**Statistical Computing and Simulation** 統碩一 106354003 林健宏  
＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

**# Question 01**. Find by calculus techniques

< Sol >

**# Question 02**. Simulation Study

(1)

Find mean and variance of importance sampling estimates using proposal and or (more proposals). Compare it with that Monte Carlo.

(Which proposal gives a smaller variance? Give arguments to support your finding with using numbers and graphs.)

< Sol >

Derive of important sampling:

So if, The smaller variance is better.

If ,

So, it is important to choose g(x).

這個方法為Important sampling，若是能找到一個適合的Important function，便能達到縮減變異數的效果。

下表我們使用Mnote Carlo、Important sampling N(0,1)以及Important sampling N(0,5)去估計

（使用1000個樣本去估計一個 ，並總共估計5000個，計算這5000個 估計量 的樣本平均與樣本變異數）

|  |  |  |  |
| --- | --- | --- | --- |
|  | Monte Carlo | Important sampling N(0,1) | Important sampling N(0,5) |
| Mean | 1.99851 | 1.98926 | 1.99872 |
| Variance | 0.03619 | 0.14084 | 0.00865 |

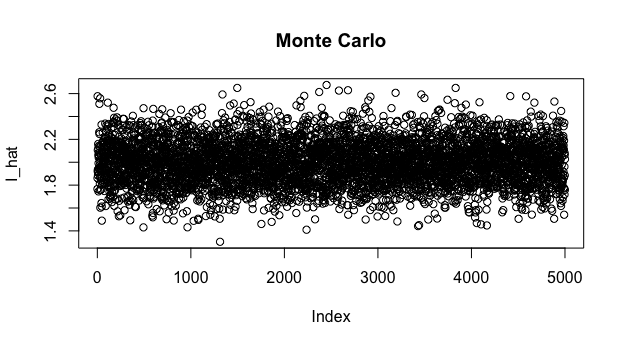
小結：

三種方法用於估計都很接近我們使用微積分機算的數值2，符合理論的推導，三種方法皆為不偏估計的方法。為了比較方法的好壞，我們比較三種方法的變異數。

由上表可以清楚看出三種方法變異數的差異，最大的為Important sampling 選擇N(0,1)當作important function時得到的變異數0.14084以及最小的為Important sampling 選擇N(0,5)當作important function時得到的變異數0.00865。可見Important sampling不一定可以降低變異數，反而會造成變異數變大。

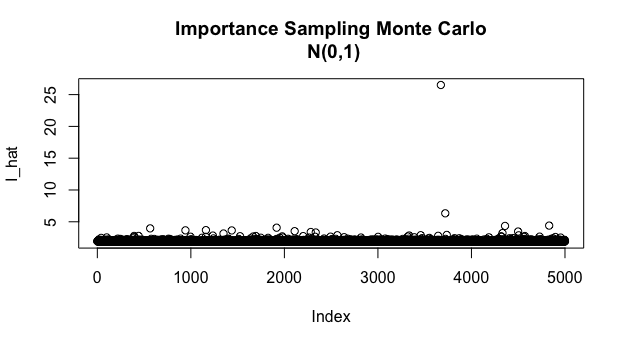
我們以圖表解釋為何差異這麼大，將5000個估計值，每一個估計值由1000個樣本構成，做成散佈圖，看三種方法的的估計值浮動範圍差異。

(a) Monte Carlo



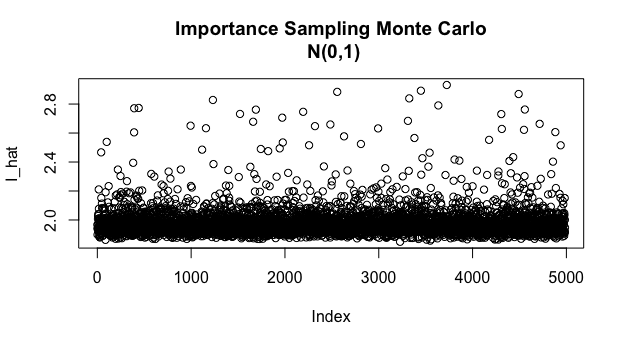
由上圖可見，使用Monte Carlo方法去估計時，估計值的範圍大約為 (1.4, 2.6)，其中大部分都集中在(1.8, 2.3)。

