{2}$$

支持(1)式(或(2)式), 就是**弱大數法則**(weak law of large numbers), 滿足(1)式(或(2)式), 即表$n \to \infty$時, n_A/n會**機率收斂**(converages in probability)至$P(A)$ 。

<u>**註1.**</u> **弱大數法則**. 設$X_1, \cdots, X_n$為一數列之**獨立且有共同分佈**(簡寫為i.i.d.)之隨機變數。令$\overline{X}_n = \sum_{i=1}^n X_i/n$, $n \geq 1$, 表**樣本平均**, 又令$\mu = E(X)$, 設$|\mu| < \infty$。則 $\overline{X}_n$會**機率收斂**至μ。即對$\forall \varepsilon > 0$,

$$\lim_{n\to\infty} P(|\overline{X}_n - \mu| < \varepsilon) = 1 \text{ 。} \tag{3}$$

即樣本平均與期望值之差會很小的機率, 隨著樣本數n之增大, 趨近至1 。

有弱大數法則當然也有**強大數法則**(strong law of large numbers), 此處不擬介紹, 可翻閱一般機率論的書 。

以這樣的方式來定義似乎有點在繞圈子。要解釋機率的意義, 卻要用到機率:

一事件A之機率為$P(A)$, 表實驗次數n很大時, A出現之
相對頻率n_A/n很接近$P(A)$的機率很大 。

看來機率的確是不太好理解 。

對於本子節一開始的A銅板 、B銅板問題, 如果投擲次數不多, B銅板尚有些贏的機會(若是只投擲1次, A銅板贏的機會為0.36, 平手的機會為0.48, B銅板贏的機會為0.16)。投擲30次, B銅板贏的機率約只有0.057; 投擲100次, B銅板贏的機率約只有0.0019。如果想贏, 要選那一銅板應是再清楚不過了。所以機率值, 是在實驗次數很多的情況下, 方能顯示其意義。如果實驗次數不多, 當然無法印證其機率值。話又說回來, 即使只做1次實驗, 譬如說投擲一出現正面機率為p之銅板, 正面出現便給你a元; 反面出現算你輸, 要收你的錢。你認為該付多少才公平合理? 即使只玩1次, 你還是會盤算一下: 輸的機率是$1-p$, 假設輸要付x元, 則一邊是pa, 一邊是$(1-p)x$, 二者應該相等, 對雙方才公平。解出

$$x = \frac{p}{1-p} a(\text{元}) \text{ 。}$$

如果你覺得該付的錢應少於上式中的x, 顯然你認為銅板出現正面之機率小於p; 如果你覺得該付的錢應大於上式中的x, 顯然你認為銅板出現正面的機率大於p; 只有當你覺得該付的錢等於上式中的x, 你才眞覺得銅板出現正面之機率等於p。以這種方式來解釋機率, 就是所謂**主觀的解釋**。以頻率來解釋機率, 由於是基於實驗的結果, 與觀察者是誰無關, 因此又稱為**客觀的解釋**。當然主觀的解釋, 有時也會根據過去客觀的事實來決定。只是即使擁有相同的資料, 不同的二人, 對同一事件, 也可能給出不同的主觀機率。

我們對機率的解釋, 常是頻率、主觀兩種意義交錯著進行。在買彩券前, 可能主觀上覺得最近運氣不錯, 這次應有很大中獎機會。如果只買一次, 不是中就是不中; 但如果買很多次(或很多張), 就還是要回到用頻率來解釋機率。氣象局預測今日降雨機率為0.9, 由於"今日"是不重覆的, 雖然機率0.9夠大, 即使沒下雨, 大家也只好作罷。還是要看長期下來, 實際下雨的日數, 與宣佈的降雨機率, 是否符合其比例, 來檢驗氣象局所宣佈降雨機率的準確性。一旦氣象局建立其公信力, 民眾就不會因一次的該帶傘沒帶, 或不需帶傘而帶, 而責怪氣象局。

在機率裡, 不是在爭一時, 而是在爭千秋。一件事終久會呈現其該有的面貌。或者說只要時間夠久, 呈現其該有的面貌之機率很大。如果只買一張彩券, 隨意挑選可能也無妨, 反正中獎機率實在太小。如果要買很多張, 投資很大, 如何選號, 以提高期望所得, 就該謹愼行之。可參考黃文璋(2003a)第十九章。

4.2 條件機率

先看一個例子。

例4. 某次選美, 最後要自A, B, C三人中挑一位給第一名。三人條件相當, 設被挑中第一名的機會一樣皆為$1/3$。在公布結果前, A悄悄地問某位一向誠實的評審: 她是不是第一名? 評審限於保密的規定, 不願告訴她。於是A改問評審: B、C二人中誰不會第一名? 該評審心想, B、C中必有一

人不是第一名, 這是大家都知道的, 告訴A也無妨, 於是告訴A“C不會第一名”。這時A覺得她會是第一名的機會增加為$1/2$。試問對不對? 另外, 若A改為問評審: C是第一名嗎? 且評審也告訴她答案, 這時 A會是第一名的機會是否會改變?

解. 我們先說明第一個情況。必須要了解的是, 若B、C皆非第一名, 評審其實是自B、C中任挑一人告訴A她不是第一名; 而若B是第一名, 評審便只能告訴 A“C 不是第一名”。由**貝氏定理**(Bayes' rule), 首先

$$\begin{aligned}
&P(\text{告訴}A\text{“}C\text{非第一”})\\
&= P(\text{告訴}A\text{“}C\text{非第一”}|A\text{為第一})P(A\text{為第一})\\
&\quad +P(\text{告訴}A\text{“}C\text{非第一”}|B\text{為第一})P(B\text{為第一})\\
&\quad +P(\text{告訴}A\text{“}C\text{非第一”}|C\text{為第一})P(C\text{為第一})\\
&= \frac{1}{2}\cdot\frac{1}{3}+1\cdot\frac{1}{3}+0\cdot\frac{1}{3}=\frac{1}{2}\text{。}
\end{aligned}$$

此處因評審為誠實地回答, 故C為第一時, 他會告訴A“C不是第一名”的機率為0。又

$$\begin{aligned}
&P(A\text{為第一}, \text{告訴}A\text{“}C\text{非第一”})\\
&= P(\text{告訴}A\text{“}C\text{非第一”}|A\text{為第一})P(A\text{為第一})\\
&= \frac{1}{2}\cdot\frac{1}{3}=\frac{1}{6}\text{。}
\end{aligned}$$

