

專題：應用上的實例

問題 1:

假設有 n 變數 x_1, x_2, \dots, x_n 皆服從標準常態 $N(0, 1)$, 令一新變數

$$x_{min} = \min(x_1, x_2, \dots, x_n)$$

請問下列何者正確?

1. x_{min} 仍服從標準常態,
 2. x_{min} 服從某個常態分配 $N(\mu, \sigma^2)$,
 3. 以上皆非。
-

要回答這個問題, 可以從理論著手,¹不過在這裡卻鼓勵利用「電腦實驗」來試著否定前兩個答案。電腦實驗並不容易證明一件對的事情, 但是否定一件錯的事情卻是可行的。請試著利用 MATLAB 程式產生適當的樣本, 觀察 x_{min} 的實驗分配 (empirical distribution),²看看能不能從實驗分配的圖形中否定前兩個選項。當然除了實驗分配外, 直方圖、boxplot 也都是可以嘗試的。除了圖形之外, 適當的統計量也是可行的, 只不過圖形是一種比較好的表達方式。

實驗分配是統計模擬常用的技巧, 適用在不知名的分配上。因為是不知名的分配, 所以找不到適當的指令來計算機率值。這時候便需要利用數值 (估計) cdf 值來估計機率值。這時候上述的 n 是否影響到實驗分配? 換句話說, 是否不同的 n 會產生不一樣的實驗分配圖 (形狀)? 試著改變 n 來觀察這個現象。

如果將 x_{min} 改成

$$x_{ave} = \frac{1}{n} \sum_{i=1}^n x_i$$

¹這個最小值 x_{min} 有其理論分配, 分配函數寫成 $F_{x_{min}}(x) = 1 - (1 - F_X(x))^n$, 其中 $F_X(x)$ 代表獨立變數 X 的分配函數。

²MATLAB 有個指令 `ecdf` 可以用來計算實驗分配的數值 (估計) cdf, 不妨試看看。另外使用 `cdfplot` 可以直接看到 empirical cdf 圖。

同樣回答上述的問題。