談統計誤差—假設檢定篇

黃文璋

國立高雄大學統計學研究所

1 12 怒漢

“12 怒漢”(12 Angry Men), 是部 1957 年出品的電影。將近 60 年前的黑白片, 參與的演員中, 除了主角亨利方達 (1905−1982) 外, 其餘今日知道他們的人, 應寥寥可數了 。亨利方達曾以“金色池塘”(On Golden Pond, 1981) 獲奧斯卡金像獎最佳男主角獎, 且於 1999 年, 被美國電影學會 (American Film Institute) 評選為百年來最偉大的男演員第 6 名。雖是這麼了不得, 但江山代有才人出, 電影界又一向新人輩出, 才剛開始引領風騷的後浪, 常便迫不急待地想推走前浪。而且排名第 6 又如何? 畢竟已過世三十餘年, 除非已有些年紀, 否則熟悉亨利方達的人, 恐怕不是很多了。這麼說來, 現今會有多少人對“12 怒漢” 有興趣? 何以我們會提這部塵封多年的電影呢?

不會過時, 歷久彌新, 這確實是一部很值得觀賞的電影。聽過 IMDb(Internet Movie Database 之縮寫) 嗎? 此為 1990 年成立的一網路電影 (包含電視) 資料庫。任何人只要上網註冊, 便能對任一部該網站上的電影, 以 1 至 10 給分。IMDb 收錄的資料極豐富, 至 2016 年 6 月, 有 1 百多年間的影片約 370 萬部, 影劇界的人物約 7 百萬位。至於註冊會員, 則多達 6 千 7 百萬位。對於每部電影, IMDb 皆給出平均分數, 且提供評分最高的 250 部影片名單。只是各影片給分的人大不相同, 怎麼能放在一起排序?

IMDb 網站設計出一套評比的辦法, 其中還用到統計裡的貝氏估計法, 看起來頗合理。在這 250 部佳片名單中, “12 怒漢” 高居第 6, 且是前 18 名中, 年代最早的一部。此片至 2016 年 6 月, 平均分數約 8.9。滿分 10, 8.9 很高嗎? 不妨給一參考值來比較。“驚爆焦點”(Spotlight, 2015), 於 2016 年 2 月底頒獎的第 88 屆奧斯卡金像獎, 獲得最佳影片獎。年度最佳影片, 該是部不錯的電影吧! 此片至 2016 年 6 月, 平均分數約為 8.1, 評分者約有 17.0 萬人, 已進入 250 部佳片名單, 排名第 187。要知會經常上網評分者, 大抵是電影的愛好者, 有一定的品味, 因此可以理解不會太輕易給高分。通常一部電影, 若能獲得 7 分以上, 且評分者夠多, 就可說是部相當好的電影。評分人數若太少, 有可能是因看過此片的人並不多, 於是得到的平均, 說不定便不太客觀。只是怎樣算是評分者夠多, 還是得因影片性質而定。“12 怒漢” 不但得到 8.9 的高分, 且評分者眾。至 2016 年 6 月, 上網為它評分者, 超過 43.8 萬人。所以, 以為 “12 怒漢” 應過氣了, 只剩典藏價值, 可就錯了。幾十年下來, 一直仍持續產生新知音, 會設法找到此片來欣賞。多年累積約 43.8 萬評分的人, 這算多嗎? 再度, 我們給一參考值。“亂世佳人”(Gone with the Wind, 1939), 雖是部比 “12 怒漢” 更老的影片, 但它大名鼎鼎, 聽過的人, 與看過的人, 相信皆不比 “12 怒漢” 少。至 2016 年 6 月, 它的平均分數約為 8.2, 在 250 部佳片名單中, 排名第 159。此片得到的評分人數, 不到 “12 怒漢” 的一半, 約有 21.2 萬人。

“12 怒漢” 不但是黑白電影, 且全片總共只有 3 個場景, 還大部分的時間在對話, 拍片成本想必不高 (IMDb 網站上估計其預算為 35 萬美元)。這樣的電影還能引人入勝, 顯然得有很好的劇本才行。向來大家較重視演員及導演, 但其實劇本是一部電影的骨髓。片子一開始, 法官對陪審團說:

> 本案已經有一人死亡，另有一人的生死，掌握在你們手上。如果你們能提出合理的懷疑，無法確認被告有罪，則應作出無罪判決；如果找不出合理的懷疑，則便須基於良知，判決被告有罪。

接著 12 位陪審員，便退庭進入一會議室。之後場景便都在此小房間，偶有工作人員拿東西進來。在商議期間，陪審員皆不能出去，直到最後 12 人離開法院。曾有跟著爸媽看此片的孩童，在過了十餘分鐘後，只見一群人一直在同一房間內講話，忍不住問“開始演了沒有？”在美國，民事案件，可採多數決。但對於刑事案件，不論有罪或無罪的決定，陪審團都需一致決。至於在房間內要待多久，就看陪審員花多長的時間，才達到共識。若一直僵持不下呢？法官可解散陪審團，然後另組新的陪審團，重新審判，這自然又是一冗長程序的開始。被告一旦被判有罪，刑罰通常交由法官決定。

採陪審團制，有若干優點，其中之一是可彌補職業法官，由於長期待在法律的象牙塔內，見識不夠多元，甚至不知民間疾苦，而有成為恐龍法官的可能。那判決的準確性呢？這是人們所關心的。逐漸大家便能理解，透過一致決高門檻的設定，是可以降低判決錯誤率。

案中一名居住在紐約的少年，被控於某日深夜在家中，以彈簧刀殺死父親。住樓下的跛腳老先生，表示親耳聽到少年大叫“我要殺了你！”並看到被告慌慌張張地逃出家門。住男孩家對街樓上的的女士，則表示她親眼目睹少年一刀插入父親的心口。附近雜貨店的老闆，更作證賣過行兇的彈簧刀，給這位少年，當然賣的時候，他並不知買刀的用途。男孩則聲稱和父親爭吵後，便怒而出門，隨即去電影院。幾小時後回到家，發現父親已經遇害。15 歲的少年有前科，曾進感化院，說去看電影，卻又無法講出片名及主要演員。幾乎所有的證據，都對少年不利。

在悶熱的的會議室裡, 有好幾位陪審員, 一開始便顯示出極度的煩躁, 只想趕快終結此一案件。有人還揚著手中的球票, 嚷嚷等下要去看的棒球賽, 將有多麼精彩。明顯是個不良少年, 為了一些小爭執, 就殺死父親, 真太凶狠了, 絕不能饒恕。既然證據如此確鑿, 就逕行投票吧! 11 比 1, 有 11 人認為被告有罪, 只有 1 人不同意。有趣的是, 在投有罪時, 有 6 人是毫不猶豫地舉手, 有 5 人則是看到 6 隻高舉的手, 才緩緩舉起手來。面對一面倒的 11 張有罪票, 亨利方達不為所動, 於投無罪時, 堅定地舉起手來, 成為唯一的無罪票。由於採一致決, 幾位極力認定少年有罪者, 大聲質疑獨排眾議的亨利方達, 彷彿他是個搗蛋鬼, 眼前這麼多無懈可擊的證據, 卻冥頑不靈, 企圖拖累大家。於是展開一場精彩又激烈的辯論。

亨利方達不疾不徐, 冷靜理性地逐一提出他的合理懷疑, 讓其他陪審員的偏見、歧視及無同理心, 逐漸融化。歷經一次又一次的投票, 有罪與無罪之比, 遂由 11 比 1, 10 比 2, 9 比 3, $\cdots$, 終於來到 1 比 11。冥頑不靈者換人了, 最後, 那位從討論之始, 便一直大聲叫喊, 斥責一個接一個 "變節者" 的陪審員, 面對 22 隻質疑的眼睛, 羞愧地說出 "無罪"。亨利方達拍拍他的肩, 替他拿西裝外套披上, 不發一語地走出會議室。

在爭論中, 所逐漸浮出的一些疑點有:

(1) 住在對面的女證人, 說她親眼看見男孩將刀舉過頭, 狠狠地往其父親胸口刺下。當時高架鐵道捷運系統, 正好有一輛 6 節的列車經過, 她表示自己是透過列車最後兩節的車窗, 看到此情景。但亨利方達說, 他曾住在高架鐵道旁, 當時若有列車經過, 所產生轟隆震耳的噪音, 將使老先生不可能清楚聽到少年說什麼 "我要殺死你! "。

(2) 老先生說他聽到少年講 "我要殺死你! " 後, 隔

了 1 秒鐘, 聽到有物體倒下 (他判斷是被害人), 再花了 15 秒, 他從臥室穿過走廊走到大門, 看見少年倉皇地逃逸。但老先生跛腳, 行動不便, 根本無法奔跑。亨利方達當場模擬, 結果發現, 以老先生緩慢的走路速度, 大約需經 41 秒, 才能走到大門, 他卻聲稱只花 15 秒。因此他既不可能聽清楚誰說“我要殺死你”, 也不可能看到少年逃出門。陪審員中年紀最大的那位, 自認很了解老人。他指出由老先生穿著破爛, 顯示他是個乏人照顧、無人在意, 且平常不會被多看一眼的人。但在本案中, 卻一躍成為主要的證人。為享受這種不曾有的被矚目之滿足感, 他說了謊。

(3) 少年說去看電影, 但卻什麼都記不得, 因此被認定是企圖製造不在場的證明, 才騙人他去戲院。但亨利方達認為, 少年與父親爭吵後, 氣呼呼地出門, 在情緒不太穩定下, 不過隨便挑部電影進去, 看時不見得很專心。回家後發現父親已身亡, 警察又守在那裡, 太過緊張下, 腦袋一片空白, 乃很自然。為了支持其論點, 亨利方達追問一位陪審員, 昨晚做什麼? 前晚做什麼? 正好這位陪審員, 前幾天也去看了電影。結果不但記錯片名, 也記錯演員的名字。連在沒有壓力下回答, 都無法講正確, 因此少年對所看電影的細節不復記憶, 亦不無可能。故不能就比斷定, 其所提供的不在場證明, 必為謊言。

(4) 亨利方達拿出一把彈簧刀, 眾人嚇一跳, 想凶刀怎會在他那兒? 亨利方達表示是他自己買的, 這種刀很普遍, 因此買刀不必然導致準備殺人。

(5) 警方指出死者是被人由上往下刺殺。但少年的

父親, 較少年高出 7 吋 (17.78 公分)。而若要殺一個比自己高的人, 一般人是不會高舉刀子往下刺。有位陪審員也說, 他幼時常目睹鄰里間的械鬥。他相當確定, 當拿彈弓刀砍人時, 為了能快速彈出刀片, 使用時必不會高舉刺下。

(6) 一個人就算對人說出 "我要殺死你!" 可能只是氣話, 不見得真想殺人。由於亨利方達一件又一件地拋出疑點, 支持他的陪審員逐漸增加, 那位最堅持有罪者, 盛怒之下, 居然脫口對他說 "我要殺死你!" 眾人一愣之下, 頓有所悟。

(7) 住對面的女證人, 說事發那晚, 她在床上輾轉難眠 1 小時, 於午夜 12 時 10 分, 正好瞥見少年行凶。但有位陪審員注意到, 女證人出庭時, 鼻樑上有很深的眼鏡壓痕, 顯然她平常需戴眼鏡才能看得清楚。而一般人睡覺時, 眼鏡會取下, 因此合理推斷, 深夜她瞥見有人行凶時, 由於並沒時間抓起眼鏡來戴, 遂只可能看到一模糊的身影, 因此豈有辦法看清凶手就是那位少年?

