

Metropolis-Hastings

Seung Ah Ha, Jaymo Kim, Wonbin Song

Metropolis-Hastings

Our target distribution is $\text{Beta}(6,4)$ and our proposal distribution is $\phi_{prop}|\phi_{old} \sim \text{Beta}(c\phi_{old}, c(1-\phi_{old}))$, for some constant c . First, we built several functions which will be used in calculating posterior distribution. Here, we defined likelihood function which gives the log-likelihood value of the parameter, and also prior function that gives the log density value of the parameter. The density used in the likelihood function is $\text{Beta}(6,4)$ since it is our target distribution, and the density used in the prior function is $\text{Uniform}[0,1]$ since our start value is from $\text{Uniform}[0,1]$. In addition, since the parameters in Beta distribution have to be larger than 0, for $c > 0$, $\phi < 1$.

For $c = 1$

Functions

```
a <- 6
b <- 4

likelihood <- function(param){
  singlelikelihoods = dbeta(param, a, b, log = T)
  return(singlelikelihoods)
}

prior <- function(param){
  pr = dunif(param, min=0, max=1, log = T)
  return(pr)
}

posterior <- function(param){
  return (exp(likelihood(param) + prior(param)))
}
```

Posterior function can be calculated by multiplying likelihood function and prior function. However, since we used log in the likelihood function and the prior function, we have used $\exp(\text{likelihood} + \text{prior})$ instead in the posterior function.

```
metropolis_MCMC <- function(startvalue, iterations, c){
  chain = array(dim = c(iterations+1,1))
  chain[1,] = startvalue
  for (i in 1:iterations){
    proposal = rbeta(1, c*chain[i,], c*(1-chain[i,]))
    probab = (posterior(proposal)*dbeta(chain[i,],c*proposal,c*(1-proposal)))/
      (posterior(chain[i,])*dbeta(proposal, c*chain[i,], c*(1-chain[i,])))
    if (runif(1) < probab){
      chain[i+1,] = proposal
    }else{
      chain[i+1,] = chain[i,]
    }
  }
}
```

```

}
return(chain)
}

```

After implementing all the functions that are required in obtaining a chain, we have finally defined metropolis_MCMC function that yields a chain. Note that

$$\begin{aligned}
\phi_{new} &= \phi_{proposal} \text{ with probability } \alpha(\phi_{old}, \phi_{new}) \\
&= \phi_{old} \text{ with probability } 1 - \alpha(\phi_{old}, \phi_{new})
\end{aligned}$$

where $\alpha(\phi_{old}, \phi_{new}) = \min(1, \frac{p(\phi_{new})q(\phi_{old}|\phi_{new})}{p(\phi_{old})q(\phi_{new}|\phi_{old})})$. Here, $p(\phi)$ is the posterior function, and $q(\phi_{new}|\phi_{old})$ is our proposal function which follows $Beta(c\phi_{old}, c(1 - \phi_{old}))$. Since Beta distribution is asymmetric, the probability form cannot be simplified to the ratio of posterior functions. Hence, we need to calculate the probability by using posterior functions and proposal functions.

Finally, we got the chain using metropolis_MCMC function with the startvalue generated from Uniform[0,1]. We also generated random values from Beta(6,4) for comparison.

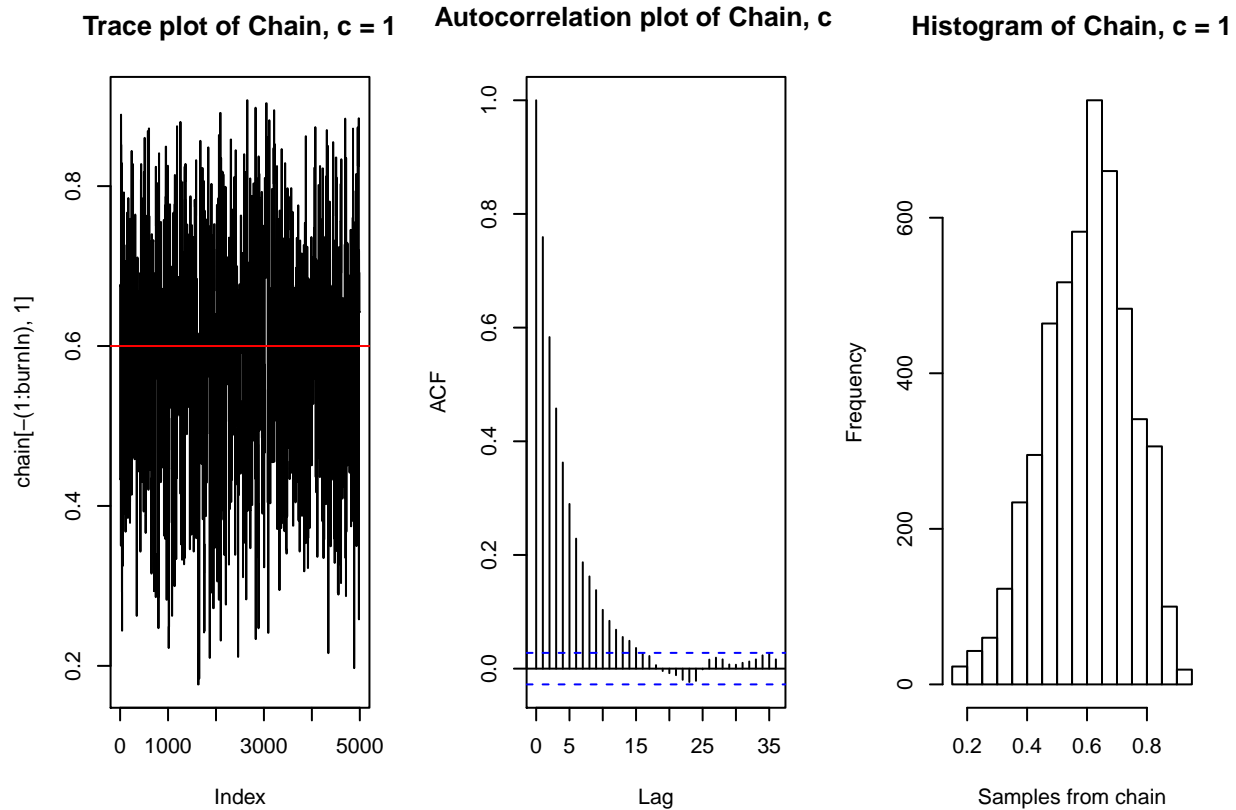
Performance of the sampler with BurnIn

```

set.seed(2)
startvalue <- runif(1, 0, 1)
chain = metropolis_MCMC(startvalue, 10000, c=1)
test <- rbeta(10000, 6, 4)

par(mfrow=c(1,3))
test2 <- test[5001:10000]
burnIn = 5000
acceptance = 1-mean(duplicated(chain[-(1:burnIn),]))
plot(chain[-(1:burnIn),1], type='l', main = "Trace plot of Chain, c = 1")
abline(h=0.6, col="red")
acf(chain[-(1:burnIn),1], main = "Autocorrelation plot of Chain, c = 1")
hist(chain[-(1:burnIn),1], xlab = "Samples from chain", main = "Histogram of Chain, c = 1")

```

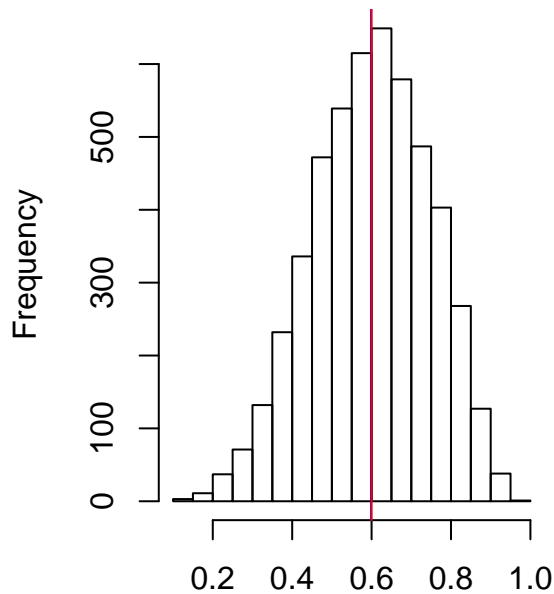


Here, we evaluate the performance of the sampler when $c = 1$ without burn in. The first plot is the trace plot, showing the trace of our chain. The specific explanation of those plots will be covered at the last part.

