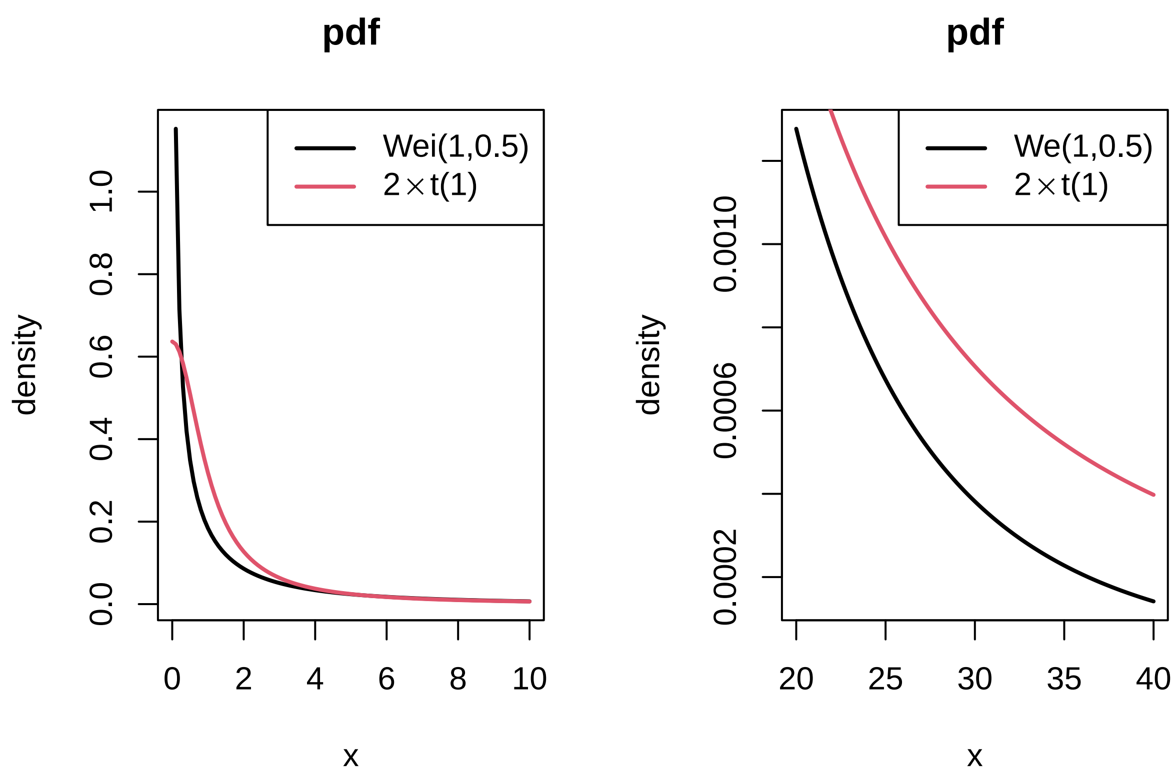
Statistical Computing Homework 2

110024516 邱繼賢

Problem 1.

$\underline{\eta = (1, 0.5)}$

使用 positive t(df=1) distribution 做為 *proposal pdf q(x)*，然後用 MH method 來抽選出樣本。



- target pdf *f(x)* 和 proposal pdf *q(x)* 有著相同的 support

- *q(x)* 的尾巴分佈比 *f(x)* 要來得厚

sampling scheme :

(i) 選定起始點 $x^{(1)} = 2$ 滿足 $f(x^{(1)}) > 0$

(ii) For $t = 2, 3, ..., 50000$

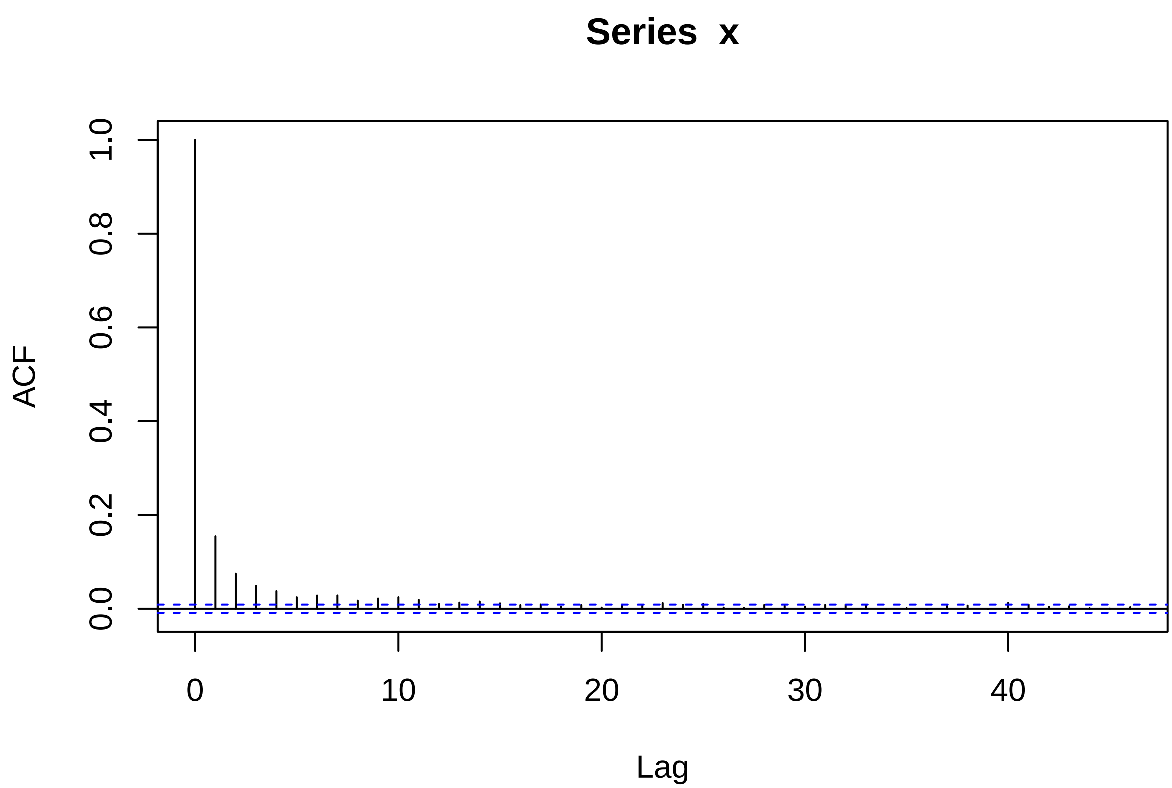
- draw x^* from proposal q(x)

