Question 1

Importance Sampling

```
n=1000
b =1/2
inv_cdf_mh_p<-function(x){
  if(x >= 0.5){
    w = -log(2-2*x)}
  if(x < 0.5){
    w = log(2*x)
    }
  x1_sample_list<-c(x1_sample_list,-w)
}
x1_sample<-c()
for (i in 1:n) {
  x1_sample<-c(x1_sample,inv_cdf_mh_p(runif(1)))
}
x2_sample<-rnorm(n, mean = 0, sd = b)

f_muti<-function(x1,x2){
  exponentional<- (1/2*exp(-abs(x1)))
  normal_dist<- (1/sqrt(2*pi*b**2)*exp(-x2^2/(2*b**2)))
  normal_dist*exponentional
}

h_muti<-function(x1,x2){
  A =as.numeric(x1^2+x2>=0)
  cos(x1+x2)*exp(-(abs(x1)+2*x2^2))*A}

mean_a<-mean(h_muti(x1_sample,x2_sample)/f_muti(x1_sample,x2_sample))
```

```
> mean_a
[1] 0.6902987
```

MCMC – Random Walker

```r
aux_p<-function(x1, x2){
  part1<- (-abs(x1))
  part2<-(-2*x2^2)
  exp(part1+part2)
}

q1_mcmc<-function(x1_init,x2_init,step_size, iteration){
  x1<-x1_init
  x2<-x2_init
  x1_all<-c(x1)
  x2_all<-c(x2)
  p_1<-aux_p(x1,x2)

  for (i in 1:iteration) {
    y1<-rnorm(1,x1,step_size)
    y2<-rnorm(1,x2,step_size)
    p_2<-aux_p(y1,y2)
    accept_prob<-min(1,p_2/p_1)
    u =  runif(1)
    if( accept_prob > u){
      x1<- y1
      x2<- y2
      p_1<-p_2
      x1_all<-c(x1_all, x1)
      x2_all<-c(x2_all, x2)
    }

    if( accept_prob <= u){
      x1<- x1
      x2<- x2
      p_1<-p_1
      x1_all<-c(x1_all, x1)
      x2_all<-c(x2_all, x2)
    }
  }

  return(cbind(x1_all,x2_all))
}
expect<-function(x1,x2){
  cos(x1+x2)*as.numeric(x1^2+x2>=0)
}

x_input1<-1
x_input2<-2
n=1000
a<-q1_mcmc(x_input1,x_input2,1,n)


z= 2 * sqrt(pi/2)
mean(expect(a[,'x1_all'],a[,'x2_all']))*z
```

```
> mean(expect(a[,'x1_all'],a[,'x2_all']))*z
[1] 0.7169394
```

Question 2

```
q2_target<-function(x,s1,s2,g1){
  part1<-exp(-x^2/(2*s1^2))
  part2<-exp(-(x-g)^2/(2*s2^2))
  part1 + part2
}

gradient_p<-function(x,s1,s2,g1){
  # outter dervative for log
  part1 <- 1/q2_target(x,s1,s2,g1)

  # inner dervative
  part2_1 <- (-x/s1^2)*exp(-x^2/(2*s1^2))
  part2_2 <- exp(-(x-g1)^2/(2*s2^2))*(-(x-g1)/s2^2)
  part2<- part2_1 + part2_2

  #together
  part1*part2
}

propose_a_given_b<- function(a,b,step_size,s1,s2,g1){
  part1<-1/sqrt(2*pi*2*step_size^2)
  inner_part2_1<- (a-(b+step_size^2*gradient_p(b,s1,s2,g1)))^2
  inner_part2_2<- 2*2*step_size^2
  part2<- exp(-inner_part2_1/inner_part2_2)
  part1*part2
}

q2_mala<-function(x1_init,s1,s2,g1,step_size,iteration){
  x1_all_2<-c(x1_init)
  x1<-x1_init
  p_1<-q2_target(x1,s1,s2,g1)

  accept_mala = 0
  for (i in 1:iteration) {
    y1<-rnorm(1,x1+gradient_p(x1,s1,s2,g1)*step_size^2, 2*step_size^2)
    p_2<-q2_target(y1,s1,s2,g1)
    accept_prob<-min(1,(p_2*propose_a_given_b(x1,y1,step_size,s1,s2,g1))/(p_1*propose_a_given_b(y1,x1,step_size,s1,s2,g1)))
    u<-runif(1)

    if( accept_prob >= u){
      accept_mala = accept_mala+1
      print(accept_mala)
      x1 <- y1
      p_1 <-p_2
      x1_all_2<-c(x1_all_2, x1)
    }
    if( accept_prob < u){
      x1 <- x1
      p_1<-p_1
      x1_all_2<-c(x1_all_2, x1)
    }

  }

  return(x1_all_2)
}
```

```r
q2_random_walker<-function(x1_init,s1,s2,g1,step_size,iteration){
  x1_all_2<-c(x1_init)
  x1<-x1_init
  p_1<-q2_target(x1,s1,s2,g1)

  for (i in 1:iteration) {
    y1<-rnorm(1,x1, step_size)
    p_2<-q2_target(y1,s1,s2,g1)
    accept_prob<-min(1,(p_2)/(p_1))
    u<-runif(1)
    if( accept_prob >= u){
      accept_random_walker = accept_random_walker+1
      print(accept_random_walker)
      x1<- y1
      p_1<-p_2
      x1_all_2<-c(x1_all_2, x1)
    }
    if( accept_prob < u){
      x1<- x1
      p_1<-p_1
      x1_all_2<-c(x1_all_2, x1)
    }}
  return(x1_all_2)}
```