Histogram of the target distribution Beta(6,4) vs. the Chain with burnIn

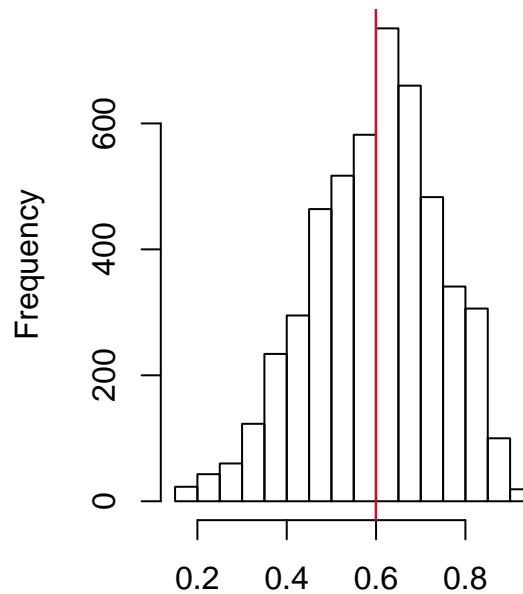
```
par(mfrow=c(1,2))
hist(test2, xlab = "Samples from Beta(6, 4)", main = "Histogram of Beta(6, 4)",
     sub = "with BurnIn")
abline(v=mean(test2), col="blue")
abline(v=0.6, col="red")
hist(chain[-(1:burnIn),1], xlab = "Samples from chain", main = "Histogram of Chain, c = 1",
     sub = "with BurnIn")
abline(v = mean(chain[-(1:burnIn),1]), col="blue")
abline(v=0.6, col="red")
```

Histogram of Beta(6, 4)



Samples from Beta(6, 4)
with BurnIn

Histogram of Chain, c = 1



Samples from chain
with BurnIn

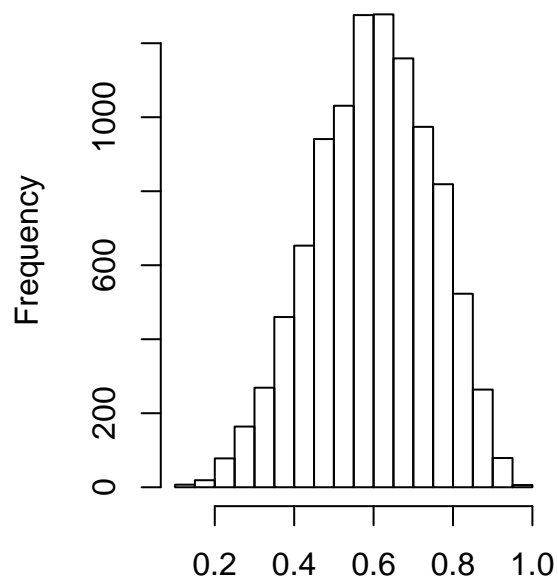
In addition, we can compare the target distribution Beta(6,4) with the histogram of the chain as we can see above. The red line indicates the real mean of the distribution, which is $\frac{6}{4+6} = 0.6$. We can check that both histogram looks similar and the mean of the chain is 0.5992045 which is very close to 0.6. This is after we remove some redundant moves from the chain using BurnIn. However, we can still compare the distribution of Beta(6,4) with the whole chain without BurnIn.

Histogram of the chain vs. the target distribution Beta(6,4) without BurnIn

```
par(mfrow=c(1,2))
hist(test, xlab = "Samples from Beta(6, 4)", main = "Histogram of Beta(6, 4)",
      sub = "without BurnIn")

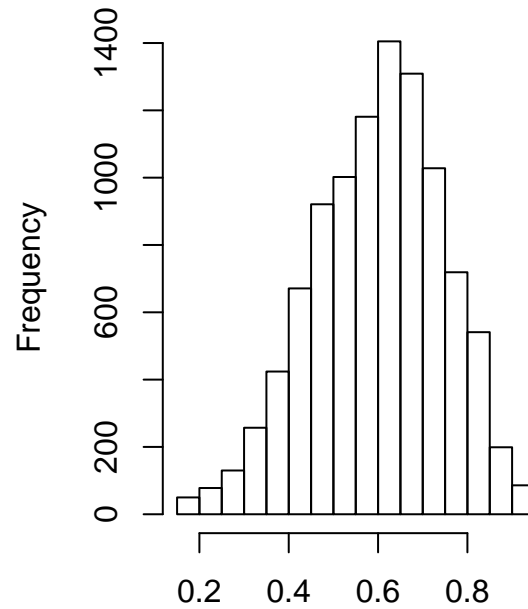
hist(chain, xlab = "Samples from Chain", main = "Histogram of Chain, c = 1",
      sub = "without BurnIn")
```

Histogram of Beta(6, 4)



Samples from Beta(6, 4)
without BurnIn

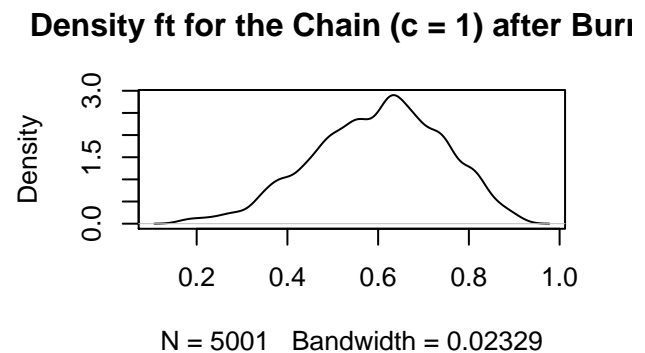
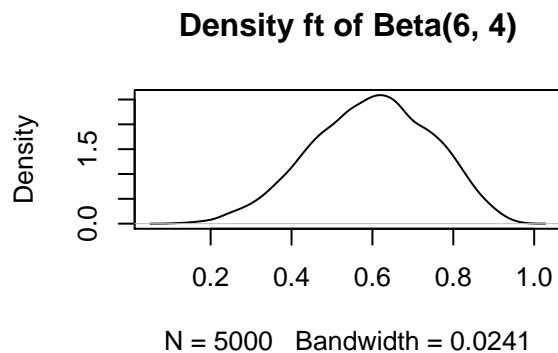
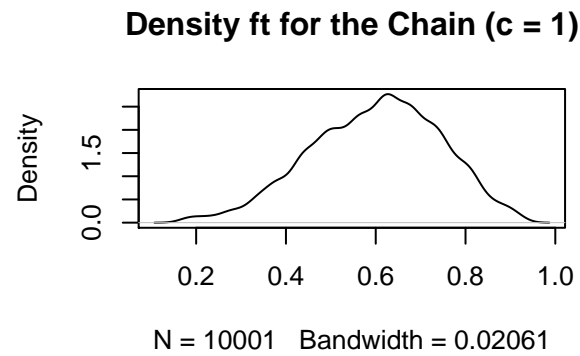
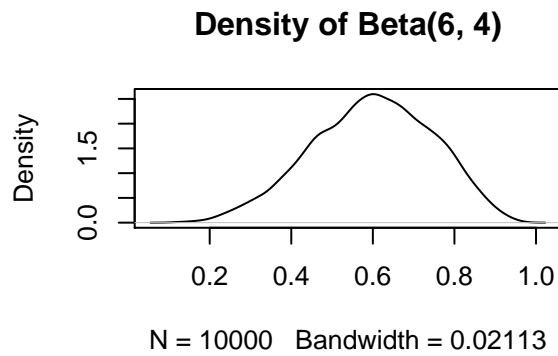
Histogram of Chain, c = 1



Samples from Chain
without BurnIn

Density function for the chain vs. Beta(6,4)

```
par(mfrow = c(2,2))
plot(density(test), main = "Density of Beta(6, 4)")
plot(density(chain), main = "Density ft for the Chain (c = 1)")
plot(density(test2), main = "Density ft of Beta(6, 4)")
plot(density(chain[-(1:burnIn),1]), main = "Density ft for the Chain (c = 1) after BurnIn")
```



Finally, we can implement Kolmogorov-Smirnov test. According to the result, we cannot say that the chain follows Beta(6, 4) distribution since p-value is smaller than 0.05.