- compute the ratio : $r = \frac{f(x^*)q(x^{(t-1)})}{f(x^{(t-1)})q(x^*)}$

- set

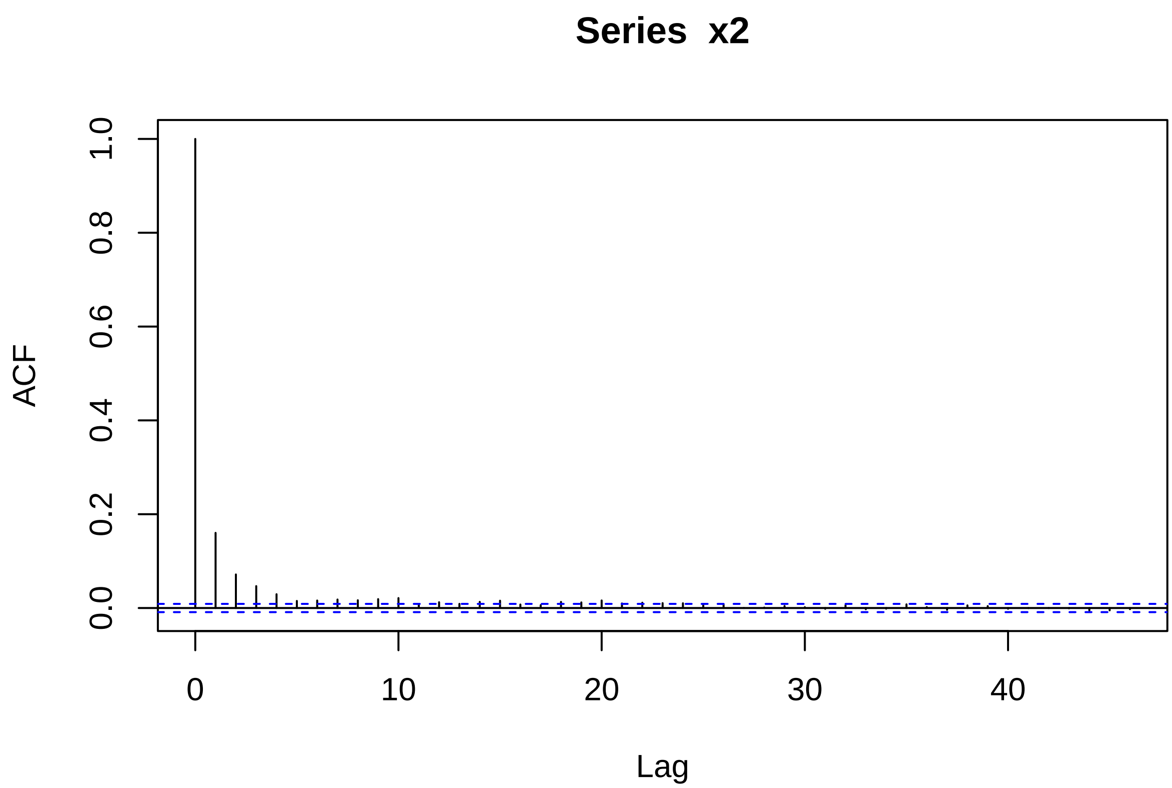
$$x = \begin{cases} x^* & , \text{ with probability } \; min\{1, r\} \\ x^{(t-1)} & , \text{ otherwise} \end{cases}$$

使用 ACF 來檢查抽出的 50000 個樣本 $\{x^{(1)}, x^{(2)}, ..., x^{(50000)}\}$ 之間的相關性：

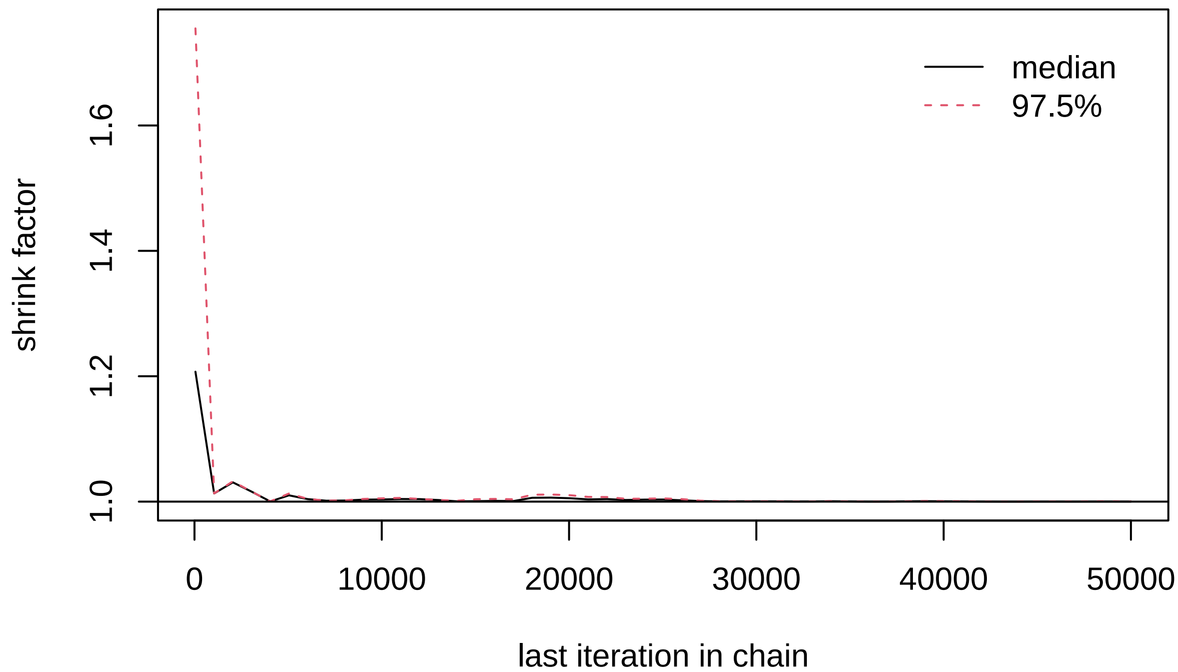



```
x2 = c()
x2[1] = 3
for (i in 2:n) {
    x2[i] = abs(rt(1,1))
    r = weibull(x2[i],1,0.5)/weibull(x2[i-1],1,0.5)*2*dt(x2[i-1],1)/(2*dt(x2[i],1))
```

```r
    succ = rbinom(1,1,min(1,r))
    if (succ == 0) x2[i] = x2[i-1]
}
acf(x2)
```