在片中亨利方達強調:

> 我並不確定這男孩無罪, 而是提出合理的懷疑。在一切證據都無法被肯定為真, 甚至可能被推翻的情況下, 我們憑什麼來定他的罪?

就是這樣, 他對原本已被認定的證據, 拋出一個又一個合理的懷疑。而有幾位陪審員受到他的啟發, 也提出自己所看到之合理的懷疑。而多數陪審員, 當自己之前相信的證據被推翻, 也就願改變立場。說到底, 都還不是真的那麼頑固。人們常說 "真理愈辯愈明",

但先決條件是有追求真理的意願, 如此才願意客觀的“辯”。而唯有客觀下的“辯”, 真理方能逐漸“明”起來。只是社會上, 仍有一些人是理盲又濫情, 寧可被他人的想法牽著鼻子走, 自己卻不願好好思考。因而人云亦云, 只知跟著搖旗吶喊。不但毫無判斷力, 還因附合的聲音大, 便自以為公義。當勢單力薄時, 不畏千夫諾諾的氛圍, 依舊實話實說, 堅持追求真理的精神, 這種人是很難得可貴的。舉世皆濁時, 常就是靠著那些很少數不屈不撓的人, 來力挽狂瀾。

“12 怒漢”, 藉一陪審團, 對少年弒父案審理的辯論過程, 讓人對什麼是公理正義, 什麼是真相, 產生省思。而亨利方達, 自始至終, 不卑不亢, 溫和而堅定, 最後成功地避免一冤案的發生。其秉持合理的懷疑之信念, 其理直而氣和之表達, 均頗值得我們學習。

2 誤差難免

基於合理的懷疑, 在“12 怒漢”中, 12 位陪審員, 從開始一面倒的認為被告有罪, 逐漸一個個看法丕變。原本千夫所指的少年, 也就被判無罪釋放。結局固然快樂, 但真相是否就此大白? 其實仍然未知。因雖不能大膽假設買刀者都準備要殺人, 但買的刀與凶刀同一款式, 還是一不容忽視的疑點, 值得小心去求證。至於跛腳老先生, 跟住對面女士的證詞, 是有瑕疵, 不能全盤接受。不過也可能是兩個缺乏社會經驗的人, 為了強化證詞之可信度, 對所見情境, 畫蛇添足下, 反而產生漏洞, 變成難以採信。這種情況, 並非不尋常。甚至還有因為多事, 造成自證己罪者。清末劉鶚 (1857−1909) 的“老殘遊記”一書, 在“初編”之第 15 至 20 回, 有一“剛弼斷案”的故事, 便為一例。

在山東齊河縣的齊東鎮, 發生一起賈府一家大小 13 口, 被人用毒藥謀害的命案, 震驚全縣。魏氏父女, 因涉嫌而被收押。山東撫

臺, 特地指派一向以清廉著稱的官員剛弼來查案。魏家管事, 見主人吃冤枉官司, 求助一胡姓舉人。胡舉人與剛弼相識, 覺得這不過小事一樁, 欣然接受委託。見到了剛弼, 胡舉人先奉上魏家管事交給他的 1 千兩銀票, 當做伴手禮。剛弼於了解來意後, 說 1 條人命本值 1 千兩銀子, 13 條共 1 萬 3 千。但看在胡舉人面上, 可減半算, 就 6 千 5 百兩。且銀子不急, 但須先寫個憑據。還要胡舉人向魏家管事說明, 6 千 5 百這數目, 是怎麼得出來的。胡舉人連連答應, 隨後寫了一封信, 連同管事提供將支付 5 千 5 百兩 (已送了 1 千兩) 的憑據, 一起送至縣衙門。

隔日剛弼升堂會審, 拿出 1 千兩銀票、5 千 5 百兩的憑據, 以及胡舉人的信, 要魏氏父女招供。一切環環相扣, 斷定 13 個人, 就是他們所殺害, 不招就用刑。待最後水落石出, 找到真凶, 還了魏氏父女清白, 剛弼尚百思不解。如果明明無辜, 為什麼要行賄? 還就同意 6 千 5 百兩銀子, 沒有爭辯根本連 1 人也未殺害, 豈來 500 乘上 13?

書中透過一位白太守來說穿此迷團, 其實並沒什麼太深奧的道理, 不過就是"鄉下人沒見識"。書上說, 魏家管事是個"愚忠的老實人", 眼見主人吃了冤枉官司, 擔心被屈打成招, 情急之下, 自做主張奔走籌款, 以營救主人。只要能救出主人, 一切聽命行事, 根本未多想是否被設計了。而胡舉人雖飽讀詩書, 卻不過是冬烘先生一個。見出面助人有效, 欣喜之下, 修書感謝, 順便結個文字緣, 不料反留下企圖買通官員的證據。

不時會聽到人們抱怨"豈有此理! "或訝異地說"怎麼可能? "顯然天下事, 並非都盡合理, 而不可能的事, 也一再發生。統計學是從所得的結果, 來驗證前提是否可接受。當看到不合理的現象發生, 的確該提高警覺, 探究到底是什麼原因, 才造成此現象出現。但卻不可如前述剛弼, 過度執著, 由所見到的不合理, 認定就只有一個原

因, 而不再想是否有其他可能性。至於所謂不可能, 其實就是小機率事件。只是不論事件的機率多小, 一旦觀測次數夠多, 就不必訝異其發生。樂透彩中頭獎的機率夠小了吧! 怎麼有人中? 作弊嗎? 未必, 賣出那麼多張, 頭獎開出來, 就幾乎是必然了。

在路易斯卡羅 (Lewis Carroll, 本名查爾斯路特維奇道奇森 (Charles Lutwidge Dodgson), 1832−1898, 亦為一數學家) 的 "愛麗絲鏡中奇遇"(Through the Looking-Glass, and What Alice Found There, 1871, 中文譯本極多, 王安琪譯, 2015, 是一較新的版本) 一書的第五章中, 愛麗絲理所當然地認為 "One can't believe impossible things. " 不能相信不可能的事, 不只是愛麗絲, 這或許是大部分人的想法。結果王后對愛麗絲說 "I dare say you haven't had much practice.⋯. Sometimes I've believed as many as six impossible things before breakfast. " 在早餐前王后就可以相信 (看到)6 件不可能的事。怎麼有辦法? 王后覺得一點都不難, 只要多練習即可。要知我們周遭有各式各樣的小機率事件, 若能細加留意, 將發現隨時都會有某一件冒出來。至於到底有那些屬於不可能, 卻會發生的事件? 事實上還真不少, 底下給個例子。

NBA(National Basketball Association) 是北美男子職業籃球組織。擁有 30 支球隊, 分屬東西兩聯盟 (conference)。每聯盟有 3 個分組 (division), 每一分組有 5 支球隊。總共 30 支球隊中, 有 29 隊位於美國, 另 1 隊來自加拿大。NBA 每年的正式賽季, 從 11 月初開始。先進行例行賽 (regular season), 每支球隊都要完成 82 場比賽。隔年 4 月底, 開始打季後賽 (playoffs)。每聯盟的 15 支球隊中, 皆有 8 隊能參加季後賽, 包括例行賽 3 支分組冠軍 (由勝率決定), 及另挑出勝率最高的 5 隊。有超過一半的球隊能打季後賽, 不算很難, 值得一拼。因此各支球隊, 在每年的例行賽中, 一個目標, 就是打進季後賽。一旦這個目標達到, 便有如那年的表現及格, 可有個

安心的暑假了。季後賽先在兩聯盟裡, 各自一強一弱抓對廝殺, 採 7 戰 4 勝淘汰制。3 輪後, 兩支聯盟冠軍產生, 爭奪年度總冠軍。能獲總冠軍, 那真是莫大的榮耀, 英勇的事蹟, 往後幾十年, 可一直傳頌下去。的確也相當不容易, 季後賽得闖過 4 關, 共贏 16 場。

2015–16 NBA 例行賽, 被公認最強的球隊, 為西區金州勇士隊 (Golden State Warriors, "金州" 是加州的別名)。勇士隊前一年才獲得睽違整整 40 年的總冠軍 (上次奪冠是 1974–75 球季)。氣勢延續下來, 這一年的例行賽, 戰績一路長紅, 備受矚目。最終取得 73 勝 9 敗, 勝率超過 8 成 9, 打破 20 年前, 1995–96 年芝加哥公牛隊 (Chicago Bulls) 所創, 72 勝 10 敗的紀錄。勇士隊這年還創下很多紀錄, 包括成為 NBA 史上, 第一支在單季例行賽裡, 沒有連敗過的球隊, 及第一支在單季例行賽裡, 投進 1,000 顆以上三分球的球隊。光是控球後衛史蒂芬柯瑞 (Stephen Curry, 1988–) 一人, 便投進 402 顆, 也是 NBA 個人於單季例行賽中, 投進最多三分球的紀錄。這樣超強的球隊, 除了查爾斯巴克利 (Charles Barkley, 1963–), 那位一向口無遮攔的電視籃球評論員外, 恐怕無人不看好他們能連莊總冠軍。

季後賽首輪, 勇士隊以 4 比 1 擊敗休士頓火箭隊 (Houston Rockets), 次輪再度以 4 比 1 終結波特蘭拓荒者隊 (Portland Trail Blazers), 順利打進西區冠軍賽, 對手是奧克拉荷馬市雷霆隊 (Oklahoma City Thunder)。首戰雷霆隊贏了, 第二戰勇士隊扳回一城, 雙方 1 比 1 平手。之後兩戰, 整年例行賽, 不曾連敗過的勇士隊, 居然都輸了, 跌破眾人眼鏡, 也讓雷霆隊取得 3 勝 1 敗的聽牌優勢。數據會說話, 這時勇士隊例行賽輝煌的戰績, 被拋到九霄雲外了, 媒體拿出過去的統計資料分析。勇士隊的衛冕之路, 不要說被看成亮起紅燈、陷入困境, 甚至有人認為已命懸一線了。球運豈能每年都那麼好? 雷霆隊的支持者興奮莫名, 勇士隊的支持者, 則難以理解, 明

