Lab 4

Shwetha, Suhani, Hoda

11/30/2020

Question 1: Computations with Metropolis Hastings

Target density :

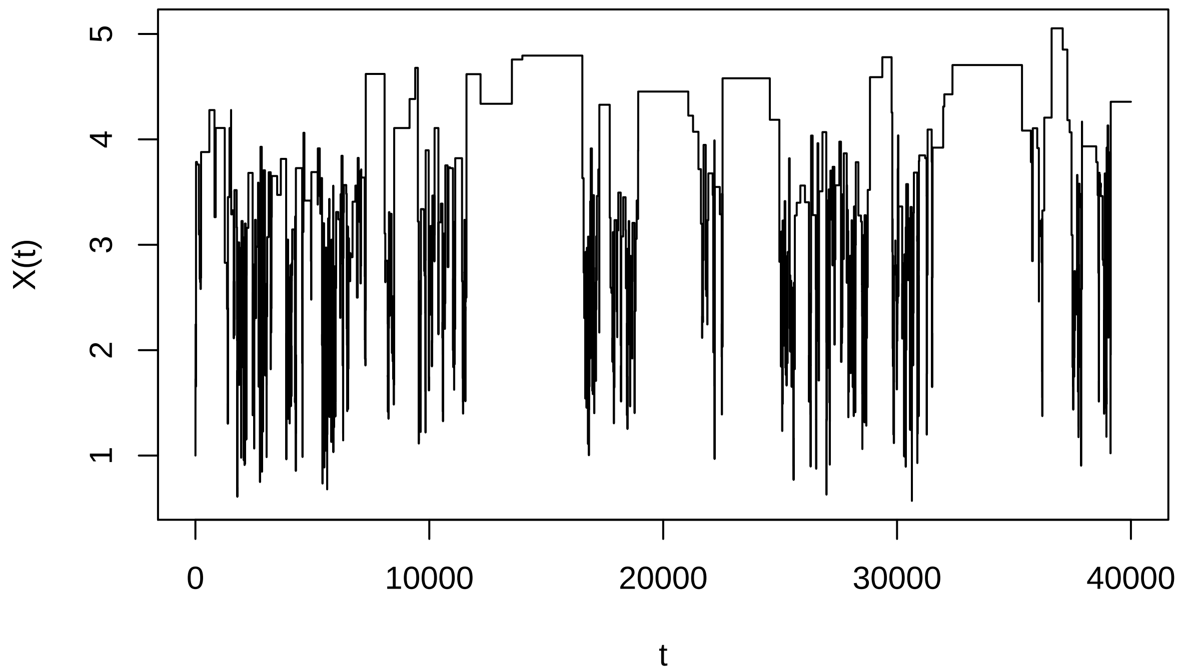
```
f = function(x){
  res = x^5 * exp(-x)
  return(res)
}
```

1

Using Metropolis Hastings algorithm to generate samples from given distribution by using proposal distribution as log normal LN(Xt; 1), take 1 as starting point.

```
propose_lognorm_MCMC<-function(nstep,X0){
  vN<-1:nstep
  vX<-rep(X0,nstep);
  for (i in 2:nstep){
    X<-vX[i-1]
    Y<-rlnorm(1,mean=X,sd=1)
    u<-runif(1)
    a<-min(c(1,(f(Y)*dlnorm(X,meanlog = Y,sdlog = 1))/(f(X)*dlnorm(Y,meanlog = X,sdlog = 1))))
    if(u<=a){vX[i] = Y} else {vX[i] = X}
  }
  return(vX)
}
set.seed(12345)
lognorm = propose_lognorm_MCMC(40000,1)

plot(1:40000,lognorm,pch=19,cex=0.3,type = "l",col="black",xlab="t",ylab="X(t)",main="")
```



