# 統計學裡無罪推定的精神

黃文璋

國立高雄大學應用數學系

被認為是現代統計學之創始者英國人費雪(R.A. Fisher), 曾提到下述故事。

在1920年代後期, 某日有位女士對一群正在喝下午茶的科學家宣稱, 奶茶的調製順序對風味有很大的影響, 把茶加進牛奶裡, 和把牛奶加進茶裡, 兩者喝起來完全不同。在座的科學家們當然對這種說法感到可笑, 他們看不出兩種混合方式的化學成分有什麼差異。但費雪卻認真地設計一個實驗步驟來對這件事做檢定, 包括要準備多少杯奶茶, 以及依照什麼順序給這位女士喝等。

民國90年12月20日, Yahoo!奇摩網站上有底下一則新聞報導。

(中央社記者郭傳信安卡拉十九日專電) 土耳其國立安卡拉大學醫學院婦科系教授庫克在專欄中表示, 早在西元前二二００至二０００年, 藥學史極為發達的古埃及人, 已能夠不靠化學藥劑即可檢驗出女性是否懷孕。

庫克說, 根據至今已發現的古埃及紙草文獻記載, 希望知道自己是否懷孕的婦女, 必須將自己清晨即起的尿液裝在一個盛有大麥種子的袋子裡, 但在此同時, 也必須要求另一位確定未懷孕的女性也將清晨即起的尿液裝在另一個有大麥的袋子裡。

庫克表示, 由於女性懷孕後, 體內會較未懷孕女性產生更多的荷爾蒙, 因此泡在懷孕女性尿液中的大麥種子容易發酵並提前發芽, 即可確定懷孕, 但如果兩袋的大麥種子同時發芽, 則證明未懷孕。

庫克最後在文中強調, 現代科學已證實這項古老的女性懷孕檢驗法「相當準確」。

在費雪的故事裡, 如果只拿一杯奶茶讓那位女士喝, 而且她說對先放茶或先放牛奶, 是否會相信她真有分辨能力? 講清楚一點好了, 有分辨能力指她每杯答對的機率大於隨機猜對的機率1/2。可能不會, 因隨機猜可有1/2的機率說對。如果給她兩杯呢? 可有1/4的機率皆說對, 不算太小的機率。如果連續10杯她皆說對呢? 1/1,024的機率算是很小的, 即使仍不太相信她有分辨的能力, 也許暫時不會排除此一可能性。但是如果20次中錯一次呢? 畢竟人難免會犯錯, 有時你叫錯朋友的名字, 但你不會承認是認錯他。那20次中錯兩次呢? 我們對犯錯是有一些忍受的程

度, 但程度究竟多大, 就因人而異, 或因情況而異。附帶一提, 如果事先知道10杯中恰5杯先放牛奶, 5杯先放茶, 則隨機猜能10杯全說對之機率為$1/\binom{10}{5} = 1/252$, 約為1/1,024的四倍。

再看檢驗女性是否懷孕的方法。俗語說"一樹之果有酸甜之別, 一母之子有賢愚之分"。即使同一批大麥, 發芽的時間可能便有快慢。就算倒入一沒有懷孕婦女的尿液, 大麥可能也會提前發芽。是否要如前述女士喝奶茶的情況, 與10位未懷孕的女性相比? "比賽"結果, 如果是倒入該婦女尿液的大麥, 以6比4提前發芽, 那算不算真的領先? 要知$A, B$兩支球隊, 即使勢均力敵(即每場$A, B$獲勝之機率均為$1/2$), 連比十場, $A$隊領先(即至少贏6場)的機率為

$$\frac{\binom{10}{6} + \binom{10}{7} + \binom{10}{8} + \binom{10}{9} + \binom{10}{10}}{2^{10}} = \frac{386}{1{,}024} \doteq 0.377,$$

