0413 作業

計財所 碩二 108071601 賴冠維

2021/4/12

Problem 1

(a)

從$Uniform(0,1)$抽樣計算積分式，所得平均數及變異數如下：

```
## [1] 0.52533583 0.05991423
```

若$f_0(x) = 1$則與原式相同，服從$Uniform(0,1)$,計算平均數及變異數如下：

$$\int_0^1 \frac{e^{-x}}{1+x^2}dx = \int_0^1 \frac{e^{-x}/1+x^2}{1} 1dx$$

```
## [1] "Mean of the values of simulation Integration: 0.524855942320723"

## [1] "Variance of the values of simulation Integration: 0.0601152584276106"
```

(b)

若$f_0(x) = exp(-x)$，則變成服從$Exp(1)$分配並且積分分配值域為[0,1]，計算其平均數及變異數如下：

$$\int_0^1 \frac{e^{-x}}{1+x^2}dx = \int_0^1 \frac{e^{-x}/1+x^2}{e^{-x}} e^{-x}dx = \int_0^1 \frac{1}{1+x^2} e^{-x}dx$$

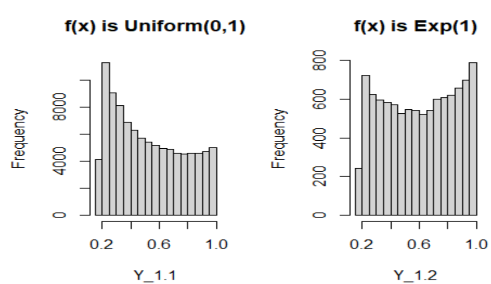
```
## [1] "Mean of the values of simulation Integration: 0.59917234298382"

## [1] "Variance of the values of simulation Integration: 0.0612878170271518
```

(c)

(a),(b)小題有何差別？

- (a)小題為計算$g(x) = Uniform(0,1)$均勻分配的積分式，而(b)小題則是計算$g(x) = Exp(1)$指數分配的積分式,可發現指數分配在機率分布上沒有均勻分配來的平穩，可能來自我們限制 x 的值域落在[0,1]之間，原指數分配的值域為$[0, \infty]$，僅擷取指數分配中的一小段,導致偏誤較大。



Problem 2

建立模擬樣本

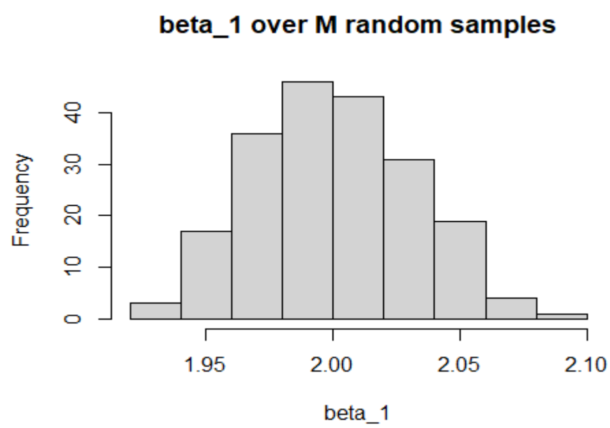
先根據$Y_i = \beta_0 + X_i\beta_1 + \epsilon_i$模擬數據

(a)

可由下圖發現β_1在$M = 200$大樣本下其估計出的 Variance 極低約為 0.001，從直方圖來看呈現鐘型曲線服從常態分配。

```
## [1] "Mean of beta_1: 2.00086840409782"

## [1] "Variance of beta_1: 0.00096465667958181"
```



(b)

可發現在估計β_1時，其值隨著每次模擬樣本的更動而有所不同，在這 200 個β_1的估計值中，全距大約 0.156 上下，而 Boostrap 後的平均數卻可讓β_1的估計值穩健地落在 2。

```
## [1] "Maximum in the estimator of beta_1: 2.08548410994247"

## [1] "Minimum in the estimator of beta_1: 1.92855244270411"

## [1] "The range of the estimator of beta_1: 0.156931667238352"
```