Kolmogorov-Smirnov Test

```
chainks <- sort(chain[-(1:burnIn),1])
ks.test(chainks, test2)

## Warning in ks.test(chainks, test2): p-value will be approximate in the
## presence of ties

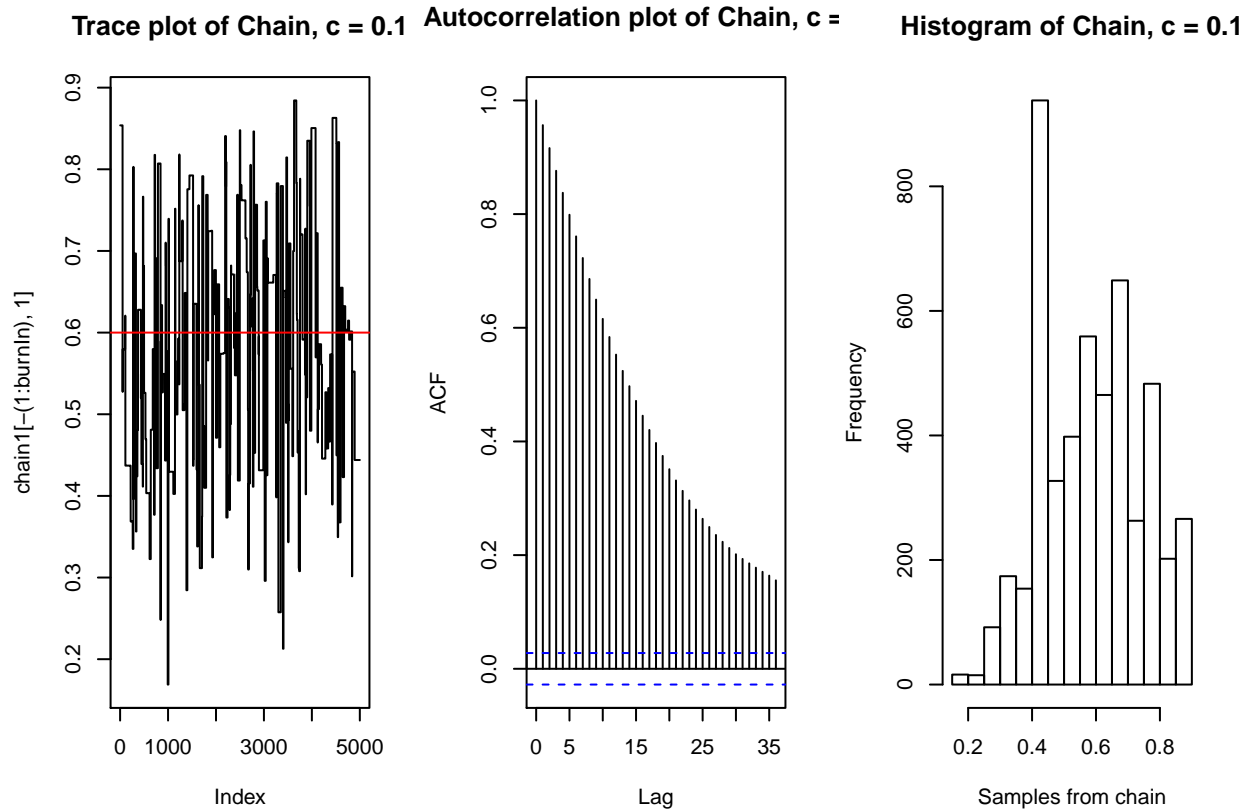
##
## Two-sample Kolmogorov-Smirnov test
##
## data: chainks and test2
## D = 0.0275, p-value = 0.04557
## alternative hypothesis: two-sided
```

For $c = 0.1$

Performance of the sampler

```
set.seed(1)
chain1=metropolis_MCMC(startvalue, 10000, c=0.1)
par(mfrow=c(1,3))
acceptance1 = 1-mean(duplicated(chain1[-(1:burnIn),]))
plot(chain1[-(1:burnIn),1], type='l', main = "Trace plot of Chain, c = 0.1")
```

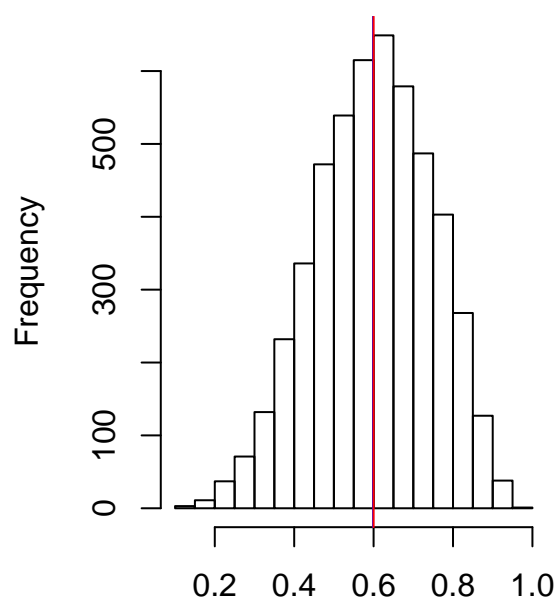
```
abline(h=0.6, col="red")
acf(chain1[-(1:burnIn),1], main = "Autocorrelation plot of Chain, c = 0.1")
hist(chain1[-(1:burnIn),1], xlab = "Samples from chain",
     main = "Histogram of Chain, c = 0.1")
```



Histogram of the target distribution Beta(6,4) vs. the Chain after burnIn

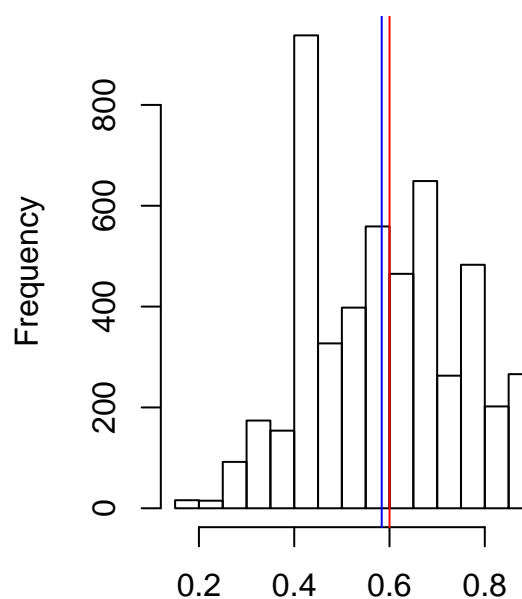
```
par(mfrow=c(1,2))
hist(test2, xlab = "Samples from Beta(6, 4)", main = "Histogram of Beta(6, 4)",
     sub = "with BurnIn")
abline(v=mean(test2), col="blue")
abline(v=0.6, col="red")
hist(chain1[-(1:burnIn),1], xlab = "Samples from chain",
     main = "Histogram of Chain, c = 0.1", sub = "with BurnIn")
abline(v = mean(chain1[-(1:burnIn),1]), col="blue")
abline(v=0.6, col="red")
```

Histogram of Beta(6, 4)