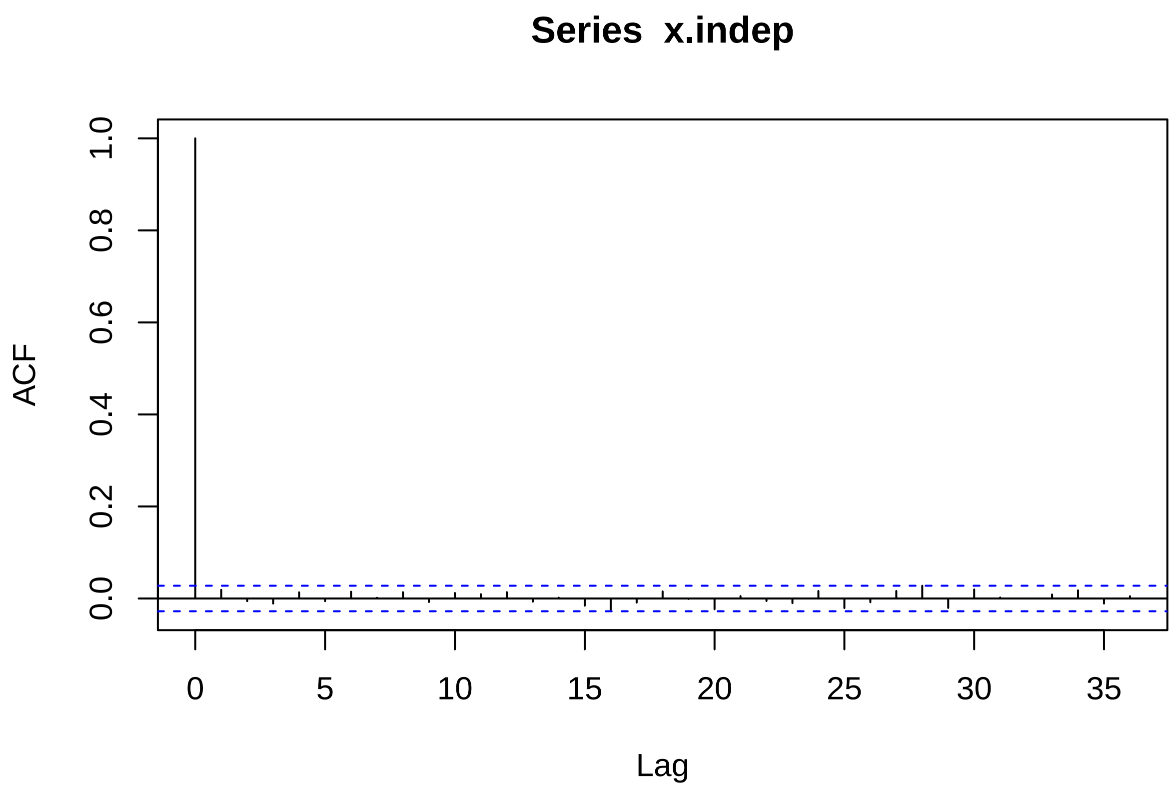



```r
library(coda)
mc.12 = mcmc.list(mcmc(x), mcmc(x2))
gelman.plot(mc.12)
```

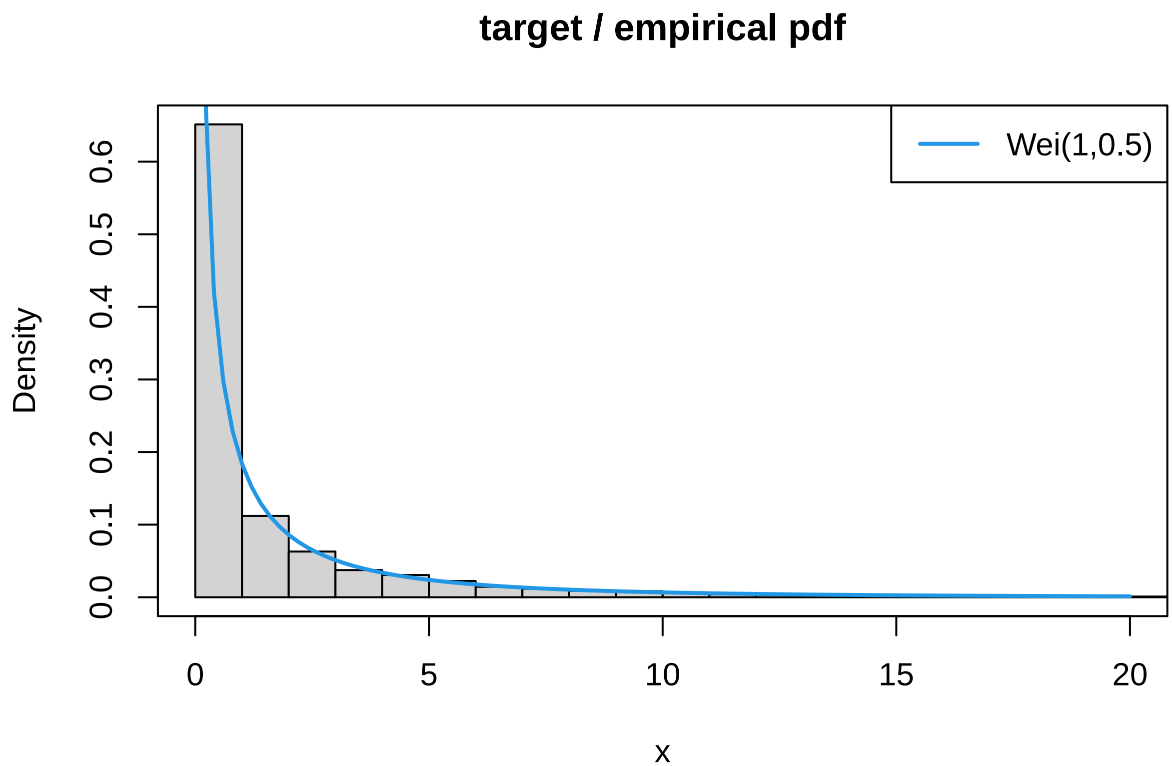


- 大概在間距 10 個樣本後，相關性就非常小，接近零了

在這 50000 筆樣本中，每 10 個就抽出一個做為新的獨立樣本，最後剩下 5000 筆樣本，一樣用 ACF 來檢查相關性：



可以看出樣本之間幾乎沒有相關性了，然後使用這 5000 筆樣本來跟 target pdf $f(x)$ 做比較：



可以看出模擬出的樣本和真實的 pdf 分佈相當一致，接下來以此 5000 筆樣本 $\left\{x^{(1)}, x^{(2)}, ..., x^{(5000)}\right\}$ 來計算以下數題：