明是一支無敵之師, 怎麼可能輸成這樣? 至於查爾斯巴克利, 該可以得意了。

NBA 分區冠軍賽史上, 之前共出現過 39 次 3 比 1 的戰局。領先的球隊, 最終有 37 隊拿下分區冠軍, 且其中有 20 次, 是只再賽 1 場, 領先者就以 4 勝 1 敗終結比賽了。僅有的兩次逆轉, 都發生在東區, 一次是 1979 年, 一次是 1981 年, 都至少是 25 年前的事了。更不妙的是, 勇士隊所隸屬的西區, 冠軍賽過去未曾有由 1 比 3 翻盤成功之例。39 分之 2, 僅比 5% 略多一點的機會。而如果單看西區, 則過去成功 0 次, 因此機會是 0。處在 1 勝 3 敗大幅落後的局面, 多少反映雙方實力有差, 若說大勢已去, 算是持平之論, 畢竟數據會說話。還有人宣稱勇士隊翻盤的機率, 不到 4%, 他們有不同的計算方式。NBA 自從將季後賽改為 7 戰 4 勝制後, 若各輪都計, 則球隊陷入 1 勝 3 敗的絕境者, 總共有 232 次, 其中只有 9 次扭轉奇蹟。232 分之 9, 約 0.03879。上述那一不到 4% 的值, 就是這麼來的。

不論機率是 0.05、0.04, 或是 0, 大部分關心 NBA 球賽的人, 都認為勇士隊不可能贏了, 只好明年再來。但世事難料, 不可能的事發生了。勇士隊有如浴火重生, 連贏 3 場, 以 4 勝 3 敗, 獲得西區冠軍, 取得與東區冠軍克里夫蘭騎士隊 (Cleveland Cavaliers), 角逐總冠軍的資格。成為 NBA 西區冠軍爭奪戰史上, 第一支在 1 勝 3 敗之劣勢下, 逆轉的球隊。也讓原本充滿希望的雷霆隊, 黯然神傷地打包回府, 結束球季。

有些較具機率背景的人, 可能覺得前述求機率的過程, 並不可取。我們常在談機率, 機率究竟是什麼? 眾所周知, 機率有幾種不同的意義。第一種是所謂古典機率, 即以相同的可能性來解釋機率。進行一項隨機試驗, 以 S 表所有可能的結果之集合, S 稱為樣本空間, 不可以是空集合, 也不可以有無限多個元素。S 的任一子

集合, 稱為一事件, S 中的元素個數以 $|S|$ 表之。想求一事件 A 的機率, 其元素個數以 $|A|$ 表之。由於假設相同的可能性, 故每一可能的結果, 發生之機率皆相同, 即為 $1/|S|$。因此事件 A 之機率, 顯然便定義為 $P(A) = |A|/|S|$。諸如玩撲克牌, 及玩樂透彩等, 都屬於古典機率適用的範圍。

重複投擲一銅板 n 次, 若得到 k 個正面, 則將 k/n 視為銅板出現正面的機率。這種以事件發生次數的相對頻率, 來解釋機率, 是一常有的作法, 稱為以頻率來解釋機率, 又稱客觀機率, 或頻率的觀點, 適用在可重複觀測的事件。所謂客觀, 乃由於就是依觀測到的結果, 沒有滲入個人的偏好。例如, 我國中央氣象局, 在其網站上, 對 "降雨機率 60%" 的說明為, 每預測 100 次, 實際降雨 (平均) 有 58.6 次。可看出其中的機率, 便是採頻率的觀點。

A 君準備出門, 其母抬頭望天, 只見烏雲密佈, 覺得下雨的機率有 8 成, 力勸 A 君帶傘。A 君卻不以為然, 他判斷烏雲應就快散去了, 下雨的機率頂多 1 成, 雨傘大可不必。這便是主觀機率, 也是一種常用來解釋機率的方式, 一般針對無法重複觀測的事件。由於各人對於同一事件的機率值, 能有不同的認定, 因此才稱之為主觀機率。諸如 "NBA 杜蘭特友人爆料: 他 90% 會留雷霆", 這是 2016 年 7 月 1 日, 中國時報一則新聞的標題。其中 0.9 的機率, 便可視為杜蘭特那位友人的主觀機率。只是留意 NBA 各支球隊動向的人, 大概都會知道, 原本是雷霆隊主將的凱文杜蘭特 (Kevin Durant, 1988–), 於 7 月 4 日 (美國時間), 也就是上述那則新聞出現的幾天後, 在 "球員論壇"(The Players' Tribune) 網站上, 宣佈他已決定加入勇士隊。勇士隊? 沒錯, 就是那支才讓雷霆隊鎩羽而歸、冠軍夢碎的球隊。這真是所謂 "不能與之為敵, 便與之為伍" (If you can't beat them, join them.)。杜蘭特友人那一高達 0.9 的預測機率, 顯然無啥大用。附帶一提, 雖說是主觀, 但往往也會依據過去一些客

觀的事實, 不見得都僅憑個人主觀好惡, 來決定機率值 。

除了上述三種對機率的解釋, 1933 年, 公理化的方式誕生了, 即以機率空間來解釋機率 。也就是在一抽象的系統上, 藉助幾條公理, 引進機率 。此後, 便不必經由一投擲銅板或骰子等情境, 才能討論機率 。但若真有一隨機試驗, 則必能在一適當的機率空間上, 定義機率 。

古典機率 、客觀機率, 及主觀機率, 常交錯著使用 。當沒有更多的資訊時, 常只能採古典機率 。如有人敲門, 是男是女? 若一無所知, 只好認為是男是女的機率相同, 即皆為 1/2 。但若平日屢有訪客, 而過去的經驗顯示, 敲門者有 7 成是女性, 則大抵便會認為男女的機率, 各為 0.3 及 0.7 。另外, 若聽到的敲門聲很大, 有人說不定便以為, 有 0.8 的機率是男生, 因而女生的機率便為 0.2 。

回到勇士隊對決雷霆隊 。於 1 勝 3 敗落居下風後, 人們依據過去的資料, 以預測勇士隊逆轉的機率, 此本為合理的統計思維 。只是對於處在 1 勝 3 敗落後的情況, 雖有過去 39 次分區冠軍賽, 及全部 232 次季後賽的數據, 但年度不同 、球隊不同 、對手不同 、裁判不同, 即使同一球隊, 不同年度的球員或教練, 也可能不盡相同 。故不論 39 或 232 次, 皆不屬重複觀測的事件 。反對者因而認為, 此處並不適用以頻率來解釋機率 。這樣說當然極有道理, 觀念正確 。正如若每次投擲不同的銅板, 則正面出現的相對頻率, 豈能用來估計下一次投擲, 會出現正面的機率? 但與其認為, 由於有逆轉及不逆轉兩種結果, 因而逆轉的機率為 1/2(即採古典機率), 倒不妨採取依據過去客觀事實之主觀機率的想法 。即基於過去數十年, 類似情況的比賽結果, 來判斷逆轉之機率 。這比起胡亂猜測, 畢竟仍客觀許多 。但由於含有主觀因素, 每人之依據可不相同, 所以才會得到 0.05, 0.04, 及 0 等, 不同的勇士隊逆轉機率 。

有趣的是, 當第 6 戰雷霆隊輸球, 與勇士隊打成 3 比 3 平手, 這

時不同的預測值出現了。過去季後賽的資料顯示, 主場球隊的第 7 戰成績是 100 勝 24 敗; 另一方面, 第 6 戰主場落敗的球隊, 第 7 戰的成績是 12 勝 24 敗。由於第 6 戰及第 7 戰的主場, 分別屬於雷霆隊及勇士隊, 因此前述兩筆數據, 便分別導致勇士隊奪得西區冠軍的機率, 為 100/124 及 24/36。二值皆遠比處於 1 比 3 落後時, 對勇士隊逆轉機率之預測, 高出不少。怎會如此? 也沒什麼, 這不過就是條件機率。當情況改變時, 機率值常隨之而變。機率值會變, 乃機率之一特色。

崇尚合理性, 棄不可能如敝屣, 通常並沒有不對。只是連再不合理, 或再不可能的事, 屢屢就是發生, 因此決策怎會永不失誤? 在隨機世界中, 錯誤就是難以避免。我們所能做的, 就是以機率的方式, 來掌握誤差的大小。

3 統計邏輯

誤差大小如何掌握? 隨機世界中不乏資料, 而只要有資料, 統計便能發揮功能, 進而掌握誤差大小。統計分析對資料的仰賴, 有如製磚需要黏土。那統計又是什麼?

統計其實有很多面相, 一種說法是, 統計與考古的原理類似。首先, 都在挖掘, 或者說在探勘。考古是挖掘古物, 統計則是挖掘資料。只是天地這麼大, 往那裡開挖, 才能挖出有價值的古物? 要先研讀文獻, 找出古物較可能之落腳處。同樣地, 資料浩如煙海, 欲有效率地收集到有用的資料, 也是得先好好規劃。另外, 考古所得, 有物無文, 是人來講故事, 說幾千年前如何如何, 將已過了很久的事, 帶到眼前。再度, 挖掘到的資料, 是冰冷的, 並沒有說話, 是人在替它們說話, 讓它們活躍起來, 告訴我們一些有意義的資訊。至於說什麼話呢?