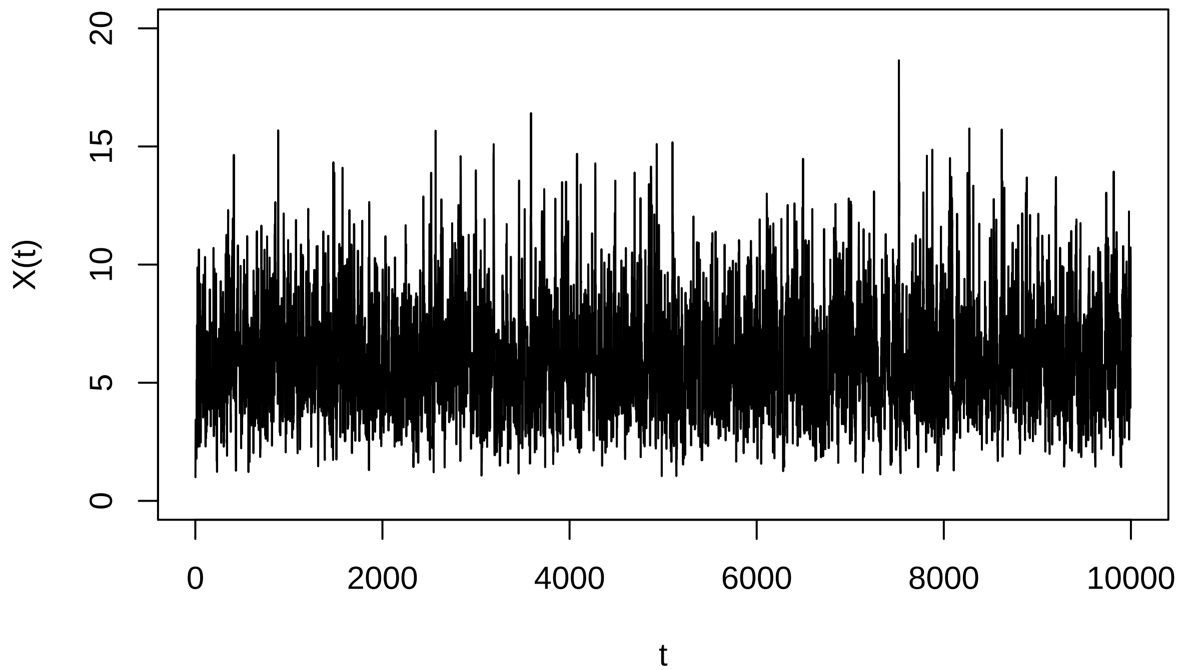
The chain takes different amount of time to converge depending on the selected starting point , this period is called as burn-in period. In the above graph we can see that the chain did not converge. Hence there is no burn in period.

2

Using the chi square distribution $\chi^2([X_t + 1])$ as a proposal distribution, where $[x]$ is the floor function.

```
propose_chi_MCMC<-function(nstep,X0){
  vN<-1:nstep
  vX<-rep(X0,nstep);
  for (i in 2:nstep){
    X<-vX[i-1]
    Y<-rchisq(1,df=floor(X+1))
    u<-runif(1)
    a<-min(c(1,(f(Y)*dchisq(X,floor(Y+1)))/(f(X)*dchisq(Y,floor(X+1)))))
    if(u<=a){vX[i] = Y} else {vX[i] = X}
  }
  return(vX)
}
set.seed(12345)
chi = propose_chi_MCMC(10000,1)

plot(1:10000,chi,pch=19,cex=0.3,type = "l",col="black",xlab="t",ylab="X(t)",main="",ylim=c(0,20))
```



In the above chain we can observe that chain converges in the very begining , ie to say it follows a similar pattern fluctuating around a mean value. Convergence seem to have achieved near the starting point , so we can consider first few iterations as burn in period.

