關於中央極限定理之圖示

黃文璋

國立高雄大學應用數學系

最近看到一篇“漫談信心水準研習心得”的文章(底下簡稱“心得”)。該文一開始提到:

> 往往我們向學生介紹完歷史背景之後, 就把常態分配的規則, 例如高中課本中的“68-95-99.7規則”, 硬生生記下來解題, 絲毫感受不到當時找到這條曲線的驚喜。研習當天, 講師用動畫展示二項分配(n,p)中, 當n愈來愈大時, 理論上出現的相對次數分配直方圖外型會逐漸逼近一條曲線, 這條曲線就是常態分配曲線, 或被稱為鐘形曲線。雖然我們無法向高中生證明這個定理, 但是電腦動畫的模擬效果令在場的老師印象非常深刻!

接著心得一文給出幾個所謂令人印象深刻的圖。但其實所給的應非模擬的圖, 也並非“直方圖”(histogram), 而是“機率質量函數”(probability mass function, 簡稱pmf), 乃描述離散型隨機變數取各可能值之機率, 因此圖形在各取值之高度和為1。

假設X_n有二項分佈$\mathcal{B}(n,p)$, 令$\overline{X}_n = X_n/n$。心得一文附上幾個$\overline{X}_n$之pmf圖, p皆取為0.5, n則分別為10, 30, 50, 70, 90, 110。隨著n的增大, 圖形“似乎”愈來愈像常態分佈曲線。可惜不知心得一文的作者是否注意到, 他所繪的那6個pmf圖, 其縱軸尺度不同, 因此並無法眞正展示圖形的外型,

會逼近那一曲線。而就算不理會各pmf圖縱軸尺度隨n而變, 如果讓n繼續增大, 恐便不易認為$\overline{X}_n$之pmf圖, 會逼近常態分佈曲線。

對X_n有$\mathcal{B}(n,p)$分佈, 仍取$p=0.5$, n則分別為10, 30, 100, 1,000, 10,000, 及100,000, 我們繪出兩種$\overline{X}_n$之pmf圖形。圖1為縱軸尺度經調整(如心得一文), 圖2為縱軸尺度皆相同。

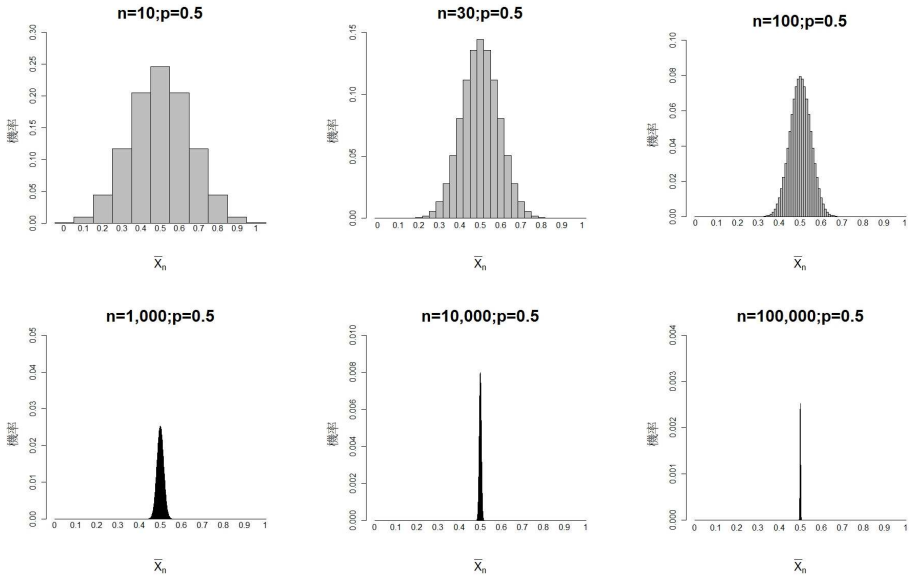

圖1. $\overline{X}_n$之pmf圖, 縱軸尺度經調整

我們看到在圖1中, 隨著n之增大, $\overline{X}_n$之pmf圖愈來愈狹窄, 最大高度(發生在$\overline{X}_n=0.5$)則逐漸下降。當$n=100,000$時, $\overline{X}_n$之pmf圖幾乎只剩一條小垂直線段了。無論如何, 對於$n=1,000$以上, $\overline{X}_n$之pmf圖大約不會令你看起來像鐘形, 因此也就不會覺得會逼近至常態分佈曲線。心得一文只繪到$n=110$便停止了, 沒有看到n再大時pmf圖之變化。因此雖作者覺得印象深刻, 產生心得, 只是這又為一郢書燕說之例。要知這些pmf圖, 對了解中央極限定理的內涵, 其實並沒什麼大幫助。