(b) Important sampling with N(0,1)



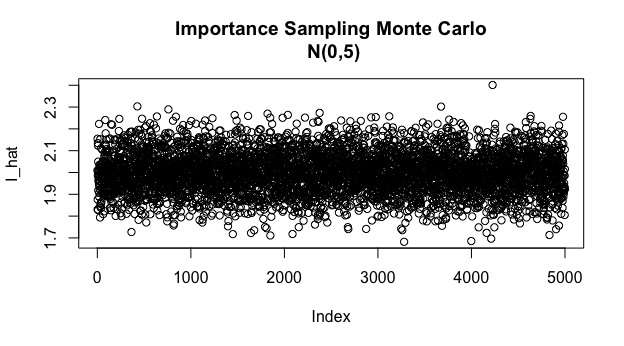
由上圖可見，使用Important sampling with N(0,1) 方法去估計時，估計值浮動範圍很大，但大部分都集中在5以下，偶而會出現極端值。

接著將極端值刪除（大於3的值），並重新繪製散佈圖



圖中顯示，雖然大部分的資料點都幾鐘在(1.8,2.2)之間，但仍有很多點出現在（2.2, 3.0）之間，甚至更大，浮動範圍比Monte Carlo的範圍(1.4, 2.6)來得大。也因此使用Important sampling非但沒有降低Monte Carlo方法的變異數，反正比Monte Carlo的變異數來得大。

(c) Important sampling with N(0,5)



這次Important sampling方法去估計時，我們改變Important function，由N(0,1)改成N(0,5)。從上圖中發現，這5000個估計值範圍為(1.7, 2.4)附近，大多都集中在(1.8,2.1)之間，比Monte Carlo的範圍(1.4, 2.6)來得小，因此變異數比Monte Carlo的變異數小，縮減量約為76.1%。

結論：

選擇適當的Important function可以使得估計效果與原本使用的方法一樣，同時變異數達到縮減的效果，但選擇不好的Important function則會造成變異數變大，雖然估計值差不多。以本題的兩個Important function 與 ，可以讓整體變異量下降76.1%，而造成整體變異量上升。

**# Question 03.**

(a)  Obtain the mean and variance of N(0,1) using importance sampling with g ∼ t3.

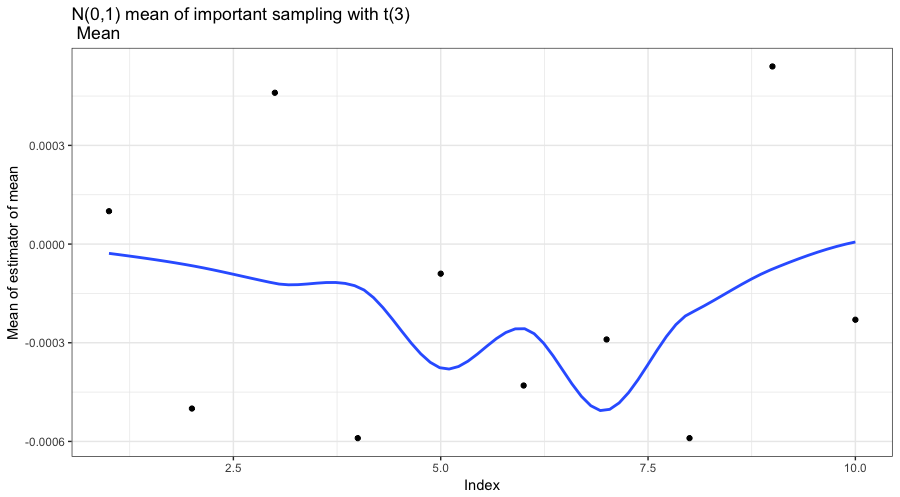
同樣取5000個估計量，每一個估計量分別取100, 200, … ,1000去探討，樣本數是否影響估計效果<