Samples from Beta(6, 4)
with BurnIn

Histogram of Chain, c = 0.1

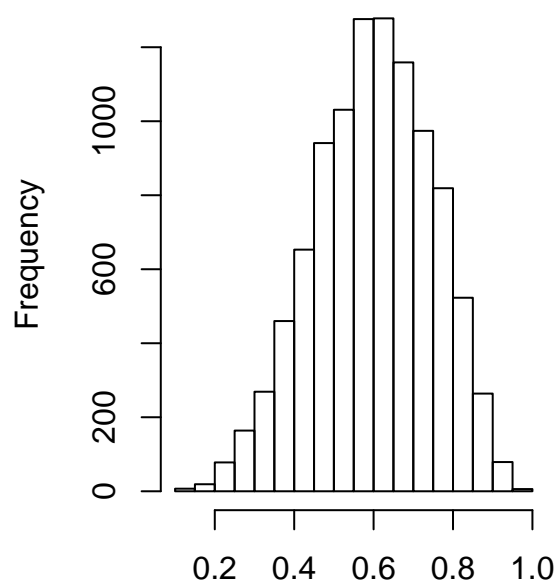


Samples from chain
with BurnIn

Histogram of the chain vs. the target distribution Beta(6,4) without BurnIn

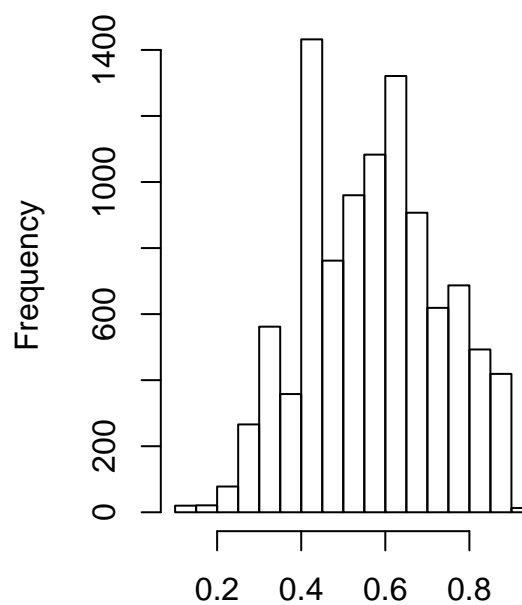
```
par(mfrow=c(1,2))
hist(test, xlab = "Samples from Beta(6, 4)", main = "Histogram of Beta(6, 4)",
      sub = "without BurnIn")
hist(chain1, xlab = "Samples from chain", main = "Histogram of Chain, c = 0.1",
      sub = "without BurnIn")
```


Histogram of Beta(6, 4)



Samples from Beta(6, 4)
without BurnIn

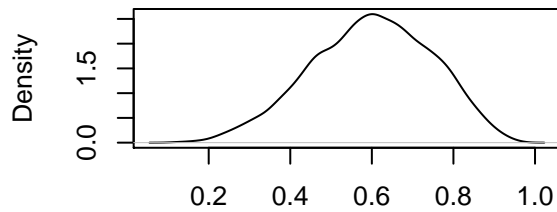
Histogram of Chain, c = 0.1



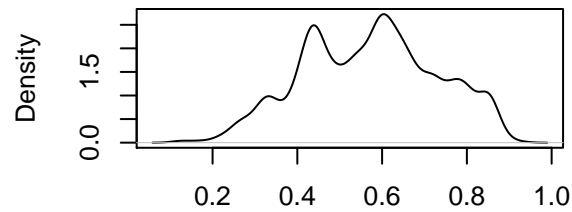
Samples from chain
without BurnIn

Density function for the chain vs. Beta(6,4)

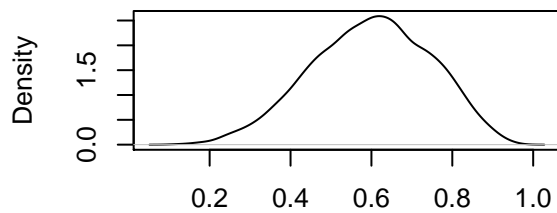
```
par(mfrow = c(2,2))
plot(density(test), main = "Density of Beta(6, 4)")
plot(density(chain1), main = "Density ft for the Chain (c = 0.1)")
plot(density(test2), main = "Density ft of Beta(6, 4)")
plot(density(chain1[-(1:burnIn),1]), main = "Density ft for the Chain (c = 0.1) after BurnIn")
```

Density of Beta(6, 4)

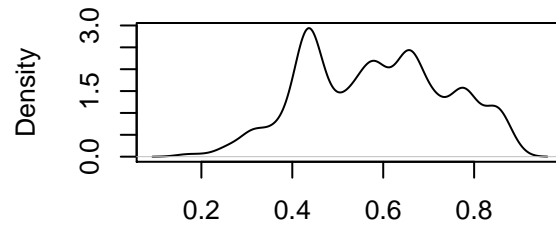
N = 10000 Bandwidth = 0.02113

Density ft for the Chain (c = 0.1)

N = 10001 Bandwidth = 0.02247

Density ft of Beta(6, 4)

N = 5000 Bandwidth = 0.0241

Density ft for the Chain (c = 0.1) after Bu

N = 5001 Bandwidth = 0.02552

As we can see in the plots, there exists significant discrepancy between the distribution of the chain when $c = 0.1$ and Beta(6,4) distribution. Moreover, the gap between the mean of the chain and Beta(6,4) distribution is larger than when $c = 1$.

Kolmogorov-Smirnov statistic

```
chain1ks <- sort(chain1[-(1:burnIn),1])
ks.test(chain1ks, test2)
```

```
## Warning in ks.test(chain1ks, test2): p-value will be approximate in the
## presence of ties
```

```
##
## Two-sample Kolmogorov-Smirnov test
##
## data: chain1ks and test2
## D = 0.11374, p-value < 2.2e-16
## alternative hypothesis: two-sided
```

For $c = 2.5$

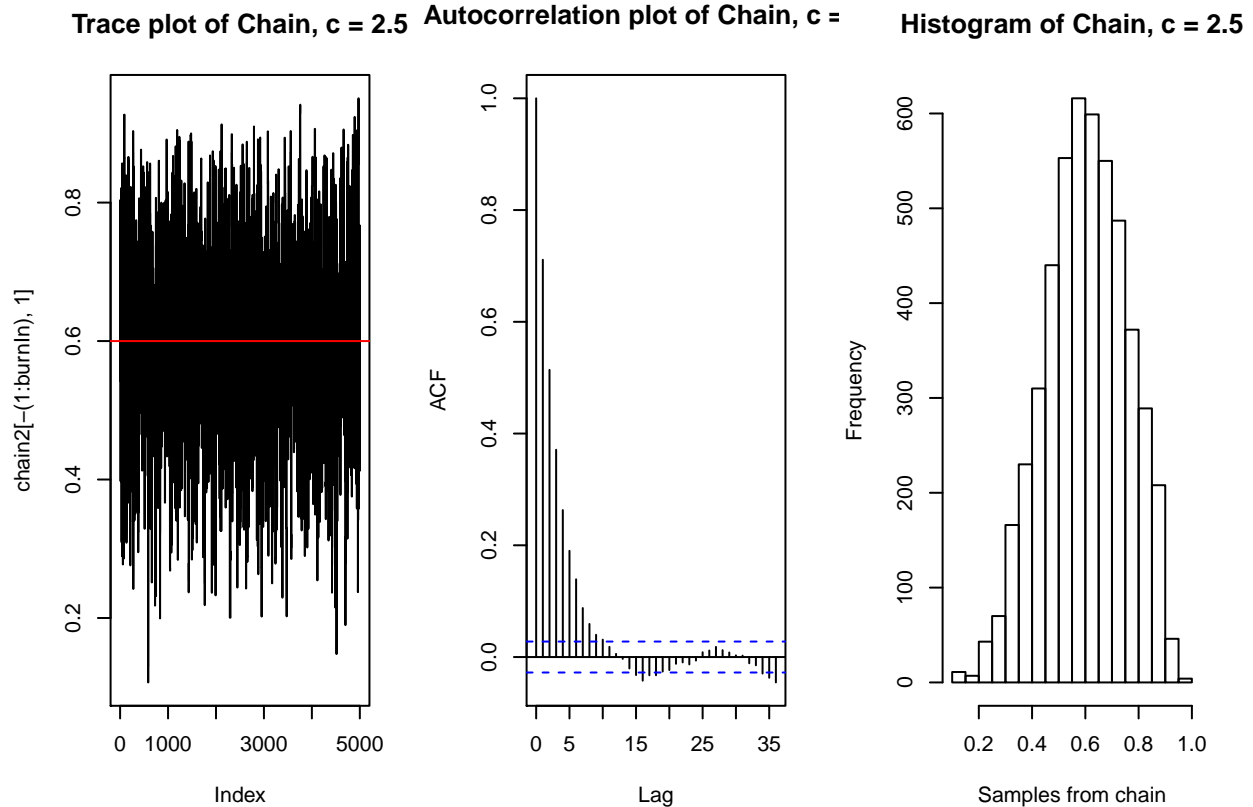
Performance of the sampler

```
set.seed(2)
chain2=metropolis_MCMC(startvalue, 10000, c=2.5)
par(mfrow=c(1,3))
acceptance2 = 1-mean(duplicated(chain2[-(1:burnIn),]))
```

```

plot(chain2[-(1:burnIn),1], type='l', main = "Trace plot of Chain, c = 2.5")
abline(h=0.6, col="red")
acf(chain2[-(1:burnIn),1], main = "Autocorrelation plot of Chain, c = 2.5")
hist(chain2[-(1:burnIn),1], xlab = "Samples from chain",
     main = "Histogram of Chain, c = 2.5")

