談信賴區間

黃文璋

國立高雄大學應用數學系

1. 前言

眾所周知，統計在現代生活中，所扮演的角色愈來愈重要。為了讓國民具備基本的統計知識，中小學的數學課程裡，近年來陸續引進一些統計的題材。自民國 95 年起，高中數學課綱，還加進信賴區間的單元。惟信賴區間並非很淺顯的概念，欲進入此主題，需要一些背景知識，因此在一般大學的統計學教科書中，信賴區間多半出現在全書的後半部，於有了足夠的鋪陳後，才開始接觸。如今高中數學在不多的篇幅中，想將此題材介紹清楚，是有些難度的。屢引起師生的困擾，乃可預期。本文便擬對信賴區間略做討論，希望能讓初學者更了解信賴區間的內涵，並藉此釐清一些疑惑。

2. 信賴區間簡介

信賴區間究竟是什麼？信賴區間是做為估計用的。人們常接觸未知的量，如某銅板出現正面的機率、某廠牌飲料的容量、台灣黑熊的數量、明年經濟成長率，及民眾對某議題之支持率等。對一未知的量，有時會想估計其值。既然是估計，便可能準或不準。有時估計不準確，影響並不太大，但

有時影響就很深遠。像是大家都知道，過去十餘年，由於執行廣設大學的政策，新設立加上升格，國內大學已多達一百六十餘所，使得博士師資需求量大幅增大。於是各校博士班，如雨後春筍般的成立。待因擴張所需的教職逐漸填滿後，博士生的就業，便開始困難了，至今已累積不少流浪博士。還有博士生中途不唸了，賣雞排去。學非所用，引起不少關注。覓一安身立命的地方不易，導致博士班的就讀意願愈來愈低。於是近幾年，很多學校的博士班報名人數，快速下降。如此一來，又引起關注了。民國 102 年 11 月 27 日，國科會提出警告，國內博士人力過多，僅是目前的短期現象，十年後，將鬧博士人才荒。由於估計誤差太大，使得高等教育人才，有時過多，有時過少，供需常處在失衡狀態。培養一名博士不易，博士“產量”高度不穩定，對國家及社會都是損失。此例顯示，藉助統計，以尋找較佳的估計法，在很多情況下，都是刻不容緩的問題。

對一隨機現象，估計值要完全準確，常極不容易。已病入膏肓者的家屬，焦慮地想知道親人能再活多久？假設醫生斬釘截鐵地告知 4 個月，則可能提早死亡，或可能活更久，就是不易剛好是 4 個月。4 個月的估計，幾乎可說一定不準。為了不讓家屬覺得醫生太不可靠，有些醫生遂回答可再活 2 至 6 個月。不以一個值，而以一個區間來估計，雖看起來沒那麼權威了，但家屬反可能認為這種估計方式，比較合理，更值得“信賴”。部分家屬覺得有些模糊，追問死亡時間會落在此區間的可能性有多大？醫生說不定答 8 成，或 9 成等。即機率 0.8，或 0.9 等。那估計 2 至 6 個月，跟估計 1 至 7 個

月，有什麼差別？就是機率的不同，後者涵蓋存活時間的機率較大。

對一未知的量，以一區間來估計，並附上該未知量會落在此區間之機率，或者說區間會涵蓋該未知量之機率，便稱為區間估計。其中的區間，稱為信賴區間，或置信區間，伴隨的機率值，則稱為信心水準、信賴水準，或信賴度等。有幾個問題立即會浮現。例如，信賴區間如何產生？它唯一嗎？信心水準又是怎麼求出？

在估計一未知量時，若採區間估計，往往是先給定信心水準，由於此乃表機率，所以為一介於 0 與 1 之間的值。然後求出一區間，即信賴區間，使該區間涵蓋前述未知量之機率，如事先所設定的信心水準。信心水準有時以百分比表示，且往往較靠近 100%。常取的信心水準有 95%，及 90%等。對一固定的信心水準，一般而言，信賴區間並不唯一，通常採取長度較短的。估計區間的長度愈短，讓人感覺估計愈精準。對一可重覆觀測的隨機現象，底下說明信賴區間產生的程序。