並觀察估計是否準確。

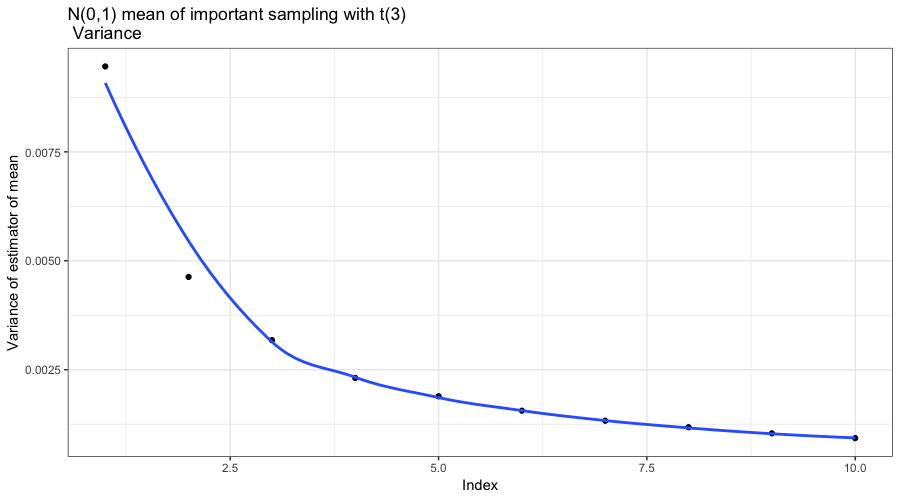
Important sampling t(3)去估計 Normal (0, 1)的平均與變異數

估計標準常態的mean的樣本平均與樣本變異數：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| n | 100 | 200 | 300 | 400 | 500 |
| Mean of | 0.0001 | -0.0005 | 0.00046 | -0.00059 | -0.00009 |
| Variance of | 0.00946 | 0.00463 | 0.00318 | 0.00231 | 0.00189 |
| n | 600 | 700 | 800 | 900 | 1000 |
| Mean of | -0.00043 | -0.00029 | -0.00059 | 0.00054 | -0.00023 |
| Variance of | 0.00156 | 0.00133 | 0.00118 | 0.00104 | 0.00093 |