```



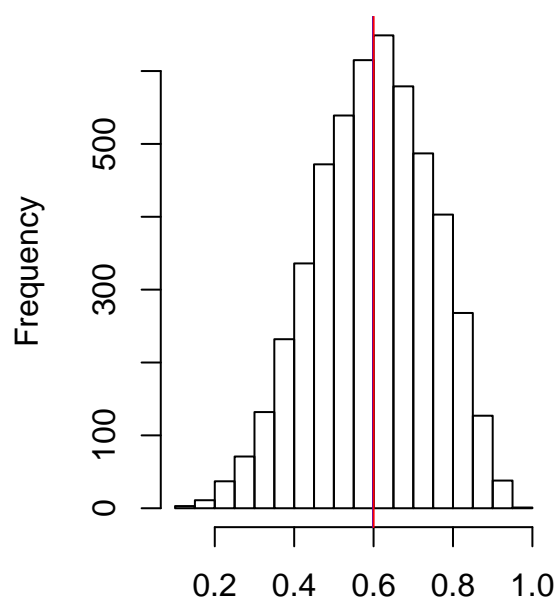
Histogram of the target distribution Beta(6,4) vs. the Chain after burnIn

```

par(mfrow=c(1,2))
hist(test2, xlab = "Samples from Beta(6, 4)", main = "Histogram of Beta(6, 4)",
     sub = "with BurnIn")
abline(v=mean(test2), col="blue")
abline(v=0.6, col="red")
hist(chain2[-(1:burnIn),1], xlab = "Samples from chain",
     main = "Histogram of Chain, c = 2.5", sub = "with BurnIn")
abline(v = mean(chain2[-(1:burnIn),1]), col="blue")
abline(v=0.6, col="red")

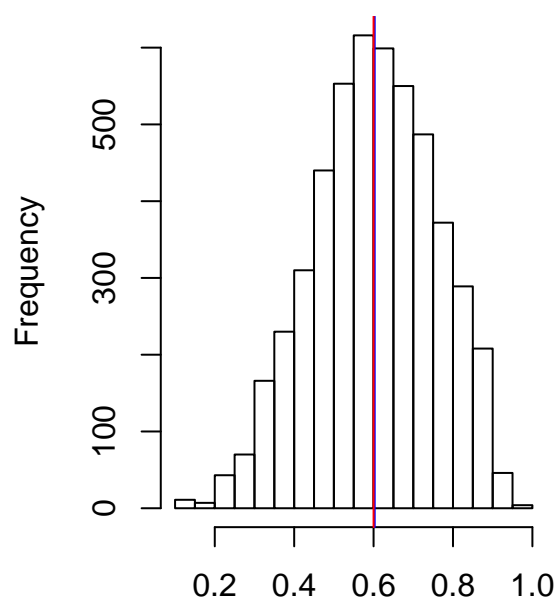
```

Histogram of Beta(6, 4)



Samples from Beta(6, 4)
with BurnIn

Histogram of Chain, c = 2.5

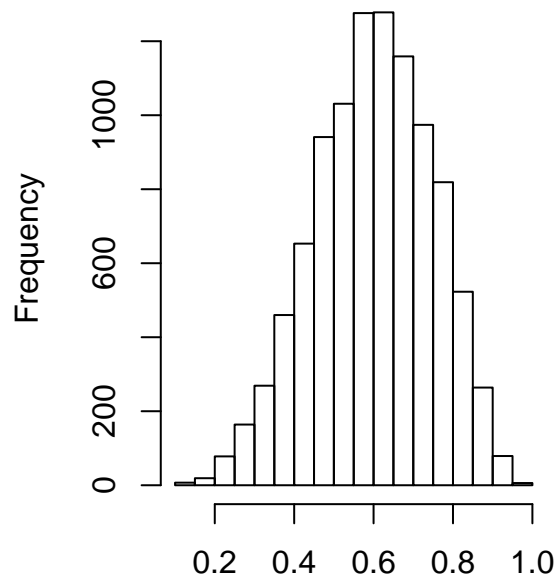


Samples from chain
with BurnIn

Histogram of the chain vs. the target distribution Beta(6,4) without BurnIn

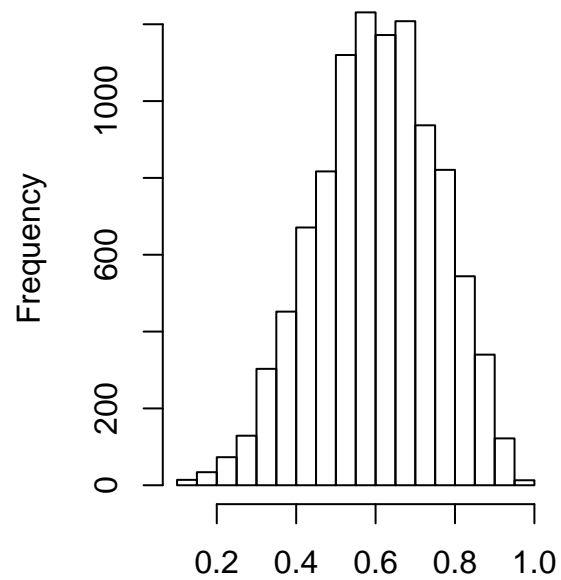
```
par(mfrow=c(1,2))
hist(test, xlab = "Samples from Beta(6, 4)", main = "Histogram of Beta(6, 4)",
      sub = "without BurnIn")
hist(chain2, xlab = "Samples from chain", main = "Histogram of Chain, c = 2.5",
      sub = "without BurnIn")
```

Histogram of Beta(6, 4)



Samples from Beta(6, 4)
without BurnIn

Histogram of Chain, c = 2.5

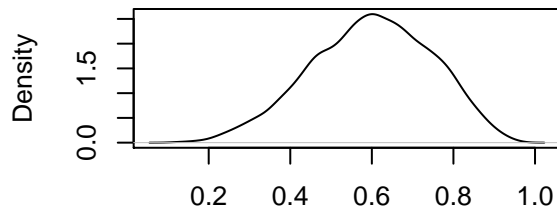


Samples from chain
without BurnIn

Density function for the chain vs. Beta(6,4)

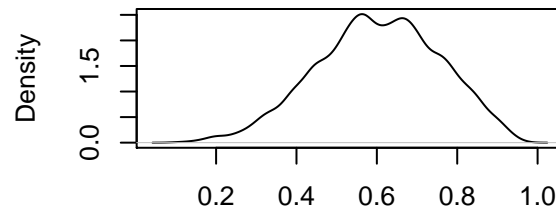
```
par(mfrow = c(2,2))
plot(density(test), main = "Density of Beta(6, 4)")
plot(density(chain2), main = "Density ft for the Chain (c = 2.5)")
plot(density(test2), main = "Density ft of Beta(6, 4)")
plot(density(chain2[-(1:burnIn),1]), main = "Density ft for the Chain (c = 2.5) after BurnIn")
```