故

$$\begin{aligned}
&P(A\text{為第一}|\text{告訴}A\text{“}C\text{非第一”})\\
&= \frac{P(A\text{為第一}, \text{告訴}A\text{“}C\text{非第一})}{P(\text{告訴}A\text{“}C\text{非第一”})}\\
&= \frac{1/6}{1/2}=\frac{1}{3}\text{。}
\end{aligned}$$

即得在評審告訴A“C不是第一名”後, A會是第一名的機率仍為$1/3$。同理, 若評審告訴A“B不是第一名”, A會是第一名的機率也還是$1/3$。

次看第二個情況, 此時A會是第一名的機率改變了。如果評審說“C是第一名”, 則A是第一名的機率為0; 而因C非第一的機率為$2/3$(此亦為評

審告訴A“C非第一名”的機率), 且A是第一名時, 評審必告訴A“C非第一名”。故得

$$P(A\text{為第一}|\text{告訴}A\text{“}C\text{非第一”}) = \frac{1/3}{2/3} = \frac{1}{2}\text{。}$$

不同的問法, 即使得到相同的回覆(“C 非第一 ”), A為第一的機率並不相同 。

大家看第一種問法, 對A並無幫助, 第二種問法則有用 。善用資訊是很重要的, 適當地發問, 將可得到較有用的資訊 。

機率值會變, 是機率裡的一大特性 。數學裡某數$a=3$, 就永遠是3 。機率裡就不是這樣, 給定某一條件後, 可能改變機率值, 此即條件機率 。在上例中, 對第一個情況, 評審所說的“C非第一 ”, 此資訊是否完全無用? 也不盡然, 如果B偷聽到A與評審的對話, 則她知自己會是第一的機率, 由1/3 增為 2/3 。

民國94年3月27日, 聯合報E4版有一則標題為“檢驗數據的迷思”一文 。其中提到由於幾乎沒有任何的檢驗項目及檢驗方法, 其結果準確率能達到百分之百, 因此檢驗結果不論是喜還是憂, 都可能有誤差存在 。我們看底下的例子 。

例5. 衛生局至高雄大學免費檢驗某疾病, 此檢驗之可靠度為90% 。即某人若真有病, 則有0.9之機率此檢驗呈正的反應(假設檢驗只有正負二反應); 若沒有病, 則有0.9的機率檢驗呈負的反應 。過去資料顯示平均每5,000人中, 有一人患有此病 。此檢驗迅速且無害, 但若檢驗呈正的反應, 則被強制至醫院住院一週, 以做進一步的檢查 。試問你是否願意接受此檢驗?

解. 我們以“正”表檢驗呈正的反應, 則

$$\begin{aligned} P(\text{有病}|\text{正}) &= \frac{P(\text{正}|\text{有病})P(\text{有病})}{P(\text{正}|\text{有病})P(\text{有病})+P(\text{正}|\text{無病})P(\text{無病})} \\ &= \frac{0.9\cdot(1/5,000)}{0.9\cdot(1/5,000)+0.1\cdot(4,999/5,000)} \\ &= \frac{9}{5,008} \doteq 0.001797\text{。} \end{aligned}$$

由上述推導知, 若檢驗呈正的反應, 而眞有病的機率其實非常小, 僅約0.001797。看到此結果後, 大部分的人可能不願接受檢驗了。在計算出此機率之前, 也許我們會以為此機率值應仍接近0.9, 但結果卻是如此小。此因樣本裡有病的很少(只有1/5,000的機會), 而一旦檢驗呈正的反應, 會有病的機率已是在檢驗前會有病的機率之大約9倍, 只是仍然很小。直觀上來看, 因檢驗有10%的錯誤會呈正反應, 因此在每5,000人中, 約有500人呈正反應, 但平均卻只有一人眞有病, 比例約為1/500=0.002, 與前述眞正的機率0.001797接近。底下我們逐漸把可靠度提高, 看此時P(有病|正)之變化。

檢驗可靠度	90%	95%	97%	98%	99%	99.9%	99.99%
P(有病|正)	0.001797	0.003786	0.006426	0.009707	0.019419	0.166556	0.666689

可以看出, 當可靠度提高到99%, P(有病|正)之值才約0.019419; 可靠度提高到99.9%, P(有病|正)之值才達到0.166556。即使可靠度高達99.99%, P(有病|正)也仍只有0.666689即約三分之二。所以, 對於那些發生率極低的罕見疾病, 許多人往往質疑其檢驗的有效性。

若不了解條件機率, 可能會常對機率錯誤的解讀。想進一步了解, 可參考黃文璋(2003a)第五章, 其中有更多例子。

4.3 隨機性

在機率裡, 一事先不能預知結果的試驗, 便稱**隨機試驗**。投擲一骰子, 不論是否為公正的骰子, 都可說是在進行一隨機試驗。即使投擲一兩個面均為正面的銅板, 也可視為是一(退化的)隨機試驗。諸如抽獎、樂透彩的開獎及電話訪問等, 都是生活裡常可見到的隨機試驗。其中**不重複的簡單隨機抽樣**(底下只稱**簡單隨機抽樣**), 常被採用以產生所要的號碼。簡單隨機抽樣中含有兩個概念: 獨立及均勻。以樂透彩開獎為例。獨立表每期開出的號碼與以前開出的不相干, 均勻表每組號碼開出的機率都一樣。

我們平常說號碼是隨機產生。隨機取樣, 往往就是指簡單隨機抽樣。所以這裡的隨機是有特定意義, 與定義較一般的隨機試驗不一樣。即

使對這特別(或說較簡單)的簡單隨機抽樣, 一般人還是不易弄明白。因而瘋狂的追求明牌, 懷疑開獎之公正性, 甚至沈迷於賭博, 就不足為奇了。

小樂透的總獎金佔總投注金額的56%, 大樂透則佔57%。大部分的賭戲對賭客並不公平, 那何以許多人一上了賭台, 就下不來? 原因有可能是如果情況不利, 就會想那有運氣那麼壞, 該轉運了, 不能就此打住。再玩若仍輸, 更加深下次該贏的想法; 若幸運贏了, 覺得眞開始翻身了。至於若情況有利, 手氣正順, 又怎可停止? 除非是一直有輸輸贏贏, 起起伏伏, 讓人覺得此賭戲沒趣, 否則不少人不論手氣好壞, 都缺乏當機立斷的決心。

