利用 R 幫助學生了解中央極限定理的概念

Using R to Help Students Understand the Concept of the Central Limit Theorem

胡學穎

Shueh-Inn Hu

銘傳大學應用統計資訊學系
Department of Applied Statistics and Information Science,
Ming Chuan University

摘要

一般的統計學課程都會提到中央極限定理，而許多統計學課本裏面都會用從某些非常態分布的母體抽樣所得之樣本平均分布的圖形來呈現出中央極限定理的結果。本文在敘述如何利用統計軟體 R 讓學生可以自己選擇不同的母體分布以及樣本大小，再利用電腦模擬出抽樣結果，以期學生能夠透過實際操作的方式，真正了解中央極限定理的概念。

關鍵字：中央極限定理、模擬

1. 前言

中央極限定理的敘述為：假設 $X_1, X_2, \cdots, X_n$ 為一個抽自某一母體的隨機樣本，且知 E(X) = μ 、Var(X) = σ^2；只要所取的隨機樣本大小 n 足夠大時，其樣本平均 $\overline{X}$ 的抽樣分布就大約會是常態，並知 $E(\overline{X}) = \mu$ 、$Var(\overline{X}) = \dfrac{\sigma^2}{n}$。本文的主旨在敘述如何利用統計軟體 R 讓學生可以自己選擇不同的母體分布以及樣本大小，再利用電腦模擬出抽樣結果，以真正了解中央極限定理的概念。

2. 幾種母體分布的樣本平均抽樣分布

2.1 指數分布

首先考慮 λ=1 的指數分布，其機率密度函數為 $f(x) = e^{-x},\ x > 0$，圖形如下。我們可以用 rexp(n,rate=1)的指令得到從 exp(1)母體所抽出大小為 n 的隨機樣本，在此分別讓 n=5, 10, 30。對於每個 n，我分別抽取了 5000 個隨機樣本並計算其樣本平均，再利用 hist(x)的指令將數據作成直方圖以呈現出其分布的形狀，所得圖形如下。由圖 1 我們可以看出此母體之分布形狀為右偏，而從圖 2 中我們能夠看出在 n=5 時，其樣本平均的分布還是很明顯的右偏，但是隨著樣本大小 n 的增加，其對應的樣本平均分布逐漸近似常態分布。

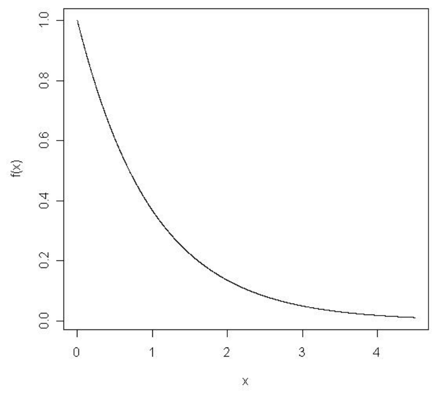

圖1 exp(1)的機率密度函數

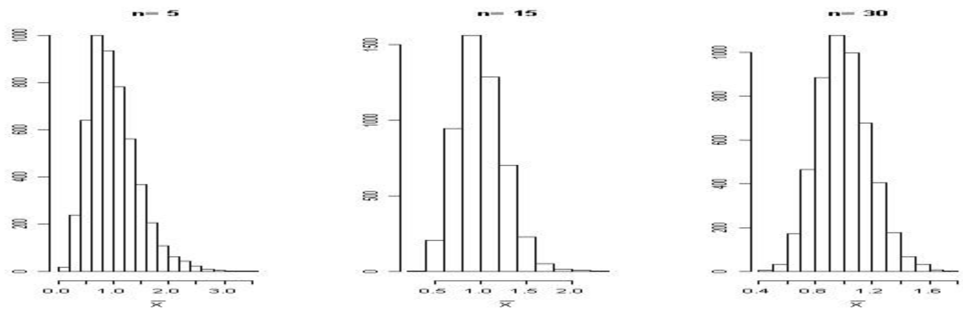

圖2 從exp(1)抽樣之樣本平均分布

2.2 均勻分布

如果 X 為均勻分布於[2,6]的隨機變數，則其機率密度函數為 $f(x)=\begin{cases}\frac{1}{4}, 2\le x\le 6\\ 0,\text{其他}\end{cases}$，見圖 3。

