

Monte Carlo error

Let's construct a table of parameters leading to different Markov chains, each having the same marginal distribution $N(0, 1)$ at the limit of large number of samples but each also having different amount of auto-correlation between the samples.

```
varTheta = 1
sigma2.1 = 1
sigma2.3 = 0.2
phi.2 = 0.5
phi.4 = 0.1
phi.1 = sqrt(1-sigma2.1/varTheta)
phi.3 = sqrt(1-sigma2.3/varTheta)
sigma2.2 = varTheta*(1-phi.2^2)
sigma2.4 = varTheta*(1-phi.4^2)
table.entries = matrix(nrow=4, ncol=4, data=c(
  varTheta, phi.1, sigma2.1, phi.1,
  varTheta, phi.2, sigma2.2, phi.2,
  varTheta, phi.3, sigma2.3, phi.3,
  varTheta, phi.4, sigma2.4, phi.4
))
table.entries <- t(table.entries) # take transpose since matrix fills in the elements in columnwise
colnames(table.entries) <- c("var(theta)", "phi", "sigma2", "corr")
print(table.entries)
```

```
##      var(theta)      phi sigma2      corr
## [1,]          1 0.0000000  1.00 0.0000000
## [2,]          1 0.5000000  0.75 0.5000000
## [3,]          1 0.8944272  0.20 0.8944272
## [4,]          1 0.1000000  0.99 0.1000000
```

Let's then construct a function to perform Markov chain sampling

```
# let's first define a function to conduct the sampling
MarkovChain <- function(phi,sigma2,initial,m){
  theta = vector(length=m)
  theta[1] = initial
  for (i1 in seq(1,m,1)){
    theta[i1+1] = phi*theta[i1] + rnorm(1,0,sqrt(sigma2))
  }
  return(theta)
}
```

For this exercise it is handy to use multidimensional arrays to store the results (not necessary but saves some lines of code). Below an example:

```

set.seed(123)
arr = array(dim=c(3,2,5))
dim(arr)

arr[1,1,] = 1
arr[1,2,] = 2
arr[3,2,] = 3
arr

```

Now we need to sample 100 independent realizations of length 2000 chains from the Markov chain defined in exercise 3.1 (that is; $\theta^{(1)}, \dots, \theta^{(2000)}$) using each of the combinations of ϕ and σ^2 in the rows of the above table.

With each of the chains we approximate $E[\theta^{(i)}]$, $\Pr(\theta^{(i)} > 0.5)$ and $\Pr(\theta^{(i)} > 2)$ using Monte Carlo with the $n = 10$, $n = 100$ and $n = 1000$ last samples. Hence, we will construct 100 independent Monte Carlo approximations for the mean and two probabilities of θ corresponding to Markov chain sample sizes 10, 100 and 1000.

For example the below rows would construct two independent Markov chains of length 2000 and calculate the Monte Carlo approximation for the mean with the last 10 samples

```

set.seed(123)

i1=1
m=2000
initial = 0
n=10
theta1 = MarkovChain(table.entries[i1,"phi"],table.entries[i1,"sigma2"],initial,m) # sample a Markov chain
theta2 = MarkovChain(table.entries[i1,"phi"],table.entries[i1,"sigma2"],initial,m) # sample a Markov chain
mean(theta1[(m-n+1):m])

```

```
## [1] 0.07540569
```

```
mean(theta2[(m-n+1):m])
```

```
## [1] -0.3171165
```

Now, we need to repeat the above steps 100 times, calculate the mean and asked probabilities for each of the 100 chains and then examine how these Monte Carlo estimates behave and match with the exact results as we vary the row of the table and n . # Answers

```

set.seed(123)
# Initial parameters for 4 different phi - sigma2 pairs chain
m=2000
initial = 0

# The n = 10, n = 100 and n = 1000 last samples
n1=10
n2=100
n3=1000

# Initial a 3x4x100 multidimensional arrays to store the results with the n = 10 last samples.
result1 = array(dim=c(3,4,100))

```

```

for (a in seq(1,100,1)){
  for (b in seq(1,4,1)){
    theta = MarkovChain(table.entries[b,"phi"],table.entries[b,"sigma2"],initial,m) # sample a Markov
    result1[1,b,a] <- mean(theta[(m-n1+1):m])
    result1[2,b,a] <- sum( theta[(m-n1+1):m] > 0.5 ) / length(theta[(m-n1+1):m])
    result1[3,b,a] <- sum( theta[(m-n1+1):m] > 2 ) / length(theta[(m-n1+1):m])
  }
}
colnames(result1) <- c("phi=0,sigma2=1", "phi=0.5,sigma2=0.75", "phi=0.89,sigma2=0.2","phi=0.1,sigma2=0.99")
rownames(result1) <- c("theta_mean", "Pr(theta>0.5)", "Pr(theta>2)")

# Print a result test
result1[, ,1]