以樂透彩為例, 雖然號碼隨機地產生, 但若檢視已開出的號碼將可發現不少"奇異"現象 (如前述大樂透29號連續56期未開出), 常讓人懷疑號碼是否眞隨機產生。其實只要有心, 總可以找到各期所開出的號碼之間, 各種看起來很特殊的現象, 而這往往是隨機下的正常現象。北銀42取6的小樂透, 自91年1月22日發行, 至94年1月21日結束, 共發行314期, 圖1給出頭獎號碼1至42各號碼出現的頻率, 出現最多的為12號, 有61次, 出現最少的為16號, 僅33次。平均應出現$e \doteq 44.857$(次)。因

$$Q = \frac{42-1}{42-6}\sum_{i=1}^{42}\frac{(o_i-e)^2}{e}$$

有近似的χ^2_{41}分佈, 其中o_i為i號出現的頻率, 算出

$$Q \doteq \frac{41\cdot 30.72}{36\cdot 44.857} \doteq 34.7615 < 56.93 \doteq \chi^2_{0.95,41}。$$

因此雖42個號碼出現頻率有多有少, 且出現最多次者為最少次者之$61/33 \doteq 1.848$倍, 在0.05的**第一型錯誤機率**(第5節會說明)下, 並無法拒絕頭獎號碼各數字出現機率相同之假設。也就是雖然各號碼出現頻率高低差不太小, 但此現象還不算太怪異。反倒是如果開了314期, 42個號碼累積之出現頻率都只有少少的1,2次之差, 才該懷疑其中有沒有搞鬼。要注意發生機率相同與出現頻率相同是不一樣的。隨機取樣下, 常會有看起來似乎不太均勻的現象產生, 這是正常的。如果出現頻率相同, 才很可能便非隨機現象。大部分的人對此點常不太了解。

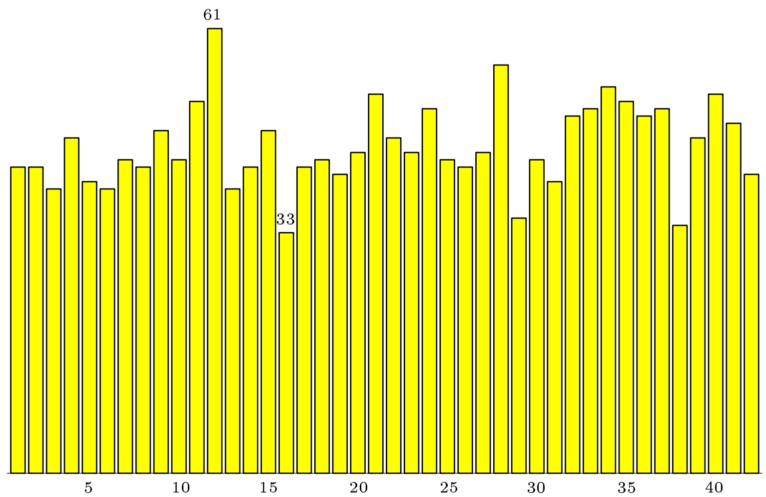

圖1. 91年1月22日至94年1月21日共314期北銀樂透彩頭獎號碼出現頻率

我們再藉投擲銅板來說明賭戲較少是輸贏交錯著進行。持續投擲一公正銅板10,000次, 每次出現正面則贏得1元, 出現反面則失去1元。令

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

表投擲n次後之淨所得, 其中X_i表第i次之所得(1或-1)。$\{S_n, 1 \le n \le 10,000\}$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$\{X_i, i \ge 1\}$為一數列之i.i.d.的隨機變數, 且

$$P(X_i = 1) = P(X_i = -1) = \frac{1}{2}。$$

以隨機過程裡**布朗運動**(Brownian motion)的結果, 來估計最大淨所得不超過a之機率:

$$P(\max_{1 \le n \le 10,000} S_n \le a) \doteq P(\max_{0 \le s \le 1} X(s) \le \frac{a}{100})$$

$$= P(|Z| \le \frac{a}{100}) \doteq \begin{cases} 0.0796, & a = 10, \\ 0.1586, & a = 20, \\ 0.3830, & a = 50, \\ 0.6826, & a = 100, \end{cases}$$

其中Z有**標準常態分佈**(standard normal distribution, 以$\mathcal{N}(0,1)$表之), $\{X(t), t \geq 0\}$表一**標準的布朗運動**。可看出最大淨所得不太大(如$a \leq 20$)的機率並不大。另外亦有 (仍以布朗運動的結果來估計):

$$P(\frac{\text{正面領先次數}}{\text{投擲次數}} \geq x) \doteq 1 - \frac{2}{\pi}\arcsin\sqrt{x}\ 。$$

當$x = 0.993$, 此機率約為

$$1 - \frac{2}{\pi}\arcsin\sqrt{0.993} \doteq 0.0533\ 。$$

即得投擲銅板10,000次,

$$P(\text{有一面領先次數} \geq 9,930) \doteq 0.0533 \times 2 = 0.1066,$$

約每10回就會觀測到1回有一面領先數至少有9,930次, 10回中有1回, 這可不算稀少。但當此事件發生時, 有些賭客可能會覺得幸運之神正照顧自己, 相對地也會有一些賭客可能會認為其中有詐。事實上 此為一隨機下的正常現象。

圖2給出六個模擬圖, 橫軸為n, 縱軸為S_n, 其中$S_i > 0$表至第i次投擲正面領先, $S_i < 0$則表反面領先。注意$E(S_n) = 0, \forall n \geq 1$。

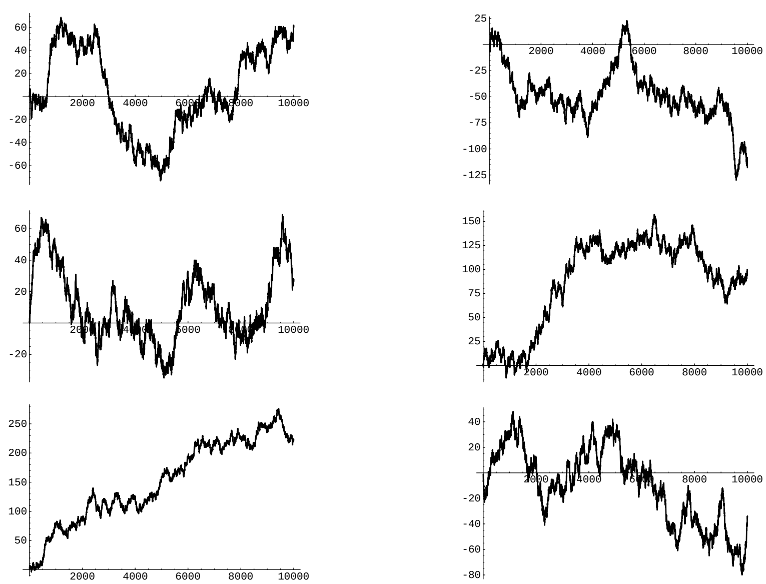

圖2. 6個$\{S_n, 1 \leq n \leq 10,000\}$的模擬圖

新聞媒體多半只報導有人樂透彩中大獎, 或在賭場大贏的新聞 。而人有選擇性記憶的傾向, 在賭之前向神明祈求, 或依據過去開出號碼來推測本期頭獎號碼, 大部分的時候當然沒有效果, 但若贏(中)了, 可能眞覺得神明聽了自己的祈求, 或眞有明牌 。