Density of Beta(6, 4)



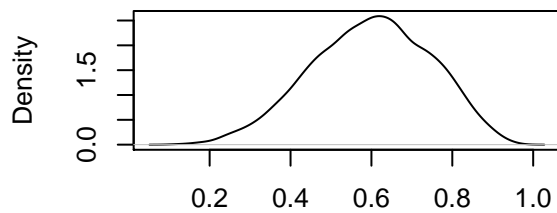
N = 10000 Bandwidth = 0.02113

Density ft for the Chain (c = 2.5)



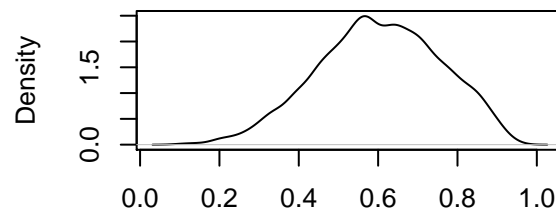
N = 10001 Bandwidth = 0.02177

Density ft of Beta(6, 4)



N = 5000 Bandwidth = 0.0241

Density ft for the Chain (c = 2.5) after BurnIn



N = 5001 Bandwidth = 0.02523

According to the plots, we can check that the distribution after BurnIn has similar distribution with Beta(6,4) compared to when $c = 0.1$. In addition, the mean of the chain and the mean of Beta(6,4), which is 0.6 are almost equal.

Kolmogorov-Smirnov statistic

```
chain2ks <- sort(chain2[-(1:burnIn),1])
ks.test(chain2ks, test2)
```

```
## Warning in ks.test(chain2ks, test2): p-value will be approximate in the
## presence of ties
```

```
##
## Two-sample Kolmogorov-Smirnov test
##
## data: chain2ks and test2
## D = 0.028967, p-value = 0.03012
## alternative hypothesis: two-sided
```

For $c = 10$

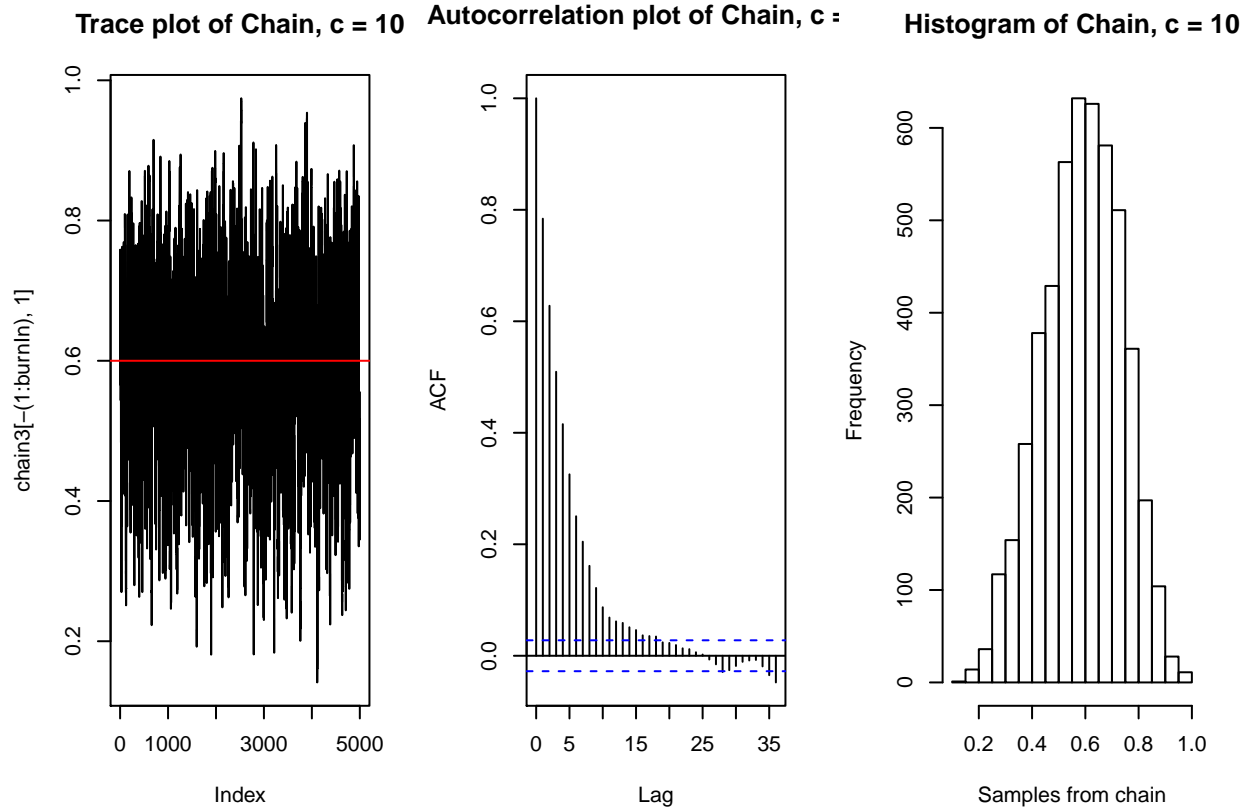
Performance of the sampler: $c = 10$

```
set.seed(3)
chain3=metropolis_MCMC(startvalue, 10000, c=10)
par(mfrow=c(1,3))
acceptance3 = 1-mean(duplicated(chain3[-(1:burnIn),]))
```

```

plot(chain3[-(1:burnIn),1], type='l', main = "Trace plot of Chain, c = 10")
abline(h=0.6, col="red")
acf(chain3[-(1:burnIn),1], main = "Autocorrelation plot of Chain, c = 10")
hist(chain3[-(1:burnIn),1], xlab = "Samples from chain",
     main = "Histogram of Chain, c = 10")

```



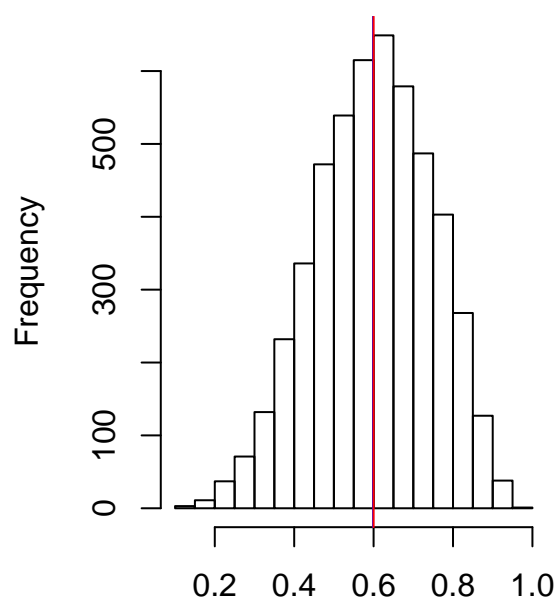
Histogram of the target distribution Beta(6,4) vs. the Chain after burnIn

```

par(mfrow=c(1,2))
hist(test2, xlab = "Samples from Beta(6, 4)", main = "Histogram of Beta(6, 4)",
     sub = "with BurnIn")
abline(v=mean(test2), col="blue")
abline(v=0.6, col="red")
hist(chain3[-(1:burnIn),1], xlab = "Samples from chain",
     main = "Histogram of Chain, c = 10", sub = "with BurnIn")
abline(v = mean(chain3[-(1:burnIn),1]), col="blue")
abline(v=0.6, col="red")

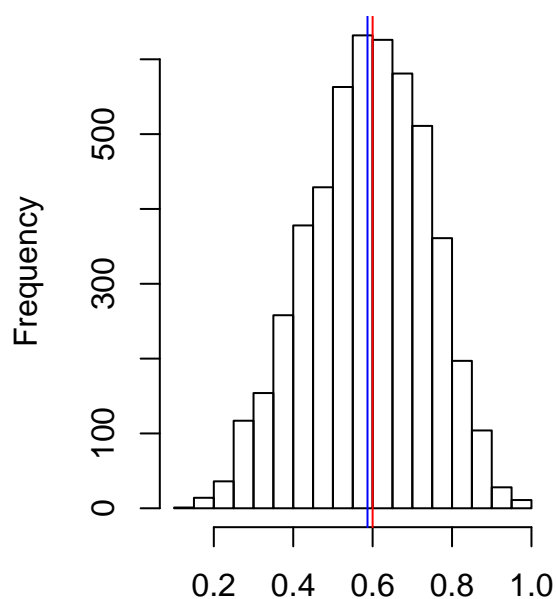
```