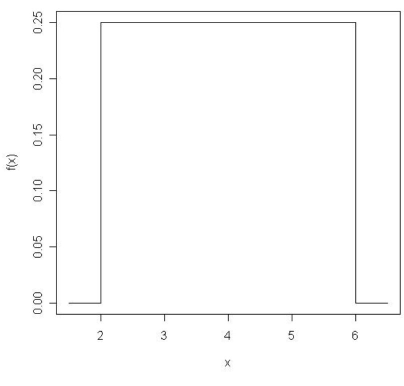

圖3 U[2,6]的機率密度函數

我們可以用 runif(n,min=2,max=6)的指令得到從 U[2,6]母體所抽出大小為 n 的隨機樣本，在此分別讓 n=5, 15。對於這兩個 n，我分別抽取了 5000 個隨機樣本並計算其樣本平均，然後作直方圖以呈現出其分布的形狀，所得圖形如下。由圖 3 我們可以看出此母體之分布形狀對稱於 x=4，而從圖 4 中我們能夠看出在 n=5 時，其樣本平均的分布已很接近常態，而 n=15 時，其樣本平均的分布更為接近常態。

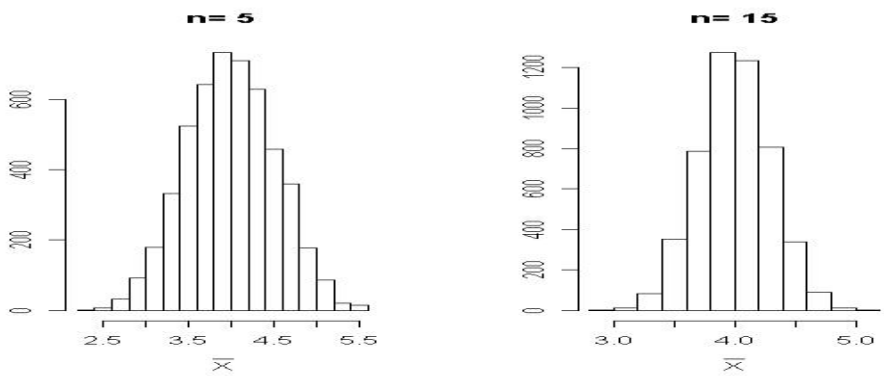

圖4 從U[2,6]抽樣之樣本平均分布

2.3 伯努利分布

假設 X 為伯努利分布的隨機變數，且其 p=0.2，則其機率函數為 $P(X = x) = 0.2^{x} \cdot 0.8^{1-x}, x = 0,1$，見圖 5。

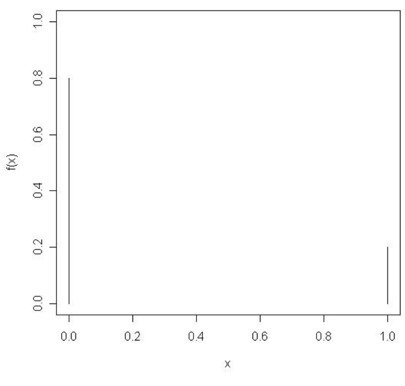

圖5 B(1,0.2)的機率函數

我們可以用 rbinom(n,1,0.2)的指令得到從 B(1,0.2)母體所抽出大小為 n 的隨機樣本，在此分別讓 n=15,50,100。對於這幾個 n，我分別抽取了 5000 個隨機樣本並計算其樣本平均，然後作直方圖以呈現出其分布的形狀，所得圖形如下。由圖 5 我們可以看出此母體之分布為離散的，而從圖 6 中我們能夠看出在 n=15 時，其樣本平均的分布呈現右偏，當 n=50 時，其樣本平均的分布較為對稱，而當 n=100，其分布則很近似常態。

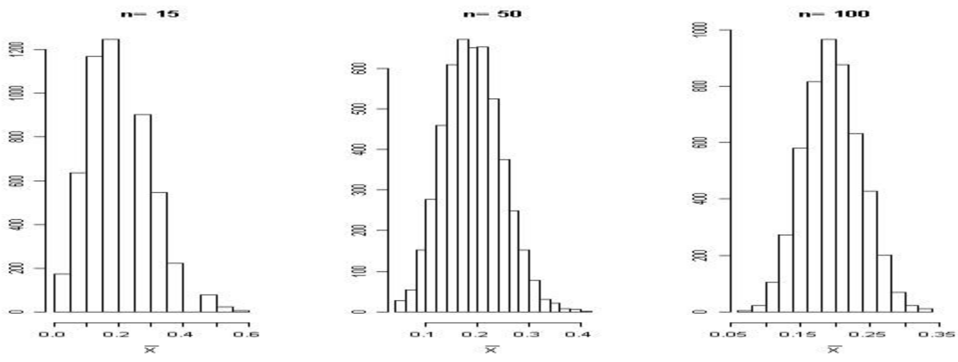

圖6 從B(1,0.2)抽樣之樣本平均分布

3. 結論

經由 R 統計軟體的模擬我們可以看出當母體分布本身為連續且對稱時，n 不需要很大其樣本平均的分布就很接近常態。但是當其母體分布為離散且不對稱時，就需要比較大的 n 其樣本平均的分布才會近似常態。

4. 參考文獻

[1] R Development Core Team (2007). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.