要知統計的目的, 並不見得是在探索真相, 因真相可能永遠未明, 真相不妨就留給上帝。統計主要的目的, 乃是在做決策。即告訴人們怎樣做最好, 其依據便是所收集到的資料。而所謂最好, 得先訂定評比的準則, 否則沒有永遠的最佳策略。舉例而言, 一隻靜止不動的錶, 與一隻慢 1 分鐘的錶, 何者較準? 大部分的人會以為是後者, 因任何時刻都只差 1 分鐘, 前者則常差很多。但也有人以為前者較準, 因每天有兩個時刻完全準確, 後者則無一時刻是準的。所以, 欲評比策略之優劣, 得先確定究竟如何評比。只是在運用統計前, 須具備一些基本的邏輯, 否則學再多方法, 及得到什麼結果都沒用, 因很可能會誤用統計, 反而比不懂統計還糟。底下來看至少有那些邏輯, 是該先弄清楚的。

首先來看因果關係。大家受數學的薰陶較久, 對數學裡 "若 p 則 q" 的因果關係較清楚, 不過我們仍扼要說明一下。假設條件 p 成立, 必導致結果 q 成立, 則稱命題 "若 p 則 q" 成立, 或說為真, 且 p 稱為此命題的充分條件, q 則稱為必要條件。一命題若為真, 則充分條件就是因, 必要條件就是果, 有因便會有果。例如, "若 $x > 3$, 則 $x^2 > 9$", 此命題為真。故若有 "$x > 3$" 這個因, 將得到 "$x^2 > 9$" 這個果。反過來就不一定對了, 由 "$x^2 > 9$", 不必然導致 "$x > 3$", 因也有可能 "$x < -3$"。數學裡經常在證明, 往往就是在確定某因果關係為真。只是在現實世界中, 兩件事是否有因果關係, 常很難判定。底下給幾個例子。

一項調查顯示 (見徐仕美譯 (2016) 頁 227), 北美洲的阿茲海默症 (Alzheimer's disease) 病人, 有三分之二是女性。聽起來很驚人, 但這是否表示女性該擔憂了, 因她們比男性更容易得阿茲海默症? 那可不必然, 看這類報導常要很謹慎, 不能率爾就自行建立一因果關係。要知通常阿茲海默症, 發生在人較年長時, 即老年是阿茲海默症的最大風險因子。而女性的平均壽命比男性長, 因此在最容易

罹病的年齡群 (其中女性佔多數) 中, 罹病者以女性居多, 乃相當合理。那些 70 歲, 沒罹患阿茲海默症, 但死於心臟病的男性, 如果能繼續活下去, 說不定便罹病了。事實上, 2014 年, 美國阿茲海默症協會 (Alzheimer's Association) 曾指出, 目前尚無證據顯示, 任一年齡層中的女性, 比男性更可能失智。所以, 雖罹病者中, 女性佔三分之二, 男性佔三分之一, 看起來比例懸殊, 卻不能就此驟下結論, 說女性較易罹病, 因其中有一不可忽視的 "干擾因子"(confounding factor)—壽命。性別與罹患阿茲海默症相關性很高, 數據提供的資訊只到這裡。但若將其解讀成女性較易罹病, 這便是人說的, 而非數據說的。除非有其他可靠的佐證, 否則千萬不能這樣說。

再看一例。2016 年 6 月 7 日, 聯合報有一則標題是 "高教育程度的人比較容易磨牙? " 之報導:

> 高教育程度的人比較容易磨牙嗎? 奇美醫院口腔顎面外科醫師林哲毅表示, 根據國外的調查確實是這樣的, 但原因並不是很清楚, 可能與壓力有關係, 但也可能是高教育者, 磨牙後就醫的比例會比較高。⋯ 。

磨牙症是指易咬緊牙齒或磨牙, 患者會不自覺地在白天咬緊牙齒, 或在夜間睡眠時磨動牙齒。雖數據顯示, 磨牙就醫的病患中, 高教育程度者, 所佔比率較高。但這不表教育程度較高的人, 有較易磨牙的傾向, 此因果關係不見得成立。真實的情況說不定是, 教育程度較高, 工作職務也可能較高, 而職務高, 壓力常隨之而來, 遂造成不自覺地以磨牙來舒緩壓力。所以, 是壓力大導致易磨牙, 而與教育程度的高低關係不大。另一可能性是, 教育水準較高的人, 常也較了解健康之重要, 當不對勁地磨牙時, 知道該去看醫生。總之, 若想知道磨牙的真實原因, 得進一步探討才行。

這種相關性高, 但其間沒有因果關係的例子處處可見。例如, 依

據統計, 當可樂銷量大時, 醫院腸胃科的門診人數也較多。難道喝可樂對腸胃有傷? 事實上, 可樂銷量大, 往往是天氣炎熱時; 當天氣炎熱, 又使食物易腐敗, 導致易吃壞肚子。氣溫為一干擾因子。

我們再提那一著名的“尿布與啤酒”事件。美國百貨連鎖店 Wal-Mart 執行購物籃分析 (market basket analysis), 他們檢視顧客的購物清單, 發現在有尿布或啤酒的帳單裡, 二者同時出現的比例很高。但兩件物品的使用對象完全不同, 怎麼會這樣? 原來在家中有嬰兒的時期, 先生大抵會較顧家, 週末儘量待在家中協助照料。居家男人, 不少以看電視來打發時間, 這時喝喝啤酒輕鬆一下, 一樂也。因而當受妻子之託, 採買尿布時, 常也順便拿些啤酒, 以犒賞自己。弄清楚原委後, Wal-Mart 遂將看起來毫不相干的尿布與啤酒放在鄰近, 結果二者的銷售量同時增加。

大家常講大數據, 藉助統計, 一般而言, 只能判定二因子間的相關性高或低, 或者說, 能找出相關性較高的因子, 以供進一步探究。只是統計畢竟不是萬能, 除非經由嚴格的實驗, 或有其他學理依據, 否則通常不易判定二因子間, 是否真有因果關係。千萬不可由兩個變數的相關性很高, 就跳到其間有因果關係的推論。

隨機性也是一不容易理解的概念。由於相對頻率, 為一種常見的對機率之解釋, 因而頻率也屢被用來詮釋機率。只是運用時若不夠嚴謹, 常會讓有些初學者, 對機率產生誤解。假設有一公正銅板, 即出現正面的機率為 $1/2$。$1/2$ 的倒數是 2, 有些人便想成“每 (投擲) 兩次出現一次正面”, 隨機性不見了。即使學過一些機率的人, 也可能會有這樣的誤解。持續投擲前述銅板, 令 X 表出現第一個正面所需之投擲數。則 X 之期望值 $E(X) = 2$。口語常便說成“平均每 (投擲) 兩次出現一次正面”, 這是正確的。只是既然是口語, 有人便覺可講得簡單些, 遂也說成“每 (投擲) 兩次出現一次正面”, 即省掉“平均”二字, 這就不對了。至於對機率較有概念的人, 有些也

會以為, 就算不是每 (投擲) 兩次出現一次正面, 只要投擲數 (偶數) 多些, 便應很可能有半數是正面。這並不正確。實際去計算, 投擲 2 次, 出現 1 正 1 反的機率為 1/2; 投擲 10 次, 出現 5 正 5 反的機率為 252/1,024, 約為 0.246, 不到 1/4; 至於投擲 100 次, 出現 50 正 50 反的機率又更小了, 約為 0.08。事實上, 投擲數 (偶數) 愈多, 愈不容易出現正反面數各半。

這種出現次數偏離預期, 當可能的結果愈多, 將會愈明顯。以 42 取 6 的樂透彩為例。每週開 2 期, 1 年大致開 104 期, 3 年就以開 312 期計。每期開出 1 組 6 個頭獎號碼, 3 年共開出 $6 \cdot 312 = 1{,}872$ 個號碼。則 1 至 42, 每號碼出現次數的期望值為 1,872/42, 約為 44.57。但觀測結果, 各號碼出現的頻率, 常會差異很大。最多與最少的, 差異高達 30 次都不稀奇, 反而較少見到每個號碼均出現 44 次左右。這本是正常的現象, 卻常有人因此質疑各號碼並非隨機產生。要知隨機下的不均勻是正常的。教室有一籤筒, 老師以抽籤的方式, 點學生回答問題。每取出一支籤後放回。若一學期下來, 每位學生被點的次數均相同, 學生可不要就誤以為老師果真是隨機抽籤, 而是該懷疑, 籤根本非隨機抽取。

條件機率亦常令人感到迷惑, 我們以測謊為例。測謊機的準確度有多高? 有人說七成, 也有人說九成。有人迷信測謊, 以為人很難能騙過測謊機; 有人則認為測謊機並沒那麼可靠, 在審判中絕不能拿來當證據。底下來略微分析, 而準確度不妨就以九成計好了。設某公司發現有某項業務機密外洩, 因無人承認, 負責安全的主管, 遂提議以測謊來找出洩密者。但這樣做妥當嗎?

假設曾接觸該項業務的人員共有 100 人, 其中有 1 人洩密。因不準的機率為 0.1, 經測謊後, 99 位無辜者中, 平均有 9.9 位會顯示洩密; 那位唯一的洩密者, 有 0.9 的機率會被測出, 有 0.1 的機率不會被測出, 即平均有 0.9 位會顯示洩密。所以測畢後, 平均共有

$10.8(=9.9+0.9)$ 位顯示洩密。但其中其實平均僅有 0.9 位洩密, 即測謊後, 每位顯示洩密的可疑者, 只有 $0.9/10.8(=1/12)$ 的機率洩密, 此值才約 0.0833, 不到一成的機率。甚至, 有 0.1 的機率, 真正洩密者, 並不在那群 (平均有 10.8 位) 被測出之可疑者中。這些都與一開始信誓旦旦之宣稱, 測謊的準確度高達九成, 差異很大。更不要說, 有時真正犯罪者, 較有經驗, 能打敗測謊機; 且有些無辜者, 較容易緊張, 反而通不過測謊, 那誤差就更大了。若不了解條件機率, 將會高估測謊的效果。最後, 如果洩密者根本不在那 100 人中, 洩密乃經由其他管道, 則測謊後, 將產生 10 位無辜的可疑者。要屈打成招嗎? 恐將難以收場了。

由機率 0.9 降至 0.0833, 便涉及條件機率。機率 0.9 的意義是

$$P(\text{顯示洩密}|\text{實際洩密})=0.9,$$

且

$$P(\text{顯示未洩密}|\text{實際未洩密})=0.9,$$

但我們有興趣的, 其實是

$$P(\text{實際洩密}|\text{顯示洩密})\text{。}$$

在本例中, 此值為 1/12, 而這與

$$P(\text{顯示洩密}|\text{實際洩密}),$$

乃不同的條件機率。機率值會變, 是機率的特性。給定某條件後, 機率值將可能會改變。即對二事件 A 與 B, $P(A|B)$ 不一定等於 $P(A)$。這與數學中強調不變性, 如 3 一直就是 3, 2 一直就是 2, 乃完全不同。而

$$P(A|B)\text{與}P(B|A),$$

此二條件機率, 其值也可能差異很大。統計裡, 若有新資訊產生, 即條件改變了, 則機率值往往隨之而變, 這是不可不知的。

4 假設檢定

數學中常在證明, 甚至還有人以為, 數學的工作, 就是證明而已。一命題若經證明為真, 便毫無例外地成立。例如, 直角三角形中, 兩股平方和等於斜邊平方。這是著名的畢氏定理, 大家中學時便學過了, 此命題早在兩千多年前便已證出。由於未曾發現證明有誤, 因此任何人若想找反例, 看是否有那一直角三角形, 兩股平方和不等於斜邊平方, 必然徒勞無功。

數學之外, 隨機世界中, 真相常難明, 既無法證明為真, 也無法證明為偽。球隊比賽前, 常以投擲銅板來決定先後, 但銅板公正嗎? 恐怕不論投擲多少次, 皆不能判定出現正面的機率, 究竟是不是 0.5。地球壽命有多長? 有天文學家估計地球已存在 46 億年, 甚至亦有人估計地球壽命剩下 75.9 億年。這些值到底有多可靠? 真的都只有天曉得。科學家遂慣常以"證實"取代"證明"。媒體上經常刊登科學上的各種證實。如

> 德國科學家證實, 喝綠茶可以減肥;
> 台大跨國團體證實, 大禹治水、建夏王朝是真的;
> 科學家證實, 曾瀕臨絕種的座頭鯨, 數量已逐漸回升;
> 一項研究證實, 喝咖啡後 24 小時內的記憶力, 會大幅提高。

除了上述這類讓人半信半疑的證實外, 亦有

> 科學家證實, 念佛持咒有神奇的力量;
> 科學家證實, 南極臭氧層破洞開始在治癒,

這類讓人無從想像, 或匪夷所思的證實 。只能當做是一個個的假設, 有人接受, 有人不接受 。生活裡便是充滿著無法辨明真偽的假設, 就看那些你能接受, 那些你不能接受, 或者說要拒絕的 。只是接受的依據為何? 除憑藉主觀外, 能否設計出一套公平合理的機制?