在圖1中, 對於$n=10$, 30, 及100, $\overline{X}_n$之pmf圖, 的確讓人看起來"像"常態分佈曲線。但這可能是視覺上的迷失。由於此處$\mathcal{B}(n,p)$分佈之$p=0.5$,

所以pmf圖為對稱, 中間最高, 向兩側下降。於是讓你覺得有如鐘形, 進而以為像常態分佈曲線。事實上, 只要是對著某點對稱的分佈, 其pmf圖往往便會讓人以為像常態分佈曲線。

其實$\overline{X}_n$之分佈為離散型, 取值$0, 1/n, 2/n, \cdots, (n-1)/n, 1$, 並非僅取整數值。而如前所述, 對離散型, pmf圖的高度表機率, 是其高度和為1, 而非pmf圖形下小長方形的面積和為1。光憑這點, 就不能說$\overline{X}_n$的pmf圖, 與常態分佈曲線有什麼太密切的關連。因常態分佈曲線下方的面積為1。大家應也注意到, 在圖1中, 當n愈來愈大, $\overline{X}_n$的pmf圖愈來愈黑。這是當然的, 因橫軸長度固定, 因此n愈大時, $\overline{X}_n$取值愈密, 擠在一起的關係, 讓圖形變黑。

在圖2裡, 我們令縱軸尺度相同, 這樣較易看出對不同的n, pmf圖眞正的變化。$n = 100,000$時, pmf圖幾乎只餘$\overline{X}_n = 0.5$上之一小黑點。這是因此時對每一$i = 0, 1, \cdots, n$, $\overline{X}_n = i/n$之機率都近乎0, 因此其pmf圖簡直與橫軸相貼, 讓肉眼幾乎看不到。

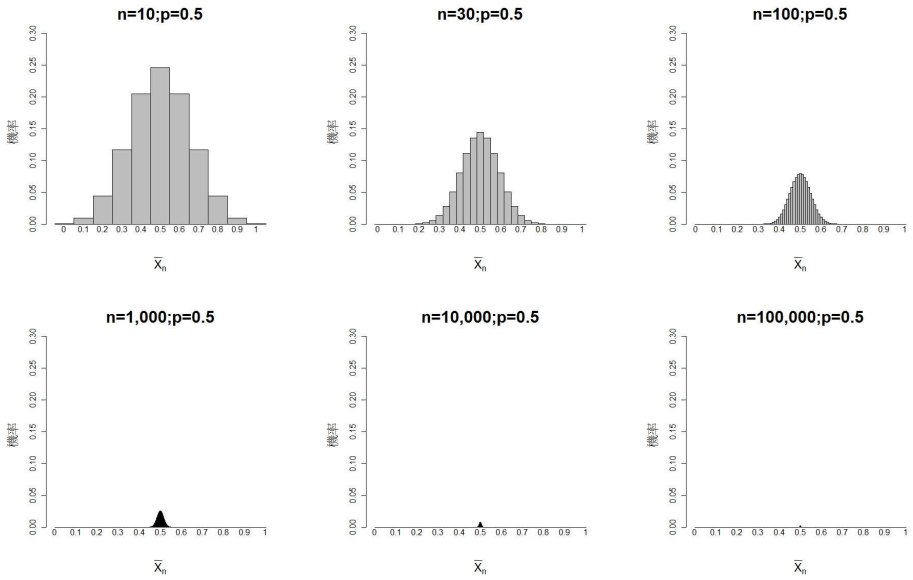

圖2. $\overline{X}_n$之pmf圖, 縱軸尺度相同

我們講的直方圖(此處為機率直方圖, 另有數據直方圖), 乃要求圖形下, 那些小長方形之面積和為1 。隨機變數若取整數值, 則因每一小長方形的底部長為1, 故以pmf圖作出的那些小長方形面積和為1 。這時pmf圖與直方圖可說是一樣的 。由於心得一文, 與圖1及圖2, 繪的都是$\overline{X}_n$的pmf圖, 所以圖形下長方形面積和為$1/n$ 。將高度變成n倍後, 才是直方圖 。