可說很容易發生。$B$隊領先的機率也一樣約為0.377。至於平手, 即各勝5場的機率則約為$1 - 2 \times 0.377 = 0.246$, 反而較小。而且怎樣算"提前"? 1秒鐘? 1分鐘? 大約少有人認為這麼小的差異, 可算做提前。庫克說現代科學已證實此法相當準確。不過古埃及人是如何操作此法, 以得到可靠的推論, 就不得而知了。

在數學裡, 一個命題, 如直角三角形兩股長之平方和等於斜邊長之平方, 三角形三邊的高相交於同一點等, 一旦被證明是對的, 就毫無疑問地成立。有時即使尚未能證實結論, 數學家知道那只是時機還沒成熟, 有如開門的鑰匙仍在找尋中。如著名的費馬最後定理:

當$n \geq 3$為一整數, 且$x, y, z$皆不為0, 則$x^n + y^n = z^n$無整數的$x, y, z$解。

此"定理"從被提出至西元1994年終於被證出, 歷經三百餘年。今後就不用想是否會運氣好, 找到一組解。在數學裡$0.6 > 0.5$, 但若銅板$M$出現正面的機率為0.6, 銅板$N$出現正面的機率為0.5, $0.6 > 0.5$, 那是否各丟10次, 各會出現6個及5個正面呢? 你知道不一定。又是否銅板$M$出現的正面數一定較銅板$N$多呢? 經驗仍告訴我們不一定。對於隨機現象往往不能有如數學中斬釘截鐵的推論。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

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

此命題的假設為$A$, 想得到的結論為$B$。但在隨機世界裡, 一件事往往很難判定真偽。到底該女士是否有分辨奶茶裡是先放牛奶還是先放茶的能力, 即使她20次皆說

對(隨機猜猜中之機率等於$1/2^{20}$約為百萬分之一), 恐怕還是有人不信她有分辨能力。因此我們不會說

試證某女士“有”分辨奶茶是先放牛奶還是先放茶的能力,

或是

試證某女士“無”分辨$\cdots$。

數學家因相信在假設$A$下, $x^n+y^n=z^n$無整數解是對的, 於是去證明它。但對奶茶那一問題, 由於該女士宣稱(也希望人家相信)她有分辨能力。因此研究人員先假設該女士無分辨能力(因此只是隨機地猜)。然後譬如說拿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之機率$\alpha$, 如0.05, 0.01, 或0.001等。再看於前述假設下(即每杯猜對之機率皆為二分之一), 會講對次數達到這麼多次的機率有多大。如果機率小於$\alpha$(也就是能有這麼多次講對是較不尋常的), 則得到結論“拒絕”原假設(即判定該女士並非無分辨能力), 否則便說“接受”原假設。

對一隨機現象, 研究人員都是先提出一**猜想**, 再將**猜想**表為**統計假設**(簡稱**假設**)的形式。而導致接受或拒絕一統計假設的步驟, 就是**統計推論**之主要工作。

統計假設與一般數學中的假設是不同的, 在數學裡我們常有下述這類敘述:

假設$x>y$。

由於並未涉及任何隨機的量, 所以此非統計假設。但如令$\mu$表北銀樂透彩頭獎號碼中, 1號出現的機率, 則

$$\mu>1/7$$

就是一統計假設了。

由於一統計假設是否為真, 常是無法確定的。所以通常的作法是, 取一組隨機樣本, 並利用此組樣本, 當做是否支持某一假設之證據。如果證據與假設所陳述的不吻合, 或更正式地講吻合的機率很低, 便拒絕該假設, 否則便接受該假設。

我們常說“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。所以在設計決策步驟時, 要考慮推論錯誤的機率。在無法避免犯錯的情況下, 只能以較好的方法儘量減小犯錯的機率, 否則所做的統計推論便不易被採信。