Histogram of Beta(6, 4)



Samples from Beta(6, 4)
with BurnIn

Histogram of Chain, c = 10

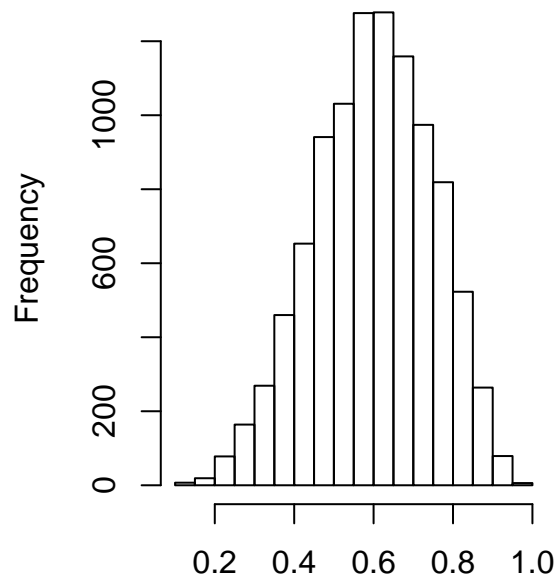


Samples from chain
with BurnIn

Histogram of the chain vs. the target distribution Beta(6,4) without BurnIn

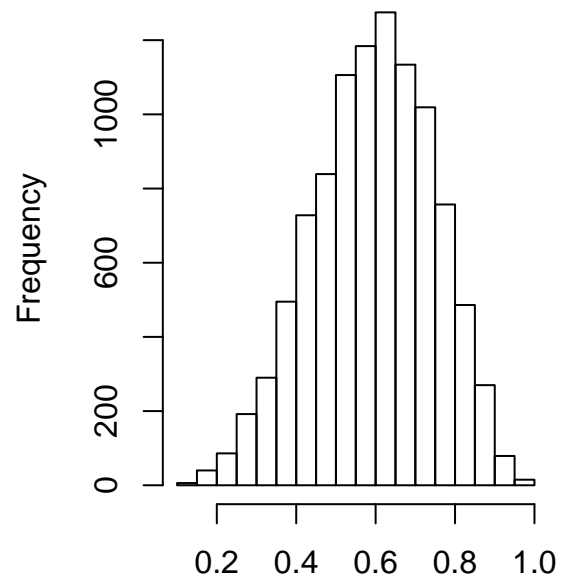
```
par(mfrow=c(1,2))
hist(test, xlab = "Samples from Beta(6, 4)", main = "Histogram of Beta(6, 4)",
     sub = "without BurnIn")
hist(chain3, xlab = "Samples from chain", main = "Histogram of Chain, c = 10",
     sub = "without BurnIn")
```


Histogram of Beta(6, 4)



Samples from Beta(6, 4)
without BurnIn

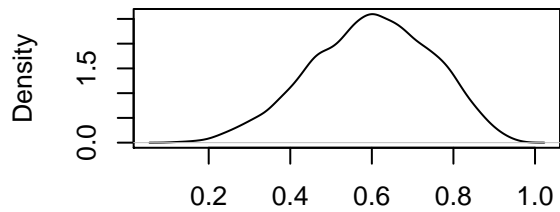
Histogram of Chain, c = 10



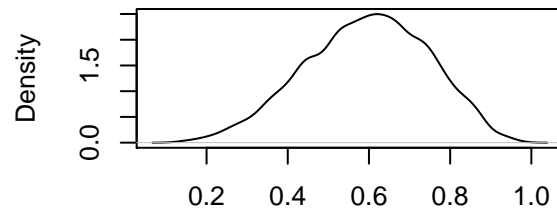
Samples from chain
without BurnIn

Density function for the chain vs. Beta(6,4)

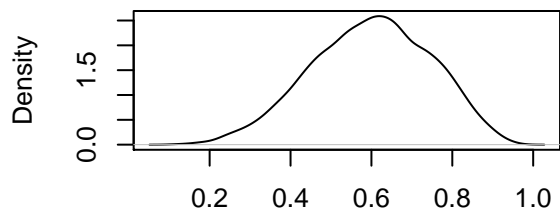
```
par(mfrow = c(2,2))
plot(density(test), main = "Density of Beta(6, 4)")
plot(density(chain3), main = "Density ft for the Chain (c = 10)")
plot(density(test2), main = "Density ft of Beta(6, 4)")
plot(density(chain3[-(1:burnIn),1]), main = "Density ft for the Chain (c = 10) after BurnIn")
```

Density of Beta(6, 4)

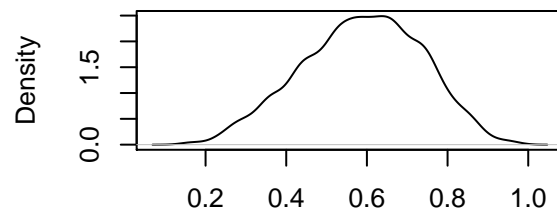
N = 10000 Bandwidth = 0.02113

Density ft for the Chain (c = 10)

N = 10001 Bandwidth = 0.02164

Density ft of Beta(6, 4)

N = 5000 Bandwidth = 0.0241

Density ft for the Chain (c = 10) after Bur

N = 5001 Bandwidth = 0.0243

According to the plot, the shape of the distributions of the chain and Beta(6,4) seem to be similar, but the gap between the mean of the chain and Beta(6,4) distribution is not negligible. That leads to the p-value of K-S test to be smaller than in $c = 2.5$.

Kolmogorov-Smirnov statistic

```
chain3ks <- sort(chain3[-(1:burnIn),1])
ks.test(chain3ks, test2)
```

```
## Warning in ks.test(chain3ks, test2): p-value will be approximate in the
## presence of ties
```

```
##
## Two-sample Kolmogorov-Smirnov test
##
## data: chain3ks and test2
## D = 0.032425, p-value = 0.01042
## alternative hypothesis: two-sided
```

Acceptance rates

```
acceptance
```

```
## [1] 0.2519496
```

```
acceptance1
```

```
## [1] 0.04219156
```

```
acceptance2
```