(a) Use Monte Carlo method to estimate $(E(X), Var(X))$ by $(\hat{\mu}, \hat{\sigma}^2)$

$$\begin{cases} \hat{\mu} &= \dfrac{1}{5000} \displaystyle\sum_{i=1}^{5000} x^{(i)} &= 1.924 \\ \hat{\sigma}^2 &= \dfrac{1}{5000-1} \displaystyle\sum_{i=1}^{5000} \left[x^{(i)} - \hat{\mu}\right]^2 &= 19.486 \end{cases}$$

estimate kurtosis $E\left(\frac{X-EX}{\sqrt{VarX}}\right)^4$ by

$$\tilde{k} = \frac{1}{5000} \sum_{i=1}^{5000} \left[\frac{x^{(i)} - \hat{\mu}}{\hat{\sigma}}\right]^4 = \frac{1}{5000} \sum_{i=1}^{5000} k_i = 82.979$$

and compute the standard error of the estimation by

$$\sqrt{\hat{Var}(\tilde{k})} = \sqrt{\frac{1}{5000 \times 4999} \sum_{i=1}^{5000} \left[k_i - \tilde{k}\right]^2} = 33.631$$

$\hat{\mu}$	$\hat{\sigma}^2$	$\tilde{k}$	$s.e.(\tilde{k})$
1.924	19.486	82.979	33.631

(b) Fisher information matrix :

$$\begin{aligned}
I(\eta=(1,0.5)) &= -E\begin{bmatrix} \frac{\partial^2}{\partial\theta^2}logf(X) & \frac{\partial^2}{\partial\theta\partial\beta}logf(X) \\ \frac{\partial^2}{\partial\beta\partial\theta}logf(X) & \frac{\partial^2}{\partial\beta^2}logf(X) \end{bmatrix}_{(1,0.5)} \\
&= -E\begin{bmatrix} \frac{\beta}{\theta^2}-\frac{\beta(\beta+1)}{\theta^2}\left(\frac{x}{\theta}\right)^\beta & \frac{-1}{\theta}+\frac{1}{\theta}\left(\frac{x}{\theta}\right)^\beta+\frac{\beta}{\theta}\left(\frac{x}{\theta}\right)^\beta(logx-log\theta) \\ \frac{-1}{\theta}+\frac{1}{\theta}\left(\frac{x}{\theta}\right)^\beta+\frac{\beta}{\theta}\left(\frac{x}{\theta}\right)^\beta(logx-log\theta) & \frac{-1}{\beta^2}-(logx-log\theta)^2\left(\frac{x}{\theta}\right)^\beta \end{bmatrix}_{(1,0.5)} \\
&= E\begin{bmatrix} \frac{3}{4}x^{\frac{1}{2}}-\frac{1}{2} & 1-\left(1+\frac{1}{2}logx\right)x^{\frac{1}{2}} \\ 1-\left(1+\frac{1}{2}logx\right)x^{\frac{1}{2}} & 4+(logx)^2x^{\frac{1}{2}} \end{bmatrix}
\end{aligned}$$

which can estimated by

$$\hat{I}(\eta=(1,0.5)) = \frac{1}{5000}\begin{bmatrix} \sum_{i=1}^{5000}\left(\frac{3}{4}x^{(i)\frac{1}{2}}-\frac{1}{2}\right) & \sum_{i=1}^{5000}\left(1-\left(1+\frac{1}{2}logx^{(i)}\right)x^{(i)\frac{1}{2}}\right) \\ \sum_{i=1}^{5000}\left(1-\left(1+\frac{1}{2}logx^{(i)}\right)x^{(i)\frac{1}{2}}\right) & \sum_{i=1}^{5000}\left(4+\left(logx^{(i)}\right)^2x^{(i)\frac{1}{2}}\right) \end{bmatrix} = \begin{bmatrix} 0.228 & -0.368 \\ -0.368 & 7.211 \end{bmatrix}$$

and its standard error

$$s.e.(\hat{I}) = \begin{bmatrix} \hat{Var}\left(\hat{I}_{11}\right) & \hat{Var}\left(\hat{I}_{12}\right) \\ \hat{Var}\left(\hat{I}_{21}\right) & \hat{Var}\left(\hat{I}_{22}\right) \end{bmatrix} = \begin{bmatrix} 0.0105 & 0.0345 \\ 0.0345 & 0.1226 \end{bmatrix}$$

(c)

$$\begin{aligned}
I(\eta=(1,0.5)) &= E\begin{bmatrix} \frac{3}{4}x^{\frac{1}{2}}-\frac{1}{2} & 1-\left(1+\frac{1}{2}logx\right)x^{\frac{1}{2}} \\ 1-\left(1+\frac{1}{2}logx\right)x^{\frac{1}{2}} & 4+(logx)^2x^{\frac{1}{2}} \end{bmatrix} \\
&= \begin{bmatrix} \int_0^\infty\left(\frac{3}{4}x^{\frac{1}{2}}-\frac{1}{2}\right)f(x)dx & \int_0^\infty\left(1-\left(1+\frac{1}{2}logx\right)x^{\frac{1}{2}}\right)f(x)dx \\ \int_0^\infty\left(1-\left(1+\frac{1}{2}logx\right)x^{\frac{1}{2}}\right)f(x)dx & \int_0^\infty\left(4+(logx)^2x^{\frac{1}{2}}\right)f(x)dx \end{bmatrix} \\
&= \begin{bmatrix} 0.25 & -0.423 \\ -0.423 & 7.295 \end{bmatrix}
\end{aligned}$$

```
cuhre(int_11, lower=0, upper = 1000)$int
```

```
## [1] 0.25
```

```
cuhre(int_12, lower=0, upper = 1000)$int
```

```
## [1] -0.4227844
```

```r
cuhre(int_22, lower=0, upper = 1000)$int
```

```
## [1] 7.294722
```

(d) Draw the samples $w^{(1)}, w^{(2)}, ..., w^{(100)} \overset{iid}{\sim} Wei(\theta = 1, \beta = 0.5)$, and compute the empirical Fisher information

$$\hat{I}(\eta = (1, 0.5)) = \frac{1}{100}\begin{bmatrix} \sum_{i=1}^{100}\left(\frac{3}{4}w^{(i)\frac{1}{2}} - \frac{1}{2}\right) & \sum_{i=1}^{100}\left(1 - \left(1 + \frac{1}{2}logw^{(i)}\right)w^{(i)\frac{1}{2}}\right) \\ \sum_{i=1}^{100}\left(1 - \left(1 + \frac{1}{2}logw^{(i)}\right)w^{(i)\frac{1}{2}}\right) & \sum_{i=1}^{100}\left(4 + \left(logw^{(i)}\right)^2 w^{(i)\frac{1}{2}}\right) \end{bmatrix} = \begin{bmatrix} 0.265 & -0.505 \\ -0.505 & 8.222 \end{bmatrix}$$

