假設與檢定

黃文璋
國立高雄大學應用數學系

1 前言

數學裡常提出命題, 然後證明該命題為眞, 或證明該命題為偽。即在某些假設下, 推導出某結果成立。命題中的假設是不必證明的。例如, 假設有一直角三角形, 則兩股平方和等於斜邊平方, 這即畢氏定理。不必懷疑該三角形是否眞有一直角, 對於所給的假設, 接受便是。而且, 即使量測一萬個直角三角形, 發現的確兩股平方和等於斜邊平方, 離完成證明, 仍差十萬八千里。事實上, 連一個直角三角形都不必去量測三邊, 所須做的, 是依數學程序, 導出對一般的直角三角形, 兩股平方和等於斜邊平方。另外, 除非證明過程中有錯誤, 否則一旦證出來, 便不可能找到反例。

現在科技發達, 拍電影時, 要什麼有什麼, 早期可沒這麼方便。劇本裡說"有片雲從山那邊飄過來", 輕描淡寫的一句話, 但拍攝時, 卻可能晴空萬里, 連一絲雲也無。為了電影中才出現一下子的那片雲, 得苦等半天, 所謂等雲到, 見吳菲及李建銓譯(2012)一書。另外, 劇本裡若有個令人不敢逼視的美女, 就得找到恰當的人來飾演。曾有某著名女星, 預定接演電影劇"神鵰俠侶"中之小龍女一角, 不料消息才剛曝光, 便頻遭網友圍剿, 炮轟她身材胖、臉又圓。還有人說"神鵰背得動這麼胖的小龍女嗎?"總之就是認為她一點都不適合擔綱此角色。製作人不得不替該演員緩

頰, 說“其實她眞的不胖, 只是比較沒下巴, 臉看起來才會圓 。”劇本乃為拍攝而寫, 要能執行, 因此不能太離譜 。至於數學裡的假設, 就相當自由了, 根本不必在意眼見為實 。既然無須被檢驗, 那可否胡亂假設? 也不是不行 。而且如果假設不眞, 則可得到任意的結論 。例如, 設$3 = 4$, 則三角形有五個邊, 此命題便為眞 。三角形怎會有五個邊? 3都可以等於4了, 還有什麼不行? 這就說明對那些指鹿為馬者, 人們何以特別防備, 因知道在那些人心中, 沒有什麼不可以 。

數學裡有時也會給定假設, 然後去解一些問題 。例如, 設汽油一公升可行駛12公里, 問20公升可行駛幾公里? 你知道當然是240公里 。只要是純數學的問題, 便不會被挑戰 。但若這是一眞實的情況, 即某廠牌汽車商, 宣稱汽油一公升可行駛12公里, 則說不定便有消費者, 因注意到每公升根本行駛不到12公里, 而懷疑廠商有誇大之嫌 。因此一公升汽油可行駛幾公里, 廠商不可胡亂假設, 一定要有所本 。

生活裡並非僅有數學, 數學裡12就是12, 永不會改變 。只是我們所處, 乃一隨機世界 。一公升汽油究竟可行駛幾公里, 不要說對同一輛車, 並非永為定值 。就算是新車, 亦不見得每部車都相同 。一公升汽油能行駛的里程, 是會變動的 。就是這樣, 在隨機世界中, 那些是眞, 那些是偽, 常難以判定 。你要眞相 ? 眞相只有天曉得 。幾乎可說, 一切都是假設, 只是有些假設能接受 。本文便是要討論如何假設, 及在怎樣的情況下, 假設可被接受, 或被棄卻的問題 。

2 證實

在蕭敬騰(1987-)那首“奮不顧身”中, 有底下一段歌詞:

> 我用我的生命愛你, 不讓尖銳的世界傷害你 。
> 不懂花言巧語的人, 我愛你, 往往來不及證明 。

寫歌詞者邏輯不差, 因其實並非來不及證明, 而是“愛”根本無法證明, 遂只好以不懂花言巧語 、來不及做為搪塞 。眾所周知, 世上很多事乃無法證

明。例如, 在莊嚴的教堂裡, 主持婚禮的牧師, 可不是問新娘“你確定這個男人是你的眞命天子嗎? ”而是問“你願意接受這個男人為你的丈夫嗎? ”至於願不願意, 不就是看標準訂多高? 標準若設太高, 便難以接受。一旦標準降低, 便容易接受了。人生乃充滿妥協, 很多時候, 雖不滿意, 卻也只能接受。