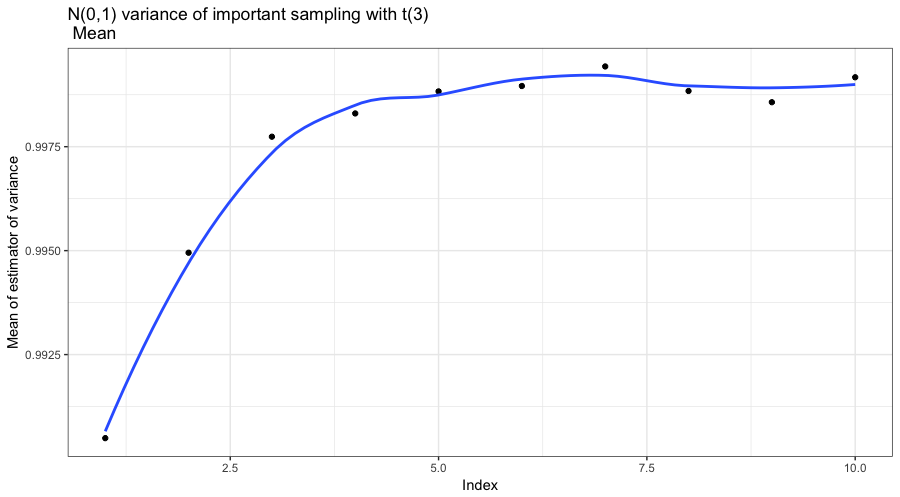
（在不同樣本數下的樣本平均）



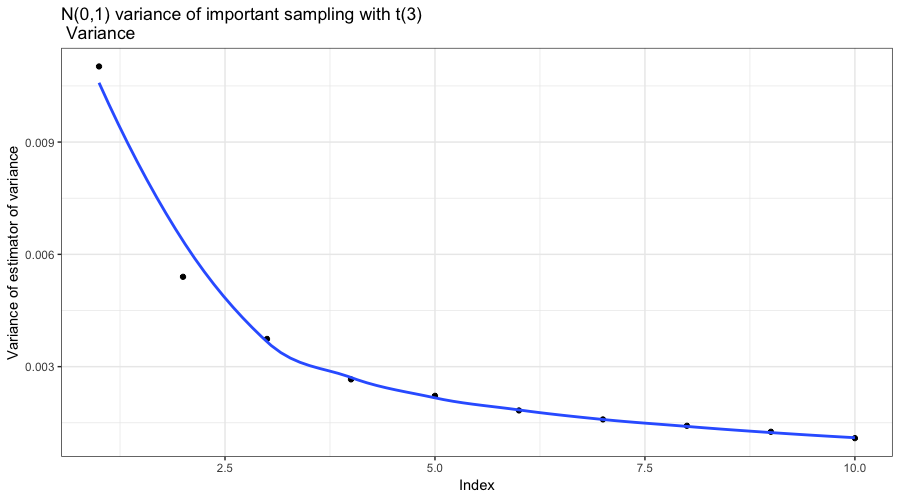
（在不同樣本數下的樣本變異數）

估計標準常態的Variance的樣本平均與變異數：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| n | 100 | 200 | 300 | 400 | 500 |
| mean of | 0.99049 | 0.99495 | 0.99774 | 0.9983 | 0.99883 |
| Variance of | 0.01102 | 0.0054 | 0.00374 | 0.00266 | 0.00222 |
| n | 600 | 700 | 800 | 900 | 1000 |
| Mean of | 0.99896 | 0.99943 | 0.99884 | 0.99857 | 0.99917 |
| Variance of | 0.00183 | 0.00159 | 0.00142 | 0.00126 | 0.00109 |



（在不同樣本數下的樣本平均）



（在不同樣本數下的樣本變異數）

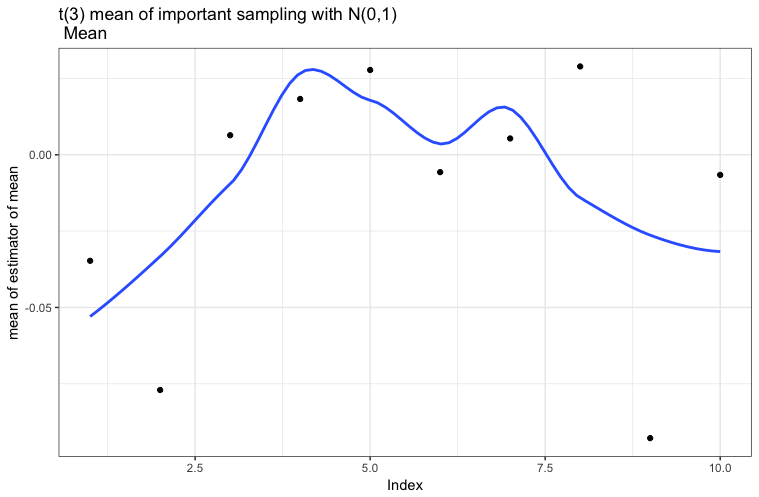
小結：

由Important sampling with t(3)去估計標準常態的mean與Variance，在不同的抽樣樣本數下都蠻接近的標準常態的meann 0以及Variance 1，特別是當抽樣樣本數逐漸變多時，估計值的變異數都逐漸下降，可知樣本數會影響準確度。