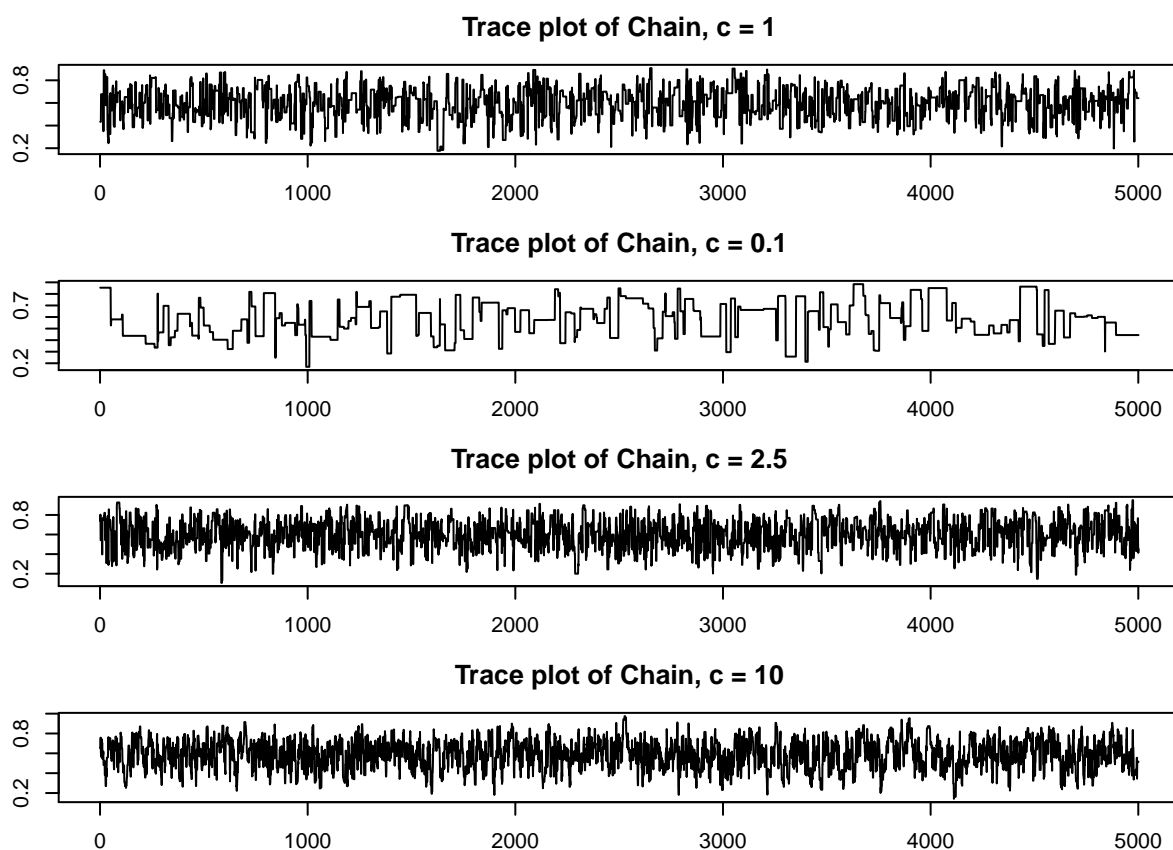
```
## [1] 0.4169166
```

```
acceptance3
```

```
## [1] 0.674865
```

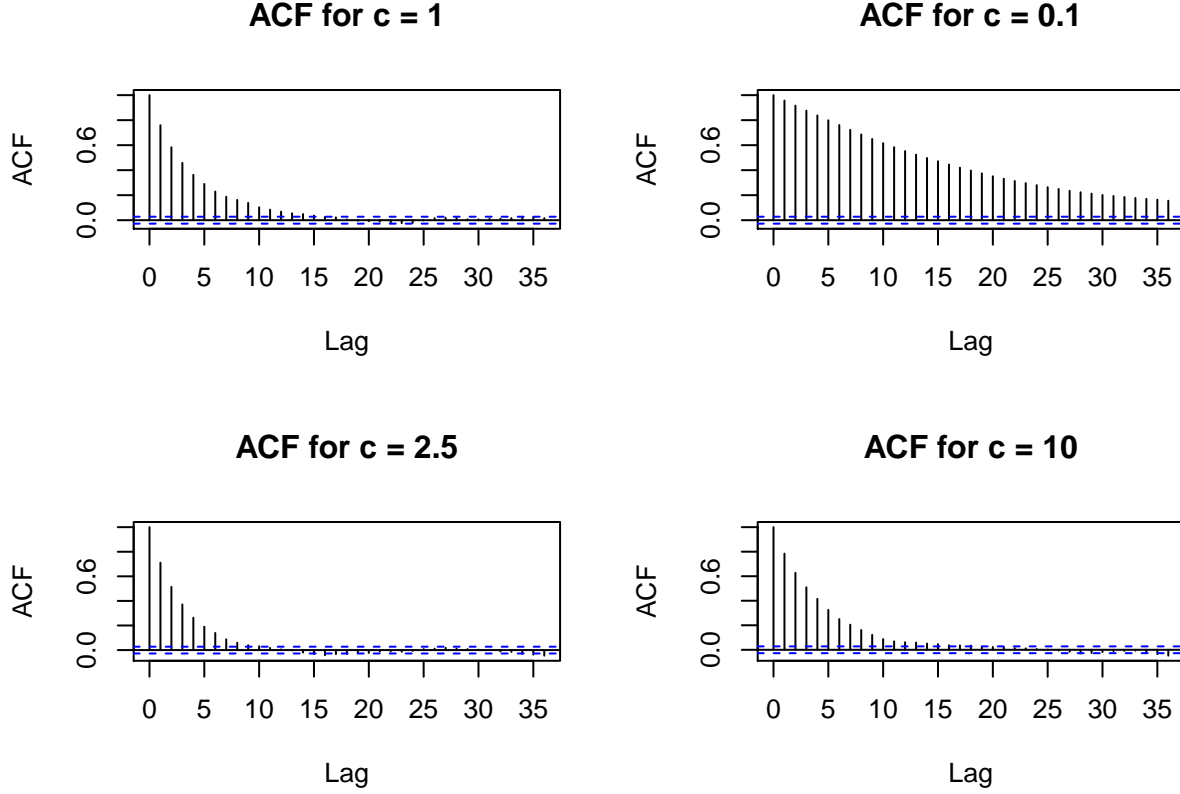
Trace plot for chains

```
par(mfrow=c(4,1), mar= c(2,3,3,2))  
plot(chain[-(1:burnIn),1], type='l', main = "Trace plot of Chain, c = 1", xlab = "")  
plot(chain1[-(1:burnIn),1], type='l', main = "Trace plot of Chain, c = 0.1", xlab = "")  
plot(chain2[-(1:burnIn),1], type='l', main = "Trace plot of Chain, c = 2.5", xlab = "")  
plot(chain3[-(1:burnIn),1], type='l', main = "Trace plot of Chain, c = 10", xlab = "")
```



ACF for chains

```
par(mfrow=c(2,2))  
acf(chain[-(1:burnIn),1], main = "ACF for c = 1")  
acf(chain1[-(1:burnIn),1], main = "ACF for c = 0.1")  
acf(chain2[-(1:burnIn),1], main = "ACF for c = 2.5")  
acf(chain3[-(1:burnIn),1], main = "ACF for c = 10")
```



As we can see in the result, as c increases, acceptance rate increases. This is because the proposal distribution $\phi_{prop}|\phi_{old} \sim \text{Beta}(c\phi_{old}, c(1 - \phi_{old}))$ is getting narrower compare to the target distribution as c increases. That is, for $c = 0.1$, the proposal distribution is wider than the target distribution. In addition, according to the trace plot for $c = 0.1$, we can check that the chain tends to stay at same place more than when $c = 2.5$ and $c = 10$. This implies that c has lower acceptance rate. Next, if we look at the autocorrelation plot, we can check that there exists high autocorrelation in the chain, and that causes the chain to stay where it was. In the trace plot for $c = 1$, the chain moves more frequently than in $c = 0.1$, but it is more stationary compared to $c = 2.5$ and $c = 10$ cases. In addition, according to the autocorrelation plot of $c = 1$ case, there exists some kind of autocorrelation after the chain goes on compared to $c = 2.5$ and $c = 10$ cases. That leads to the result of the acceptance rate for $c = 1$ case to be larger than when $c = 0.1$ but smaller than $c = 2.5$ and $c = 10$ cases. For $c = 2.5$ and $c = 10$, we can see in the autocorrelation has dropped faster than in $c = 0.1$, so it implies that the chain tends to move faster as we can check in the trace plot. This indicates that the acceptance rate is higher than $c = 0.1$ for both cases. However, as we can see in the acceptance rates of 3 cases, the acceptance rate when $c = 10$ is the highest, implying that the chain for $c = 10$ moves frequently than when $c = 2.5$. However, very high acceptance rate is not preferable, and optimal acceptance rate is 20 ~ 30%, as we obtain in $c = 1$ (which is 0.2519496). If we compare among $c = 0.1, 2.5, 10$, the case when $c = 2.5$ seems to be the most optimal. This is comparable to the Kolmogorov-Smirnov test p-value, since among those three cases, p-value in $c = 2.5$ is the highest, implying that it is less likely to reject the null hypothesis that the chain distribution has the same distribution as the target distribution, $\text{Beta}(6,4)$. (However, the results of all cases show that they reject the null hypothesis.)