(c)

以$\hat{\theta} \sim N(\theta_0, n^{-1}\hat{\Sigma})$計算其變異數後平均，如下：

```
## [1] "Mean of the Asymptotic Variance of  beta_1: 0.0040090830103712"
```

計算$Empirical\ Variance$，算出 200 組$\widehat{\beta_1}$的變異數，如下：

```
## [1] "Mean of the Empirical Variance of  beta_1: 0.00096465667958181"
```

可發現兩者在樣本數很大的情況下，變異數都非常小。

```
## [1] "Mean of the Asymptotic Variance of  beta_1: 0.0040090830103712"

## [1] "Mean of the Empirical Variance of  beta_1: 0.00096465667958181"
```

(d)

第一種$Boostrap$為$Random\ X$的$Boostrap$，先從 200 組資料中，對每組的 500 筆資料重抽樣，可得新的 200 組樣本，對這 200 組$Boostrap\ Sample$計算$\widehat{\beta_1}$，計算其變異數，如下：

```
## [1] "Mean of the estimator of beta_1 by Random x boostrap: 2.00753959826958"

## [1] "Variance of the estimator of beta_1 by Random x boostrap: : 0.00353469030345969"
```

第二種$Boostrap$為$Fixed\ X$的$Boostrap$，先將原 200 組資料配適可得殘差ϵ，接著將殘差做 200 組重抽樣後，取出對應的 X 與其配適值，再將殘差加上配適值後得到新的Y^*，以(X, Y^*)計算$\widehat{\beta_1}$，接著計算其變異數，如下：

```
## [1] "Mean of the estimator of beta_1 by Fixed x boostrap: 1.99604171567323"

## [1] "Variance of the estimator of beta_1 by Fixed x boostrap: 0.00318864235815397"
```

(e)

$Perturbation\ Boostrap\ Sampling$為在$Likelihood$中多乘上一個機率分配$X$,並且$E(X) = 1, Var(x) = 1$，可得$\widehat{\beta_1}$的$MLE$為$\frac{YXG^{(b)}}{XX^T}$，因此將$Y$乘上$G^{(b)}$，計算如下：

(a)小題加上擾動項

```
## [1] "Variance of Perturbation_beta_1 (a): 0.0268585508738842"
```

(d)小題加上擾動項

```
## [1] "Variance of Perturbation_beta_1 (d) Random x: 0.0760376069539746"

## [1] "Variance of Perturbation_beta_1 (d) Fixed x: 0.0630605144545954"
```

觀察各題結果，發現使用$Perturbation\ Sampling$會使得變異數明顯大於其他方法所求，而其他方法在樣本數足夠大下，其變異數皆非常小。

```
## [1] "Variance of beta_1 in (a): 0.00096465667958181"

## [1] "Mean of the Asymptotic Variance of  beta_1 in (c): 0.0040090830103712"

## [1] "Mean of the Empirical Variance of  beta_1 in (c): 0.00096465667958181"

## [1] "Variance of the estimator of beta_1 by Random x boostrap in (d): 0.0035346903034"

## [1] "Variance of the estimator of beta_1 by Fixed x boostrap in (d): 0.00318864235815"

## [1] "Variance of Perturbation_beta_1 for (a) in (e): 0.0268585508738842"

## [1] "Variance of Perturbation_beta_1 for (d) Random x in (e): 0.0760376069539746"

## [1] "Variance of Perturbation_beta_1 for (d) Fixed x in (e): 0.0630605144545954"
```