Functions to graphs

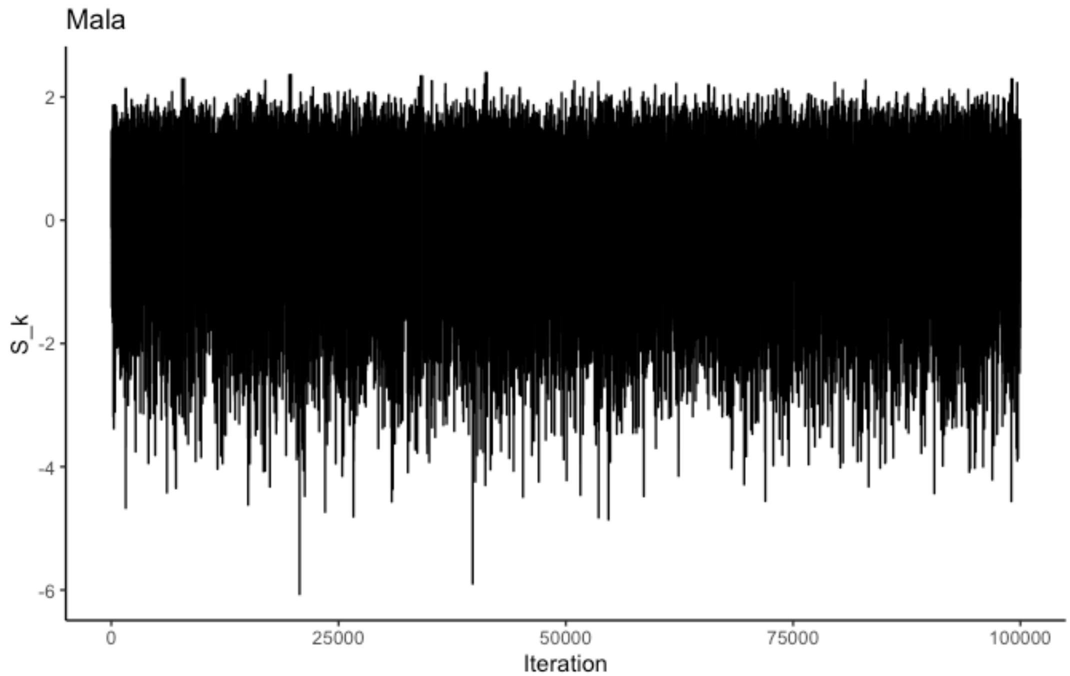
```r
b<-q2_mala(init_point,s1,s2,g,step_size, n2)
c<-q2_random_walker(init_point,s1,s2,g,step_size, n2)
b<-as.data.frame(b)
c<-as.data.frame(c)
```

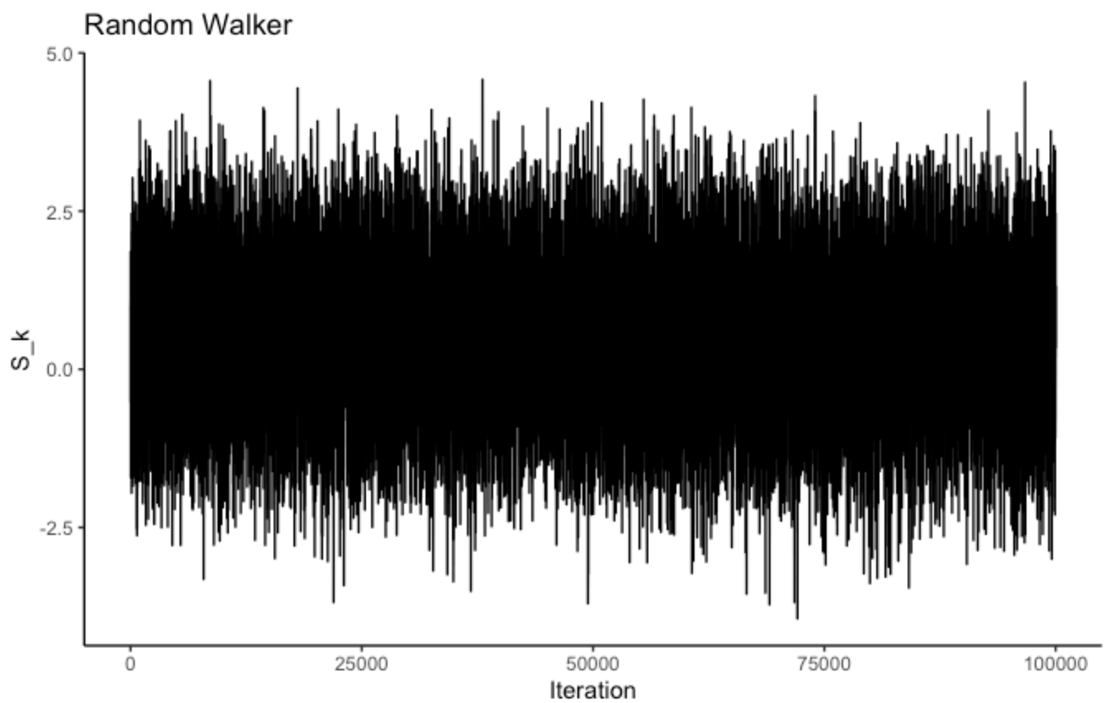
```r
library(ggplot2)
ggplot(data=b, aes(x=1:length(b), y=b, group=1)) +
  geom_line()+ theme_classic()+
  labs(title="Mala",x="Iteration", y = "S_k")

hist(b$b,main = "Mala",breaks = 50)


ggplot(data=c, aes(x=1:length(c), y=c, group=1)) +
  geom_line()+ theme_classic()+
  labs(title="Random Walker",x="Iteration", y = "S_k")

hist(c$c,main = "Random Walker",breaks = 50)
```