在“愛麗絲鏡中奇遇”一書的第七章中, 獅子對國王與獨角獸說“蛋糕要分得公平”(fair play with the cake) 。做決策公平合理是必要的 。兩人分一個蛋糕, 如何讓雙方都覺得不吃虧? 這是著名的分蛋糕問題 。有一簡單的解法, 就是由一人分, 然後另一人先挑 。當然你可以抬槓, 說負責分的那位, 會謹慎地分, 不讓有某一塊較大, 而因他後拿, 因此所拿到的, 乃他認為的一半; 但先挑的那位, 會以為他在兩塊中, 選了比較大, 或至少一樣大的一塊, 因此他拿到的, 乃他認為會大於或等於一半的一塊 。但蛋糕明明就是一個, 不會兩人各自拿到的, 加起來超過一個 。無論如何, 這種分法, 使兩人都認為自己得到的, 沒有比較小, 應能說是一種公平的方法 。

被宣判死刑者, 一旦執法, 就起手無回, 因此不可不慎 。唐宋八大家之一的歐陽修 (1007–1072), 在追念其父母的“瀧岡阡表”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”。對已被判死刑者, 歐陽修之父, 會再仔細檢視案子, 看是否有對該死囚有利的證據, 之前卻被忽視者 。若真的找不到才作罷, 判決定讞 。這時死囚也能理解歐陽修之父, 已盡力為他找一線生機卻不可得, 於是“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學裡所重視之“ 無罪推定”的精神:

> 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 。

這是我國刑事訴訟法第 154 條, 所謂“無罪推定原則”。被檢察官起訴的被告, 即使眾人皆曰可殺, 法庭上仍先假設被告是無辜的 。在

無罪的前題下, 若證據夠充分, 方能判定有罪。反之, 法官只要認定檢察官起訴的內容, 罪證不足, 即可宣判無罪, 並不必去調查。必須一提的是, 前述條文裡的"證明"一詞, 與數學裡的證明意義並不同, 而與"證實"較同義。

在我國刑事案件經檢察官起訴後, 由法院進行審理。若最終被判無罪, 檢察官可說是灰頭土臉。因不但白忙一場, 長期下來, 若定罪率若太低, 檢察官難免被認為不嚴謹、濫行起訴。因此對起訴的案子, 總是檢察官認為已蒐集了相當夠的證據, 有信心可使被告被定罪。但法官並不買帳, 仍假設被告無罪, 逐一檢視檢察官提供的證據, 是否夠強。法官倒也不是懷疑檢察官存心整被告, 隨便弄些資料來充數。而是只有站在被告立場, 從對其有利的觀點來審理, 才能避免冤屈。否則一旦定罪執行刑罰後, 若日後發現其實是誤判, 即使有冤獄賠償, 但人生不能倒帶, 當事人可能牢也坐了, 名譽也毀了, 甚至還家破人亡, 這些都不是金錢所能彌補的。

對於被起訴者, 檢察官的目的, 毫無疑義, 是要定他的罪。但審判的機制, 是先假設被告無罪。科學家若相信喝綠茶可以減肥, 想做個檢定, 該先假設什麼呢? 有如無罪推定, 也應是先假設喝綠茶不能減肥, 然後若有夠高比率的人, 喝後變瘦了, 便能推翻原假設, 即證實喝綠茶可以減肥。而為了避免干擾因子, 並非隨意找些人來進行實驗, 實驗過程也有若干規範, 這屬於統計學裡實驗設計 (experimental design) 的範疇, 在此不擬討論。

為了檢定某假設能否接受, 統計學裡, 設計了一套合理的"假設檢定"(hypothesis testing) 之流程。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, 或傾向推翻 (或說拒絕) 的情況; 至於對立假設, 則通常表現況外之一可能性, 且傾向接受的。雖想推翻虛無假設, 卻儘量保護它, 不讓它輕易被推翻。如此一旦推翻, 才有說服力。所

以, 當虛無假設被拒絕時, 乃有相當的信心該假設不成立 。反之, 若虛無假設被接受時, 通常並不表就相信它為真了, 而是表證據尚不足以推翻它, 就繼續觀察 。那些被法院宣判無罪開釋者, 常高興地說 "司法還我清白" 。其實司法並未宣佈他清白, 司法只是沒判他有罪而已 。假設他明明有罪, 在被判無罪後, 若自此改邪歸正, 那便還好; 若心存僥倖, 以為本領高明, 能騙過司法, 之後仍繼續幹些不法勾當, 則夜路走多, 總有出大紕漏, 因而被定罪的一日 。

想要推翻, 卻極力保護, 與人們常講的 "我不同意你的觀點, 但我誓死捍衛你說話的權利", 其邏輯是類似的 。事實上, 不輕易將現況推翻, 乃較具科學的精神 。朝令夕改並不該被鼓勵 。惟有秉持朝令不輕易夕改的精神, 在訂定各種規章辦法時, 才會更謹慎、更斟酌, 因知一旦通過後, 想更改就很困難了 。

附帶一提, 屢有人好奇, 統計與機率的差別何在? 現在剛好可以來說明 。先看第一個情況 。假設有某公正銅板, 獨立地投擲 20 次, 想求出現 17 個正面的機率 。眾所皆知, 不必管是否真有一個公正銅板, 也不必實際去投擲, 就能求出所要的機率 。此為機率問題, 即在給定的前提下, 去推導結果 。不論給什麼前提, 接受就是, 不必多問 。再看第二個情況 。有人原本以為某銅板為公正, 結果獨立地投擲 20 次後, 出現 17 個正面, 由於正面數過多 (比期望值 10 高出 3 個多的標準差), 遂懷疑該銅板並非公正, 出現正面的機率可能大於 0.5 。既然有此懷疑, 想要釐清, 就去執行一假設檢定 。這便是統計問題, 由觀測到的結果, 來檢驗原本的前提是否可接受 。

對隨機現象做決策, 不誤判乃幾乎不可能 。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 $p = 0.5$, 對立假設取為 $p > 0.5$ 。只是即使銅板實際出現正面的機率為 0.55, 比 0.5 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, 但誤判就發生了 。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。就是因不相信銅板為公正, 才去進行檢定, 大費周章後, 卻仍接受銅板為公正, 可說白忙一場。英文 null 之意義為空的。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, 若宣佈某食品的成份, 經檢驗後合格, 這些都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了, 破壞店家商譽。

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 (Type I error); 其二是對立假設為真卻拒絕, 稱此為第二型錯誤 (Type II error)。以法院審判為例。在無罪推定之原則下, 以 H_0 表被起訴者無罪, H_a 表被起訴者有罪。若被起訴者明明無罪, 卻被判有罪, 便犯了第一型錯誤; 而若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。因無罪被判有罪的後果, 可能得坐牢、受處罰, 或至少名譽受損, 這種錯誤較難彌補。一般會先設定一第一型錯誤的機率之上限 α, 稱為顯著水準 (significance level), α 該是一較小的值, 常取成 0.05, 0.01, 或 0.001 等, 也可以是其他值。在 α 給定後, 決定何時拒絕 H_0, 即決定拒絕域 (rejection region)。拒絕域的選取, 若能使第二型錯誤的機率 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。不過, 往往憑直觀給出的拒絕域, 並不會太壞, 甚至可能就是最佳拒絕域。

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, 當 α 愈小, β 將愈大, 無法兩全。仍以法庭審判為例。若悲天憫人, 寧可錯放 1 千不願錯罰 1 人 (想讓 α 儘量小), 對證據的審核高度嚴格, 則很多實際有罪者, 將連自己都難以置信地被判無罪釋放了 (β 將很大)。所以 α 取得過小, 不見得是好事。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 $H_0 : p = 0.5$, $H_a : p \neq 0.5$ 。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 (以 $\mathcal{B}(n, p)$ 表之) 。當 H_0 為真, 即 $p = 0.5$, 則 X 較可能落在期望值 $n/2$ 的附近 。所以直觀上, 當 X 較偏離 $n/2$, 便該拒絕 H_0 。於是取拒絕域為

$$\left\{\left|X - \frac{n}{2}\right| \geq c\right\} = \left\{X \geq \frac{n}{2} + c, \text{或} X \leq \frac{n}{2} - c\right\},$$