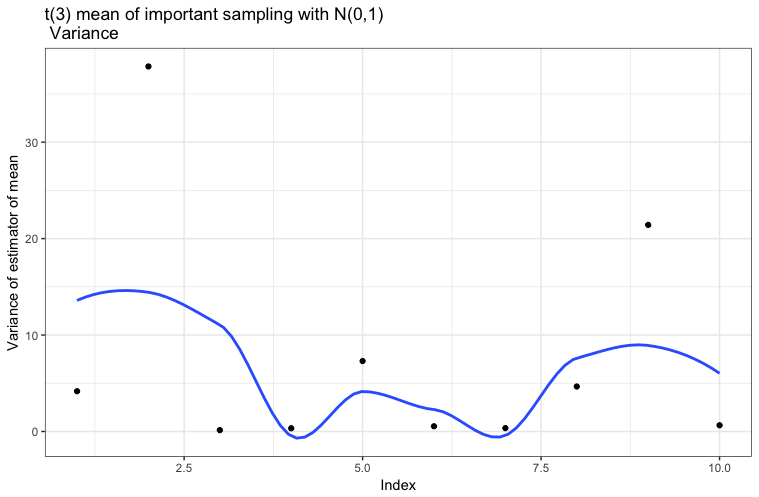
(b)  Obtain the mean and variance of t3 using importance sampling with g ∼ N(0,1).

估計t(3)的mean的樣本平均與樣本變異數：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| n | 100 | 200 | 300 | 400 | 500 |
| Mean of | -0.03472 | -0.07704 | 0.0064 | 0.01828 | 0.02777 |
| Variance of | 4.17662 | 37.85109 | 0.14083 | 0.33347 | 7.30422 |
| n | 600 | 700 | 800 | 900 | 1000 |
| Mean of | -0.00569 | 0.00535 | 0.02893 | -0.09281 | -0.00659 |
| Variance of | 0.54173 | 0.35007 | 4.66431 | 21.42051 | 0.64627 |