Setting 1

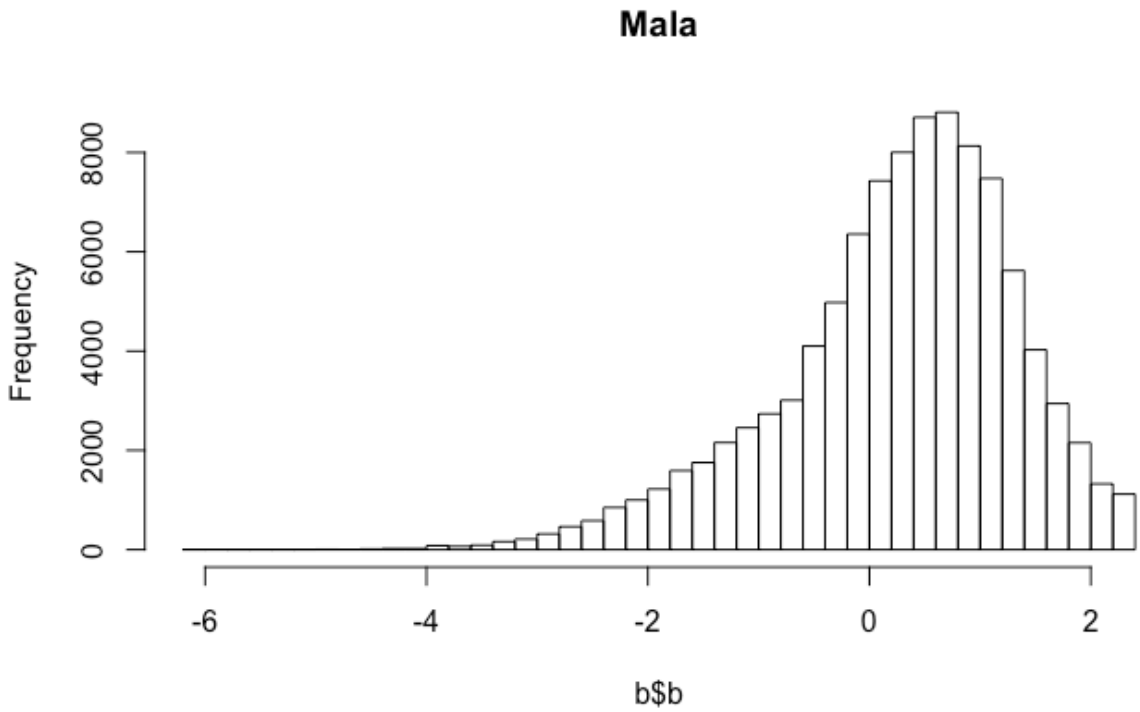
```r
s1<-1
s2<-1
g<-1
step_size<-1
n2<-10^5
init_point<-1
b<-q2_mala(init_point,s1,s2,g,step_size, n2)
# accpted points: 33452
c<-q2_random_walker(init_point,s1,s2,g,step_size, n2)
# accpted points: 73306
```