其中 c 將由 n 及 α 來決定 。至於當 X 落在上述集合的餘集, 即若

$$\left\{\frac{n}{2} - c + 1 \leq X \leq \frac{n}{2} + c - 1\right\},$$

便接受 H_0 。

現取 $n = 100$, $\alpha = 0.05$ 。由於 n 較大時, 二項分佈的機率值並不好計算, 故以常態分佈來近似, 而 c 須為整數, 得 c 為 11 。如此

$$\text{拒絕域} = \{X \geq 61, \text{或} X \leq 39\},$$

此時

$$\alpha = P(X \geq 61, \text{或} X \leq 39),$$

在 H_0 下, 即當 $p = 0.5$, 其值約為 0.0358, 比 0.05 小 。對離散型分佈, 有時無法取到剛好能達到所給 α 之拒絕域, 但實際的 α 較小些並無妨 。對上述拒絕域, 第二型錯誤的機率

$$\beta = P(40 \leq X \leq 60),$$

乃與 p 有關 。以 $p = 0.6$ 為例, 且仍以常態分佈來近似, 則得 β 約為 0.5446, 可說是一相當大的誤差 。若縮小 α, 如取 $\alpha = 0.01$, 則得 $c = 14$, 即此時

$$\text{拒絕域} = \{X \geq 64, \text{或} X \leq 36\},$$

且實際的 α 約為 0.007, 再度較原先給定的 0.01 小。上述拒絕域比 $\alpha = 0.05$ 時的小, 這是合理的。因在相同的樣本數 n 之下, α 值取得愈小, 表 H_0 愈被保護, 於是拒絕域將愈小, 即愈不容易拒絕 H_0。那此時第二型錯誤的機率是否該變大? 因之前曾說過, 當 α 減小, β 將增大。仍以 $p = 0.6$ 為例, 可求出

$$\beta = P(37 \leq X \leq 63),$$

其值約為 0.7625, 的確較 $\alpha = 0.05$ 時的 β 大。

另外, 當 $\alpha = 0.01$, 於投擲銅板 100 次後, 若得到 63 個正面, 比在 H_0(銅板為公正) 下的期望值 50 多了 13, 即超過 26%, 感覺上很偏差, 但由於 63 並未落在拒絕域, 因此仍得接受此銅板為公正。沒辦法, 那是 α 取得太小的關係, 如果 α 取大一點, 如 $\alpha = 0.05$, 則同樣得到 63 個正面時, 就要拒絕銅板為公正了。

在相同的 α 下, 如何才能避免得到的正面數看起來很偏差, 卻不能拒絕銅板為公正? 解決之道是加大 n。假設取 $n = 10,000$, 則在 H_0 下, X 之期望值 $= 5{,}000$。則當 $\alpha = 0.01$ 時, 將得 $c = 130$, 因而

$$\text{拒絕域} = \{X \geq 5{,}130, \text{或} X \leq 4{,}870\},$$

此時正面數 X, 只要比 5,000 偏離逾 $130/5{,}000 = 2.6\%$, 就得拒絕 H_0 了。又當 $p = 0.6$, 此時 β 將小到差不多是 0 了。這表樣本數夠大後, $p = 0.5$ 與 $p = 0.6$ 將很容易區隔。即實際 $p = 0.6$ 時, 觀測到的正面數, 將有較大的機率偏離5,000, 使幾乎不會誤判為 $p = 0.5$, 換句話說, 此時幾乎不會犯第二型錯誤。若取較靠近 0.5 的 p, 如取 $p = 0.52$, 則 β 增大至約為 0.0777, 犯第二型錯誤的機率便比 0 大多了。不過相較之前 $n = 100$, 且 $p = 0.6$ 的情況, β 仍小很多。

我們再介紹 p-值。由於不同的人所取之 α 可能不同, 在實務裡, 於得到一觀測值後, 人們常會給出 p-值 (p-value)。所謂 p-值, 即在

H_0 為真下, 會得到比觀測值, 至少同樣極端之事件的機率值。對前述檢定銅板是否公正之例, 設 $n = 100$, 且觀測到 $X = 63$。所得正面數, 至少偏離期望值 50 有 13 之事件為

$$\{X \geq 63, 或 X \leq 37\},$$

則 p-值便為上述事件當 $p = 0.5$ 時之機率, 約為 0.0124。得到此 p-值後便知, 只要給定的 α 比 0.0124 小, 就不能拒絕 H_0, 而若給定的 α 比 0.0124 大, 就得拒絕 H_0。

曾有某公司的劣油案, 一審被判無罪, 引起輿論嘩然, 有些對判決不滿意的人, 遂提出"法律上無罪推定原則, 不該適用食品和藥物等受管制行業"的建議。只是我們已反覆說明, 無罪推定原則, 乃是在現實社會裡, 於做決策時, 一較合理的原則。同理, 在假設檢定裡, 現況仍是該被優先保護的。

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真的發生了, 不過稀鬆平常, 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? 須視情況而定, 不能一概而論。所以得先訂個標準, 亦即先給定顯著水準, 依此以決定拒絕域。若觀測值落在拒絕域, 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。

5 假設檢定之進一步探討

49 取 6 的樂透彩, 頭獎要 6 碼全吻合, 不計順序。因此中頭獎的機率為

$$\frac{1}{\binom{49}{6}} = \frac{1}{13{,}983{,}816},$$

將近 1 千 4 百萬分之 1 的機率, 可說微乎其微。A 君之前從不買彩券, 某日心血來潮買了 1 張, 居然就中了頭獎。A 君的朋友 B 君, 每期都買好幾張, 卻最多只中些小獎。他半開玩笑地對 A 君說, "你一定是買通樂透彩公司的員工, 或能算牌之類的, 否則豈可能一買就中? "A 君當然否認。兩人都學過統計, B 君說 "我們來做一假設檢定。"A 君欣然同意, 但強調 "虛無假設須取成我是清白的。"B 君說 "那當然。" 虛無假設是被保護的, A 君遂安心地等著看 B 君證實他清白。現令

$$H_0 : \text{A 君沒作弊},$$
$$H_a : \text{A 君作弊},$$

分別表虛無假設及對立假設。此處作弊的意思很廣, 有預測能力、求來明牌, 或任何作假等都算。至於拒絕域要取成什麼? 只要是合理的拒絕域, 便該包含 A 君中頭獎, 這自然是一有可能觀測到的結果。如今在 H_0 為真下, 觀測到 A 君中頭獎, 此機率即

$$p\text{-值} = \frac{1}{13{,}983{,}816},$$

B 君認為實務上極少有這麼小的顯著水準 α。所以在任一合理的 α 之下, 皆該拒絕 H_0, 而接受 H_a, 也就是接受 A 君作弊。怎會這樣? A 君愣住了。

A 君倒也不必覺得太委曲, 因這不過是一正常假設檢定的結果, 類似的情境不少。被認為是現代統計學的創始者費雪 (Sir Ronald Aylmer Fisher, 1890−1962, 他是位爵士), 在他的 "實驗設計"(The Design of Experiments, 1935) 一書中, 曾講了一 "淑女品茶"(the lady tasting tea) 的故事, 我們稍加修改如下。一日無事小神仙, 某日一群朋友, 正悠閒地享用下午茶。其中有位男士, 大談他喝遍天下好奶茶之經驗。講個不停, C 女士聽了有些不耐, 忍不住說, 奶茶

的調製順序, 深深影響其風味, 先加奶後加茶, 與先加茶後加奶, 喝起來完全不同。眾人被吸引過來了, 只是也對 C 女士的說法, 感到存疑, 成分都一樣, 豈能喝出差別? 有位統計學家說, 那就來檢定一下。顯然 H_0 與 H_a, 可分別取成:

$$H_0: \text{C 女士無法分辨奶茶是先放奶或先放茶,}$$
$$H_a: \text{C 女士能分辨奶茶是先放奶或先放茶。}$$

請餐廳製作 20 杯奶茶, 有些先放奶, 有些先放茶, 所有賓客皆不知順序, 然後讓 C 女士依序飲用並辨別, 最後由餐廳宣佈答案。令人驚奇的是, C 女士全部講對。不論拒絕域取成什麼, 顯然都該包含 20 杯全對。因而

$$p\text{-值} = \left(\frac{1}{2}\right)^{20} = \frac{1}{1{,}048{,}576},$$

即使遠大於 1/13,983,816, 但 H_0 想必會被拒絕。

樂透彩是公開開獎, 就算得到前述 A 君作弊的推論, 大部分的人, 恐怕依舊相信 A 君就只是好運而已, 而未作弊。否則用相同的方法, 將得到每位樂透彩的頭獎得主, 都作弊之推論。甚至, 那些員工有幾千人以上的大公司, 歲末尾牙餐會裡常有的摸彩, 中頭獎者, 豈不也都是靠作弊? 處處皆作弊? 這世界固然沒那麼美好, 倒還不至於充滿著作弊。但若不相信有弊端, 難道是假設檢定的推論不值得採信? 另一方面, C 女士品茶的能力, 的確令人嘖嘖稱奇, 但即使通過檢定, 她是否真具備分辨先放奶或先放茶的特異功能, 恐怕眾人仍莫衷一是。有人可能相信了, 有人可能只視為魔術, 就算 100 杯都講對, 還是一點都不信。究竟這是怎麼一回事?

我們先對假設檢定裡的顯著性, 再做些闡釋。某公司研發出一種新飲料, 市面上飲料品牌眾多, 如何才能脫穎而出? 該公司遂設計出一套實驗流程, 找了 25 所中學進行實驗, 每所中學的飲料配方

略微有些不同。各校均執行一檢定, 取

$$H_0: \text{喝此飲料無法提高記憶力},$$
$$H_a: \text{喝此飲料能提高記憶力}。$$

中學生升學壓力大, 飲料若對提高記憶力有幫助, 就有賣點。在 $\alpha = 0.05$ 下, 其中有一所學校得到顯著的結果。於是該公司以此校的結果, 完成一份研究報告, 宣稱經嚴格的統計檢定, 證實喝該配方的飲料, 能提高記憶力, 且大力促銷。一瓶才 20 元左右的飲料, 會有那麼驚人的功能? 不少人嗤之以鼻。但報告看起來, 卻不像有造假。

其實只要想想設定 $\alpha = 0.05$ 的涵義, 就知上述飲料提高記憶力之檢定, 根本不必造假。在虛無假設為真下, 會犯第一型錯誤的機率為 0.05。也就是當實際上, 喝該飲料無助提高記憶力, 觀測值仍有 0.05 的機率, 會落在拒絕域, 因而會接受 H_a。也就是每 20 所參與實驗的中學, 平均會有 1 所"證實"喝該飲料能提高記憶力。如今共有 25 所中學參與實驗, 若其中有 1 所得到顯著的結果, 不過合理而已。你現在知道了, 拿 1 公正銅板, 讓 20 人依序各投擲 100 次去檢定, 雖是相同的銅板, 在 $\alpha = 0.05$ 之下, 若有人得到銅板不公正的推論, 並不足為奇。甚至對一般的檢定, 不論 α 取得多小, 只要檢定做得夠多回, 便不難在其中發現有幾回的結果為顯著。由於有上述這種現象, 科學上的觀測, 不宜一得到顯著的結果, 便很興奮, 迫不及待地拿去發表。實驗的結論, 要能重複才行。即不可只有你能得到某新藥對治療某疾病有效, 別人重做實驗都無效。否則學術刊物上, 將經常發表一些光怪陸離的研究報告。

要知不論再小的機率, 只要碰到大樣本, 其發生便都不稀奇。先舉一常見到的例子來說明。每逢過年期間, 台灣各地廟宇, 屢舉行擲筊比賽的活動。擲筊是民間一種求神問卜的儀式, 將兩片用木頭

