

Question 01

Find $\int_{-20}^{20} e^{|x|} dx$ by calculus techniques.

$$\int_{-20}^{20} e^{|x|} dx = \int_{-20}^0 e^x dx + \int_0^{20} e^{-x} dx = 2(1 - e^{-20}) \approx 1.9999999958776928$$

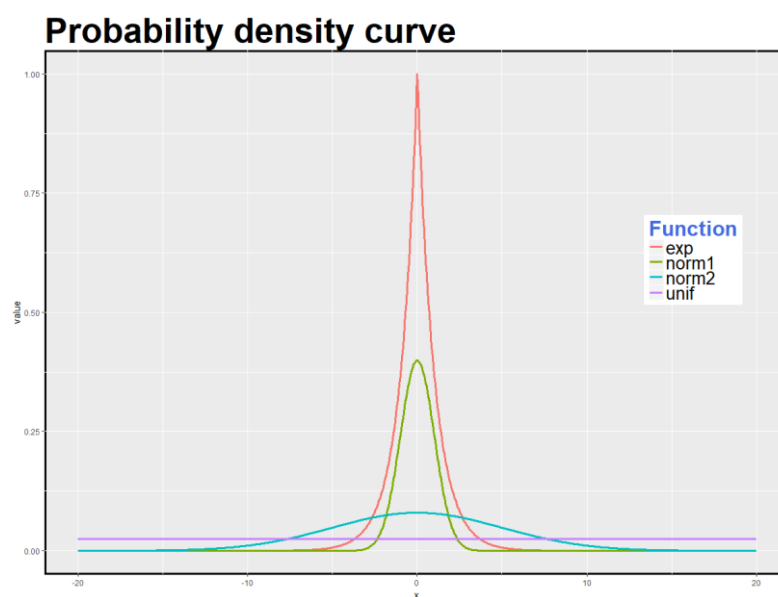
Simulation study:

- Take 5000 samples of size 1000.
- Find mean and variance of importance sampling estimates using proposal $\phi(x, 0, 1)$ and $\phi(x, 0, 5)$ (or more proposals).
Compare it with that by MC method.
- Which proposal gives a smaller variance? Give arguments to support your findings (use numbers and graphs).

每次抽取 1000 個樣本帶入估計，並進行 5000 次模擬
各方法結果如下：

	Monte Carlo	Important sampling N(0,1)	Important sampling N(0,5)
MEAN	2.0022490524053609	1.996765208848214	1.9991554029402279
Variance	0.037462248962476456	0.096102303949433782	0.0087440212903155488

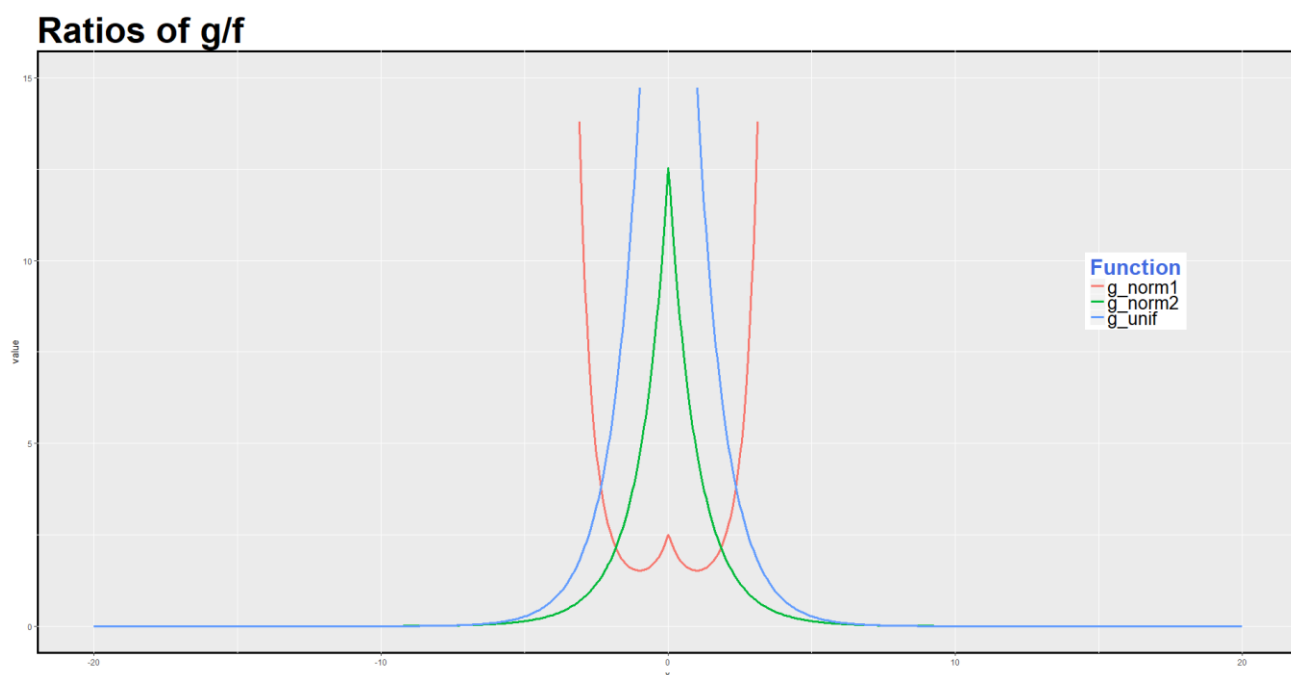
從上述表格我們發現當使用 Important sampling 的方法時，其變異數會因為選取重要性函數不同，而有不同的變異大小，接著我們將各函數繪製在同一張圖進行比較。