```

```

##                phi=0,sigma2=1 phi=0.5,sigma2=0.75 phi=0.89,sigma2=0.2
## theta_mean      0.07540569      -0.5700395      1.172491
## Pr(theta>0.5)    0.30000000      0.1000000      0.700000
## Pr(theta>2)      0.00000000      0.0000000      0.200000
##                phi=0.1,sigma2=0.99
## theta_mean      -0.1026297
## Pr(theta>0.5)    0.2000000
## Pr(theta>2)      0.1000000

```

```

set.seed(123)

# Initial a 3x4x100 multidimensional arrays to store the results with the n = 100 last samples.
result2 = array(dim=c(3,4,100))
for (a in seq(1,100,1)){
  for (b in seq(1,4,1)){
    theta = MarkovChain(table.entries[b,"phi"],table.entries[b,"sigma2"],initial,m) # sample a Markov
    result2[1,b,a] <- mean(theta[(m-n2+1):m])
    result2[2,b,a] <- sum( theta[(m-n2+1):m] > 0.5 ) / length(theta[(m-n2+1):m])
    result2[3,b,a] <- sum( theta[(m-n2+1):m] > 2 ) / length(theta[(m-n2+1):m])
  }
}
colnames(result2) <- c("phi=0,sigma2=1", "phi=0.5,sigma2=0.75", "phi=0.89,sigma2=0.2","phi=0.1,sigma2=0.99")
rownames(result2) <- c("theta_mean", "Pr(theta>0.5)", "Pr(theta>2)")

# Print a result test
result2[, ,2]

```

```

##                phi=0,sigma2=1 phi=0.5,sigma2=0.75 phi=0.89,sigma2=0.2
## theta_mean      -0.04686311      0.2365227      0.233663
## Pr(theta>0.5)    0.30000000      0.3600000      0.490000
## Pr(theta>2)      0.04000000      0.0200000      0.040000
##                phi=0.1,sigma2=0.99
## theta_mean      -0.0617175
## Pr(theta>0.5)    0.2800000
## Pr(theta>2)      0.0200000

```

```

set.seed(123)

# Initial a 3x4x100 multidimensional arrays to store the results with the n = 1000 last samples.
result3 = array(dim=c(3,4,100))
for (a in seq(1,100,1)){
  for (b in seq(1,4,1)){
    theta = MarkovChain(table.entries[b,"phi"],table.entries[b,"sigma2"],initial,m) # sample a Markov
    result3[1,b,a] <- mean(theta[(m-n3+1):m])
    result3[2,b,a] <- sum( theta[(m-n3+1):m] > 0.5 ) / length(theta[(m-n3+1):m])
    result3[3,b,a] <- sum( theta[(m-n3+1):m] > 2 ) / length(theta[(m-n3+1):m])
  }
}
colnames(result3) <- c("phi=0,sigma2=1", "phi=0.5,sigma2=0.75", "phi=0.89,sigma2=0.2","phi=0.1,sigma2=0.99")
rownames(result3) <- c("theta_mean", "Pr(theta>0.5)", "Pr(theta>2)")

# Print a result test
result3[, ,3]

```

```

##                phi=0,sigma2=1 phi=0.5,sigma2=0.75 phi=0.89,sigma2=0.2
## theta_mean      -0.01416871      -0.07839907          0.02549539
## Pr(theta>0.5)    0.29700000          0.29900000          0.29000000
## Pr(theta>2)      0.02500000          0.01600000          0.03300000
##                phi=0.1,sigma2=0.99
## theta_mean              0.02459249
## Pr(theta>0.5)           0.31600000
## Pr(theta>2)             0.02600000