我們常在講隨機性, 但其實一般人並不眞正了解隨機性的意義 。由所獲得之數據, 往往也不易正確判斷這些數據是否為隨機產生, 甚至常得到相反的判斷 。可參考黃文璋(2003a)第四章及第八章 。

5 決策

對於隨機現象, 所做的決策, 常要用到機率統計的方法 。給一銅板, 估計其出現正面的機率 。直觀上, 可以投擲n次後, 出現正面的相對頻率來估計 。此問題也可先化為統計的語言 。假設有一數列之i.i.d.的隨機變數$\{X_i, i \geq 1\}$, 又設$P(X_i = 1) = \theta = 1 - P(X_i = 0)$ 。則樣本平均$\overline{X}_n$為一直觀的θ之估計量 。統計的理論指出此估計量有很多優點 。但你也可以認為此銅板一定是公正, 投擲所得正面數與反面數不等, 只是隨機下的必然結果, 如我們上節所述 。也許即使投擲100次, 得到高達90個的正面, 仍堅持認為$\theta = 0.5$, 是有矇對的時候 。但機率畢竟太小, 以0.9來估計θ還是一較明智的選擇 。

用$\overline{X}_n$來估計銅板出現正面的機率θ, 弱大數法則為其中的理論依據 。因

$$E(X) = 1 \cdot P(X = 1) + 0 \cdot P(X = 0) = \theta,$$

而$n \to \infty$時, $\overline{X}_n$會機率收斂至$E(X) = \theta$ 。所以我們知道, 只要樣本數夠多, $\overline{X}_n$多少會接近θ 。這種因期望值($E(X)$)等於參數(θ), 便以樣本平均($\overline{X}_n$)來估計參數, 便發展出**動差估計法**(method of moments estimator), 為一常見的估計法 。所依據的為弱大數法則, 或說基於相對頻率的想法 。支持這種想法的, 被稱為古典學派, 有時也被稱為**頻率學派**(frequentist) 。在十九世紀, 這派想法可說是主導統計的應用 。

再看一例子。

例6. 設一盒子中有不少白球及黑球, 假設只知數目比為3比1, 但不知是白球多還是黑球多。依序隨機地抽取3球, 每次取出後放回。令X表所得之黑球數。則由假設知X有**二項分佈**(binomial distribution), 參數為3, p(以$\mathcal{B}(3,p)$表之), 其中$p=1/4$或$3/4$。即X之機率密度函數為

$$f(x|p)=P(X=x)=\binom{3}{x}p^x(1-p)^{3-x}, x=0,1,2,3,$$

且$p=1/4$或$3/4$。

我們想根據觀測到之黑球數x來估計p。此估計並不算困難, 因只有二選擇: $p=1/4$或$3/4$。抽取後, x有四種可能, 其機率值為

x	0	1	2	3
$f(x\|\frac{1}{4})$	$\frac{27}{64}$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{1}{64}$
$f(x\|\frac{3}{4})$	$\frac{1}{64}$	$\frac{9}{64}$	$\frac{27}{64}$	$\frac{27}{64}$

如果$x=0$, 就宜選$p=1/4$, 因為27/64的機率值較1/64為大。也就是$p=1/4$比$p=3/4$更會使$x=0$發生。所以$x=0$出現時, 認為$p=1/4$才較合理。如果$x=0$出現, 還硬要選$p=3/4$, 除了頑固(死忠?), 大約很難有什麼較好的解釋。$x=1$時, 也宜選$p=1/4$。至於若出現$x=2$或3, 則$p=3/4$當然是較合理的選擇。故p之估計值$\hat{p}$為

$$\hat{p}=\begin{cases}1/4 & ,\text{若}x=0,1,\\ 3/4 & ,\text{若}x=2,3\text{。}\end{cases}$$

上述估計法便是根據所觀測到的樣本x, 看p為何值時, 會使x較易出現。在只能對θ以一值來估計下, 此法當然有其道理。

現將機率密度函數$f(x|\theta)$視為一未知參數θ之函數, 假設$\theta\in\Omega$, 而將x固定, 且以$L(\theta|x)$表此函數, 並稱此為**概似函數**(likelihood function)。**最大概似法**(method of maximum likelihood), 就是找參數θ之估計值$\hat{\theta}(x)$,

使得在$\theta = \hat{\theta}(x)$之下, 會最可能(most likely)產生數據x。本來要先給出θ, X的分佈才完全決定。現由於觀測值x的出現, 我們倒回去想, 怎樣的θ, 才會使此x在諸多可能的x中拔得頭籌? 即當$X = x$, 尋找$\hat{\theta}(x)$, 使其滿足

$$
\begin{aligned}
L(\hat{\theta}(x)|x) &= f(x|\hat{\theta}(x)) = \max\{f(x,\theta), \theta \in \Omega\} \\
&= \max\{L(\theta, x), \theta \in \Omega\}。
\end{aligned}
$$

若如此的$\hat{\theta}(x)$存在, 便稱為θ之**最大概似估計值**。至於$\hat{\theta}(X)$則為**最大概似估計量**。

最大概似法為一極廣泛被採用的估計法。在日常生活裡我們也常以此方法來做決策。例如, 教室的玻璃窗破了, 老師第一個猜是小明打破, 為什麼? 因他平常最喜歡亂丟東西, 班上同學裡就屬他最愛搗蛋, 因此最有可能是他。警方辦案, 從有前科者開始調查, 也是同樣的道理。

在倚天屠龍記裡, 張無忌於練成九陽眞經的神功後, 輕易地鑽過一個洞穴, 朱九齡也想鑽進去, 卻被四周岩石卡住, 動彈不得。書上說他心中卻兀自在想“這小子比我高大, 他既能過去, 我也必能過去。為什麼我會擠在這裡? 當眞豈有此理!”可是世上確有不少豈有此理之事, 這個文才武功俱臻上乘、聰明機智算得是第一流人物的高手, 從此便嵌在這窄窄的山洞之中, 進也進不得, 退也退不出。

也許天下豈有此理之事就是不少, 有時發生的是機率較小的事件(高大的人反而過得去), 但若沒有其他資訊(如知道張無忌已練成神功), 以最大概似法來做決策, 不獨在統計裡, 也是生活中一普遍的方法。