數學中處處是證明, 科學裡則講“證實”。例如, 不時會看到標題如下的新聞報導“德國科學家證實, 喝綠茶可以減肥。”科學家並未宣佈喝綠茶可以減肥已得到證明, 這大概不易證明, 而是說證實可以減肥。類似的原理, 科學家若經一套合理的取樣過程, 及一合理的判定過程, 逕予宣佈“證實直角三角形兩股平方和等於斜邊平方”, 並無不可。但證實畢竟不是證明, 數學家對這類證實, 即使不嗤之以鼻, 通常不會有太大興趣。

有些證實的確為眞, 有些後來有人提出結論迥異的證實, 有些則一直令人半信半疑。底下將看, 科學家到底如何證實。

3 假設與檢定

兩人分一塊蛋糕, 該如何分, 才會讓雙方都覺得公平, 或者說都覺得自己拿到的並非較小? 由一人負責切, 然後讓另一人優先挑選, 是一合理的方法。有一男孩不知該如何做, 才能取悅女孩。有人建議他秉持民之所欲常在我心的精神, 揣摩女孩喜歡什麼, 他就使命必達。

很多時候, 就是會遇到上述這類如何讓雙方都較服氣, 或較不至於心有不甘的問題。唐宋八大家之一的歐陽修(1007-1072), 在追念其父母的“岡阡表”一文中, 提到其父為死囚“求其生而不得, 則死者與我皆無恨也”。早在1千年前, 歐陽修之父, 可說便已具備現代法律裡所重視之“無罪推定”的精神。我國刑事訴訟法第154條:

> 被告未經審判證明有罪確定前, 推定其為無罪。犯罪事實應依證據認定之, 無證據不得認定犯罪事實。

這就是所謂“無罪推定原則”。我國最高法院, 於民國25年立下有罪推定原則的判例後, 經過65年, 民國90年廢止該判例。自此被告原則上是無罪

的, 法官只要認定罪證不足, 即可宣判無罪, 而不必窮調查之途。在無罪的前題下, 若證據夠充分, 就不得不判定有罪。必須一提的是, 前述條文裡的“證明”一詞, 應與數學裡的證明意義不同, 而與前一節裡的“證實”同義。

依無罪推定原則, 如果德國科學家相信喝綠茶可以減肥, 那該如何假設? 要先假設喝綠茶不能減肥, 然後若證據夠強, 便能推翻此假設, 即證實喝綠茶可以減肥。統計學裡, 設計了一套“假設與檢定”的流程。一開始要先確定二假設, 即虛無假設(null hypothesis)與對立假設(alternative hypothesis)。虛無假設通常表現況, 或傾向想推翻的; 而對立假設則表傾向接受的。雖想推翻虛無假設, 卻很有風度, 儘量保護, 不讓它輕易被推翻。如此一旦推翻, 才能減少不服。也就是須儘量做到“求其生而不得, 則死者與我皆無恨也”。

對隨機現象做決策, 要不誤判幾乎是不可能。例如, 若懷疑某銅板較易出現正面, 令p表銅板出現正面的機率, 則可將虛無假設取為$p = 1/2$, 對立假設取為$p > 1/2$。只是即使銅板實際出現正面的機率為0.55, 比1/2大, 投擲100次, 也可能正反面各得50次。這時合理的推論, 當然是接受虛無假設, 但這就誤判了。至此, 讀者也許可以明白, 虛無假設何以名之為“虛無”了。天下本無事, 庸人自擾之。不相信銅板為公正, 進行檢定, 大費周章後, 卻接受銅板為公正, 可說白忙一場。英文null意義為空的。接受虛無假設, 乃接受一空的假設。試想, 法官若判定被檢察官起訴的嫌犯無罪, 消費者保護會若宣佈經檢驗某食品成份合格, 表示整個過程沒有建設性, 多此一舉。檢察官得重新偵辦案子了, 而消費者保護會說不定會被認為擾民。

為了簡便, 我們以H_0及H_a, 分別表虛無假設及對立假設。有兩種可能的誤判。其一是虛無假設為真卻拒絕, 稱此為第一型錯誤; 其二是對立假設為真卻拒絕, 稱此為第二型錯誤。在無罪推定之原則下, 被起訴者明明無罪卻被判有罪, 便犯了第一型錯誤; 實際有罪卻被判無罪, 便犯了第二型錯誤。通常犯第一型錯誤比較嚴重。因無罪被判有罪, 可能坐牢、受處罰, 或至少名譽受損, 這種錯誤較難彌補。至於有罪被判無罪, 犯罪者僥