```

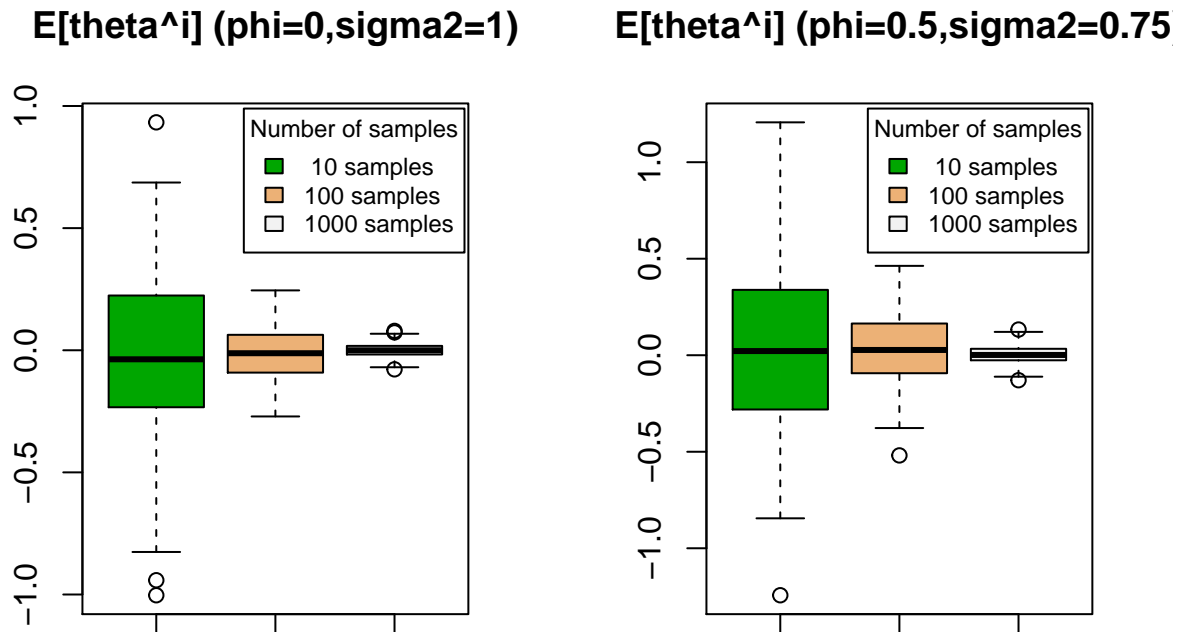
1) How does the Monte Carlo estimate of $E[\theta^{(i)}]$ behave with respect to the number of samples and with respect to the autocorrelation of the Markov chain?

a) The Monte Carlo estimate of $E[\theta^{(i)}]$ behave with respect to the number of samples.

```

par(mfrow=c(1,2))
boxplot(result1[1,1,],result2[1,1,],result3[1,1,],main="E[theta^i] (phi=0,sigma2=1)",col=terrain.colors(3),
legend("topright", inset=.01, title="Number of samples",
c(" 10 samples","100 samples","1000 samples"), fill=terrain.colors(3),cex = 0.75)
boxplot(result1[1,2,],result2[1,2,],result3[1,2,],main="E[theta^i] (phi=0.5,sigma2=0.75)",col=terrain.colors(3),
legend("topright", inset=.01, title="Number of samples",
c(" 10 samples","100 samples","1000 samples"), fill=terrain.colors(3),cex = 0.75)