在數學裡, 對於一命題, 有真或偽兩種結論。但在統計學裡, 我們不說一假設成立或不成立。而是說**接受**一假設, 或**拒絕**一假設。要注意的是, 當我們拒絕一假設, 並不表該假設為**不可能**(impossible, 英文解釋有not capable of existing or happening, unacceptable等)發生, 而是表該假設“不像會發生”, “似不可信的”(improbable, 英文解釋有unlikely to take place or to be true, doubtful)。接受一假設也並不表認為該假設必定成立。

<u>**註1**</u> 口語裡的"不可能", 有時並非眞的表不可能, 而是指發生的可能性極低。在不可能的任務(Mission Impossible)那部電影裡, 主角湯姆克魯斯還是將任務完成了。

有了一統計假設, 下一步就是要去**檢定**此假設是否要接受或拒絕。這整個過程稱為**假設檢定**。假設檢定的理論及架構是波蘭人**奈曼**(J. Neyman)及英國人**皮爾生**(E.S. Pearson), 在西元1933年, 給出著名的**奈曼-皮爾生引理**所奠定的。在奈曼-皮爾生的架構裡, 有一**虛無假設**(null hypothesis), 及一**對立假設**(alternative hypothesis)。虛無假設通常表現況, 而對立假設表我們傾向相信的。例如, 對北銀樂透彩頭獎號碼中1號出現的機率$\mu$是否大於1/7的問題, 若研究人員認為答案是肯定的, 則會將虛無假設取為"$\mu = 1/7$", 而對立假設當然就是取為"$\mu > 1/7$"。在奶茶問題, 若以$\mu$表每杯奶茶答對的機率, 則將虛無假設取為"$\mu = 1/2$", 而對立假設取為"$\mu > 1/2$"。

<u>**註2**</u> 奈曼出生於俄國, 後來移民到美國, 在加州大學柏克萊分校任教。近悅遠來, 將該校統計系, 建立成一世界著名的統計研究重鎮。

虛無假設是被保護的, 除非證據夠強, 否則不輕易推翻。但這是有道理的, 如果宣佈樂透彩1號出現的機率大於1/7, 可能將引起不小的震憾, 之後即使做了更正, 宣佈實驗有誤, 已造成的損失將難以彌補。所以先相信虛無假設, 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。在虛無假設下, 若會出現這種或比這種還極端的數據之機率(此機率稱為***p*-值**)很小, 便認為不合理。有時我們會給出**拒絕域**(rejection area), 也就是所獲數據如果落在拒絕域便拒絕虛無假設。對於現況不輕易推翻, 會使人們在做決策(如宣佈某產品之規格, 制定某項辦法等)時更謹愼。因一旦宣佈後, 便很難被更改, 如此會使大家下決定前, 能更考慮周全。古人批評"朝令夕改", 今人說"朝令有錯, 夕改何妨?"想想古人還眞有智慧。

某公司宣稱其產品之不良率為$p = 0.1$, 某消費者協會懷疑不良率超過0.1, 想做檢定。先相信公司的宣稱, 此處的虛無假設就是"$p = 0.1$", 對立假設則為"$p > 0.1$"。隨機地取100個樣本來檢驗, 且發現其中有10個不良品, 則(不經計算)直觀上很可能會接受虛無假設。如果將虛無假設改為"$p = 0.095$", 且仍得到10個不良品, 則此假設大約也會被接受。所以, 務必要了解的是, 接受虛無假設, 僅表示樣本未提供充分的證據以拒絕該假設。另一方面, 拒絕虛無假設, 則表樣本提供的證據夠強, 足以推翻該假設(但仍有可能犯錯)。以另一方式來說, 拒絕虛無假設, 表當該假設為眞時, 會產生所獲得的樣本之機率很小。例如, 對上述情況, 若