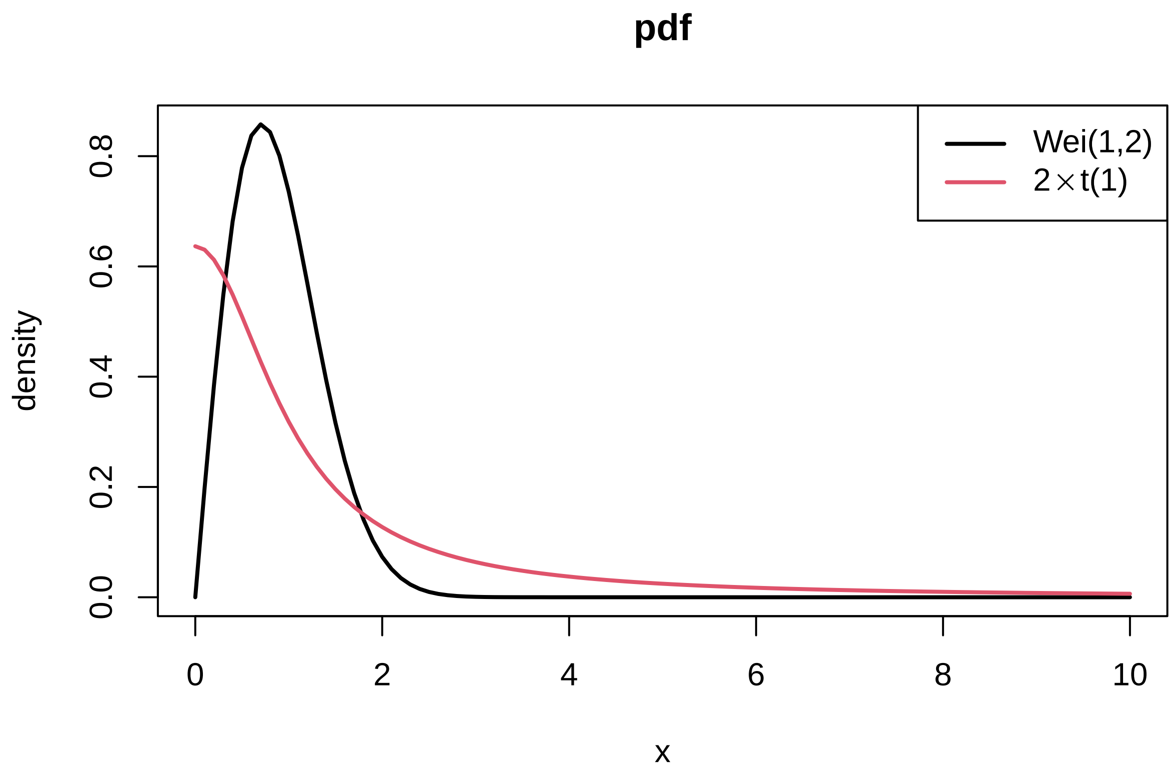
its standard error

$$s.e.(\hat{I}) = \begin{bmatrix} \hat{Var}\left(\hat{I}_{11}\right) & \hat{Var}\left(\hat{I}_{12}\right) \\ \hat{Var}\left(\hat{I}_{21}\right) & \hat{Var}\left(\hat{I}_{22}\right) \end{bmatrix} = \begin{bmatrix} 0.086 & 0.316 \\ 0.316 & 1.382 \end{bmatrix}$$

可以發現，雖然 $w^{(i)}$ 是直接從 Weibull distribution 中所抽出的樣本，但是他所計算出估計值的 standard error 和 bias 並沒有比起我們使用 MH method 所抽出的樣本計算出的來得小，其中最可能的原因是因為 $w^{(i)}$ 的樣本數比較小，只有 100 個，不像我們使用 MH method 時抽出了 5000 個樣本，由於 WLLN 和 CLT 的性質，使得 MH method 所計算出估計值的 standard error 和 bias 較為收斂。

$\eta = (1, 2)$

使用 positive t(df=1) distribution 做為 *proposal pdf q(x)*，然後用 MH method 來抽選出樣本。



- target pdf *f(x)* 和 proposal pdf *q(x)* 有著相同的 support

- *q(x)* 的尾巴分佈比 *f(x)* 要來得厚

sampling scheme :

(i) 選定起始點 $x^{(1)} = 1$ 滿足 $f(x^{(1)}) > 0$

(ii) For $t = 2, 3, ..., 50000$

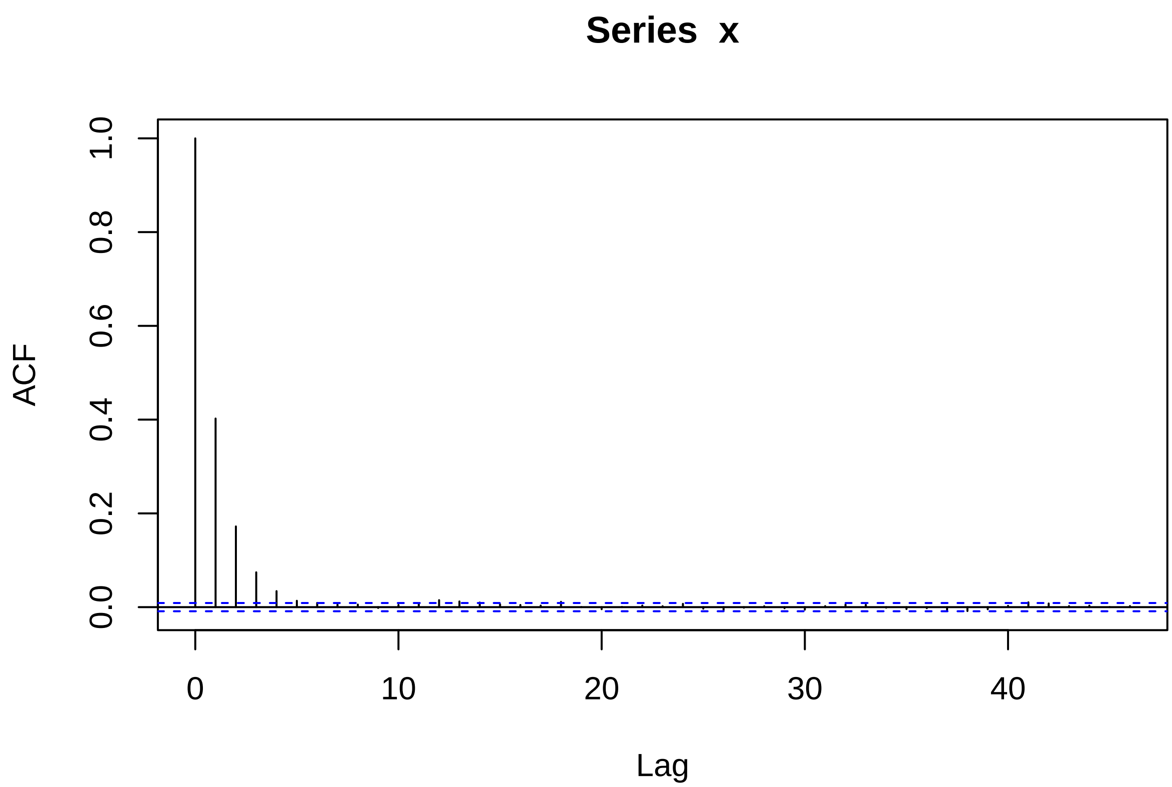
- draw x^* from proposal q(x)