動差估計法及最大概似法, 為兩種主要的點估計方法, 各基於不同的想法, 都有很多優點。當然也還有很多其他的估計法。例如, 另一常見的作法, 就是**貝氏作法**(Bayesian approach), 持此主張者, 也稱**貝氏學派**(Bayesian)。此學派認為對所欲估計的參數, 事先都該有些看法。此有如雖然每個人的情況不同, 但我們常先入為主的認為台大的畢業生如何, 清華的畢業生又如何。在此作法裡, 參數θ不再視為定值, 而是假設有某一機率分佈, 即所謂**事前分佈**(prior distribution)。此分佈可視為**主觀的分佈**(subjective distribution), 乃是基於做實驗者個人主觀的看法。當然主

觀的看法, 也可以依據過去的資料, 而做出客觀的推斷。由於此分佈是在觀測到資料之前就已建立的, 因此才稱為事前分佈。取樣後, 依據觀測值, 修正對θ之分佈的看法, 便得到θ之**事後分佈**(posterior distribution)。“貝氏”此一名稱的由來, 就是因這一修正後的分佈, 是利用**貝氏定理**而得到。

總之, 每種估計法都是基於某種原理, 這些原理往往與我們的思維方式相對應。換句話說, 我們潛意識裡的某種判斷方式, 用統計的語言來描述, 常就是統計中之一種重要的決策方法。

估計參數θ時, 有時亦會考慮**損失函數**(loss function)。再依不同的損失函數, 及不同的原理來做決策。例如, 有所謂**極大極小原理**(minimax principle)。這是一種保守的決策方式, 即依各種策略裡, 針對不同的θ, 所產生之最大損失中之最小者, 挑選為決策。細節在此不多討論。極大極小原理, 當然也是我們的一種思維方式。

以一值來估計一未知的參數θ, 是所謂**點估計**。有時會想以一區間來估計θ, 稱為**信賴區間**(confidence interval), 並伴隨一機率值p。在得到觀測值前, 預知此隨機的區間有p的機率包含θ, 其中p為一較接近1的值, 如0.90, 0.95, 0.99等, 希望此區間有一不小的機率會包含θ。取樣後得到一固定的區間, 便只有包含或不包含θ, 此時便不再說此區間有多少的機率包含θ。

所欲估計的未知量, 並非都是如銅板出現正面機率一般單純, 有較複雜的情況。例如, 想估計這批觀測值, 究竟是源自何種分佈的母體。估計論可說是統計裡的一大學問, 有很多值得探討的題材。

覆水難收, 做決策有時是不能回頭的, 我們給一例如下。

例7. 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? 你應知道, 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。就是一開始面試的幾位都放棄, 而自約0.368比例起的面試者, 若有比前面都更佳的人, 便挑選他。如此會挑中最佳者之機率約為$1/e(\doteq 0.368)$。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, 挑中最佳者之機率也才約0.368。換句話說, 約有0.632的機率, 你得不到最想要的。

假設檢定(testing hypothesis), 這是隨機世界的很多情況中, 做決策之一重要的依據。本節最後我們略做討論。

有位女士宣稱她能分辨奶茶是把茶加進牛奶裡, 還是把牛奶加進茶裡。若只拿一杯奶茶讓那位女士喝, 她說對先放茶或先放牛奶, 會相信她眞有分辨能力嗎? 可能不會吧! 兩次皆說對呢? 大約也還不會吧! 那連續20次皆說對呢? 可能會有點信了。但20次中錯一次呢? 這可還是相當不容易呢! 20次中錯兩次呢? 在此分辨能力, 不妨寬容一些, 指她每杯講對的機率大於隨機猜的機率 1/2。我們對犯錯是有一些容忍度, 但程度究竟多大, 就因人、因情況而異。而且同樣多的試驗成功次數, 也不一定導致相同的結論。例如, 假設有個音樂家宣稱只要任意彈奏一頁樂譜, 他能分辨是海頓(Haydon)或莫札特(Mozart)的曲子。如果10次試驗都講對, 我們大約便信了。但若有人宣稱他能預知投擲銅板是正面或反面朝上, 而且10次試驗都講對, 我們可能還是不信他有此能力。

我們自國中起, 習慣了數學中的證明。在數學裡, 一命題一旦被證明是對的, 就毫無疑問地成立, 如**費馬最後定理**(Fermat's last theorem)。在數學上我們可以寫

假設

$$n \geq 3\text{為一整數, 且}x, y, z\text{皆不為 }0(\text{假設}A),$$

試證

$$x^n + y^n = z^n\text{無整數解}(\text{結論}B)。$$

但在隨機現象裡, 一件事往往自始至終不知其究竟是眞還是僞。到底該女

士能否分辨奶茶是先放奶還是先放茶, 即使她20次皆說對(亂猜猜中之機率$= 1/2^{20} \doteq$百萬分之一), 恐怕還是有人不信她有此能力 。此因

平均一百萬人約有一人20次皆猜對 。

如果對台灣2,300萬人每人測試, 約有23人可全答對 。因此我們不會說:

試證某女士"有"分辨奶茶是先放奶還是先放茶的能力,

或是：

試證某女士"無"分辨$\cdots$ 。

數學家因相信在條件A下, $x^n + y^n = z^n$無整數解是對的, 於是去證明 。對奶茶問題, 我們相信什麼呢? 由於該女士宣稱(也希望人家相信)她有分辨能力 。因此先假設該女士無分辨能力 。然後拿20杯讓她分辨, 觀察她講對幾次 。先設定一能忍受的推論錯誤機率α, 然後求在無分辨能力的假設下, 講對次數會這麼多的之機率有多大? 如果機率小於α(即這麼多次講對是較不尋常), 則拒絕原假設, 否則接受原假設 。

對一隨機現象, 先提出猜想, 然後將猜想表為統計假設, 最後決定接受或拒絕統計假設 。此過程稱為假設檢定, 所得到的推論, 稱為**統計推論**(statistical inference) 。統計假設與一般數學中的假設不同 。在數學裡"假設$x > y$", 並未涉及任何隨機的量, 此非統計假設 。令θ表北銀小樂透彩, 1號出現在頭獎號碼中的機率 。則$\theta > 1/7$, 便為一統計假設 。數學中的假設不須去檢驗是否為真 。至於對一統計中的假設, 並不是要"證明"是否為眞, 而是要判定該"接受"或"拒絕"此一假設 。先取一組隨機樣本, 並利用此組樣本, 當做是否接受某一假設之證據 。如果證據與假設所陳述的不合, 或者說吻合的機率很低, 便拒絕該假設, 否則便接受該假設 。數據會說話, 但不論方法多好, 對一統計假設所做的推論, 是可能有錯的 。數學證明不能有錯, 統計推論則允許犯錯! 換一組樣本, 結論可能便相反 。在無法避免犯錯下, 只能以較好的方法減小犯錯的機率 。