做成的半月形狀筊杯, 投擲至地面後, 若二筊杯呈現一正一反, 便稱得到聖筊, 代表向神明祈求或請示的事, 獲得應允或認為可行 。在擲筊比賽裡, 獲第一名的, 往往連續擲出十餘個聖筊 。有如投擲 1 公正銅板得正面, 得 1 聖筊之機率亦為 1/2 。不妨試試, 實際拿一銅板投擲, 看能否連得 10 個正面? 恐怕不容易, 因 (假設銅板為公正) 機率僅有

$$\left(\frac{1}{2}\right)^{10} = \frac{1}{1{,}024}\ 。$$

那怎麼可能有人得到 13 個聖筊? 喔! 原來報名者有 1 萬多人, 這就難怪了 。由於

$$\left(\frac{1}{2}\right)^{13} = \frac{1}{8{,}192},$$

故只要報名有 8,192 人以上, 平均便至少能有 1 位連得 13 個聖筊 。媒體上出現的標題, 常是醒目的奪冠者擲出多少個聖筊, 讓人驚嘆萬分 。其實若知道該活動的報名人數, 大約就不會訝異了 。諸如公司尾牙餐會的摸彩, 或樂透彩開獎等, “有人” 中頭獎, 都屬上述小機率碰到大樣本之例 。對某位 “特定” 的人, 要中頭獎的確很難 。但要 “有人” 中頭獎, 則往往不算難事, 有時還成為必然會發生之事 。事實上, 任一期只要銷售量夠大, 有人中頭獎並沒什麼, 此處不過剛好就是 A 君而已 。但若下一期 A 君又只買 1 張彩券, 且又中頭獎, 那就真會令人起疑, 連樂透彩公司, 可能都會請警方進行調查 。

一般而言, 對很多事件, 人們心中各有不同的事先信賴程度, 或者事前機率 (prior probability) 。對堅定相信樂透彩頭獎號碼為隨機產生的人, 即使 A 君在彩券開獎前, 自信地說, “我拜過菩薩了, 菩薩應許我, 這張必中頭獎”, 而開獎後 A 君果然美夢成真, 恐怕仍不認為 A 君對菩薩的祈求有效 (或 A 君有預知能力, 或 A 君作弊

等)。而且不少買彩券者, 每次都自以為會獲幸運之神的眷顧, 因此各個人那些喃喃自語的信心, 向來不太會有人在乎。但如果 A 君中頭獎之事二度發生, 原本堅定相信他怎麼拜都沒用的人, 信心可能便開始動搖了; 若三度發生, 將有如三人成虎, 或視 A 君如神, 或不再相信樂透彩的公正性。但是, 除非出現可靠的佐證, 否則光由 A 君接二, 甚至連三的中頭獎, 雖大部分的人, 會懷疑其中必有名堂, 不可能僅是憑運氣, 卻無法這樣就認定 A 君一定作弊, 須有其他佐證才行。簡單講, 假設檢定接受某一假設, 與該假設確實為真, 完全是兩回事。底下給一實例。

2015-16 NBA 的季後賽, 可說高潮迭起。之前已指出, 季後賽都是採 7 戰 4 勝制。先是在西區冠軍賽, 打完 4 場後, 雷霆隊以 3 勝 1 敗, 領先勇士隊聽牌。在這一年的例行賽, 創下許多歷史紀錄的勇士隊, 眼看就要被淘汰了。因西區冠軍賽, 之前從未有從 1 比 3 落後, 而能奪冠之例。雷霆隊的球迷, 已等不及要慶祝了。不料韌性十足的勇士隊, 彷彿覺得這一年創的紀錄尚不夠多, 他們連勝 3 場, 登上西區冠軍寶座。原本心情有如洗三溫暖的支持者狂喜, 且信心高漲。因動心忍性, 曾益其所不能, 勇士隊經此鍛鍊, 還有打不敗的球隊嗎? 勇士隊過關後, 接著要與已等候在那裡的東區冠軍騎士隊, 進行殊死戰, 以決定今日之域中, 竟是誰家天下?

總冠軍賽打完 4 場, 勇士隊取得 3 比 1 領先, 看來勝券在握了。因 NBA 總冠軍賽史上, 3 比 1 領先卻被逆轉, 之前從未發生過。何況, 在西區冠軍賽, 勇士隊已完成一次不可能的任務, 球迷又不是前面所提 “愛麗絲鏡中奇遇” 書中的王后, 在同一季後賽裡, 豈會接連看到兩件類似之不可能的事? 要相信機率! 結果騎士隊壓根兒不理機率, 硬是連贏 3 場封王。勇士隊在前一輪, 才成為西區冠軍賽史上, 從 1 比 3 落後下逆轉的首例, 下一輪便成為總冠軍賽史上, 第一支從 3 比 1 領先, 卻被逆轉的苦主, 只能徒呼負負。但對上

天實在也無從抱怨起, 因大幸與大不幸, 都發生在勇士隊身上。看來愛麗絲現在會相信了, 早餐前的確可有 6 件不可能的事發生。

見到連續兩回這麼峰迴路轉的比賽, NBA 的球迷, 興奮之餘, 會懷疑是為了票房或其他原因, NBA 暗中搞鬼所造成的嗎? 猜想大部分的人不會。要知假設檢定雖是現代做決策之一重要依據, 但我們已數度強調, 見到顯著事件發生, 是該睜大雙眼, 仔細檢視。卻不表所懷疑的事, 果真就該被推翻。甚至, 若盲目依賴統計, 過度重視顯著事件, 認為由看到這個 “果”, 必然就是那個 “因” 所造成, 有時會產生 “檢察官的謬誤”(prosecutor's fallacy)。

設有某法院審理某刑事案件。在無罪推定的前提下, 虛無假設自然取為 H_0: 被告無罪。假若檢查官提出不少疑點, 且找某學者算出, 如果 H_0 為真, 這些疑點會發生之機率, 只有百萬分之 1。這的確是一很小的 p-值, 看起來對被告很不利。但被告的律師, 若有些概念, 所該提出的反駁, 乃是指出應在意的, 並非僅有此機率, 而是當這些疑點發生時, H_0 為真之機率。大家現在大概都知道了, 此其實是另一條件機率。這有時得要有更多的資訊才能計算, 與那百萬分之 1 的機率, 完全是兩回事。二條件機率值可能差很多, 不妨以第 3 節那一有關測謊的例子來說明。首先, 依假設

$$P(\text{顯示洩密}|\text{實際未洩密}) = 0.1\text{。}$$

另一方面, 利用已求出的 $P(\text{實際洩密}|\text{顯示洩密}) = 1/12$, 得

$$\begin{aligned} P(\text{實際未洩密}|\text{顯示洩密}) &= 1 - P(\text{實際洩密}|\text{顯示洩密}) \\ &= 1 - \frac{1}{12} = \frac{11}{12} > 0.91\text{。} \end{aligned}$$

0.1 與 0.91, 差異的確不小。只是在法庭上, 屢會將二者混淆, 有時冤獄便這樣產生的, 因此才會常被稱為檢察官的謬誤。由於受不了統計有時成了冤獄的幫凶, 還有人提出 “靠統計數字定罪的

危險” 之警告。底下以著名的 “莎莉克拉克案”(The Sally Clark case), 來說明法庭若要引用統計來當證據, 須格外謹慎。資料取自 Wikipedia(維基百科)。

莎莉克拉克 (Sally Clark, 1964–2007, 本名 Sally Lockyer, Clark 為夫姓), 是家中的獨生女, 她父親是位資深警官, 母親是美容師。1990 年莎莉結婚, 夫妻都是初級律師 (solicitor)。家庭及事業, 看起來都很美好。1996 年 9 月, 莎莉生了個男孩。一個健康的嬰兒, 卻在當年 12 月, 11 週大時在家中猝死 (Sudden cardiac death, 縮寫 SCD, 指突然的死亡)。莎莉好不容易才從悲傷中復原, 於 1997 年 11 月, 再度生了一個兒子。黑暗已去, 可開始過新生活了。豈料 8 週後, 1998 年 1 月, 愛子又在家中猝死, 兩次都只有莎莉一人在家。喪子顯然還不是最大的痛, 莎莉面臨殺嬰的控訴。

檢察官憑其直覺, 完全不相信嬰兒猝死症 (Sudden infant death syndrome, 縮寫 SIDS, 指嬰兒突然死亡, 不論從其病史、身體檢查, 或研究調查, 都無法發現死因), 就是莎莉孩子死亡的真正原因。雖沒有莎莉行凶的直接證據, 但他就是認為, 接連有兩個嬰兒猝死, 極不尋常。為了說服陪審團, 這絕非猝死, 檢察官找來梅鐸 (Sir Roy Meadow, 1933–, 他亦是位爵士) 作證。梅鐸是位夙負盛名的小兒科醫生 (paediatrician), 且上法庭作證的經驗豐富。

梅鐸不愧是行家, 他以簡易的方式, 向陪審團說明, 一家有兩個嬰兒接連猝死的機率有多小。他說同一家庭有兩個小孩死於 SIDS 的機率, 為 7,300 萬分之 1。7,300 萬分之 1 的機率, 並不表示事件不會發生, 這點梅鐸倒是承認。但他指出, 這種意外, 每 1 百年才會有 1 次。搞不清楚 7,300 萬分之 1, 到底有多小的人, 聽到 1 百年 1 次, 立即就懂了。人生不滿百, 百年 1 次的事, 怎會見得到? 顯然不是猝死! 梅鐸還說, 英國有兩個小孩的家庭, 差不多有 1,500 萬個。對照 7,300 萬分之 1 的機率, 一聽又更明白了。不會發生! 既

然不可能是猝死, 結論就浮出了。

我們先來看梅鐸的數據之由來。梅鐸宣稱, 對一如克拉克這種富裕且不抽煙的家庭, 會發生一件嬰兒猝死 (cot death) 的機率為 1/8,543。因此會發生兩件嬰兒猝死的機率, 為前一值的平方, 即

$$\left(\frac{1}{8{,}543}\right)^2 = \frac{1}{72{,}982{,}849},$$

7,300 萬分之 1 的機率, 便是這樣產生的。那百年 1 次又是如何來的? 梅鐸說, 全英國每年約有 70 萬新生兒, 他將 7,300 除以 70, 得到約 104, 近似 100。

陪審團接受了梅鐸的證詞。1999 年, 莎莉被判無期徒刑 (life imprisonment), 並於 2000 年入獄。直到 2003 年 1 月, 經第二次上訴, 基於新出爐的死嬰之病理報告, 最高法院改判莎莉無罪。只是遲來的正義, 對莎莉的幫助已不大了。出獄後, 莎莉一直處於精神不佳的狀態, 有如槁木死灰, 終日酗酒, 茫然度日。2007 年 3 月, 她因酒精中毒, 死於家中。

陪審團及檢察官的數學及統計, 可能都不會太好, 那梅鐸的功力如何呢? 被封為爵士, 總不至於浪得虛名吧! 那 1/8,543 嬰兒猝死的機率, 我們就先接受好了。但將兩個 1/8,543 相乘, 就毫無根據了。同一家庭的兩個嬰兒猝死, 是獨立事件嗎? 這一對未曾謀面的兄弟, 由於遺傳之關係, 兩人說不定會有類似的基因缺陷。再加上照顧方式, 及生長環境皆相同等因素, 二猝死事件, 絕不可理所當然地視為獨立。無論如何, 不管三七二十一, 就將兩機率相乘, 是很輕率的。不該是一個受過水準以上機率訓練的人, 所會犯的錯。因此 7,300 萬分之 1 的機率, 應是極被低估的。至於因機率 7,300 萬分之 1, 且每年約 70 萬新生兒, 將二者相除, 就得到同一家庭兩件嬰兒猝死案, 百年才會有一樁, 這更是莫名其妙了。只是既然有這麼多統計上的缺失, 英國眾多統計學者, 難道都不吭聲嗎?