隨著n之增大, $\overline{X}_n$pmf圖之最高點愈來愈低的原因是什麼? 當n為偶數, $\overline{X}_n = 0.5$之機率為$\binom{n}{n/2}/2^n$ 。利用近似$n!$的Stirling公式(見黃文璋(2010)頁128-130), 可得此機率近似相等(asymptotically equal)於

$$\frac{\sqrt{2}}{\sqrt{n\pi}} \text{ 。} \tag{1}$$

即$\overline{X}_n$pmf圖之最大高度, 以$\sqrt{2}/\sqrt{n\pi}$的速度下降至0 。這可以圖1來印證 。對於$n = 100$, 與$n = 10,000$, 二pmf圖之最大高度, 分別大約為0.08與0.008, 後者約為前者的$\sqrt{100}/\sqrt{10,000} = 1/10$ 。

另外, n愈大時, $\overline{X}_n$之pmf圖愈來愈窄的原因是什麼? 這正說明大數法則(更明確的說是弱大數法則) 。對X_n有$\mathcal{B}(n,p)$分佈, 當$n \to \infty$時, $\overline{X}_n$會機率收斂至p(此處p為0.5) 。也就是$\overline{X}_n$會聚集在p附近一很小範圍內的機率很大 。以數學式子來表示, 即對每一正數a, 當$n \to \infty$時,

$$P(|\overline{X}_n - p| < a) \to 1 \text{ 。}$$

有某高中數學課本, 說“機率裡的期望值就是統計試驗中大量數據的平均值”, 這當然是錯誤的講法 。以二項分佈為例, 大數法則是說, n很大(大量數據)時, $\overline{X}_n$(平均值)與期望值(p)之差異很小的機率很大 。絕對不是說n很大時, $\overline{X}_n = p$ 。對常數數列$\{a_n, n \geq 1\}$, 當$n \to \infty$時, $a_n \to a$, 人們不會以為其含意為n很大時, $a_n = a$ 。但對隨機數列$\{X_n, n \geq 1\}$, 有些人雖大數法則能琅琅上口, 卻不只沒能正確掌握其中收斂的含意, 連原本常數數列裡, 不會弄錯的“收斂不必然導致相等”都忘了 。

事實上由$\overline{X}_n$之pmf圖, 對了解大數法則幫助很大 。但若欲藉此了解中央極限定理, 便可能是緣木求魚了 。那是否仍有辦法藉由圖示來說明中央極限定理?

當X_n有$\mathcal{B}(n,p)$分佈, 由中央極限定理, $n \to \infty$時,

$$\frac{\overline{X}_n - p}{\sqrt{p(1-p)/n}} = \frac{\sqrt{n}(\overline{X}_n - p)}{p(1-p)} \tag{2}$$

之分佈, 即將原本幾乎很小的$\overline{X}_n - p$適當地放大(減去期望值後除以標準差, 稱為將$\overline{X}_n$標準化), 會趨近標準常態分佈$\mathcal{N}(0,1)$。對$p = 0.5$, 即得$n \to \infty$時,

$$Z_n = 2\sqrt{n}(\overline{X}_n - 0.5)$$

之分佈趨近$\mathcal{N}(0,1)$分佈。底下來看如何將上述過程圖示。

分別將圖2中的6個pmf圖, 高度先乘上n, 以得直方圖, 然後向右平移0.5單位(表$\overline{X}_n - 0.5$)。最後將縱軸單位長改為$2\sqrt{n}$, 橫軸單位長改為$1/(2\sqrt{n})$, 此步驟使直方圖面積維持為1, 乃反映將$\overline{X}_n - 0.5$乘上$2\sqrt{n}$。如此得$2\sqrt{n}(\overline{X}_n - 0.5)$之直方圖, 我們給在圖3。