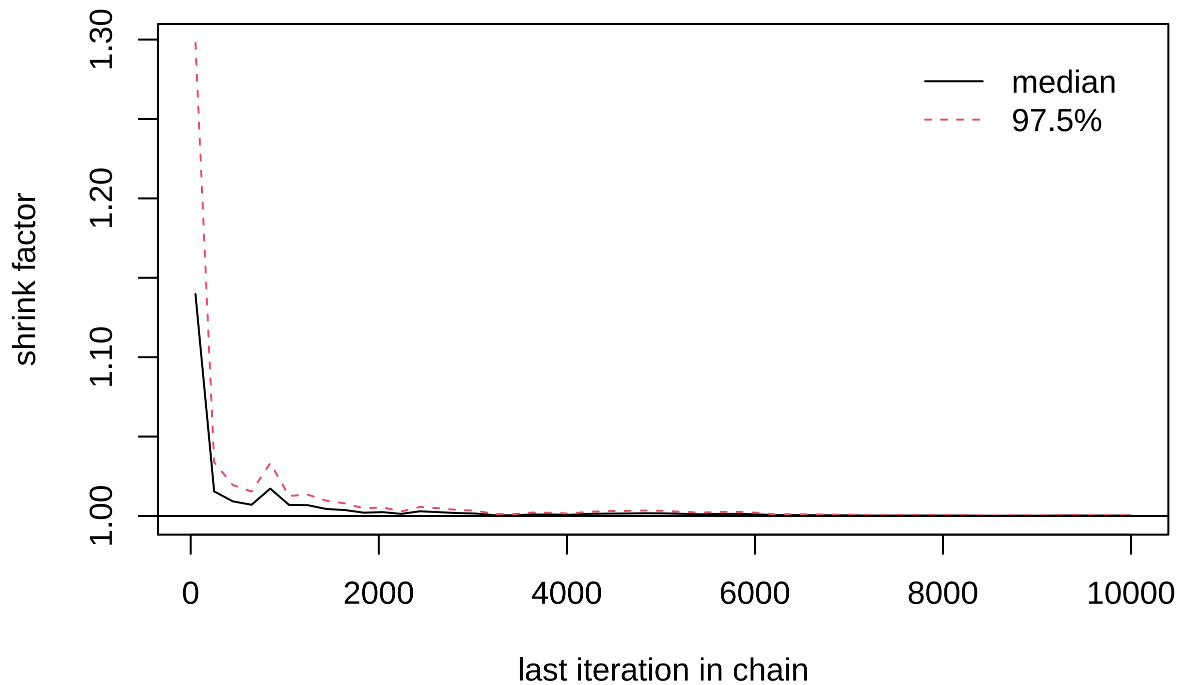
3

In conclusion we can say that chi squared is the better choice as a proposal distribution for the given target distribution, this can be justified by observing the time series plot for both proposal distributions, we could not find a proper convergence for log normal distribution , where as in sampling with chi squared convergence was very fast with a small burnin period.

4

```
# reference : http://ugrad.stat.ubc.ca/R/library/coda/html/gelman.diag.html
library(coda)
set.seed(12345)
mcmc_data = mcmc.list()
for(i in 1:10){
  mcmc_data[[i]] = as.mcmc(propose_chi_MCMC(10000,i))
}

gelman.plot(mcmc_data)
```




```r
gelman.diag(mcmc_data)
```

```
## Potential scale reduction factors:
##
##      Point est. Upper C.I.
## [1,]          1          1
```

Two ways to estimate the variance of stationary distribution 1. The mean of the emperical variance within each chain W 2. The emperical variance from all chain combined, this can be expressed as

$$\sigma^2 = \frac{(n-1)W}{n} + \frac{B}{n}$$

where n is number of iteration and $\frac{B}{n}$ is emperical between chain variance.

Assumption : Target distribution is normal. A bayesian probaility interval can be constructed using t distribution with

$$\hat{\mu} = samplemeanofallchainscombined$$

variance

$$\hat{V} = \sigma^2 + \frac{B}{mn}$$

degrees of freedom as

$$d = \frac{2\hat{V}}{var(\hat{V})}$$

The convergence diagnostic iteself is given by :

$$R = \sqrt{\frac{(d+3)\hat{V}}{(d+1)W}}$$

The value substantially above 1 indicates that the lack of convergence , we have obtained this factor as 1 confirming the convergence.

5

Estimate $\int_0^\infty xf(x)dx$

Step 1 : Mean of generated value in step 1 is given by

$$\frac{1}{n}\sum_i x_i$$

where x is generated values with log normal as proposal density.

```r
mean(lognorm)
```

```
## [1] 3.854309
```

Step 2 : Here x is generated values with chi square as proposal density.

```r
mean(chi)
```

```
## [1] 6.000143
```

6

The probability density function of gamma distribustion is given by

$$f(x;\alpha,\beta) = \left[\frac{1}{\Gamma(\alpha)\beta^\alpha}\right] x^{\alpha-1} exp^{\frac{-x}{\beta}}$$

$$0 < x < \infty$$

Comparing given density function and gamma function , we can find alpha and beta as follows

$$\alpha - 1 = 5$$

ie

$$\alpha = 6$$

$$\frac{1}{\beta} = 1$$

ie

$$\beta = 1$$

Mean or expected value for gamma distribution is given by

$$\mu = \alpha\beta$$

ie 6*1 = 6

Our estimation in step 2 is 6.0001431 which is almost 6. This indicates that chi square is a good choice for proposal distribution for the given target distribution.