波蘭人奈曼(J. Neyman)及英國人**皮爾生**(E.S. Pearson), 於1933年, 給出**奈曼-皮爾生引理**(Neyman-Pearson lemma) 。在其架構中,

虛無假設(H_0):通常表現況, 而我們對其正確性已產生懷疑 。

對立假設(H_a):表我們傾向相信, 或希望它是對的 。

對於北銀樂透彩頭獎號碼中, 每期1號出現的機率θ是否大於1/7的假設, 若傾向相信答案是肯定的, 則取

$$H_0\colon\ \theta = 1/7,\quad H_a\colon\ \theta > 1/7\ 。$$

在奶茶問題, 令θ表每杯奶茶該女士能正確分辨先放奶或先放茶之機率 。則取

$$H_0\colon\ \theta = 1/2,\quad H_a\colon\ \theta > 1/2\ 。$$

虛無假設是被保護的, 除非證據夠強, 否則不輕易推翻 。想想若宣佈樂透彩1號較易出現, 或該女士有分辨能力, 將造成多大的新聞轟動 。對於現況不輕易推翻, 會使人們在做決策時更謹慎 。要知朝令夕改, 絕非統計學裡假設檢定的精神!

假設檢定裡, 多大的錯誤機率可忍受? 這其中有兩種錯誤的機率: 虛無假設為眞卻拒絕(**第一型錯誤**), 虛無假設不眞卻接受(**第二型錯誤**) 。以圖3來表示 。第一型錯誤及第二型錯誤, 又分別稱**第一型誤差**及**第二型誤差** 。

	虛無假設為眞	虛無假設不眞
接受虛無假設	正確	第二型錯誤
拒絕虛無假設	第一型錯誤	正確

圖3. 假設檢定裡會犯的錯誤

一般而言, 當樣本數固定, 兩型錯誤的機率, 有一減小另一必增大 。因第一型錯誤較嚴重, 所以通常的作法是, 先控制第一型錯誤的機率不要超過某一事先給定的α值, 然後設法使第二型錯誤的機率愈小愈好 。依此原則給出一**拒絕域** 。即若觀測值落在此區域便拒絕H_0, 否則便說"不能拒絕 "H_0 。"不能拒絕 "有時簡化一點, 直接說"接受" 。假設檢定裡的"接受 "H_0, 是"雖不滿意, 但可接受"的意義 。"不能拒絕 "是一種較保守的講法 。94年4月4日台灣某政黨主席, 率該黨一些成員赴日參拜靖國神社, 引起國內媒體討論有關靖國神社參拜的問題 。到底適不適合去參拜呢? 報上刊出1980年11月17日, 日本鈴木內閣曾就首相暨內閣閣員參拜靖國神社, 發表日本政府對此問題之統一見解, 認為"尚不能否認正式參拜有違反

憲法嫌疑”。又是尚不能否認, 又是嫌疑, 也是對“參拜靖國神社乃違反憲法”的一很保守的講法。

在樂透彩號碼的隨機性檢定裡, 先假設各號碼出現的機率相同(H_0), 再看於此假設下, 會出現這麼異常的結果之機率是否夠小, 以判定該不該推翻機率相同的假設。通過檢定並不表號碼眞的是隨機產生, 只是説尚無不合; 但若不通過, 大約便不相信號碼為隨機產生。

常取的α值為0.10, 0.05, 0.01等。其他α或再小的α當然也可以, 但α太小並不適合, 何以故? α愈小, 表示H_0愈不容易被拒絕。虛無假設若被過度保護, 很難拒絕它, 也會造成偏差。不過如果是較重大的事件, 如牽涉到人命關天, α當然要取得小些。

統計假設的架構, 與刑事訴訟法中的“**無罪推定原則**”(第154條)類似:

被告未經審判證明有罪確定前, 推定其為無罪。

即先相信虛無假設, 然後看實驗出現的數據合不合理。以機率大小來判定合理性。即若會出現這種數據的機率很小, 便認為不合理, 於是拒絕虛無假設。對於隨機現象, 統計裡常無法如數學裡, “證明 ”何者為眞。醫學上、法庭上, 所宣稱的“證實”那一看法, 往往與假設假定裡“接受”那一假設的意義類似。就算用了多強的語氣, 如“百分之兩百正確”, 仍可能有誤差存在。舉二例如下。

例8. 假設投擲一銅板100次, 皆得到正面。在合理的α下(只要$\alpha > 1/2^{100}$), 會拒絕 H_0: 此為公正銅板。但公正銅板還是有可能出現如此極端之結果(只是機率很小為$1/2^{100}$)。

例9. 某大公司有一萬名員工, 年終摸彩, 總經理最貼近的秘書中了首獎, 引發員工們認為有作弊的猜測。現作一檢定:

H_0:沒有作弊, H_a:有作弊。

在H_0之下, 隨機地抽獎, 該秘書中首獎之機率為0.0001。只要$\alpha > 0.0001$(這是通常的情況), 會拒絕H_0。但此推論合理嗎? 依此程序, 將會得到全公司每一員工都不該中獎, 否則便有作弊之嫌。

此例再度顯示, 假設檢定所得的推論, 並非數學上的證明。但若次年, 該秘書再度得首獎, 員工們必然“更加懷疑”其中有弊, 此時總經理大約很難給個“合理”的解釋。

如前所述, 雖然機率很小的事件不表示絕無可能發生, 但在很多情況下, 還是會依此來判定某些事的對與錯。見下例。

例10. 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。曾有一案例, 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百分之99.9993, 及百分之99.988, 均“不能排除”兩人有親子關係。由於誤差均小於0.001, 法官遂判定兩人有親子關係。

雖然有誤差存在, 但兩家醫院的鑑定結果很一致, 法官做此判定, 當事人即使不服, 考慮再上訴要翻案並不易, 只好就接受了。

最後來看“虛無”二字的由來? 對於影視界, 讀者有興趣的是某對明星夫婦婚姻出現裂痕, 並非一切傳聞都是子虛烏有。同樣地, 在樂透彩裡, 若做出的推論是北銀樂透彩每期頭獎號碼的出現頻率符合隨機性, 此推論誰有興趣? 接受虛無假設, 表實驗失敗。此假設彷彿是空的, 不具價值的。大家有興趣的是拒絕虛無假設! 宣佈頭獎號碼是有公式可以算出, 或宣佈明牌存在, 或宣佈有某幾個號碼較容易出現, 這種推論才有新聞價值。