Trace plot

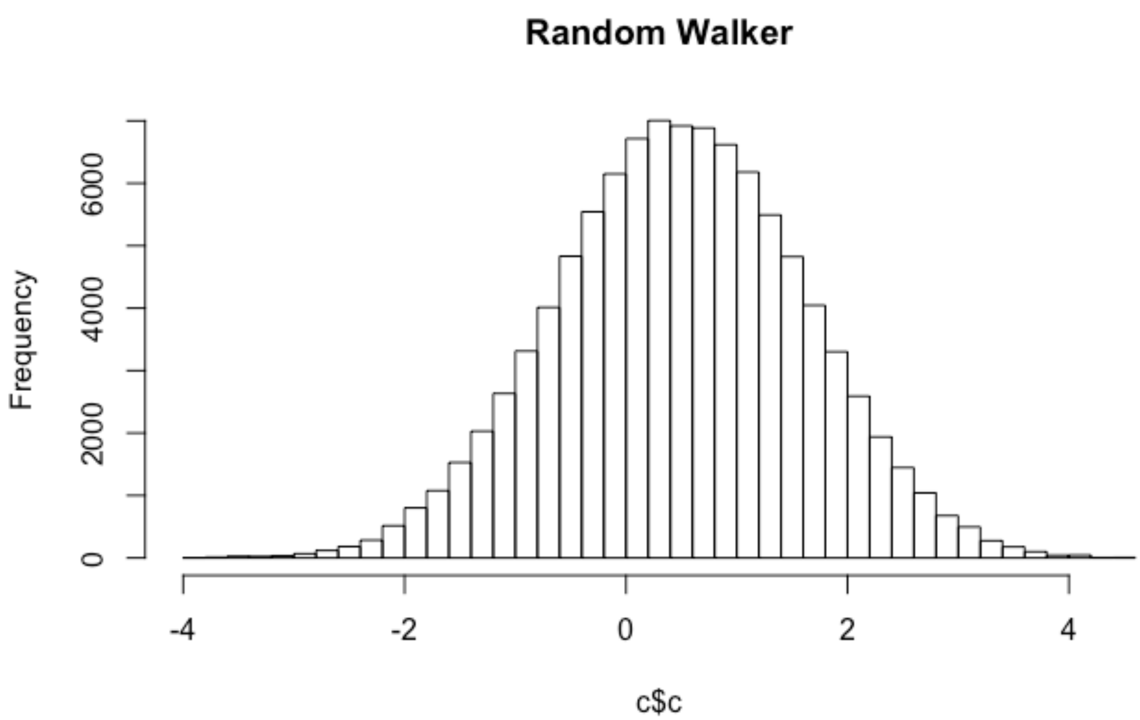

Mala
2
0
-2
-4
-6
S_k
0
25000
50000
75000
100000
Iteration


Random Walker
5.0
2.5
0.0
-2.5
S_k
0
25000
50000
75000
100000
Iteration

Histogram



Mala
Frequency
0
2000
4000
6000
8000
-6
-4
-2
0
2
b$b

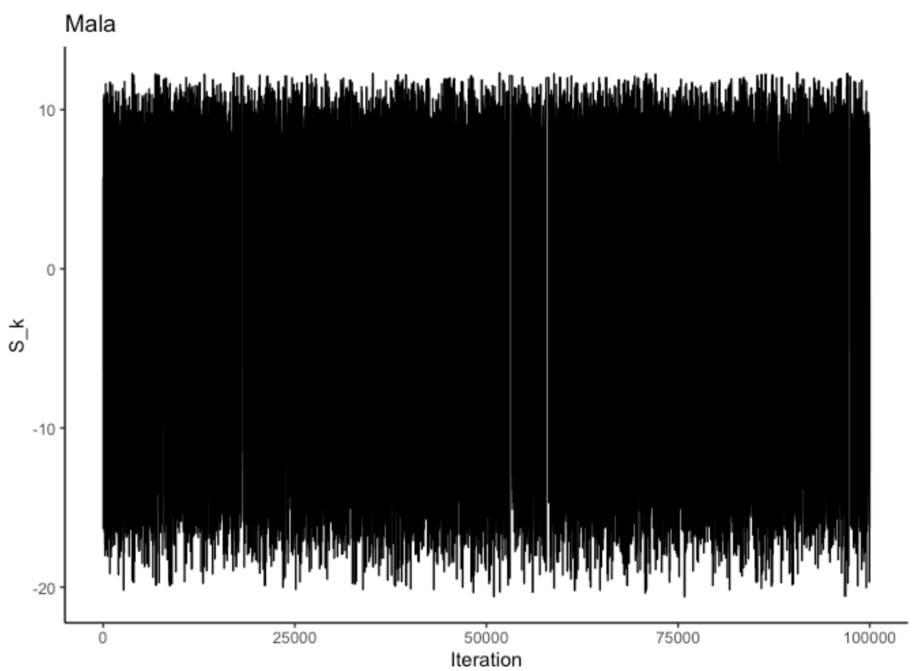


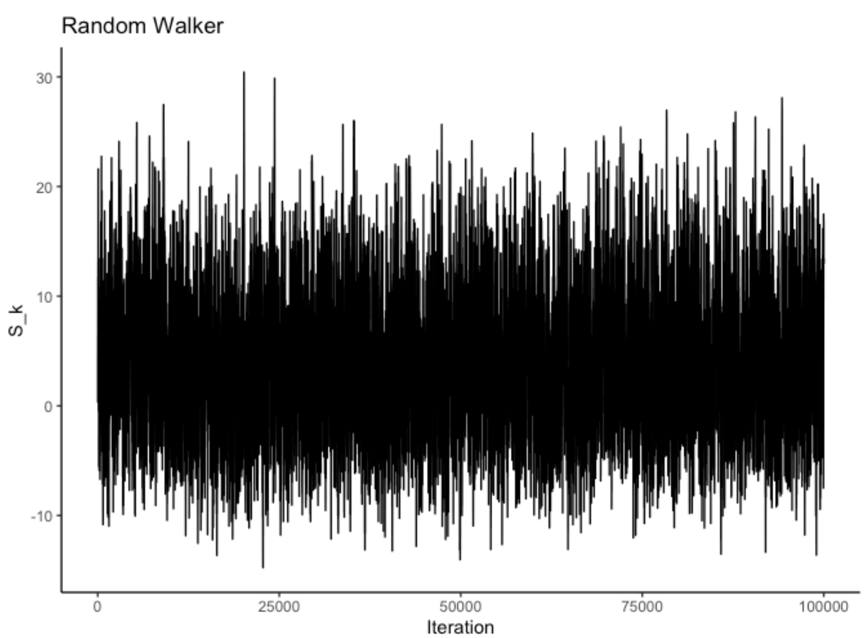
Random Walker
Frequency
0
2000
4000
6000
-4
-2
0
2
4
c$c

Setting 2

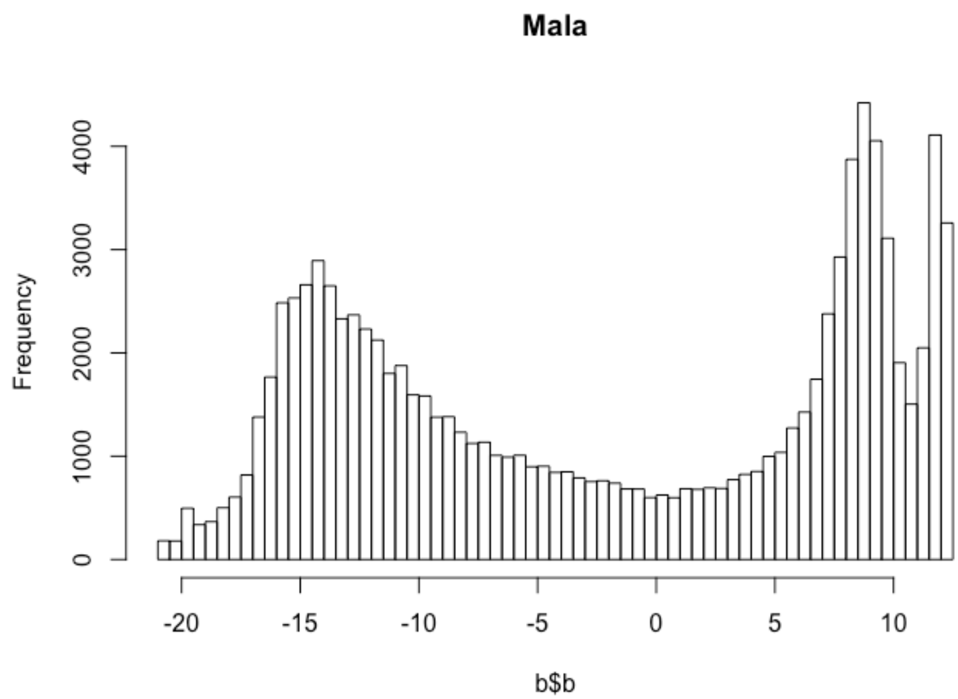
```
s1<-4
s2<-6
g<-7
step_size<-2
n2<-10^5
init_point<-0.3
b<-q2_mala(init_point,s1,s2,g,step_size, n2)
# accpted points: 41342
c<-q2_random_walker(init_point,s1,s2,g,step_size, n2)
# accpted points: 89875
```