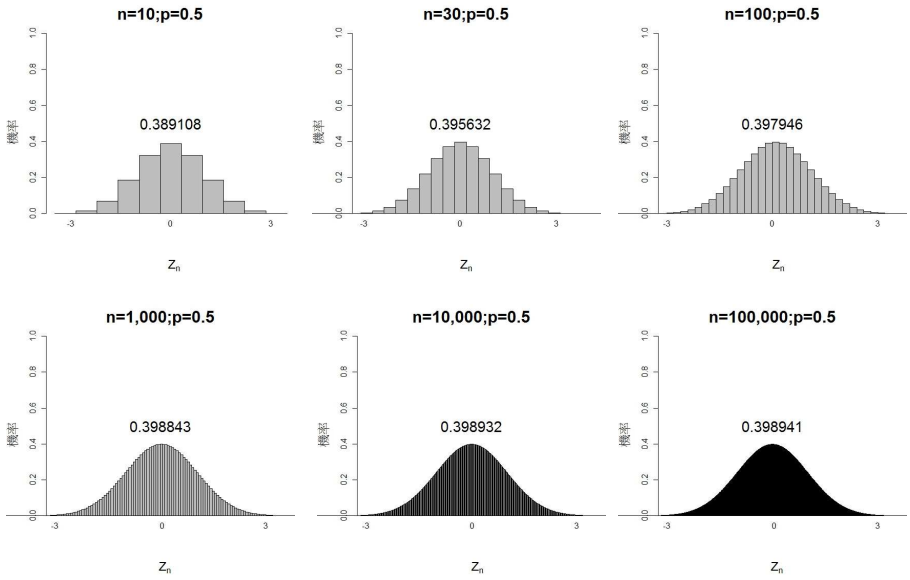

圖3. $\overline{X}_n$標準化後之直方圖

改變直方圖之縱軸及橫軸的單位長, 有如將原本又高又窄的直方圖向下擠壓, 使其往兩側散開。奇妙的是, 經由這樣的擠壓, $\mathcal{N}(0,1)$分佈的曲線, 隨著n之增大, 便逐漸浮出了。不妨來檢驗一下。因$\mathcal{N}(0,1)$分佈之機率密度函數為

$$f(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2},\ x \in R,$$

故其函數圖形之最大高度發生在$x = 0$, 即最大高度為$1/\sqrt{2\pi}$。而由(1), $\sqrt{2}/\sqrt{n\pi}$乘上n, 得$\sqrt{2n}/\sqrt{\pi}$, 此為n很大時, $\overline{X}_n$直方圖近似相等的最大高度。又高又窄的直方圖, 縱軸單位長改為$2\sqrt{n}$後,

$$\frac{\sqrt{2n}}{\sqrt{\pi}} \cdot \frac{1}{2\sqrt{n}} = \frac{1}{\sqrt{2\pi}},$$

原最大高度便下降了, 近似於$1/\sqrt{2\pi}$, 與$\mathcal{N}(0,1)$分佈曲線之最大高度吻合。在圖3中, 對各n, 我們特別標示出圖形之最大高度, 以供讀者與$1/\sqrt{2\pi}$(約為0.3989422)比較。大家可看到, 即使n僅為10, 標準化後$\overline{X}_n$之直方圖的最大高度

$$\frac{\binom{10}{5}}{2^{10}} \cdot 10 \cdot \frac{1}{2\sqrt{10}} = \frac{126\sqrt{10}}{1,024} \doteq 0.389108,$$

便與$1/\sqrt{2\pi}$差異不大了。對$n = 10$及30, 我們將圖3中的圖形, 與$\mathcal{N}(0,1)$分佈曲線分別比較, 見圖4。可看出即使對不太大的n, 二者已算接近了。

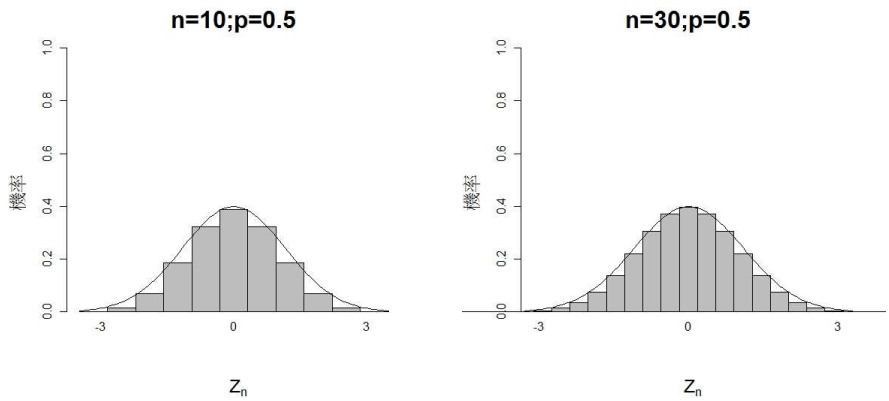

圖4. $\overline{X}_n$標準化後之直方圖與$\mathcal{N}(0,1)$分佈曲線比較

上述中央極限定理的圖示過程, 可能要對機率理論有些基礎的人, 才能明白其中原委。而這樣的大費周章, 對高中生理解二項分佈趨近常態分佈, 卻不知能有多大的幫助?