我們發現當選取重要函數為 $N(0, 1)$ 時估計的變異數比 Monte carlo 的變異數來的大，考慮下列式子：

$$\text{Var}_f(h(x)) - \text{Var}_g\left(\frac{h(x)f(x)}{g(x)}\right) = E_f(h^2(x)) - E_g\left(h^2(x)\frac{f^2(x)}{g^2(x)}\right) = E_g(h^2(x)(1 - \frac{f(x)}{g(x)}))$$

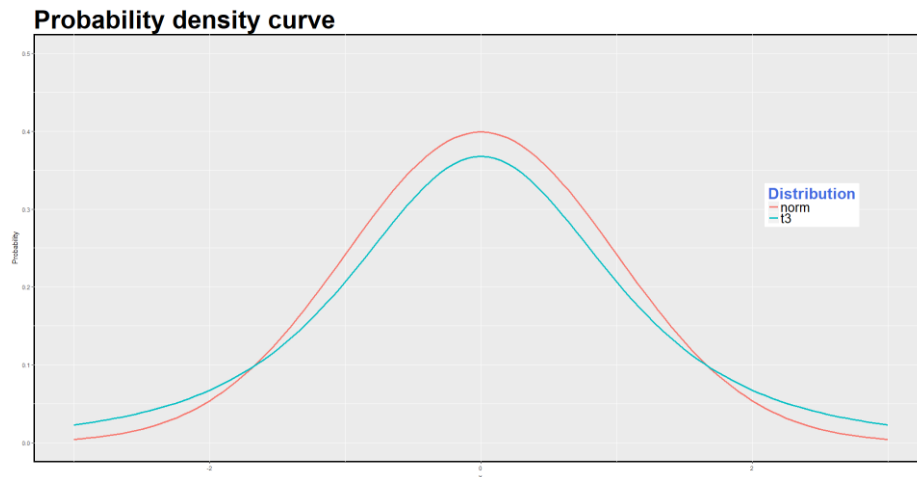
其中 $h(x) = e^{-|x|} - 20 < x < 20$, $f(x) = \frac{1}{20 - (-20)} = 0.025$, $g(x) \sim N(0, 1)$



在上圖中，我們可以看到原函數除以各個重要函數的圖形，上式變異數降低會發生在 $(1 - \frac{f(x)}{g(x)}) > 0$ 的時候，而當 $g(x) \sim N(0, 1)$ 時，從 Ratios of g/f 可以觀察 $g(x)$ 的曲線呈現 W 的形式，所以用此重要函數變異數不太會下降，而選擇重要函數為 $N(0, 5)$ 時，其 $g(x)$ 會呈現較穩定且平滑的曲線， $\frac{f(x)}{g(x)}$ 的比值是否小於 1，會影響到變異數是否能下降，而從此題中我們觀察到綠色線段較為穩，其值小於 1 的範圍最多，故能達到較好的變異數下降。

Question 02

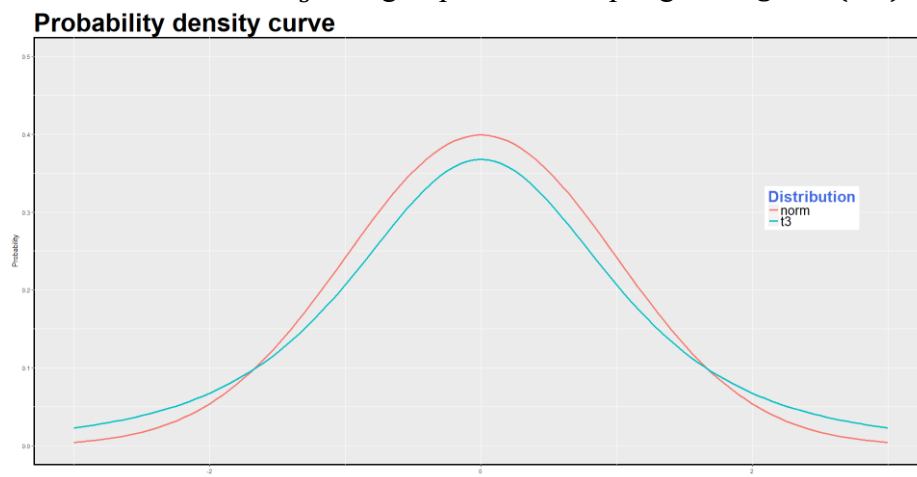
- (a) Obtain the mean and variance of $N(0,1)$ using importance sampling with $g \sim t_3$.



與上題所使用的方法一致，模擬後得到的表格如下：

	Important sampling $N(0,1)$	真實
MEAN	0.000450174	0
Variance	0.999723	1

- (b) Obtain the mean and variance of t_3 using importance sampling with $g \sim N(0,1)$



與上題所使用的方法一致，模擬後得到的表格如下：

	Important sampling t_3	真實
MEAN	-0.0216617	0
Variance	1.84003	3

(c) Comment on the result.



(d) 結論

當以 $N(0,1)$ 為重要性函數來估計 t_3 的期望值與變異數時，我們發現估計值的變異數會不穩定，而從兩者函數的ratios圖表觀看，我們也會發現右圖(b)的線段為較不平穩的線段，原因為 $f(x)/g(x)$ 的比值到一定的範圍後會變很大，而這也就代表我們在第2題使用 $N(0,1)$ 作為重要性函數估計時，估計的變異數會有不穩定的情形發生，而第1題ratio圖中的線段較為平穩，估計值的變異數就不容易發生不穩定的情形，適合使用此重要函數作為估計。