想進一步了解假設檢定者, 可參考黃文璋(2003b)第八章。

6 誤差

除非是退化的隨機現象, 如骰子6個面均為1(此時投擲所得點數必為1, 期望值為1, 變異數為0), 否則不論樣本數n多大, 都無法保證樣本平均與期望值的差一定很小。由第4節六個模擬投擲銅板累積正面數的圖便可得知。但弱大數法則告訴我們, n很大時差異會很小的機率很大。一般而言,

決策的誤差不見得都好求出來。底下給一簡單的情況。

誤差理論是高斯對機率論的主要貢獻。統計裡常要做估計, 估計會有誤差, 一項決策少有萬無一失的, 需掌握誤差的大小。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。

中央極限定理(central limit theorem)導致當樣本數n很大時, 樣本平均$\overline{X}_n$滿足

$$
\begin{aligned}
&P(\overline{X}_n \in [\mu - \sigma/\sqrt{n},\ \mu + \sigma/\sqrt{n}]) \doteq 0.6827,\\
&P(\overline{X}_n \in [\mu - 2\sigma/\sqrt{n},\ \mu + 2\sigma/\sqrt{n}]) \doteq 0.9545, \qquad (4)\\
&P(\overline{X}_n \in [\mu - 3\sigma/\sqrt{n},\ \mu + 3\sigma/\sqrt{n}]) \doteq 0.9973,
\end{aligned}
$$

其中μ為樣本之期望值, σ為標準差。上式又等價於

$$
\begin{aligned}
&P(\sum_{i=1}^n X_i \in [n\mu - \sigma\sqrt{n},\ n\mu + \sigma\sqrt{n}]) \doteq 0.6827,\\
&P(\sum_{i=1}^n X_i \in [n\mu - 2\sigma\sqrt{n},\ n\mu + 2\sigma\sqrt{n}]) \doteq 0.9545, \qquad (5)\\
&P(\sum_{i=1}^n X_i \in [n\mu - 3\sigma\sqrt{n},\ n\mu + 3\sigma\sqrt{n}]) \doteq 0.9973,
\end{aligned}
$$

由(4)式, 當樣本數n愈大, 對相同的機率, $\overline{X}_n$會落在一愈窄的區間。由(5)式, 對相同的機率, 當n愈大, 樣本和$S_n = \sum_{i=1}^n X_i$會落在一愈寬的區間。是樣本平均$\overline{X}_n$才會接近期望值$E(X) = \mu$, 樣本和S_n倒不會接近$E(S_n) = n\mu$。而且隨著n的增大, S_n與$n\mu$的差異有變大的趨勢。又當σ已知, 由(4)式即可分別給給出μ的68.27%, 95.45%, 99.73%信賴區間:

$$
\begin{aligned}
&P(\mu \in [\overline{X}_n - \sigma/\sqrt{n},\ \overline{X}_n + \sigma/\sqrt{n}]) \doteq 0.6827,\\
&P(\mu \in [\overline{X}_n - 2\sigma/\sqrt{n},\ \overline{X}_n + 2\sigma/\sqrt{n}]) \doteq 0.9545, \qquad (6)\\
&P(\mu \in [\overline{X}_n - 3\sigma/\sqrt{n},\ \overline{X}_n + 3\sigma/\sqrt{n}]) \doteq 0.9973,
\end{aligned}
$$

<u>註2</u>. 中央極限定理. 設$X_1,\cdots,X_n$為i.i.d.之隨機變數, 又設$\mu = E(X)$, $\sigma = \sqrt{\mathrm{Var}(X)} > 0$皆存在。令$\overline{X}_n = \sum_{i=1}^n X_i/n$, $n \geq 1$。則對$\forall z \in R$,

$$P(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq z) \doteq \Phi(z), \tag{7}$$

其中

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \text{ 。} \tag{8}$$

誤差到底算大還是小, 乃與樣本數n有關。見下例。

<u>例11</u>. 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$X_1,\cdots,$ X_n為i.i.d.之隨機變數,且設$P(X=1) = p = 1 - P(X=0)$。則$E(X) = p$, $\mathrm{Var}(X) = p(1-p)$。欲檢定

$$H_0\colon p = 0.5, \quad H_a\colon p > 0.5 \text{ 。}$$

取拒絕域為$\{\overline{X}_n > c\}$, 其中c為一常數。由中央極限定理,

$$\frac{\sqrt{n}(\overline{X}_n - p)}{\sqrt{p(1-p)}} \approx \mathcal{N}(0,1) \text{ 。}$$

依題意觀測到$\overline{X}_n = 0.50114$。

(1) 設$n = 13,000,000$。則

$$\begin{aligned}
&P(\overline{X}_n \geq 0.50114 | H_0\text{為真}) \\
&\doteq P(Z \geq \frac{\sqrt{13,000,000}(0.50114 - 0.5)}{\sqrt{0.5 \cdot 0.5}}) \\
&\doteq P(Z \geq 3605.55 \cdot 0.00228) \\
&\doteq P(Z \geq 8.22) \\
&\doteq 1.03 \cdot 10^{-16} \text{ 。}
\end{aligned}$$

上述值稱為***p*-值**(p-value)。p-值表在H_0為真時, 會得到比觀測值更極端的機率。p-值愈小, 愈該拒絕H_0。因此處所得之p-值微乎其微, 故拒絕H_0, 即認為銅板並非公正。

(2) 設$n = 1,000,000$ 。

$$\begin{aligned} p\text{ -值} &= P(Z \geq \frac{\sqrt{1,000,000}(0.50114-0.5)}{\sqrt{0.5\cdot 0.5}}) \\ &= P(Z \geq 1,000\cdot 0.00228) \\ &= P(Z \geq 2.28) \\ &\doteq 0.0113 \text{ 。} \end{aligned}$$

若$\alpha > 0.0113$, 則拒絕H_0, 否則接受H_0 。

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

$$\begin{aligned} p\text{ -值} &= P(Z \geq \frac{\sqrt{10,000}(0.50114-0.5)}{\sqrt{0.5\cdot 0.5}}) \\ &= P(Z \geq 100\cdot 0.00228) \\ &= P(Z \geq 0.228) \\ &\doteq 0.40978 \text{ 。} \end{aligned}$$

由於p-值夠大, 因此對大部分的情況下(只要$\alpha < 0.40978$), 皆會接受H_0 。

舉上三情況為例, 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 。

註3. $x \to \infty$時,

$$1-\Phi(x) \sim \frac{\phi(x)}{x}, \tag{9}$$

其中

$$\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2},\ x \in R \text{ 。} \tag{10}$$