附錄(程式碼)

```
library(tidyverse)

## Problem 1

### (a)

set.seed(123)

X = runif(100000,0,1)

Y = (exp(-X)/(1+X^2))

c( mean(Y), var(Y) )

w <- function(x) (exp(-X)/(1+X^2))/dunif(x, 0, 1)

f <- function(x) dunif(x,0,1)

X = runif(100000,0,1)

Y_1.1 = w(X)*f(X)

paste0("Mean of the values of simulation Integration: ",mean(Y_1.1))

paste0("Variance of the values of simulation Integration: ",var(Y_1.1))

### (b)

w <- function(x) 1/(1+x^2)

f <- function(x) dexp(x,rate = 1)

X = matrix(NA,nrow = 10000)

i = 1

while (i< 10001) {

 e = rexp(1,rate = 1)

 if (e<1) {

  X[i] = e

  i = i+1

 }

}

Y_1.2 = w(X)*f(X)

paste0("Mean of the values of simulation Integration: ",mean(Y_1.2))

paste0("Variance of the values of simulation Integration: ",var(Y_1.2))

par(mfrow=c(1,2))

hist(Y_1.1,main = "f(x) is Uniform(0,1)")

hist(Y_1.2,main = "f(x) is Exp(1)")
```

```
## Problem 2

set.seed(1234)

Simulation_Data = lapply(1:200, function(a){

  X = rnorm(n = 500,mean = 0,sd = sqrt(2))

  Y = 1+2*X+rnorm(n = 500,mean = 0,sd = 1)

  list(X,Y)

})

names(Simulation_Data) = paste0("Trial_",1:200)

### (a)

beta_1 = sapply(1:200, function(a){

  trial = Simulation_Data[[a]]

  trial_lm = lm(trial[[2]]~trial[[1]])

  beta_1 = trial_lm$coefficients[[2]]

  beta_1

})

paste0("Mean of beta_1: ",mean(beta_1))

paste0("Variance of beta_1: ",var(beta_1))

hist(beta_1,main = "beta_1 over M random samples")

paste0("Maximum in the estimator of beta_1: ",max(beta_1))

paste0("Minimum in the estimator of beta_1: ",min(beta_1))

paste0("The range of the estimator of beta_1: ",max(beta_1)-min(beta_1))

### (c)

Var_MLE = sapply(1:200, function(a){

  x = Simulation_Data[[a]][[1]]

  y = Simulation_Data[[a]][[2]]

  (sum(y*x)/sum(x*x))/length(y)

})

paste0("Mean of the Asymptotic Variance of  beta_1: ",mean(Var_MLE))

a = sapply(1:200, function(a){

  LM = lm(Simulation_Data[[a]][[2]]~Simulation_Data[[a]][[1]])

  beta_1 = LM$coefficients[[2]]

  beta_1
```

```
})
paste0("Mean of the Empirical Variance of  beta_1: ",var(a))
paste0("Mean of the Asymptotic Variance of  beta_1: ",mean(Var_MLE))
paste0("Mean of the Empirical Variance of  beta_1: ",var(a))
### (d)
set.seed(12345)
Bst_Data = lapply(1:200, function(a){
 index = sample(1:500,200,replace = T)
 d = cbind(X = Simulation_Data[[a]][[1]][index],
       Y=Simulation_Data[[a]][[2]][index]) %>%
  as.data.frame()
})
Obs_Bst_Var = sapply(1:200, function(a){
 L = lm(Bst_Data[[a]]$Y~Bst_Data[[a]]$X)
 var = L$coefficients[[2]]
 var
})
paste0("Mean of the estimator of beta_1 by Random x boostrap: ",mean(Obs_Bst_Var))
paste0("Variance of the estimator of beta_1 by Random x boostrap: : ",var(Obs_Bst_Var))
set.seed(12345)
LM = lapply(1:200, function(a){
 Origin_LM = lm(Simulation_Data[[a]][[2]]~Simulation_Data[[a]][[1]])
 list(res = Origin_LM$residuals,