以投擲銅板為例。假設想估計某一銅板出現正面的機率 p。直觀上就是先持續投擲銅板，此過程即收集資料。每次出現正面的機率都假設是 p，且各次投擲間，假設相互獨立。投擲 n 次後，以 S_n 表共得之正面數，則由排列組合知，S_n 有參數 n 及 p 之二項分佈，此分佈以 $B(n, p)$表之。則相對頻率 S_n/n，便常用來做為 p 之估計量。投擲前，S_n/n 為一隨機變數，投擲後，得到 S_n 的觀測值，也就有了估計值。設 $n = 10$，

$S_{10} = 6$，則 p 之估計值為 0.6。若另一天投擲數 n 仍為 10，而 $S_{10} = 5$，則 p 之估計值便為 0.5。因 S_n 是一隨機變數，即使用同一銅板，及採同樣的 n，每次所得的 S_n 不盡相同，估計值是會變的。

$S_n=k$ 之機率有如下表示

$$P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0,1,2,\cdots,n,$$

其中二項式係數

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \text{ 。}$$

給定一信心水準，如 95%，然後找一 $a>0$，a 表誤差，使得以 S_n/n 為中心點，半徑為 a 的區間，即$(S_n/n - a, S_n/n + a)$，涵蓋 p 的機率為 0.95。則區間$(S_n/n - a, S_n/n + a)$便是 p 之 95%(或 0.95)信心水準之信賴區間，或簡單地稱之為 95%信賴區間。若無法找到 a，使得前述涵蓋機率剛好是 0.95，則找出最小的 $a>0$，使得涵蓋機率至少是 0.95。

靈敏的讀者可能已看出，由於 p 未知，所以 a 根本無法找出；而且如果 n 大些，$n!$便不太好計算。利用著名的中央極限定理，當 n 夠大，將 S_n 標準化，即 S_n 減去期望值 np，再除以標準差 $\sqrt{np(1-p)}$，所得之$(S_n-np)/\sqrt{np(1-p)}$，將近似標準常態分佈，此分佈以 $N(0,1)$表之。再經一些不太複雜的計算，將得到一個 p 之近似的 95%信賴區間。怎可有近似？這

樣不是不夠準確嗎？這乃沒有辦法中的辦法。在統計估計的過程裡，常會採取近似，而只要樣本數充分大，誤差便能小至可以忍受。

給定一信心水準，且採同樣的估計方法，直觀上若樣本數愈大，則信賴區間的長度將愈短。付出更多代價(取樣較多)，總該有些收穫(估計較精準)。有時會先給定信心水準，如 0.95(即 95%)，及信賴區間的半徑 a，如 0.03(即 3%)，然後看樣本數 n 究竟該多大？

3. 實際應用時遭遇的困惑

媒體上常會公佈各種民調的結果。於分析討論後，底下為新聞報導裡一典型的結論：

> 這次調查於 12 月 16 至 19 日晚間進行，成功訪問了 1,023 位設籍台北市的成年民眾，另有 353 人拒訪；在 95%的信心水準下，抽樣誤差在正負 3.1%以內。調查以台北市住宅電話為母體作尾數兩位隨機抽樣，並依據台北市成年民眾之性別及年齡結構進行加權。

除給出成功訪問之樣本數、拒訪數、信心水準，及抽樣誤差等數據外，會說明樣本如何產生、抽樣的日期，以及有做加權等，即對調查的過程須交待清楚。高中數學引進信賴區間，主要是為了講抽樣調查，因這已成為生活統計了。而為了近似，又引進中央極限定理，這即使在大學的機率論裡，都不算是簡單的定理，如今卻堂而皇之地進入高中數學了。

數學中的定理，多少需要些條件。高中數學裡所給的中央極限定理，乃最基本的版本，即假設各樣本相互獨立，且有共同分佈。對於銅板，它可是毫無怨言，讓人想投擲幾次都可以，且可合理的假設各次投擲間相互獨立，每次有相同出現正面的機率。現欲估計某地區成年選民，對某特定候選人之支持率 p。乍看之下，此有如想估計銅板出現正面的機率，差別只是以隨機抽樣，收集支持與否的資料，來取代投擲銅板。只是民調與投擲銅板畢竟大不相同，要能成功訪問，得花點功夫，要有些技巧，否則易被拒訪，因此怎可有人二次被抽中？這樣必被抱怨不已。故民調裡的抽樣，一般是“取出後不放回”。但取出後不放回，將導致各樣本間不獨立，中央極限定理因而就不適用了。