且“$\sim$”表近似相等, 即二者之比值當$x \to \infty$時, 趨近至1 。

例12. (取材自Harris(1988)). 某次選舉, 有A, B二候選人, 選民共投了N張有效票 。選舉結果A以些微票數落敗 。由於採人工計票, A認為有可能會

誤計, 要求重新計票。假設選舉沒有弊端, 且每一張票會被誤計之機率為p。又假設計票過程形成一i.i.d.的試驗, 而重新計票可更正所有錯誤。我們想了解重新計票後, 會翻盤之機率。

解. 設n_1, n_2為A, B二人實際該得之票數, $n_1+n_2=N$。令

$$X_1 \text{表應屬} A \text{的票數被計為} B \text{之票數},$$
$$X_2 \text{表應屬} B \text{的票數被計為} A \text{之票數}。$$

假設$X_1 \sim \mathcal{B}(n_1, p)$, $X_2 \sim \mathcal{B}(n_2, p)$, 即$X_1, X_2$分別有參數$n_1, p$及$n_2, p$之二項分佈。

在重新計票前, A所得之票數為

$$V_1 = n_1 + X_2 - X_1,$$

B所得之票數為

$$V_2 = n_2 + X_1 - X_2,$$

$V_1 + V_2 = N$。重新計票後, 更正所有錯誤, A得n_1票, B得n_2票。令D表重新計票前, B領先的票數。則

$$D = V_2 - V_1 = n_2 - n_1 - 2(X_2 - X_1)。\tag{11}$$

又由中央極限定理(只要N夠大)

$$\frac{D - E(D)}{\sqrt{\mathrm{Var}(D)}} \approx \mathcal{N}(0,1)。$$

其中(注意$n_2 = N - n_1$)

$$E(D) = n_2 - n_1 - 2(n_2 - n_1)p = (N - 2n_1)(1-2p),\tag{12}$$

$$\mathrm{Var}(D) = 4(n_1 p(1-p) + n_2 p(1-p)) = 4Np(1-p)。\tag{13}$$

故

$$\begin{aligned} P(D \geq d) &= P\Big(\frac{D - E(D)}{\sqrt{\mathrm{Var}(D)}} \geq \frac{d - E(D)}{\sqrt{\mathrm{Var}(D)}}\Big) \\ &\doteq P\Big(Z \geq \frac{d - (N-2n_1)(1-2p)}{2\sqrt{Np(1-p)}}\Big) \\ &= 1 - \Phi\Big(\frac{d - (N-2n_1)(1-2p)}{2\sqrt{Np(1-p)}}\Big)。\end{aligned}\tag{14}$$

p應不太大，不妨設$p \leq 0.5$。則在實際A獲勝的假設下(即$n_1 > n_2$)，原先計票B領先超過d票之機率

(15) $$P(D \geq d) \leq 1 - \Phi(\frac{d}{2\sqrt{Np(1-p)}})。$$

此處d應為正，否則A也不會想重新計票。現設$N = 13,000,000$, $d = 30,000$。則翻盤(即重新計票後A獲勝)之機率小於

(16) $$1 - \Phi(\frac{30,000}{2\sqrt{13,000,000p(1-p)}}) \doteq 1 - \Phi(\frac{4.16}{\sqrt{p(1-p)}})。$$

對一些不同的p，表4給出上述機率。可看出翻盤之機率微乎其微。

表4 $d = 30,000$, $N = 13,000,000$, $1 - \Phi(z)$之值, $z = 4.16/\sqrt{p(1-p)}$

p	z	$1 - \Phi(z)$
0.01	41.8096	2.50×10^{-382}
0.025	26.6453	1.25×10^{-156}
0.05	19.0874	1.61×10^{-81}
0.10	13.8667	5.07×10^{-44}
0.20	10.4	1.25×10^{-25}
0.25	9.6071	3.77×10^{-22}
0.30	9.0779	5.60×10^{-20}
0.40	8.4916	1.03×10^{-17}
0.50	8.32	4.46×10^{-17}

其次對一給定的p，於觀測到$D = d$後，由(12)式，令$d = E(D)$，可得n_1之估計值$\widehat{n}_1$滿足

(17) $$d = (N - 2\widehat{n}_1)(1 - 2p),$$

解出

(18) $$\widehat{n}_1 = \frac{1}{2}(N - \frac{d}{1-2p})。$$

當$d = 30,000$, $N = 13,000,000$, 表5給出一些不同的p之下$\widehat{n}_1$之值。令$\widehat{d} = \widehat{n}_2 - \widehat{n}_1 = N - 2\widehat{n}_1$表重新計票後, A落後的票數。結果發現, 在我們的假設下, 重新計票, 原先落後者將落後更多!

表5 當$d = 30,000$, $N = 13,000,000$, 對不同的p, $\widehat{n}_1$, $\widehat{d}$之值

p	$\widehat{n}_1$	$\widehat{d}$
0.01	6,484,694	30,612
0.025	6,484,211	31,578
0.05	6,483,333	33,334
0.10	6,481,250	37,500
0.20	6,475,000	50,000
0.25	6,470,000	60,000
0.30	6,462,500	75,000
0.40	6,425,000	150,000
0.45	6,350,000	300,000

7 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。數學裡的每一數字力求精準, 誤差的存在是難以忍受的。但話說回來, 在這隨機的世界, 不依據統計給出的結果來做判斷, 顯然不是明智之舉。

習題

1. 某醉漢被送到警察局, 吵鬧間隨意講了5個號碼, 有個警察照他講的號碼(另外自選一碼)去買大樂透彩, 結果中了3個號碼及特別號, 得到陸獎獎金1000元。某報當做新聞報導。試對此事件加以評述。(**解.**隨意猜, 猜中之機率為5/6,486)

2. A, B二人各投擲一出現正面機率為0.6及0.4之銅板n次, 誰得到較多正面便獲勝。試估計當$n =$(i) 30, (ii) 100時, A獲勝之機率。(**解.**(i)

0.9430, (ii) 0.9981)

3. 試回答例11最後所提之問題。(**解.**仍與樣本數n有關)

參考文獻

1. 黃文璋(2003a). 隨機思考論。華泰文化事業股份有限公司, 台北。
2. 黃文璋(2003b). 數理統計。華泰文化事業股份有限公司, 台北。
3. 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。
4. Freedman, D., Pisani, R., Purves, R. and Adhikari, A.(1991). *Statistics*, 2nd ed. W.W. Norton & Company, New York.
5. Simpson, E.H.(1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society Series B*, 13, 238-241.