- compute the ratio : $r = \frac{f(x^*)q(x^{(t-1)})}{f(x^{(t-1)})q(x^*)}$

- set

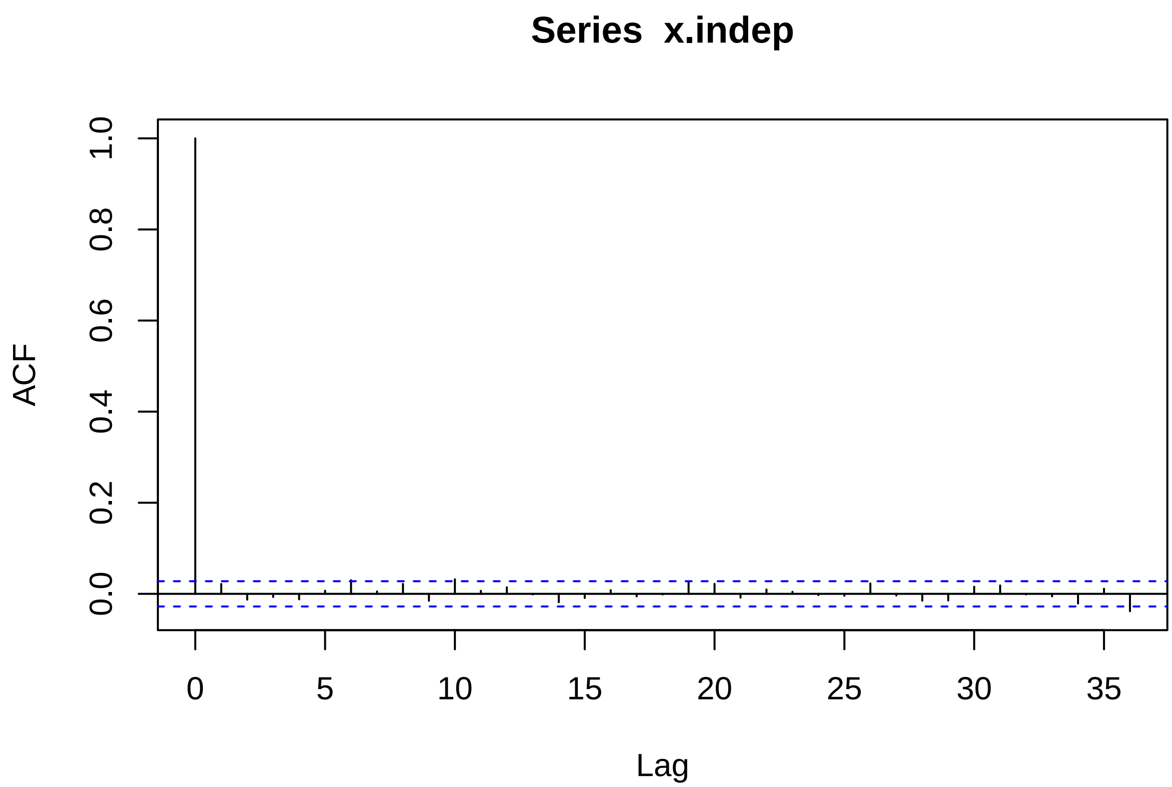
$$x = \begin{cases} x^* & , \text{ with probability } min\{1, r\} \\ x^{(t-1)} & , \text{ otherwise} \end{cases}$$

使用 ACF 來檢查抽出的 50000 個樣本 $\{x^{(1)}, x^{(2)}, ..., x^{(50000)}\}$ 之間的相關性：

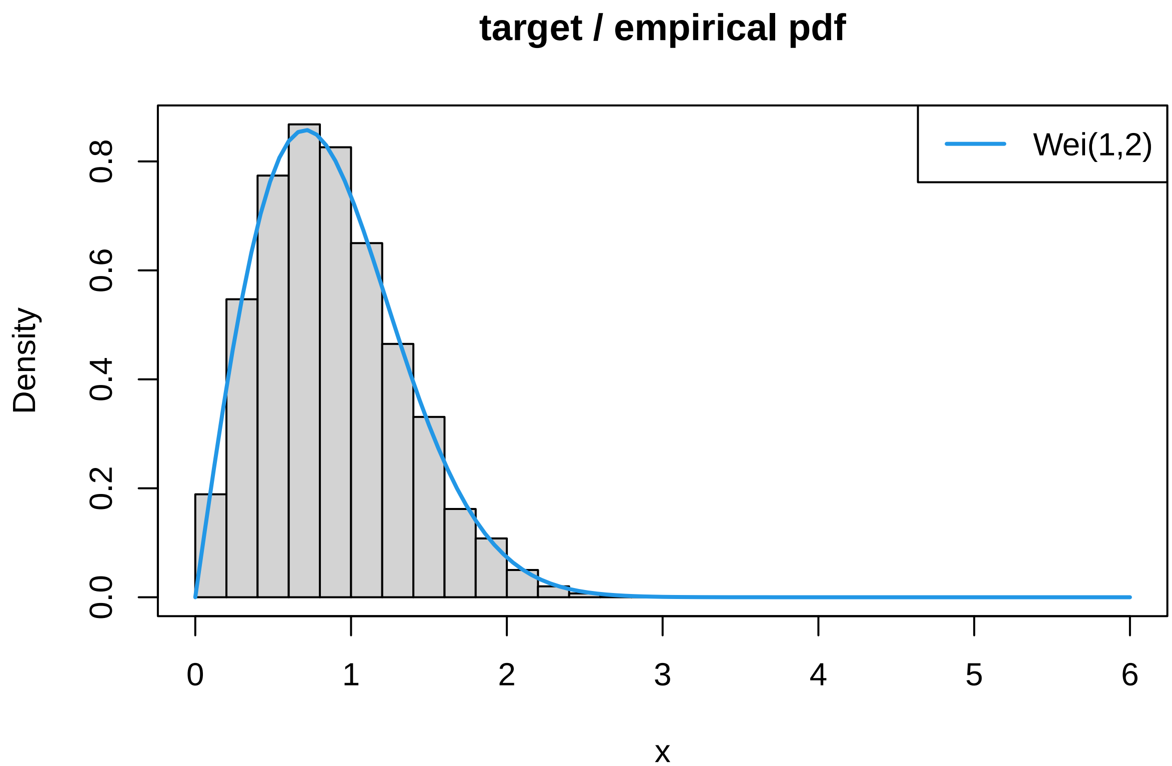


- 大概在間距 10 個樣本後，相關性就非常小，接近零了

在這 50000 筆樣本中，每 10 個就抽出一個做為新的獨立樣本，最後剩下 5000 筆樣本，一樣用 ACF 來檢查相關性：



可以看出樣本之間幾乎沒有相關性了，然後使用這 5000 筆樣本來跟 target pdf *f(x)* 做比較：



可以看出模擬出的樣本和真實的 pdf 分佈相當一致，接下來以此 5000 筆樣本 $\left\{x^{(1)}, x^{(2)}, ..., x^{(5000)}\right\}$ 來計算以下數題：