```

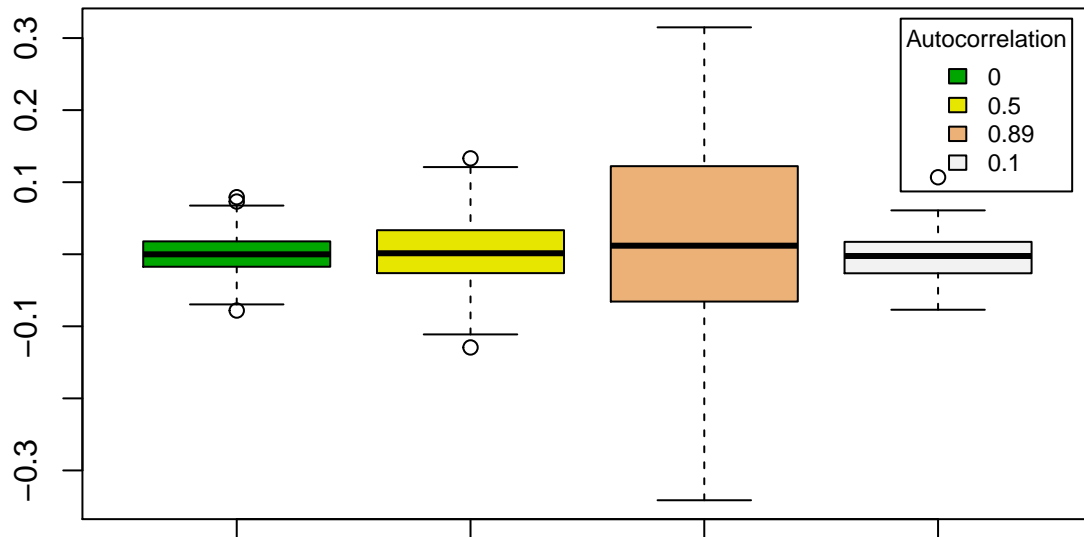


$E[\theta^{(i)}]$ will be closer to the exact answers for $\theta \sim N(0,1)$ as the number of samples increases, and the distribution will be more concentrated.

b) The Monte Carlo estimate of $E[\theta^{(i)}]$ behave with respect to the autocorrelation of the Markov chain.

```
boxplot(result3[1,1,],result3[1,2,],result3[1,3,],result3[1,4,],main="E[theta^i] with 1000 samples",col=
legend("topright", inset=.02, title="Autocorrelation",
c("0","0.5","0.89","0.1"), fill=terrain.colors(4),cex = 0.75)
```

E[θ^i] with 1000 samples



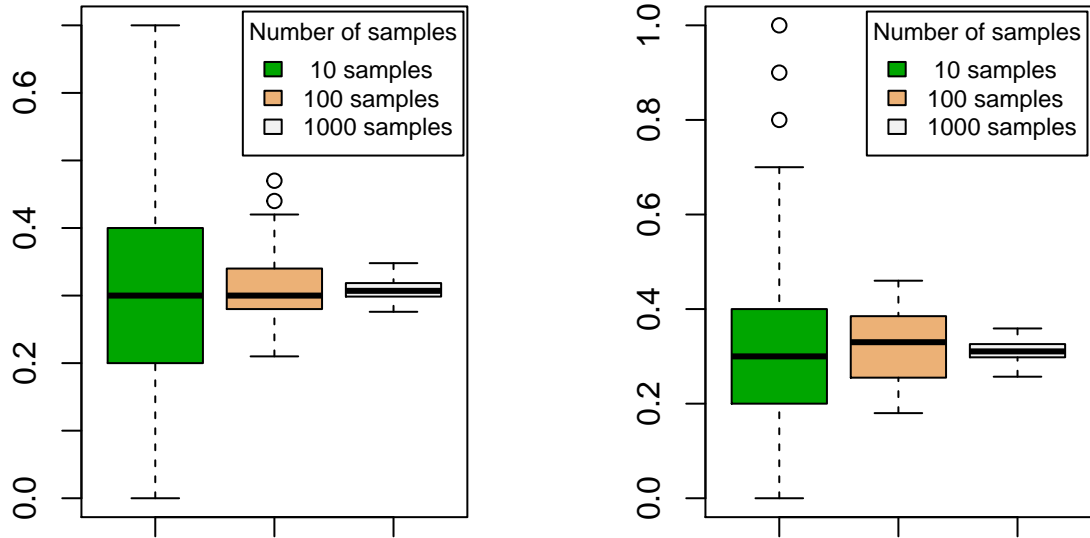
$E[\theta^{(i)}]$ will be more distributed as Autocorrelation increases.

2) How does the Monte Carlo estimate of $\Pr(\theta^{(i)} > 0.5)$ behave with respect to the number of samples and with respect to the autocorrelation of the Markov chain?

a) The Monte Carlo estimate of $\Pr(\theta^{(i)} > 0.5)$ behave with respect to the number of samples.

```
par(mfrow=c(1,2))
boxplot(result1[2,1,],result2[2,1,],result3[2,1,],main="Pr(theta>0.5) (phi=0,sigma2=1)",col=terrain.col,
legend("topright", inset=.01, title="Number of samples",
c(" 10 samples","100 samples","1000 samples"), fill=terrain.colors(3),cex = 0.75)
boxplot(result1[2,2,],result2[2,2,],result3[2,2,],main="Pr(theta>0.5) (phi=0.5,sigma2=0.75)",col=terrain.col,
legend("topright", inset=.01, title="Number of samples",
c(" 10 samples","100 samples","1000 samples"), fill=terrain.colors(3),cex = 0.75)
```

Pr(theta>0.5) (phi=0,sigma2=1) Pr(theta>0.5) (phi=0.5,sigma2=0.7)