X_n有$\mathcal{B}(n,p)$分佈, 經(2)式的標準化後, 趨近$\mathcal{N}(0,1)$分佈。$p=0.5$的情況為de Moivre (1667-1754)在1733年證出; 一般的p之證法, 則是Laplace (1749-1827)在1812年發表的。他們的作法, 是將$\mathcal{B}(n,p)$分佈的機率相加, 也就是將pmf圖之高度相加, 然後求其極限。此大工程, 可參考黃文璋(2010)頁269-276。隨著機率理論的發展, 後來不限二項分佈, 有更一般版本的中央極限定理, 且證法很簡潔。由於其一般性, 使中央極限定理, 成為一強而有力的工具, 與大數法則, 並列機率中兩個極重要的定理。

附帶一提, 對X_n有$\mathcal{B}(n,p)$分佈, 說"n夠大時, X_n的機率分配會近似期望值np, 標準差$\sqrt{np(1-p)}$的常態分配"; 或說此時其"機率圖是對稱且呈鐘形, 接近常態分配"; 或說"機率分配圖與期望值np, 標準差$\sqrt{np(1-p)}$的常態分配曲線, 兩者頗為密合"(這是一些高中數學課本上的講法), 都不很恰當。要知隨著n之增大(而非只停留在100左右), X_n的pmf圖(因X_n取整數值, 所以其pmf圖也是直方圖)會愈來愈像一條水平線, 根本不像鐘形; 而期望值np, 標準差$\sqrt{np(1-p)}$的常態分佈, 當n很大時, 期望值跑到很遠(np很大), 而因標準差很大, 故其曲線也極像一條水平線。指出兩者很密合, 並未突顯中央極限定理的威力。有如說n很大時, $1/n$, $1/(n+1)$及$1/n^2$很接近, 數學上雖正確, 對極限的描述, 卻不夠精準。$n\to\infty$時, 三者皆趨近0, 因此的確很接近。但$1/n$與$1/(n+1)$的比值趨近1, 而$1/n$與$1/n^2$的比值卻趨近∞。

類似地, 有高中數學課本說, n很大時, "$\overline{X}_n$的分配會近似平均數為p, 變異數為$p(1-p)/n$的常態分配", 或說"$\overline{X}_n$的直方圖與期望值為p, 標準差為$\sqrt{p(1-p)/n}$的常態分配曲線, 兩者頗為密合", 也非精準的講法。因一旦n真的夠大, 此時前者的直方圖, 與後者常態分佈的曲線(變異數極小), 皆有如一條垂直線段。

對X_n有$\mathcal{B}(n,p)$分佈, 中央極限定理所強調的, 乃標準化後的$\overline{X}_n$,

如(2)式, 或標準化後的X_n, 即

$$\frac{X_n - np}{\sqrt{np(1-p)}},$$

當$n \to \infty$時, 其分佈會趨近$\mathcal{N}(0,1)$分佈。高中數學課本裡, 卻往往去強調$\overline{X}_n$或X_n的極限分佈。這樣的引申, 已非中央極限定理的本意。補充一點, 將X_n的直方圖向右平移np單位, 再將縱軸單位長改為$1/\sqrt{np(1-p)}$, 橫軸單位長改為$\sqrt{np(1-p)}$, 則n不斷增大時, 呈現出來的圖形, 將愈來愈像$\mathcal{N}(0,1)$分佈的曲線。

我們的結論是, 要在高中數學裡講解中央極限定理, 還說做"通識性的介紹", "建立學生對於中央極限定理的直觀"(高中數學課綱裡的用語), 實在是高難度。我們須肯定負責制定課綱者, 對機率統計的重視。只是他們似乎太低估中央極限定理的深度。

誌謝辭

感謝郭美惠、銀慶剛及羅夢娜等三位教授, 對本文提供許多寶貴意見, 使本文更臻完善。

參考文獻

1. 黃文璋(2010). 機率論, 第二版。華泰文化事業股份有限公司, 台北。