(a) Use Monte Carlo method to estimate $(E(X), Var(X))$ by $(\hat{\mu}, \hat{\sigma}^2)$

$$\begin{cases} \hat{\mu} & = \dfrac{1}{5000}\displaystyle\sum_{i=1}^{5000} x^{(i)} & = 0.88 \\ \hat{\sigma}^2 & = \dfrac{1}{5000-1}\displaystyle\sum_{i=1}^{5000} \left[x^{(i)} - \hat{\mu}\right]^2 & = 0.208 \end{cases}$$

estimate kurtosis $E\left(\frac{X-EX}{\sqrt{VarX}}\right)^4$ by

$$\tilde{k} = \frac{1}{5000}\sum_{i=1}^{5000}\left[\frac{x^{(i)} - \hat{\mu}}{\hat{\sigma}}\right]^4 = \frac{1}{5000}\sum_{i=1}^{5000} k_i = 3.213$$

and compute the standard error of the estimation by

$$\sqrt{\hat{Var}(\tilde{k})} = \sqrt{\frac{1}{5000 \times 4999}\sum_{i=1}^{5000}\left[k_i - \tilde{k}\right]^2} = 0.243$$

$\hat{\mu}$	$\hat{\sigma}^2$	$\tilde{k}$	$s.e.(\tilde{k})$
0.88	0.208	3.213	0.243

(b) Fisher information matrix :

$$
\begin{aligned}
I(\eta=(1,2)) &= -E\begin{bmatrix} \frac{\partial^2}{\partial\theta^2}logf(X) & \frac{\partial^2}{\partial\theta\partial\beta}logf(X) \\ \frac{\partial^2}{\partial\beta\partial\theta}logf(X) & \frac{\partial^2}{\partial\beta^2}logf(X) \end{bmatrix}_{(1,2)} \\
&= -E\begin{bmatrix} \frac{\beta}{\theta^2}-\frac{\beta(\beta+1)}{\theta^2}\left(\frac{x}{\theta}\right)^\beta & \frac{-1}{\theta}+\frac{1}{\theta}\left(\frac{x}{\theta}\right)^\beta+\frac{\beta}{\theta}\left(\frac{x}{\theta}\right)^\beta(logx-log\theta) \\ \frac{-1}{\theta}+\frac{1}{\theta}\left(\frac{x}{\theta}\right)^\beta+\frac{\beta}{\theta}\left(\frac{x}{\theta}\right)^\beta(logx-log\theta) & \frac{-1}{\beta^2}-(logx-log\theta)^2\left(\frac{x}{\theta}\right)^\beta \end{bmatrix}_{(1,2)} \\
&= E\begin{bmatrix} 6x^2-2 & 1-(1+2logx)\,x^2 \\ 1-(1+2logx)\,x^2 & \frac{1}{4}+(logx)^2x^2 \end{bmatrix}
\end{aligned}
$$

which can estimated by

$$
\hat{I}(\eta=(1,0.5)) = \frac{1}{5000}\begin{bmatrix} \sum_{i=1}^{5000}\left(6x^{(i)2}-2\right) & \sum_{i=1}^{5000}\left(1-\left(1+2logx^{(i)}\right)x^{(i)2}\right) \\ \sum_{i=1}^{5000}\left(1-\left(1+2logx^{(i)}\right)x^{(i)2}\right) & \sum_{i=1}^{5000}\left(\frac{1}{4}+\left(logx^{(i)}\right)^2x^{(i)2}\right) \end{bmatrix} = \begin{bmatrix} 3.889 & -0.371 \\ -0.371 & 0.444 \end{bmatrix}
$$

and its standard error

$$
s.e.(\hat{I}) = \begin{bmatrix} \hat{Var}\left(\hat{I}_{11}\right) & \hat{Var}\left(\hat{I}_{12}\right) \\ \hat{Var}\left(\hat{I}_{21}\right) & \hat{Var}\left(\hat{I}_{22}\right) \end{bmatrix} = \begin{bmatrix} 0.0823 & 0.0336 \\ 0.0336 & 0.0074 \end{bmatrix}
$$

(c)

$$
\begin{aligned}
I(\eta=(1,0.5)) &= E\begin{bmatrix} 6x^2-2 & 1-(1+2logx)\,x^2 \\ 1-(1+2logx)\,x^2 & \frac{1}{4}+(logx)^2x^2 \end{bmatrix} \\
&= \begin{bmatrix} \int_0^\infty(6x^2-2)\,f(x)dx & \int_0^\infty\left(1-(1+2logx)\,x^2\right)f(x)dx \\ \int_0^\infty\left(1-(1+2logx)\,x^2\right)f(x)dx & \int_0^\infty\left(\frac{1}{4}+(logx)^2x^2\right)f(x)dx \end{bmatrix} \\
&= \begin{bmatrix} 4 & -0.423 \\ -0.423 & 0.456 \end{bmatrix}
\end{aligned}
$$

```r
cuhre(int_11, lower=0, upper = 100)$int
```

```
## [1] 4
```

```r
cuhre(int_12, lower=0, upper = 100)$int
```

```
## [1] -0.4227843
```

```r
cuhre(int_22, lower=0, upper = 100)$int
```

```
## [1] 0.4559202
```

(d) Draw the samples $w^{(1)}, w^{(2)}, ..., w^{(100)} \overset{iid}{\sim} Wei(\theta = 1, \beta = 2)$, and compute the empirical Fisher information

$$\hat{I}(\eta = (1,2)) = \frac{1}{100}\begin{bmatrix} \sum_{i=1}^{100}\left(6w^{(i)2} - 2\right) & \sum_{i=1}^{100}\left(1 - \left(1 + 2logw^{(i)}\right)w^{(i)2}\right) \\ \sum_{i=1}^{100}\left(1 - \left(1 + 2logw^{(i)}\right)w^{(i)2}\right) & \sum_{i=1}^{100}\left(\frac{1}{4} + \left(logw^{(i)}\right)^2 w^{(i)2}\right) \end{bmatrix} = \begin{bmatrix} 3.527 & -0.224 \\ -0.224 & 0.414 \end{bmatrix}$$

its standard error

$$s.e.(\hat{I}) = \begin{bmatrix} \hat{Var}\left(\hat{I}_{11}\right) & \hat{Var}\left(\hat{I}_{12}\right) \\ \hat{Var}\left(\hat{I}_{21}\right) & \hat{Var}\left(\hat{I}_{22}\right) \end{bmatrix} = \begin{bmatrix} 0.539 & 0.211 \\ 0.211 & 0.036 \end{bmatrix}$$

