談高中數學裡的交叉分析

黃文璋
國立高雄大學應用數學系

有鑑於統計日益重要, 九十五年開始實施的高中數學課綱(稱為九五暫綱), 增加一些統計題材。如在高二下, 增加信賴區間與信心水準的解讀, 高三選修增加了交叉分析和二項分配。雖自己在大學從事統計教學多年, 但我一直不贊成在高中數學, 教太多統計, 因很難對高中生講清楚。即使在大學裡, 一般認為統計比微積分難教。後者有物理意義, 諸如極限、連續、微分, 及積分, 都可以圖形來說明其涵意。但對於統計, 連最基本的, 諸如機率及期望值, 究竟是什麼意思, 都很難解釋清楚。像是投擲一公正的骰子, 所得點數之期望值為3.5, 這3.5是指什麼?

首先"期望"二字就很令人迷惑, 因這與平常大家對"期望"一詞之理解並不一致。骰子一擲之下, 得到的點數, 必為1至6中的某一個, 不論如何投擲, 都得不到所"期望"的3.5。好吧! 你說應將"期望值"想成平均。但想藉由投擲多次, 來說明點數平均會逐漸接近3.5, 恐怕仍非易事。

信賴區間與交叉分析, 在一般大學統計教科書中, 是放在較後面的章節。如在本人所著"數理統計"一書, 全書共十一章, 信賴區間及交叉分析, 分別在第九章, 及第十章。學生是在已有相當多的機

率及統計之預備知識後, 才接觸此二題材。如今卻堂而皇之地進入高中數學, 不引起師生困擾才怪。其實諸如信賴區間, 及交叉分析等, 就算僅是初步介紹, 仍不易理解, 歸根結底, 還是"隨機"的概念沒有掌握。甚至由於將統計放在說1是1, 說2是2, 一向斬釘截鐵, 不會有變的數學中教, 隨機性常有意無意便被忽略了。例如, 在九九課綱伍附錄(數學甲I 3.3節, 數學乙I 5.3節)"常態分布、信賴區間與信心水準的解讀", 有底下一段:

> 高中程度的統計推論只做隨機變數期望值的估計, 它的背後理論是中央極限定理。要介紹中央極限定理, 就需要引入常態分布。此部分僅做通識性的介紹, 以活動方式建立學生對於中央極限定理的直觀。對一固定的信心水準, 給出信賴區間公式, 再讓學生以亂數表模擬或實驗投擲正面出現機率為p的銅板n次, 代入信賴區間公式, 以說明信心水準的意涵; 並以此解讀, 何以大多數的學生所得的信賴區間都會涵蓋p?

"大多數的學生所得的信賴區間都會涵蓋p"? 這句話九五暫綱中就有類似的。我曾藉"統計裡的信賴"一文, 指出這種講法很不妥, 但九九課綱卻仍保留著, 只是略微修改。這就是典型缺乏隨機的概念之講法。

首先多少比率的信賴區間涵蓋p, 才算"大多數"? 課綱寫得有些含混。要知每個人對"大多數"一詞的理解, 可能差異不小。以95%的信心水準為例, 既然要學生以實作來理解信心水準的意涵, 則合理而言, 總要95%左右的信賴區間涵蓋p, 才會視為大多數。若未超過85%, 就不太會視為大多數了。設一班有40人, 每人得一p之95%信賴區間。則涵蓋p的信賴區間數(有二項分配), 不超過34個(85%)之機率約為0.01388。只要參與實驗的班級數夠多, 要

找到幾班, 有15%以上的信賴區間, 沒有涵蓋p, 並不足為奇 。

仍以投擲銅板為例 。投擲一銅板2次, 若沒有出現1正1反, 是否就不相信銅板為公正? 不盡然, 因即使銅板公正, 出現1正1反的機率不過是1/2, 仍有1/2機率出現2正, 或2反 。那投擲數增大, 是否就愈容易出現正反面數各半? 恰好相反, 愈不容易 。例如, 投擲一公正銅板10次, 則出現的正面數有二項分配 。而出現5正5反的機率為

$$\binom{10}{5}(\frac{1}{2})^{10} = \frac{252}{1{,}024},$$

略小於1/4 。至於投擲一公正銅板$2n$次, 出現n正n反的機率, 為

$$\binom{2n}{n}(\frac{1}{2})^{2n} \text{ 。}$$

當n很大時, 此機率變得很小, 大約可以$1/\sqrt{n\pi}$來近似, 其中π為圓周率 。隨著n的增大, 此值趨近至0 。因此投擲一銅板$2n$次, 若沒有恰出現n正n反, 並不能就此下結論“銅板非公正” 。事實上, 銅板究竟是否公正, 只有“天曉得” 。就算$2n$次都出現正面, 也只能“強烈懷疑”, 銅板非公正 。原因是此機率為

$$(\frac{1}{2})^{2n} > 0 \text{ 。}$$

而機率為正的事件, 便都可能發生 。另一方面, 若有人每次投擲, 都恰出現n正n反, 你會因此堅定相信銅板為公正? 還是反而認為其中必有詐?