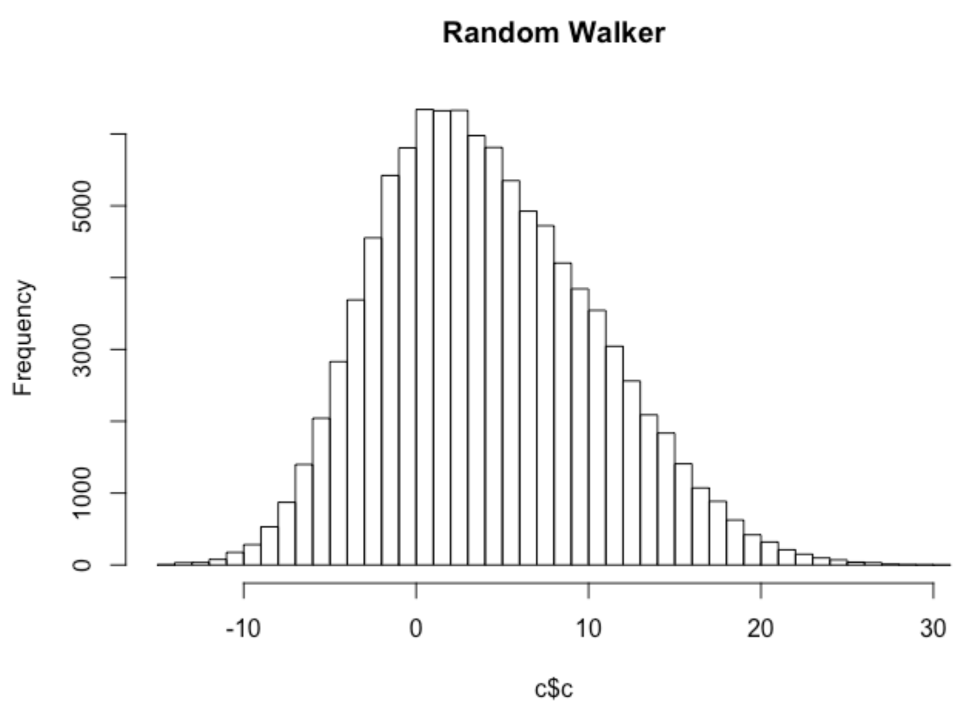
Trace plot





Histogram

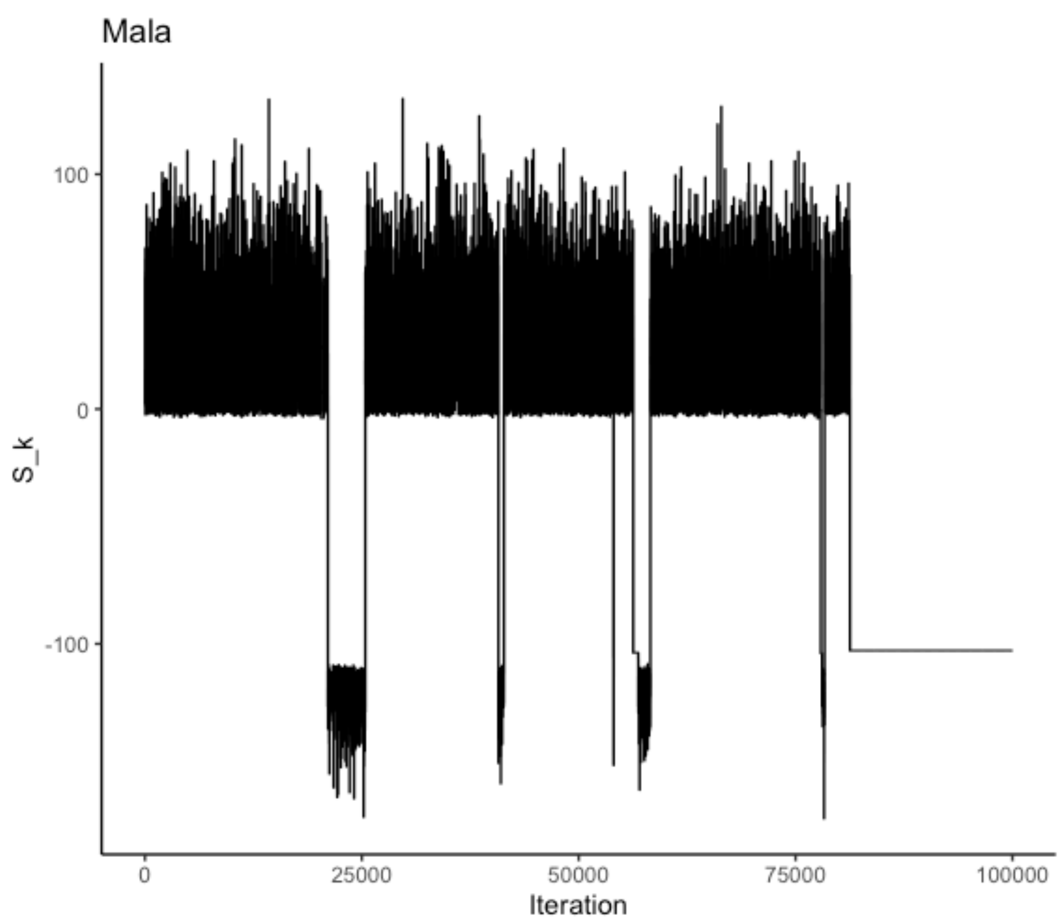

Mala
Frequency
0
1000
2000
3000
4000
-20
-15
-10
-5
0
5
10
b$b

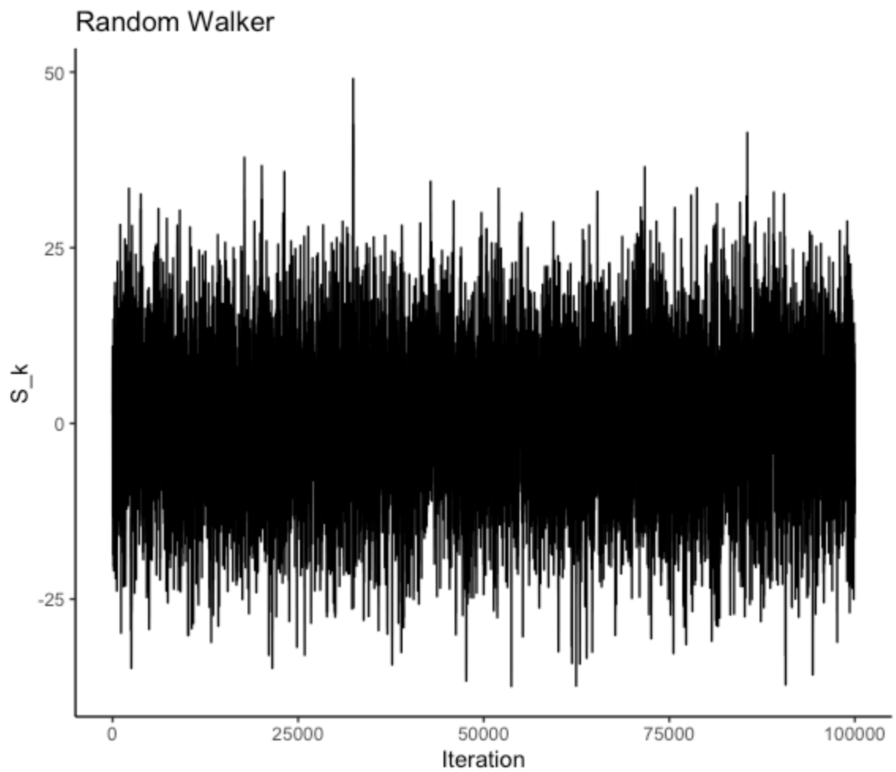

Random Walker
Frequency
0
1000
3000
5000
-10
0
10
20
30
c$c

Setting 3

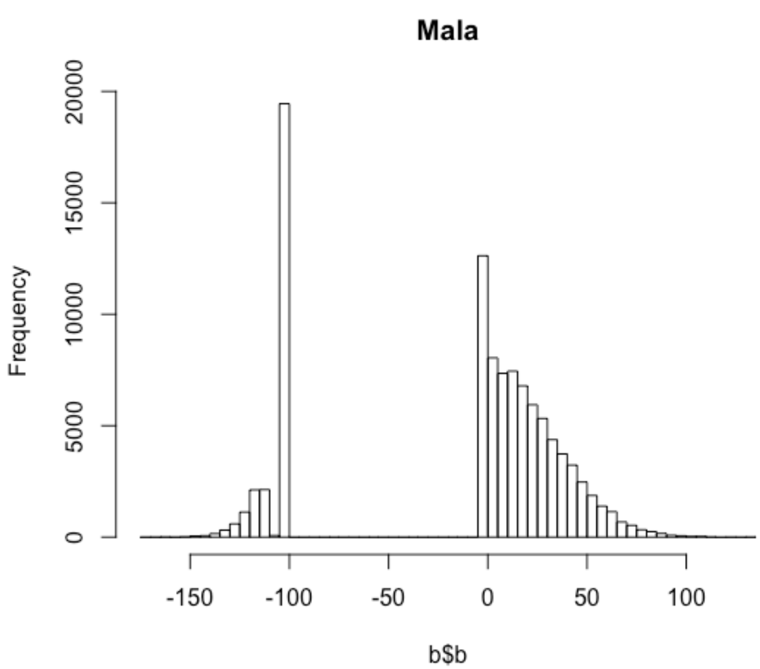
```
s1<-10
s2<-1
g<-7
step_size<-4
n2<-10^5
init_point<-2
b<-q2_mala(init_point,s1,s2,g,step_size, n2)