可以發現，雖然 $w^{(i)}$ 是直接從 Weibull distribution 中所抽出的樣本，但是他所計算出估計值的 standard error 和 bias 並沒有比起我們使用 MH method 所抽出的樣本計算出的來得小，其中最可能的原因是因為 $w^{(i)}$ 的樣本數比較小，只有 100 個，不像我們使用 MH method 時抽出了 5000 個樣本，由於 WLLN 和 CLT 的性質，使得 MH method 所計算出估計值的 standard error 和 bias 較為收斂。

Problem 2.

選擇使用 Alternative Model II : $t(\nu)$ with df $\nu = \frac{1}{\theta}$, $0 < \theta \leq 1$

For $n = 20$ $(or\ 100)$, $M = 10000$, $\nu_i = 1, 2, 3, ..., 10, \infty$ $(\Leftrightarrow \theta_i = 1, \frac{1}{2}, \frac{1}{3}, .., \frac{1}{10}, 0)$

Algorithm :

1. 抽取 $\{X_k\}_{k=1}^{n} \overset{iid}{\sim} t(\nu_i)$，並計算 JB statistic

$$JB = \frac{S^2}{6/n} + \frac{(K-3)^2}{24/n}$$

where

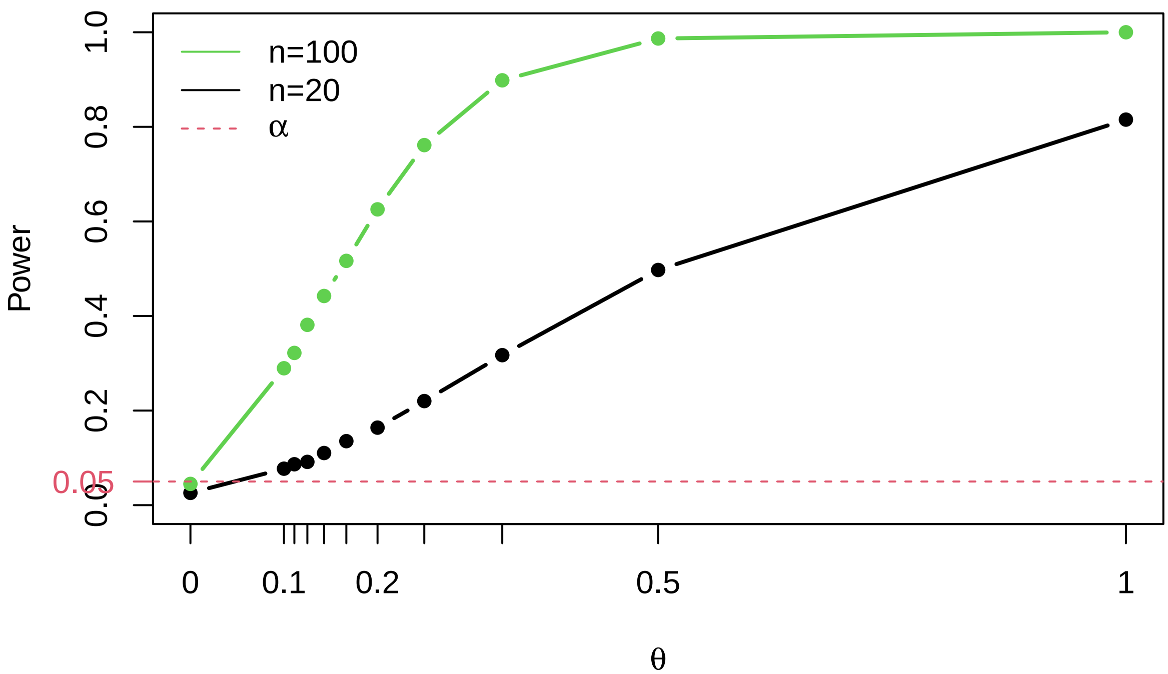
$$S = \frac{\widehat{\mu}_3}{\widehat{\sigma}^3} = \frac{\frac{1}{n}\sum_{k=1}^{n}(x_k - \bar{x})^3}{\left(\frac{1}{n}\sum_{k=1}^{n}(x_k - \bar{x})^2\right)^{3/2}}$$

$$K = \frac{\widehat{\mu}_4}{\widehat{\sigma}^4} = \frac{\frac{1}{n}\sum_{k=1}^{n}(x_k - \bar{x})^4}{\left(\frac{1}{n}\sum_{k=1}^{n}(x_k - \bar{x})^2\right)^{4/2}}$$

2. 重複以上步驟 M(10000) 次，並記錄下每次的 JB statistic 數值
3. 計算所有的 JB statistics 中，大於 $\chi^2_{1-\alpha, df=2}$ ，也就是 reject 的比例，即為 Monte Carlo estimator $\hat{\pi}(\theta_i)$
4. 計算 Monte Carlo s.e. $\sqrt{\frac{\hat{\pi}(\theta_i)(1-\hat{\pi}(\theta_i))}{M}}$
5. 對每個參數 $\nu_i = \frac{1}{\theta_i}$ 重複以上四個步驟

θ_i	MC.est(n=20)	MC.se(n=20)	MC.est(n=100)	MC.se(n=100)
0.000	0.026	0.002	0.045	0.002
0.100	0.077	0.003	0.290	0.005
0.111	0.086	0.003	0.322	0.005
0.125	0.092	0.003	0.381	0.005
0.143	0.110	0.003	0.442	0.005
0.167	0.135	0.003	0.517	0.005
0.200	0.164	0.004	0.625	0.005
0.250	0.220	0.004	0.761	0.004
0.333	0.317	0.005	0.898	0.003
0.500	0.497	0.005	0.987	0.001
1.000	0.815	0.004	1.000	0.000

Plot $\hat{\pi}(\theta_i)$ v.s. θ_i to sketch the power function



- 隨著 θ 的數值逐漸遠離 0，也就是 alternative distribution 越來越不像 *N(0,1)*，我們能檢定出這兩個分配不同的機率也越大，*power* 值就越大
- 將每次實驗抽取樣本數 n 從 20 增加到 100，我們能從 alternative distribution 中獲得的資訊也越多，能檢定出這兩個分配不同的機率也越大，*power* 值就越大