再看一例 。教室裡放個籤筒, 班上n個人各有一支籤 。老師上課若要問問題, 就“隨機”抽一支籤, 每次取出後放回 。假若n次後, 恰好每個人都被抽中一次, 學生是否相信老師眞的隨機地抽? 極可能不會 。此機率為

$$\frac{n!}{n^n} \text{ 。}$$

當$n=10$, 其值為0.00036288, 即約萬分之3.6 。當一班有40人, 即$n=40$, 則其值約為6.749×10^{-17} 。每個人恰被抽中一次, 看起來很隨機, 反而讓人不信此為隨機地抽 。因這事件太不可能發生了! 附帶一提, 機率裡常提到隨機, 它有一些不同的涵意 。此處隨機抽, 指每支籤被抽中的機率皆相同 。

隨機與隨意是不同的 。前陣子我擔任某縣市中小學科學展覽高中組數學的評審 。有一組學生口頭報告完後, 為展示他們的結果, 請評審自桌上分別寫有1至10的紙牌中, 隨機抽一張 。我跟他們確認幾次, 是否真的是隨機抽? 他們皆表示肯定 。我說這是隨意抽, 而非隨機抽 。前者是依你高興抽選, 後者是每張牌被抽中的機率為1/10 。這樣將牌攤開在你面前, 怎還能算是隨機抽? 那幾位高二學生, 雖學過機率, 顯然仍不了解隨機的意義 。

統計的工作, 就是由觀察到的結果(或說收集到的數據), 做出推論 。要知在隨機世界中, 一切都是假設, 就看你接受那一個 。而如何拒絕或接受? 是屬於統計裡"假設檢定"的問題 。簡單講, 乃在"無罪推定"的原則下, 依發生機率的大小, 決定原先假設, 該不該接受 。如果懷疑銅板不是公正, 出現正面的機率p不為1/2, 則先假設$p=1/2$, 稱為虛無假設 。虛無就是空的意思, 因這不是你想接受的假設 。接受此假設, 表示你做虛功, 本來無一物, 何處惹塵埃, 白忙一場 。另外有一假設, 稱為對立假設, 是你傾向相信的 。可以是

$$p\neq1/2,\ p<1/2, \text{或}\ p=0.6$$

等 。虛無假設是要被保護的, 不能輕率地推翻 。有如在法庭上, 要有足夠強的證據, 才會判被起訴者有罪 。事先設一可以容忍的錯誤機率, 常以α表之, 為一比較小的值 。如$\alpha=0.05$, 或$\alpha=0.01$等 。如果在法庭, α可能還得更小些, 因冤獄畢竟難以彌補 。對隨機現象, 所做的決策, 是難以不犯錯的 。我們只能以更好的方法, 降低犯錯

機率。例如, 實際上對立假設為眞, 卻接受虛無假設(有如明明有罪卻被無罪開釋), 這種錯誤的機率, 也要控制住。根據投擲的次數n及α值, 可得到拒絕虛無假設的區域(簡稱拒絕域, 如何得到此處不說了)。

以

$$\text{虛無假設 } p = 0.5, \text{ 對立假設} p \neq 0.5$$

為例。在$n = 100$, 及$\alpha = 0.05$下, 拒絕域為正面數落在0至39間, 或61至100間。所以出現的正面數, 只要落在40至60間, 都接受$p = 0.5$。絕不是正面數不等於50, 就認為銅板非公正。

令銅板出現正面的機率$p = 1/2$, 然後要你去求一些事件(如投擲10次得6正4反)的機率值, 這是數學問題, 不用去質疑p是否眞為1/2。但由投擲後得到幾次正面, 來推測$p = 1/2$是否成立, 就是統計問題了。投擲100次, 這次得58個正面, 下次可能得48個正面, 第3次說不定得61個正面, 是會有變異的, 因此各次推論當然有可能不同。只是不論投擲數再大, p值究竟為何, 仍是"天曉得"。而我們所能做的, 是控制犯錯機率, 不超過所能容忍的值。

其實對隨機現象,允許誤差的概念,我們平常就有。99年4月19日, 台大公布甄選榜單, 報載醫學系20個名額中, 有"近半名額"9位為女生, 寫下歷史紀錄。台大教務長表示, 女生錄取比例增加是"剛好", 台大招生的立場, 向來就是找最適合、最有能力的學生, 不會考慮學生性別。