首先, 如我們之前所一再強調的, 這裡犯了檢察官的謬誤。因就算 7,300 萬分之 1 的機率為正確, 也不表在兩個嬰兒猝死下, 莎莉無辜的機率也是 7,300 萬分之 1。這是另一條件機率, 要有更多的資訊, 才能估算。另外, 英國皇家統計學會 (Royal Statistical Society, 縮寫 RSS), 倒也沒有袖手旁觀。事實上, 他們於 2001 年 10 月, 發表一公開的聲明, 對本案裡的 “法庭誤用統計”(misuse of statistics in the courts), 表示關切。並說 “7,300 萬分之 1 的機率毫無統計根據”(no statistical basis for the 1 in 73 million)。2002 年 1 月, RSS 還寫信給上議院大法官 (Lord Chancellor), 明確指出 7,300 萬分之 1 的計算是錯的 (the calculation leading to 1 in 73 million is invalid)。

2005 年, 英國醫學總會 (General Medical Council, 縮寫 GMC), 鑑於梅鐸曾多次在法庭上擔任專家證人, 卻提供錯誤資訊, 因而數度入無辜者於罪, 撤銷了他的醫師執照。雖經上訴後, 隔年梅鐸重新拿回其執照, 但名聲已毀了一大半了。

回到 1/8,543 那一機率值。2004 年, 索爾福德大學 (University of Salford) 的數學教授希爾 (Ray Hill), 在期刊 Paediatric and Perinatal Epidemiology 上, 發表一篇論文。他依據英國的統計資料, 推導出嬰兒猝死的機率約 1/1,300, 而非 1/8,543。並且估計出, 一家庭若已有一嬰兒猝死, 則會發生第二個嬰兒猝死的機率, 提高到 5 至 10 倍。看來梅鐸醫生自以為能善用統計, 其實從頭到尾犯了不少錯, 卻一直毫無所覺。統計! 多少人假統計之名!

統計是門入世的學問。統計學裡提供一套假設檢定的程序, 以為做推論時來遵循。此程序並非在判定事情的真偽, 而是用來當做決定對策時之指引。限於篇幅, 本文僅對假設檢定, 給一粗淺的介紹, 想進一步認識此題材者, 可參考一般統計學教科書。另外, 黃文璋 (2016) 一文中, 有一些關於假設檢定之簡單的例子, 可供有興趣

者練習用。了解假設檢定的內涵後, 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。因此該算是統計學裡, 一個較容易接受的主題。

6 結語

從媒體上, 不時獲知警方利用比對 DNA(deoxyribonucleic acid, 縮寫 DNA, 去氧核醣核酸) 來破案。福爾摩斯若復出, 也必嘆為觀止。在柯南道爾 (Sir Arthur Ignatius Conan Doyle, 1859−1930, 又是一位爵士) 的偵探小說裡, 對那些棘手的案子, 福爾摩斯之所以常能比警方快一步, 並不僅是因他善於觀察與邏輯推演, 他還很會利用科學儀器。DNA 比對, 顯然是極其科學的。底下先引用兩則新聞。

2016 年 7 月 5 日, 中國時報有一則標題是"16 年前性侵案 DNA 揪惡狼"之報導。2000 年, 有件發生在台南市的持刀搶劫, 並性侵女子的案子。雖採集到 DNA, 只是當時的比對儀器, 尚不是很精良, 遂讓犯人逍遙法外。不過盡職的警方, 將證據都相當完整地保留下來。今日儀器的精準性, 與往昔自然不可同日而語。刑事警察局經 DNA 比對, 查出嫌犯。當年那位受害女子, 被警方找來指認時, 一眼便確認嫌犯。心靈創傷難忘, 16 年前那段經歷, 彷彿惡夢, 一直如影隨形, 揮之不去。女子憤慨地說"化成灰我都認得他。"全案遂交由台南市的警方重啟調查。這麼多年了, 嫌犯以為早就船過水無痕。當警方舊案重提, 且證據確鑿, 簡直難以置信。DNA 比對, 加上被害人指認, 嫌犯這回將閃躲不掉了。

2016 年 7 月 8 日, 聯合報亦有一則有關 DNA 比對的報導, 標題是"警方如何鎖定爆炸案嫌犯? 根據乘客傷勢"。台北市松山車站, 於 2016 年 7 月 7 日晚間, 發生一起驚人的台鐵車廂爆炸案, 造成

25 人受傷。新聞中說:

> 案發前, 並未發現有可疑人士做出拋投, 衝跑脫離現場的舉動, 那說明炸裂物一直在嫌犯手上。既然手持爆裂物, 則爆炸瞬間, 嫌犯勢必遭波及。因此檢警從昨晚就認定, 嫌犯就在 25 名傷者中。經訪查目擊者、過濾傷者傷勢, 其中 55 歲林姓傷者前胸嚴重燒灼、加上爆炸衝擊力有 35% 鈍挫傷, 顯然是最直接被炸傷的人, 懷疑他涉重嫌。此外, 檢警在車廂廁所查扣的紅色登山包包發現, 有兩個與炸彈相關連的工具, 且採集 DNA 比對, 證實與林嫌相符。

合理推測, 加上 DNA 比對, 迅即鎖定嫌犯。

DNA 比對, 成了今日破案的利器。而既然能讓 16 年前的犯案者, 無所遁形, 則不足為奇, 也當然有平反冤獄的時候。有些人已含冤坐了多年牢, 再也不相信什麼老天有眼。結果拜 DNA 比對之賜, 終於還了清白。由於有關 DNA 比對立功的報導, 一再出現, 使人們不禁以為, 一旦祭出 DNA 比對, 涉案者便莫不俯首認罪。大致是這樣沒錯, 只是在運用時, 如前兩則新聞內容所顯示的, 仍需有其他佐證。即不能光是比對吻合, 就立即以為案子破了, 找到真相。若過度執著於 DNA 比對的高吻合率, 有時難免犯下"檢察官的謬誤"。

2016 年 6 月號的 Scientific American(Volume 314, Issue 6), 有一篇標題為"When DNA Implicates the Innocent"的文章, 作者是 Peter Andrey Smith。由林雅玲譯的"DNA 證據牽連無辜? "一文, 則刊登在 2016 年 7 月號 (No.173) 的"科學人"。在該文中, 先舉一 DNA 比對, 引導錯誤的案例。2012 年 12 月, 美國警方根據 DNA 的比對, 指控一位名叫安德森 (Lukis Anderson) 的遊民, 涉嫌矽谷富豪庫姆拉 (Raveesh Kumra) 的謀殺案。安德森若被判有罪, 最重

是死刑。只是他雖是個遊民, 處在美國社會底層, 與此案卻根本毫不相干, 他有不在場的證明。在 11 月庫姆拉被謀殺的那晚之前, 安德森便因醉酒且近乎昏迷 (drunk and nearly comatose), 一直待在醫院接受治療。但安德森不在場的證明, 真的無懈可擊嗎? 會不會他的昏迷是假的? 藉故躺醫院, 然後趁醫護人員沒注意時, 溜出醫院行凶? 嗯! 像是電影裡的情節。不過, 畢竟天下事無奇不有, 不能說絕對不會發生, 想像力要豐富些。而且, 如果安德森真的無辜, 那現場的 DNA, 又是怎麼來的? 結果安德森的辯護律師團發現, 他的 DNA, 是在事發之後, 隨著醫護人員進入庫姆拉家中。那幾位醫護人員, 在當天稍早曾治療過安德森, 然後在 3 個多小時後, 不經意地把他的 DNA, 帶到命案現場, 並留在那裡, 成為安德森涉案的 “鐵證”。居然有這種事! 有些人感到不可思議。其實倒也不必太驚訝, 因我們已舉出太多不可能卻發生的事件了。

上述文章中指出:

> DNA 分析以統計模型做出預測, 比起其他法醫技術, 更為明確與客觀。(DNA analysis is more definitive and less subjective than other forensic techniques because it is predicated on statistical models.)

一般而言, 利用統計, 當然比其他方法更明確且更客觀, 這是無庸置疑的。但該文要講的重點是:

> DNA 檢驗和其他證據一樣, 透露的僅僅是完整案件的其中一個面向。··· 。如果我們過度依賴這個方法, 以及在檢視 DNA 證據時, 缺乏適度的質疑, 誤判案例早晚會發生。例如, 生物樣本可能分解, 或遭到污染, 而法官和陪審團可能曲解統計機率 (statistical

probability)。甚至, 就像安德森一案所揭露的, 皮膚細胞也可以“移動”。··· 。一個人攜帶的衣物, 就算只是碰到另一人的脖子, 都可能把後者的 DNA, 轉移到此人從未接觸過的物體上。··· 。DNA 的轉移究竟有多常引發錯誤的指控, 目前仍未知。

與 DNA 的比對類似, 統計裡的假設檢定, 此一科學的思維與方法, 的確常能讓人們將事情看得更清晰, 因而有助於提高決策品質。但切記它並非萬無一失。若“過度依賴, 而缺乏適度的質疑”, 則誤判的發生, 便絕非偶然。失之毫釐, 可能產生差之千里的誤差, 這是運用假設檢定來做決策時, 必須銘記在心的。

參考文獻

1. 徐仕美譯 (2016)。記憶的盡頭 (The End of Memory：A Natural History of Aging and Alzheimer's, Jay Ingram 原著)。遠見天下文化出版股份有限公司, 台北市。

2. 黃文璋 (2016)。機率與統計在高中。中國統計學報, **54**(2): 43-61。