    fit = Origin_LM$fitted.values,
    x = Origin_LM$model$`Simulation_Data[[a]][[1]]`)
})
Res_Bst_Data = lapply(1:200, function(a){
 BS_index = sample(1:500,200,replace = T)
 index = sample(1:500,200,replace = F)
 res = LM[[a]]$res[BS_index]
 x = LM[[a]]$x[index]
 Y_1 = LM[[a]]$fit[index]+res
```

```
  cbind(x,Y_1) %>% as.data.frame()
})
Res_Bst_Var = sapply(1:200, function(a){
  L = lm(Res_Bst_Data[[a]]$Y_1~Res_Bst_Data[[a]]$x)
  var = L$coefficients[[2]]
  var
})
paste0("Mean of the estimator of beta_1 by Fixed x boostrap: ",mean(Res_Bst_Var))
paste0("Variance of the estimator of beta_1 by Fixed x boostrap: ",var(Res_Bst_Var))
# (e)
set.seed(12345)
per_beta_1 = sapply(1:200, function(a){
  trial = Simulation_Data[[a]]
  x = rexp(n = 500,rate = 1)
  per_x = trial[[1]]
  per_y = trial[[2]]*x
  trial_lm = lm(per_y~per_x)
  beta_1 = trial_lm$coefficients[[2]]
  beta_1
})
paste0("Variance of Perturbation_beta_1 (a): ",var(per_beta_1))
set.seed(12345)
Per_Bst_Data = lapply(1:200, function(a){
  index = sample(1:500,200,replace = T)
  d = cbind(X = Simulation_Data[[a]][[1]][index],
          Y=Simulation_Data[[a]][[2]][index]*rexp(n = 200,rate = 1)) %>%
    as.data.frame()
})
Per_Obs_Bst_Var = sapply(1:200, function(a){
  L = lm(Per_Bst_Data[[a]]$Y~Per_Bst_Data[[a]]$X)
  var = L$coefficients[[2]]
  var
```

```
})
paste0("Variance of Perturbation_beta_1 (d) Random x: ",var(Per_Obs_Bst_Var))
set.seed(12345)
LM = lapply(1:200, function(a){
  Origin_LM = lm(Simulation_Data[[a]][[2]]~Simulation_Data[[a]][[1]])
  list(res = Origin_LM$residuals,
       fit = Origin_LM$fitted.values,
       x = Origin_LM$model$`Simulation_Data[[a]][[1]]`)
})
Per_Res_Bst_Data = lapply(1:200, function(a){
  BS_index = sample(1:500,200,replace = T)
  index = sample(1:500,200,replace = F)
  res = LM[[a]]$res[BS_index]
  x = LM[[a]]$x[index]
  Y_1 = (LM[[a]]$fit[index]+res)*rexp(n = 200,rate = 1)
  cbind(x,Y_1) %>% as.data.frame()
})
Per_Res_Bst_Var = sapply(1:200, function(a){
  L = lm(Per_Res_Bst_Data[[a]]$Y_1~Per_Res_Bst_Data[[a]]$x)
  var = L$coefficients[[2]]
  var
})
paste0("Variance of Perturbation_beta_1 (d) Fixed x: ",var(Per_Res_Bst_Var))
paste0("Variance of beta_1 in (a): ",var(beta_1))
paste0("Mean of the Asymptotic Variance of  beta_1 in (c): ",mean(Var_MLE))
paste0("Mean of the Empirical Variance of  beta_1 in (c): ",var(a))
paste0("Variance of the estimator of beta_1 by Random x boostrap in (d): ",var(Obs_Bst_Var))
paste0("Variance of the estimator of beta_1 by Fixed x boostrap in (d): ",var(Res_Bst_Var))
paste0("Variance of Perturbation_beta_1 for (a) in (e): ",var(per_beta_1))
paste0("Variance of Perturbation_beta_1 for (d) Random x in (e): ",var(Per_Obs_Bst_Var))
paste0("Variance of Perturbation_beta_1 for (d) Fixed x in (e): ",var(Per_Res_Bst_Var))
```