# accpted points: 31761
c<-q2_random_walker(init_point,s1,s2,g,step_size, n2)
# accpted points: 83864
```

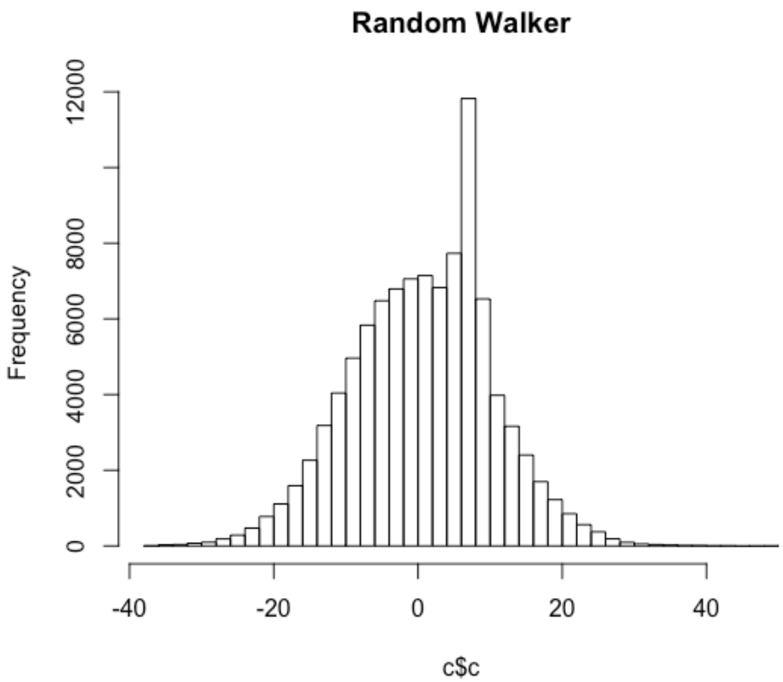
Trace plot




Random Walker
S_k
50
25
0
-25
0
25000
50000
75000
100000
Iteration

Histogram


Mala
Frequency
0
5000
10000
15000
20000
-150
-100
-50
0
50
100
b$b



How should your histograms look like?

There should be two peaks. S1 and S2 are similar to variance and g is how separate those two peaks should be.