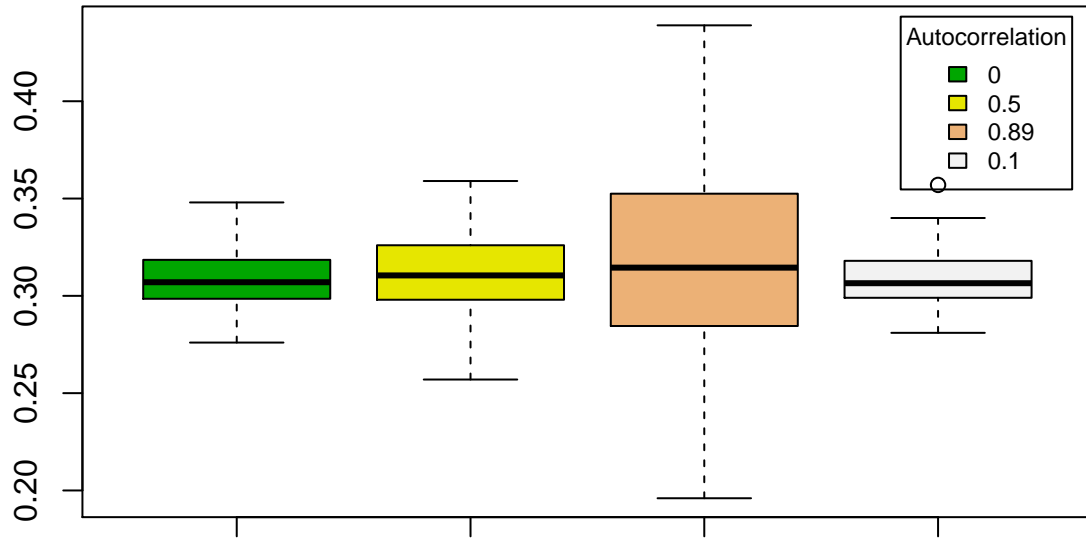


$\Pr(\theta^{(i)} > 0.5)$ will be closer to the exact answers for $\theta \sim N(0, 1)$ as the number of samples increases, and the distribution will be more concentrated.

b) The Monte Carlo estimate of $\Pr(\theta^{(i)} > 0.5)$ behave with respect to the autocorrelation of the Markov chain.

```
boxplot(result3[2,1,],result3[2,2,],result3[2,3,],result3[2,4,],main="Pr(theta>0.5) with 1000 samples",
legend("topright", inset=.02, title="Autocorrelation",
c("0","0.5","0.89","0.1"), fill=terrain.colors(4),cex = 0.75)
```

Pr($\theta > 0.5$) with 1000 samples



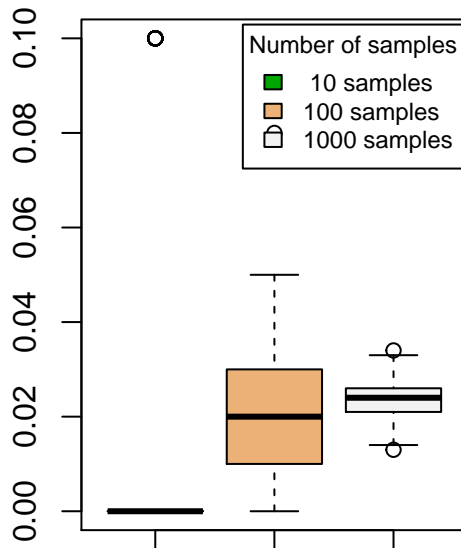
$\Pr(\theta^{(i)} > 0.5)$ will be more distributed as Autocorrelation increases.

3) How does the Monte Carlo estimate of $\Pr(\theta^{(i)} > 2)$ behave with respect to the number of samples and with respect to the autocorrelation of the Markov chain?