你看, 只要"近半", 並不需女生錄取正好一半, 才認定未考慮學生性別。當然此報導是有瑕疵的, 應由各性別的錄取率, 而非由各性別的錄取人數, 來判定兩性錄取是否有異。

目前高中選修數學(I), 在交叉分析那節, 常以考試錄取率為例。先給出所謂的列聯表, 一個典型的例子是,

	錄取人數(A)	未錄取人數(F)	合計
男生(B)	24	36	60
女生(G)	36	54	90
合計	60	90	150

由此得到

$$男生錄取率為P(\mathrm{A}|\mathrm{B}) = 40,$$

$$女生錄取率為P(\mathrm{A}|\mathrm{G}) = 40\text{。}$$

然後就說男生女生錄取率沒有差異。有些書會加類如下述的一句“至於比例不相等時, 是否就代表男女生錄取率有差異, 留待日後再學習。”

這其中有兩點必須指出。其一就是我們已反覆指出, 由男女錄取率的“相等與否”, 來判定錄取與男女性別是否有關, 是錯的, 這並非統計思維。除非事先設定男女錄取率一定要相同(這時男女的“錄取標準”, 就很難相同了), 否則即使用抽籤(這時錄取與否總該跟性別無關了), 來決定錄取名單, 都不能保證抽出的男女錄取率相同。其二是不應將觀測值視為機率。$P(\mathrm{A}|\mathrm{B}) = 40\%$是錯的概念。這點人們平常其實大都了解。例如, 投擲銅板100次, 出現52次正面, 並不會將$52/100 = 0.52$當做正面出現的機率(只會視為估計值); 也不會將一次民調的支持率, 當做候選人的得票率。但不知何以統計只要一擺進高中數學課本中, 就連常識都失去了。

最後我們給一有名的例子。交叉分析並不只能用來檢定各類之比率相同與否, 如檢定樂透彩1至42, 42個號碼出現之頻率是否相同。也可用來檢定各類之出現, 是否符合一組給定的比率(當然用途還不僅於此)。著名的遺傳學家孟德爾(Gregor Johann Mondel, 1822-1884), 有一關於豌豆生長的實驗。他將圓黃(round yellow)種子的豌豆, 與縐綠(wrinkled green)種子的豌豆雜交。依其

理論, 會生長出圓黃、圓綠、縐黃及縐綠種子的後代之比率, 應分別為$9/16 = 56.25\%$, $3/16 = 18.75\%$, $3/16 = 18.75\%$及$1/16 = 6.25\%$。經由一組有556個樣本的實驗, 他得到如下的觀測比率與預期比率。

	圓黃	圓綠	縐黃	縐綠	合計
後代數	315	108	101	32	556
觀測比率	56.65%	19.42%	18.17%	5.76%	100%
預期比率	56.25%	18.75%	18.75%	6.25%	100%

乍看之下, 4種豌豆觀測到的後代比率, 與預期比率有些差異。但經過統計檢定(這要用到所謂卡方檢定), 即使α值大到0.90, 都無法拒絕

虛無假設: 孟德爾的理論為正確。

但你知道嗎, 由於此實驗結果與預期太吻合(fit too well), 曾引起著名統計家費雪(Ronald Alymer Fisher, 1890-1962)的懷疑, 認為孟德爾可能是重覆做實驗, 直到結果看起來很好才停止, 然後只公佈結果較好的那組數據。這就是我們前面所述, 對於隨機實驗, 若結果與理論值過於一致, 反而會讓人懷疑做假。

還是如本文一開頭便強調的, 交叉分析並不適合放進高中數學的題材。既講不清楚, 也無法讓學生學到任何正確的統計概念。幸好九九課綱將此題材拿掉了。

參考文獻

1. 黃文璋(2003)。數理統計。華泰文化事業股份有限公司, 台北。

2. 黃文璋(2005)。統計顯著性。數學傳播季刊, 29(4): 29-38。

3. 黃文璋(2006)。決策的誤差。數學傳播季刊, 30(3): 66-79。

4. 黃文璋(2006)。統計裡的信賴。數學傳播季刊, 30(4): 48-61。

5. 黃文璋(2007)。統計裡的關係。數學傳播季刊, 31(1): 49-67。

6. 黃文璋(2007)。統計裡的估計。數學傳播季刊, 31(20: 3-20。

7. 黃文璋(2009)。統計思維。數學傳播季刊, 33(4): 30-46。

8. 黃文璋(2010)。機率應用不易。數學傳播季刊, 已接受。

9. 葉偉文譯(2001)。統計, 改變了世界(Salsburg, D. 原著: The Lady Tasting Tea)。天下遠見出版股份有限公司, 台北。