Which algorithms seem to be mixing faster?

It seems like Mala mixing faster because the acceptance rate is lower.

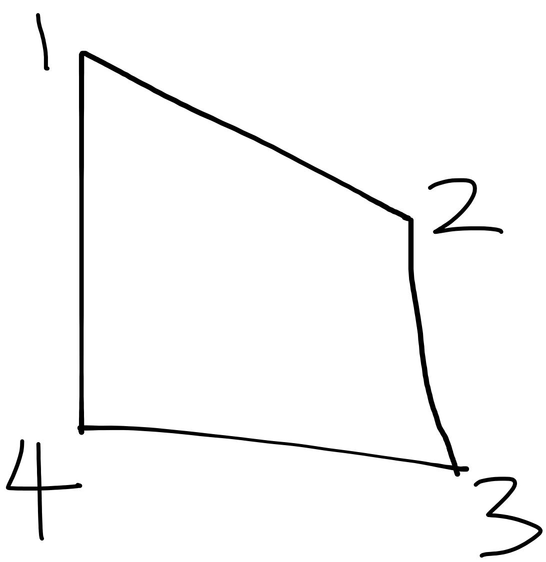
Question 3

Part 1

```
total = c()

for (x1 in 1:2) {
  for (x2 in 1:2) {
    for (x3 in 1:2) {
      for (x4 in 1:2) {
        t<-((x1**x2)*(x2**x3)*(x3**x4)*(x4**x1))
        total<-c(total,t)
      }
    }
  }
}

sum(total)
 > sum(total)
 [1] 433
```

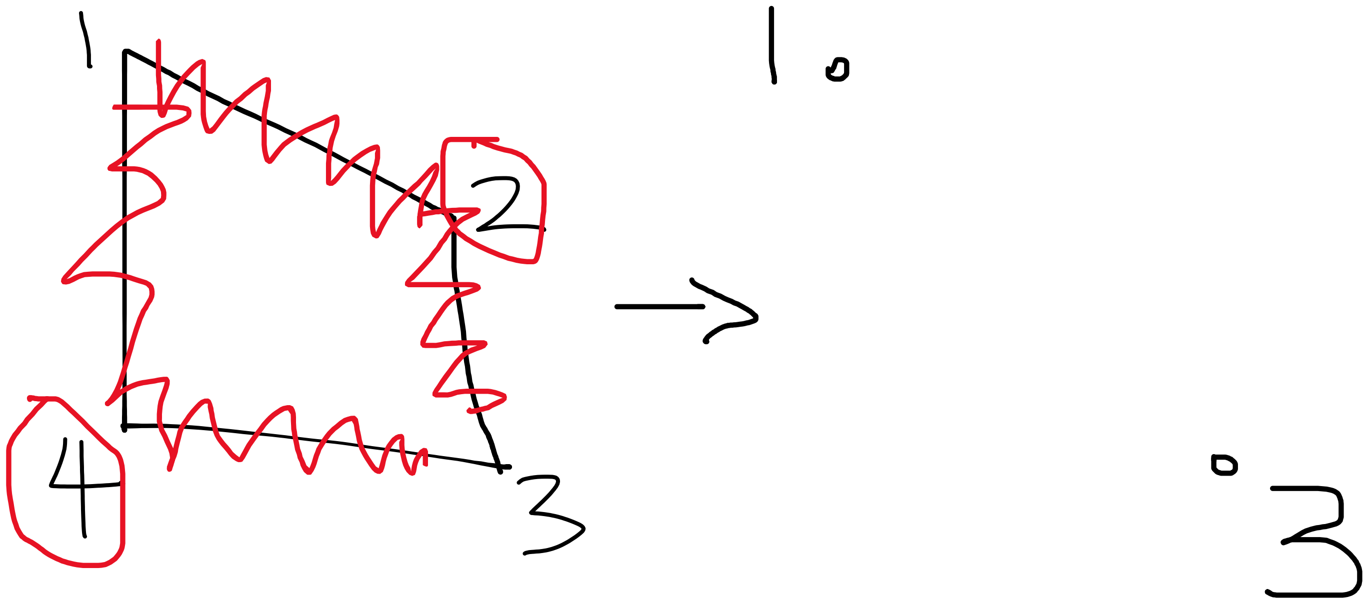
Z = 433; The purpose of Z is for 1/Z*PMF = 1

Part 2



Part 3

Yes, they are independent as we can use global Markov property to deduce this result that X2 and X4 forms a separating set which allows conditional independence among X1 and X3. Please refers to the picture below for visual.



Part 4

(1) $P(X_1 = x_1, X_3 = x_3 \mid X_2 = 1, X_4 = 2)$

$= (2)\ P(X_1 = x_1 \mid X_2 = 1, X_4 = 2)\, P(X_3 = x_3 \mid X_2 = 1, X_4 = 2)$

→ If they are conditionally independent

$$(1) \Leftrightarrow \frac{P(X_1 = x_1, X_3 = x_3, X_2 = 1, X_4 = 2)}{P(X_2 = 1, X_4 = 2)} = \frac{\frac{1}{433} X_1 X_3^2 \cdot 2}{\frac{1}{433} \sum_{x_1} \sum_{x_3} X_1 X_3^2 \cdot 2}$$

$$= \frac{X_1 X_3^2}{1 \cdot 1 + 1 \cdot 2^2 + 2 \cdot 1^2 + 2 \cdot 2^2} = \frac{X_1 X_3^2}{15}$$

$$(2) \Leftrightarrow \frac{P(X_1 = x_1, X_2 = 1, X_4 = 2)}{P(X_2 = 1, X_4 = 2)} \cdot \frac{P(X_3 = x_3, X_2 = 1, X_4 = 2)}{P(X_2 = 1, X_4 = 2)}$$

$$= \frac{\sum_{x_3} X_1 X_3^2}{15} \cdot \frac{\sum_{x_1} X_1 X_3^2}{15} = \frac{5X_1}{15} \cdot \frac{3X_3^2}{15} = \frac{X_1 X_3^2}{15}$$