得到20個不良品, 則應足以拒絕$p=0.1$的假設。何以故? 若$p=0.1$, 利用排列組合, 得到20個, 或20個以上(比20個還極端), 也就是至少20個不良品之機率為

$$\sum_{i=20}^{100}\binom{100}{i}0.1^i0.9^{100-i}。$$

這個機率值雖不太好算, 但若利用機率論裡重要的**中央極限定理**(central limit theorem), 可得近似值約為0.0008。也就是若$p$真等於0.1, 會得到至少20個不良品之機率是很小的, 僅約萬分之八。故此時拒絕"$p=0.1$"的假設, 會犯錯的機率很低。由於如前所述, 拒絕虛無假設, 表我們認為該假設極可能不真。但接受虛無假設, 倒並不排除其他可能性。因此對一虛無假設, 有些人認為以"**不能拒絕**"的說法, 取代"**接受**"的說法較適宜。這是一種較保守的講法, 有點像"不能說不喜歡"不見得等同於喜歡。不過一般在實際應用時, 往往並不那麼謹慎。反正只要理解這是一隨機現象, 在這組數據下"接受", 在另一組數據下可能便拒絕, 誤判是難以避免的。因此在文字上那麼在乎, 似無必要。

假設檢定裡, 多大的推論錯誤機率是可以忍受, 乃視情況而定。譬如說依誤判的損失會有多大。事實上這中間有兩種錯誤的機率, 其一為虛無假設為真卻拒絕(此稱第一型錯誤), 其二為虛無假設不真卻接受(此稱第二型錯誤)。理想的狀況當然是兩型錯誤機率皆為0, 但通常不會有這種情形。當樣本數固定, 一般而言, 兩型錯誤的機率, 有一減小另一必增大。

| | 虛無假設為真 | 虛無假設不真 |
|---|---|---|
| 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 |

由於虛無假設為真卻誤判它不真之後果往往較嚴重, 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設定的值, 然後使第二型錯誤的機率愈小愈好。

統計假設的架構, 與刑事訴訟法中的無罪推定原則(第一百五十四條: 被告未經審判證明有罪確定前, 推定其為無罪)是類似的。我國最高法院, 於民國二十五年立下的有罪推定原則判例, 於經過六十餘年後, 終於在民國九十二年修正為無罪推定。此後被告原則上是無罪的, 不必證明自己無罪。法官只要認為被告罪證不足, 即可宣判無罪, 不必窮調查之途, 才可以宣判被告無罪。宋朝歐陽修在追述其父母生前言行事蹟的瀧岡阡表一文中, 提及其父為死囚"求其生而不得, 則死者與我皆無恨也"。也是這種無罪推定的精神。

以往檢察官若認為法官未窮盡調查之途徑, 便宣判被告無罪, 會不服而提起上訴。那是因長期以來法院是採有罪推定原則。但法官判決時, 採用無罪推定原則是較有道理的。設一嫌犯因證據不足而被釋放, 如果他是無辜的, 那當然最好; 如果他其實是有罪, 但被釋放後洗心革面, 再也不犯罪, 那也很好; 如果他因心存僥倖或其他原因, 又犯了罪, 則第二次以後, 就不見得每次都有那麼好的運氣了, 夜路走多總是沒有好下場。若採有罪推定原則, 由於此假設不易被推翻, 被起訴者容易被判有罪, 一旦執行刑罰(如死刑), 日後如果真相大白, 錯誤將如何挽回?

讀者大約也會明白"虛無"二字的由來。如果研究結果, 宣佈北銀樂透彩每期頭獎號碼的產生符合隨機性, 這種結論可能沒有幾個人有興趣。社會大眾有興趣的結論是拒絕虛無假設! 要嘛宣佈頭獎號碼是有公式可以算出, 要嘛宣佈明牌存在, 或宣佈的確有某幾個號碼較容易出現。此正如一般人有興趣的是電影明星的婚姻有問題(對立假設), 影迷們對其經紀人一再宣稱的該明星夫妻恩愛如常(虛無假設), 是不感興趣的。很多政治人物在被檢察官起訴時常說"被迫害", 等法院判決無罪時, 又改口稱許"司法還他清白"。如果明白法院可能是因證據不夠充分, 不得已之下而做出接受"虛無"假設(無罪)的判決, 就不用眞認為司法還他什麼了。