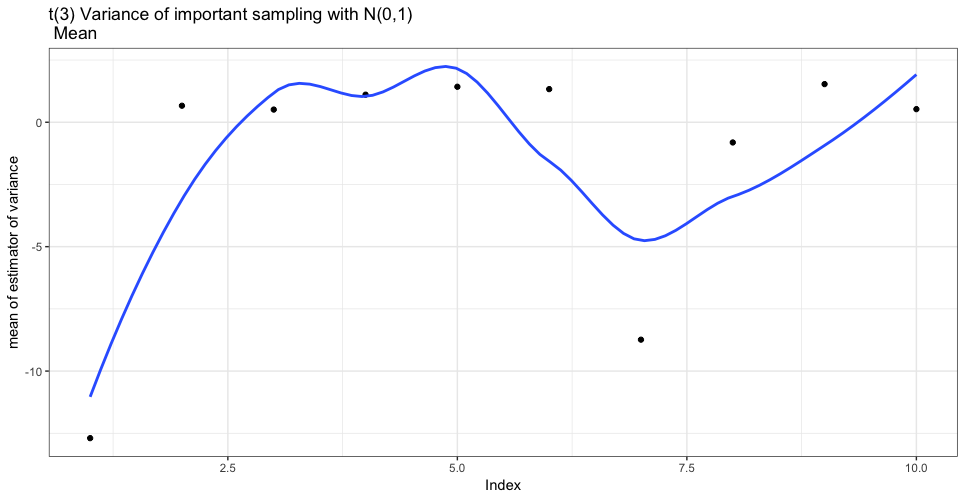
（在不同樣本數下的樣本平均）



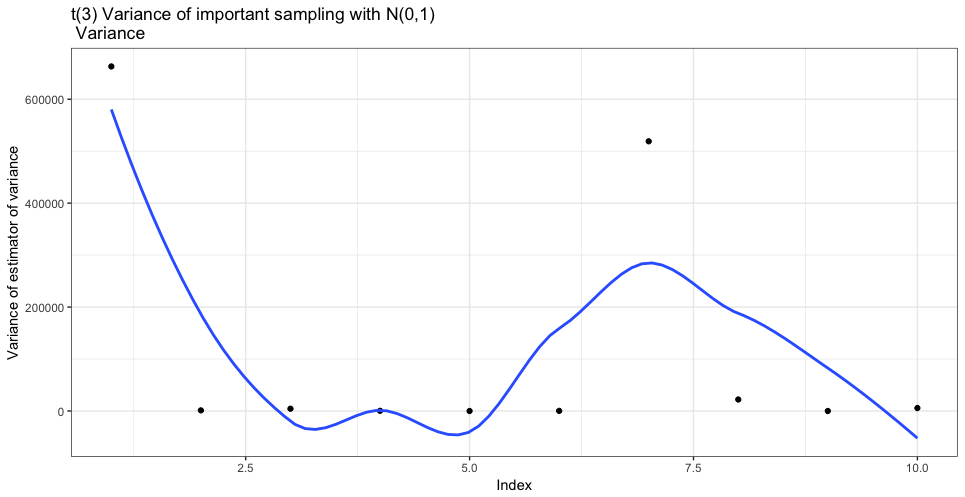
（在不同樣本數下的樣本變異數）

估計t(3)的variance的樣本平均與樣本變異數：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| n | 100 | 200 | 300 | 400 | 500 |
| mean of | -12.69518 | 0.66387 | 0.50941 | 1.10809 | 1.42652 |
| Variance of | 663024.834 | 1238.09275 | 4433.13618 | 387.9221 | 16.91081 |
| n | 600 | 700 | 800 | 900 | 1000 |
| Mean of | 1.33154 | -8.73801 | -0.81247 | 1.53318 | 0.52701 |
| Variance of | 272.9102 | 518980.652 | 22146.2786 | 19.25389 | 5697.17336 |



（在不同樣本數下的樣本平均）



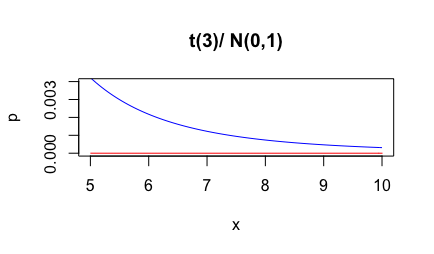
（在不同樣本數下的樣本變異數）

小結：

由important sampling with N(0,1)去估計t(3)的平均與變異數，在不同的抽樣樣本數下，產生極大的差異，估計平均值的部分相對於估計變異數來得準確，估計平均值大多落在0的上下，與t(3)實際的母體平均0蠻接近的。再來看估計變異數的部分，跳動範圍蠻大的，甚至出現變異數小於0的部分，與t(3)實際的母體變異數相差甚大，同時估計變異數的變異數跳動範圍也是很大。因此我認為用標準常態去估計t(3)並不是可行的方法。

(c)  Comment on the results.

比較(a)與(b)的部分，由t(3)去估計N(0,1)與N(0,1)去估計t(3)兩個差異很大，最主要差異性來自於兩個分配的尾巴的部分，t(3)尾巴較厚而N(0,1)尾巴較薄（下圖），而important sampling要估計的好時，要小於1(f(X)為原分配，g(x)為important function))，所以當t(3)/N(0,1)，尾巴的部分會因為N(0,1)的值太小，而讓比值過大，因而讓Variance變過大。因此，用important sampling with t(3)去估計N(0,1)是一個好的方法，但相反用important sampling with N(0,1)去估計t(3)則不是一個好的方法。



(上圖只截取一段尾巴，藍線為t(3)分配，紅線為N(0,1)分配)

**# Appendix**

**Code:** [**http://rpubs.com/YaPi/388887**](http://rpubs.com/YaPi/388887)