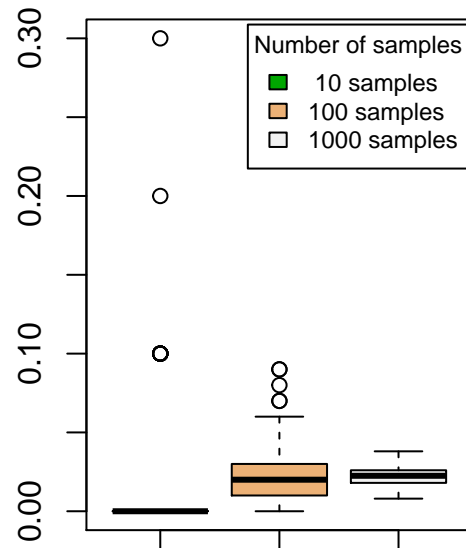
a) The Monte Carlo estimate of $\Pr(\theta^{(i)} > 2)$ behave with respect to the number of samples.

```
par(mfrow=c(1,2))
boxplot(result1[3,1,],result2[3,1,],result3[3,1,],main="Pr(theta>2) (phi=0,sigma2=1)",col=terrain.colors(3),
legend("topright", inset=.01, title="Number of samples",
      c(" 10 samples","100 samples","1000 samples"), fill=terrain.colors(3),cex = 0.75)
boxplot(result1[3,2,],result2[3,2,],result3[3,2,],main="Pr(theta>2) (phi=0.5,sigma2=0.75)",col=terrain.colors(3),
legend("topright", inset=.01, title="Number of samples",
      c(" 10 samples","100 samples","1000 samples"), fill=terrain.colors(3),cex = 0.75)
```


Pr(theta>2) (phi=0,sigma2=1)



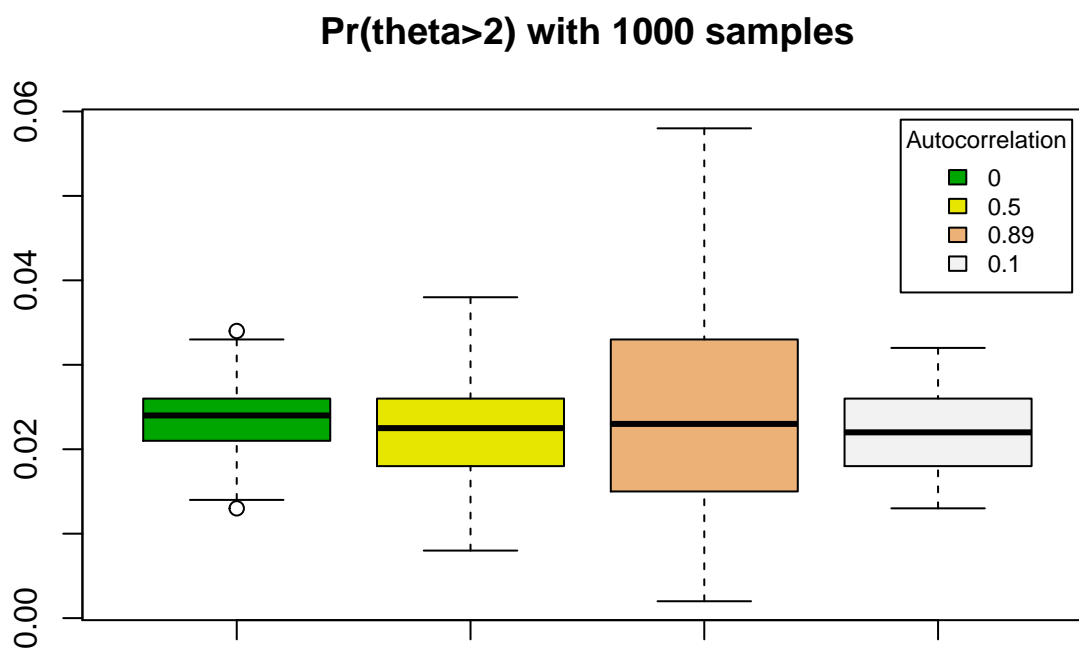
Pr(theta>2) (phi=0.5,sigma2=0.75)



$\Pr(\theta^{(i)} > 2)$ will be closer to the exact answers for $\theta \sim N(0, 1)$ as the number of samples increases, and the distribution is different from before, 10 samples have the mean of $\Pr(\theta^{(i)} > 2) = 0$, but some values are very scattered, as the number of samples increases, the distribution becomes more concentrated.

b) The Monte Carlo estimate of $\Pr(\theta^{(i)} > 2)$ behave with respect to the autocorrelation of the Markov chain.

```
boxplot(result3[3,1,],result3[3,2,],result3[3,3,],result3[3,4,],main="Pr(theta>2) with 1000 samples",col=
legend("topright", inset=.02, title="Autocorrelation",
c("0","0.5","0.89","0.1"), fill=terrain.colors(4),cex = 0.75)
```



$\Pr(\theta^{(i)} > 2)$ will be more distributed as Autocorrelation increases.

4) What kind of general conclusions can you make based on these results?

Monte Carlo error will relatively decrease as the number of samples increases, on the other hand, Monte Carlo error will increase with the increase of autocorrelation.