倖逃過制裁, 若因此改邪歸正, 那便還好; 但若得意忘形, 則夜路走多, 疏漏愈來愈大, 總有被定罪的一日。通常會先設定一個第一型錯誤機率值之上限α, 稱為顯著水準(significance level), α為一較小的值, 常取成0.05, 0.01, 或0.001等。在α給定後, 決定何時拒絕H_0, 即決定拒絕域。拒絕域的選取, 若能使第二型錯誤的機率值β最小, 當然最好。此時的拒絕域, 稱為顯著水準不超過α下之最佳拒絕域。對某些情況, 統計學裡有一套找到最佳拒絕域的方法。

既然犯第一型錯誤較嚴重, 那α是否取得愈小愈好? 一般而言, α愈小β將愈大, 無法兩全。以法庭審判為例, 若悲天憫人, 寧可錯放1千不錯罰1人, 對證據的審核高度嚴格, 則很多實際有罪者, 將可興高采烈地等著被判無罪了。所以α取得過小, 不見得就好。宜視不同狀況, 斟酌取適當大小的α。

現考慮檢定銅板出現正面的機率p。設$H_0 : p = 1/2$，$H_a : p \neq 1/2$。即擬檢定此是否為一公正銅板。持續投擲銅板n次, 以X表所得正面數, 則X有參數n, p之二項分佈, 即$\mathcal{B}(n, p)$。當H_0為真, 即$p = 1/2$, 則X較可能落在期望值$n/2$的附近。所以直觀上, X偏離$n/2$較大時該拒絕H_0。由是取拒絕域為$\{|X - n/2| \geq c\} = \{X \geq n/2 + c, 或X \leq n/2 - c\}$, 其中$c$由$n$及$\alpha$決定。現取$n = 100$, $\alpha = 0.05$。由於n較大時, 二項分佈的機率值不太好算, 以常態分佈來近似, 得c約為9.8。取$c = 10$, 因c該是整數。如此拒絕域 $= \{X \geq 60, 或X \leq 40\}$, 且實際的$\alpha$值約為0.0456。對離散型分佈, 有時無法取到剛好能達到所給α值之拒絕域。若$\alpha = 0.01$, 則c約為12.88, 所以取$c = 13$。如此拒絕域 $= \{X \geq 63, 或X \leq 37\}$, 且實際的$\alpha$值約為0.0094。在同樣的$n$之下, α值愈小, 表H_0愈被保護, 因此拒絕域將愈小, 即愈不容易拒絕H_0。投擲銅板100次, 得到62個正面, 比在H_0下(銅板為公正)的期望值50多了12, 即超過24%, 感覺上很偏差, 卻仍得接受此銅板為公正。沒辦法, 那是α太小的關係。解決之道是加大n。假設取$n = 10,000$, 則在H_0下, X之期望值$= 5,000$。當$\alpha = 0.01$時, c約為128.8, 故取$c = 129$, 且拒絕域 $= \{X \geq 5,129, 或X \leq 4,871\}$。此時正面數$X$, 只要比$5,000$偏離逾$129/5,000 = 2.58\%$, 就得拒絕$H_0$了。至於

有關求第二型錯誤的機率, 比較複雜些, 在此不討論 。

最後來看"顯著"一詞的由來 。顯著與否, 乃依發生機率的大小 。發生機率較大的事件若發生, 乃屬稀鬆平常, 無須大驚小怪 。但若小機率事件發生, 此事件便屬顯著, 顯著事件自然引人注意 。至於怎樣的機率算小? 0.05或0.01? 當然視情況而定, 不能一概而論 。所以得先訂個標準, 亦即給定顯著水準, 依此決定拒絕域 。因而觀測值若落在拒絕域, 便稱檢定結果為顯著, 且接受H_a; 否則便是不顯著, 且接受H_0 。

4 結語

統計是門入世的學問 。統計學裡提供一套假設與檢定的程序, 以為做選擇, 或給推論時之依據 。此程序並非在判定事情的眞偽, 而是用來做為擬採對策之指引 。本文僅對假設與檢定給一粗淺的介紹, 想進一步認識此題材者, 可參考黃文璋(2004)及(2005)二文 。了解假設與檢定的內涵後, 將發現它不過是將人們平常做抉擇的思維, 有系統地給出一執行的流程 。設計出此流程, 算不上是一顯著事件 。因就是有那麼多極具洞察力的統計學家, 他們能透視人們的某種思維, 並據此設計出好的方法 。

參考文獻

1. 黃文璋(2004). 統計學裡無罪推定的精神. 科學發展月刊, 383期(2004年11月號): 68-73.
2. 黃文璋(2005). 統計顯著性. 數學傳播季刊, 29(4): 29-38.
3. 吳菲及李建銓譯(2012). 等雲到: 與黑澤明導演在一起的美好時光(野上照代原著) 。漫遊者文化事業股份有限公司, 台北市 。