

## 以中央極限定理選才？

黃文璋

國立高雄大學統計學研究所

九合一選舉，終於結束了。在投票前，有位候選人，拋出以遴選委員會，來挑選首長的主張。報載“他解釋，在統計學上， $N > 25$ ，就會接近大數法則，也就是中央極限理論，不太容易出現偏頗的情況。雖然他準備設置的遴選委員會成員不到 25 人，但以過去經驗來看，‘只要 15 個人就會蠻準確的’。”又是大數法則，又是中央極限理論，還說蠻準確。一付極有根據的樣子，只是究竟在說些什麼？

著名的大數法則及中央極限定理，其起源都可追溯到十七世紀末至十八世紀初期，原本是針對二項分佈。什麼是二項分佈？重複觀測一有兩個結果的試驗，其中一個結果不妨稱為成功，另一稱為失敗，成功的機率假設為  $p$ ，又假設各次觀測間為獨立。則  $n$  次後所得成功數  $S_n$ ，便有有參數  $n, p$  之二項分佈，此分佈常以  $B(n, p)$  表之。就算原本試驗之所有可能的結果超過兩個，若將有興趣的稱為成功，其餘全稱為失敗，即轉換成一有兩個結果的試驗，因此二項分佈可說無所不在。當  $S_n$  有  $B(n, p)$  分佈，由排列組合可求出， $S_n = k$  的機率  $P(S_n = k) = C(n, k)p^k(1-p)^{n-k}$ ，其中  $C(n, k) = n! / (k!(n-k)!)$ ，而  $n! = n(n-1) \cdots 1$ ， $0! = 1$ 。數學家證出來，只要  $n$  夠大，則  $S_n/n$

與  $p$  任意接近的機率，將很接近 1。這就是早期版本的大數法則，可用來解釋機率的涵義。例如，某銅板出現正面機率為  $p$ ，是什麼意思？持續反覆地投擲，當投擲數  $n$  很大後，平均所得之正面數  $S_n/n$ ，與  $p$  便很可能會很接近，或者說  $S_n/n$  大致在  $p$  的附近。有時人們會籠統地說， $n$  很大時， $S_n/n$  與  $p$  很接近。這樣講其實並不對，因這是隨機現象，不能保證必然會很接近。至於以為  $n$  很大時， $S_n/n$  就等於  $p$ ，當然更是錯的。另外，中央極限定理，則能更進一步表示出， $S_n$  經標準化後，極限分佈為標準常態。更明確地說，對每一常數  $a > 0$ ，當  $n$  很大時，機率  $P(|(S_n - np)/[np(1-p)]^{1/2}| \leq a)$ ，可以一標準常態分佈的機率值來近似。此定理能用來表示出， $n$  很大時， $S_n/n$  落在一個以  $p$  為中心之區間的近似機率。

其後上述二結果一再被推廣，在一些簡單的條件下，針對一數列獨立且有共同分佈的隨機變數，大數法則及中央極限定理便皆能成立。但首先，讀者該可看出了，“大數法則，也就是中央極限理論(定理)”此講法並不正確，二者內容其實不一樣。其次，此法則及定理的用途，與“不太容易出現偏頗”乃毫不相干。又實務上  $n$  到底要多大？對中央極限定理，有些教科書會說  $n$  在 30 以上，便大致適用。但所謂適用，乃看你能容忍的誤差有多大？一般在做民調時，其間的理論，有用到中央極限定理。但因希望誤差不要超過 3%，一開始設定的成功樣本數為 1,068，而不是 30 而已。即使 1 千個左右隨機產生的樣本，都還常被批評做出的調查偏頗，未能反映真相。至於為特定目的而成立的遴選委員會，其委員當然不可能隨機產生，因此從委員會成立之始，便難以避

免主觀了。而不同委員的立場迥異，更不會是“有共同分佈”，因此不論大數法則或中央極限定理，在此皆不適用。舉一例子來看。民國 103 年 11 月落幕的第 51 屆金馬獎，得獎名單由總共 17 位委員共同討論決定。但未獲最佳女主角獎的著名演員鞏俐，事後透過經紀人表示，金馬獎不專業、不公正，不會再參加。由此可看出，像遴選這類評比，有相當程度的主觀成分，容易產生爭議，並無所謂“只要 15 個人就會變準確的”這種推論。

大數法則及中央極限定理，乃機率中二極重要的結果，用途廣泛，但卻也沒那麼無所不包到連選才都用的上。否則一旦所選非人，豈不砸了中央極限定理的招牌。