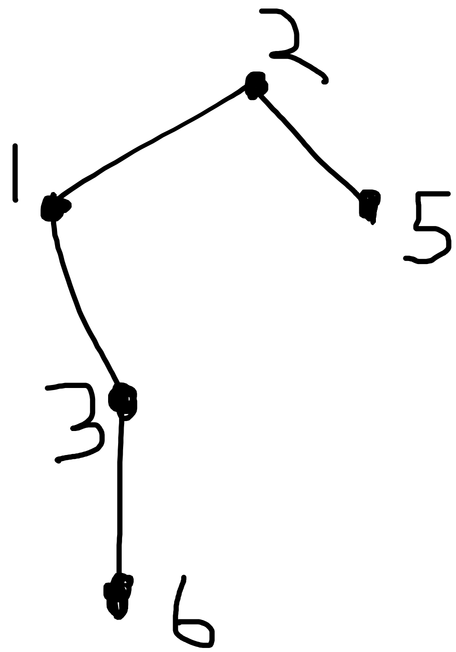
$\boxed{\text{Thus}}$ we proved (1) = (2)

$\boxed{\text{So}}$ This means X_1 and X_3 is conditionally independent given X_2 and X_4

Question 4

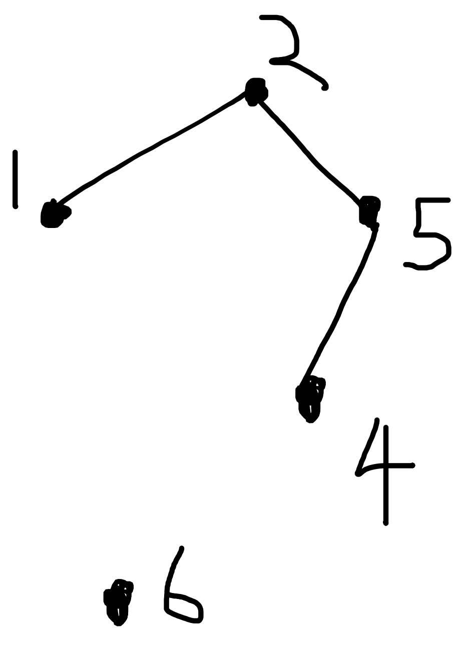
Part A

No, X5 and X6 is not independent given X4 as X4 is not a separating set for X5 and X6.



Part B

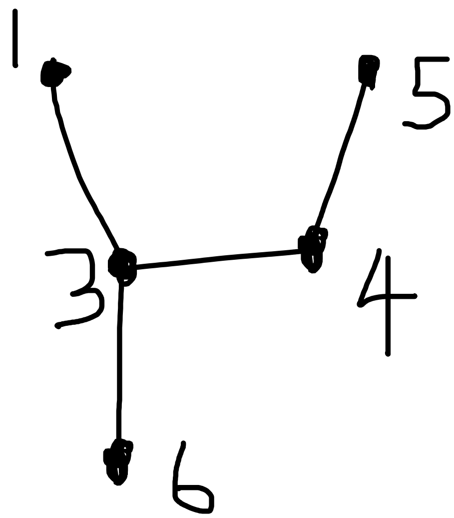
Yes, X5 and X6 is independent given X3 according to the Global Markov property. As X3 is a separating set for X5 and X6.



Part C

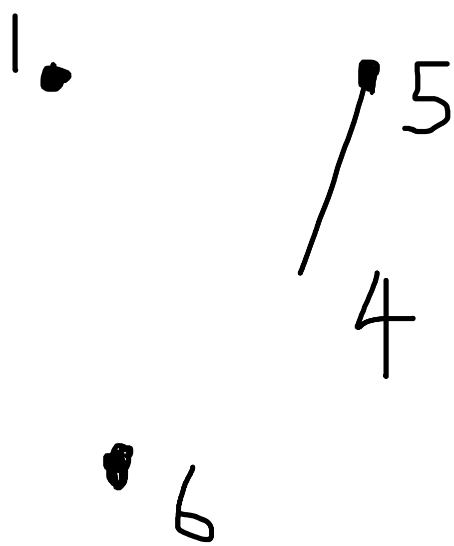
No, we cannot make statement regarding conditionally dependence with the Global Markov property.

`



Part D

Yes, as X2 and X3 is a separating set for X1 and X4. We can use Global Markov property.

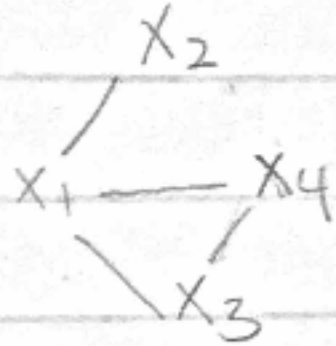


Question 5

$x = \{X_1, X_2, X_3, X_4\}$

1) $X_2 \perp\!\!\!\perp X_3 \mid X_4\ X_1$

2) $X_2 \perp\!\!\!\perp X_4 \mid X_3 X_1$



Appendix

Calculation for gradient log(p)

$$p(x) \propto \exp(-x^2/2\sigma_1^2) + \exp(-(x-r)^2/2\sigma_2^2)$$

$$\rightarrow \nabla \log(p(x))$$

$$\overbrace{\frac{1}{\exp(-x^2/2\sigma_1^2) + \exp(-(x-r)^2/2\sigma_2^2)}}^{\boxed{\text{part 1}}} \cdot \overbrace{\frac{d}{dx} \exp(-x^2/2\sigma_1^2) + \exp(-(x-r)^2/2\sigma_2^2)}^{\boxed{\text{part 2}}}$$

$$\text{part 1} \times \cdot \left(\frac{-x}{\sigma_1^2} \exp(-x^2/2\sigma_1^2) + \exp(-(x-r)^2/2\sigma_2^2)\left(\frac{-(x-r)}{\sigma_2^2}\right)\right)$$