取出後不放回，於取了 n 個樣本後，其中支持某特定候選人之人數，將有超幾何分佈，而非二項分佈。有些中學教師，試圖證明當母體總數 N(如前述報導中設籍台北市的成年民眾數)趨近至無限大時，其中超幾何分佈的機率值，會趨近至二項分佈的機率值。所謂“超幾何分佈的極限是二項分佈”，只是其證明大都是錯的，可參考黃文璋(2011a) 一文之說明。事實上，應說若抽取的樣本數 n，與母體總數 N 相比很小，則將取出後不放回，視之為取出後放回，然後將支持的樣本數 S_n 之分佈，捨超幾何分佈，而用較容易處理的二項分佈來取代，所造成的誤差便不至於太大。這裡面原本便用到近似了，即以常態分佈來近似標準化後的 S_n，現在不過是多了一個近似。

在民調裡，信賴區間的半徑 a，常稱為抽樣誤差。雖事先設定為 95%的信心水準，及 3%的抽樣誤差，換算出樣本數 n 該有 1,068 個。若採用電訪，多隻電話同時進行，待結束時，剔除不合格的樣本，則成功樣本數，很難恰好為 1,068 個。由獲得之成功樣本數，再換算出對應的抽樣誤差。若成功樣本數多於 1,068，則抽樣誤差將小於 3%；若成功樣本數少於 1,068，則抽樣誤差便大於 3%。那何以不增加樣本數，以減小抽樣誤差？要知抽樣誤差若降為原設定的 1/3，即 1%，則樣本數要增至 9 倍，即 9,612 份。在相同設備下，花的時間將成為 9 倍。成本大幅增加不說，民意如流水，一項調查若不盡快完成，則因種種事件的發生，將使得民眾對某議題的意見受到干擾。一方面是調查過程中，已用到數個近似，本來就不是那麼精準。另一方面是人未必說真話，且會改變想法，因此訪問對象是人，與投擲銅板，是截然不同的情況。故與其將精力花在增加樣本數，還不如設法提高調查品質，包含問卷設計、訪員訓練、抽樣方法、加權方式，及資料分析等流程，如此才能真正有效降低抽樣誤差。

信賴區間裡信心水準的涵意，也常令人感到困惑。取樣前信賴區間是一隨機區間，取樣後得到一固定的區間。此固定區間要嘛包含所欲估計的未知量，要嘛就不包含。說有 0.95 的機率包含，令不少人難以理解。要知機率從來不是僅看少數幾次結果。對同一信心水準，如 95%，及同一樣本數，若反覆觀測，得到很多不同的信賴區間。在這些信賴區間中，將大約有 95%個，會包含所欲估計的未知量。就如一銅板出現正面的機率若為 0.6，表投擲多次後，其中將約有 60%次

是出現正面。但若只投擲一次，當然不是出現正面便是反面，不會有 0.6 次為正面。至於所謂信心(或信賴)，乃指對獲得信賴區間的過程有信心，而非針對所得那一特定區間而言。若將信心水準 95%，想成一事件發生的機率為 0.95，大致便可了解 95%的意義了。至於對不能重覆觀測的事件，如前述醫生說某病人能再活 2 至 6 個月的機率為 0.8，如果長期來說，差不多有 80%的病人，的確在醫生所講 0.8 的區間過世，那就表示醫生夠專業，估計的不錯，並非信口開河。

4. 結語

信賴區間之較詳細的討論，及其中牽涉的一些推導，可參考黃文璋(2007)一文。有關抽樣調查的初步介紹，可翻閱黃文璋(2010) 一文。至於中央極限定理之闡釋，可見黃文璋(2011a)一文。民國 98 至 101 年，那 4 年學測及指考的數學考題中，有幾題是關於信賴區間的，不論題目敘述或解答有爭議者，我們皆挑出來討論，並彙整成黃文璋(2011b)一文。

參考文獻

1. 黃文璋(2007). 統計裡的信賴。數學傳播季刊，30(4)：48-61。
2. 黃文璋(2010). 抽樣調查(1)–(9)，高瞻計畫資源平台。
3. 黃文璋(2011a). 庶民中央極限定理。黃家小館(http：//huang.nuk.edu.tw/cindex.htm)。
4. 黃文璋(2011b). 機率統計考題探討。黃家小館(http：

//huang.nuk.edu.tw/cindex.htm)。