Stats202C-HW2

Juan Piao

5/22/2020

Problem1.

Calculate the mean of a Gamma(4.3, 6.2) random variable using (a) Accept-Reject with a Gamma(4, 7) candidate. (b) Metropolis- Hastings with a Gamma(4, 7) candidate. (c) Metropolis- Hastings with a Gamma(5, 6) candidate.

```r
set.seed(123456)
b<-4.3/4
#M<- b^(-4)*((4.3-4)/((1-b)*exp(1)))^(4.3-4)

M <- gamma(4) / gamma(4.3) * b**4 * (4.3 * exp(-1))**(4.3 - 4) * 6.2^4.3/7^4

g47=rgamma(5000,4,1/7)
u=runif(5000,max=dgamma(g47,4,1/7))
x=g47[u<dgamma(g47,4.3,1/6.2)/M]

ylim <- c(0, 0.04)

par(mfrow=c(1,3),mar=c(4,4,1,1))
hist(x,freq=FALSE,xlab="",ylab="",col="wheat2", ylim = ylim,
main="a: Accept-Rej with Ga(4.7) prop")
curve( dgamma(x,4.3,1/6.2),lwd=2,col="sienna",add=T)
a <- length(x)/5000
mean(x)
```

```
## [1] 26.61735
```

```r
###### Met-Hast with Ga(5,6) prop
X=rep(0,5000)

X[1]=rgamma(1,4.3,1/6.2)

for (t in 2:5000){
  rho=(dgamma(X[t-1],4,1/7)*dgamma(g47[t],4.3,1/6.2))/(dgamma(g47[t],4,7)*dgamma(X[t-1],4.3,1/6.2))

  X[t]=X[t-1]+(g47[t]-X[t-1])*(runif(1)<rho) #move to g47[t] if it is accepted, otherwise stay
}
hist(X,freq=FALSE,xlab="",ylab="",col="wheat2",  ylim = ylim,
main="b: Met-Hast with Ga(4,7) prop")
curve(dgamma(x,4.3,1/6.2),lwd=2,col="sienna",add=T)

b <- length(unique(X))/5000
mean(X)
```
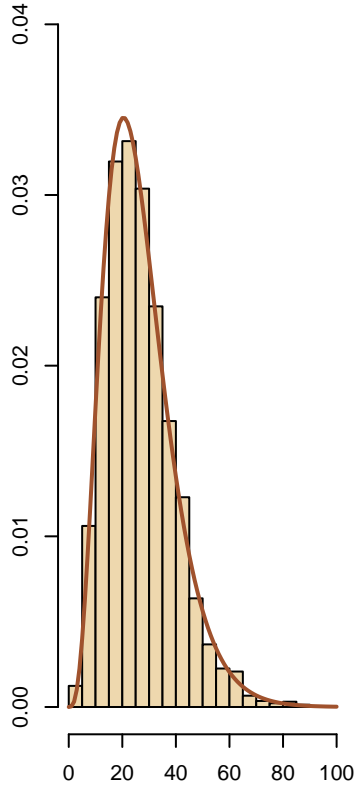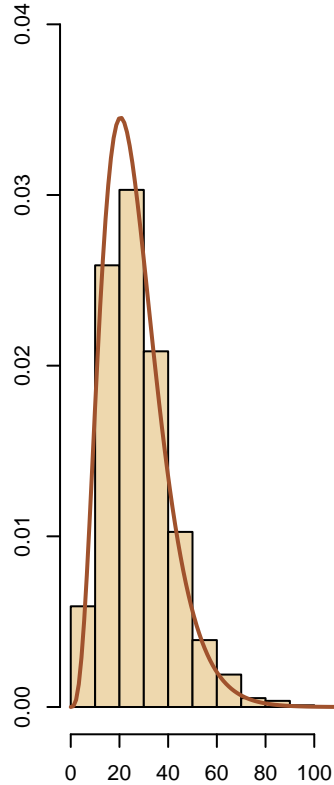
```
## [1] 27.8449
```

```
############ Met-Hast with Ga(5,6) prop
X2=rep(0,5000)

g56=rgamma(5000,5,1/6)
X2[1]=rgamma(1,4.3,1/6.2)
for (t in 2:5000){
rho=(dgamma(X2[t-1],5,1/6)*dgamma(g56[t],4.3,1/6.2))/
(dgamma(g56[t],5,1/6)*dgamma(X2[t-1],4.3,1/6.2))
X2[t]=X2[t-1]+(g56[t]-X2[t-1])*(runif(1)<rho)
}
c <- length(unique(X2))/5000
hist(X2,freq=FALSE,xlab="",ylab="",col="wheat2",  ylim = ylim,
main="c: Met-Hast with Ga(5,6) prop")
curve(dgamma(x,4.3,1/6.2),lwd=2,col="sienna",add=T)
```
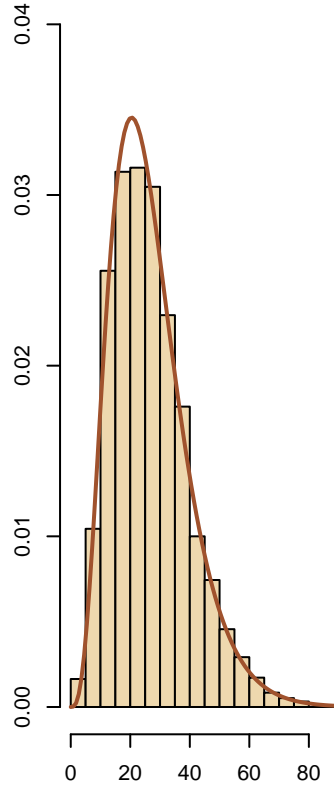

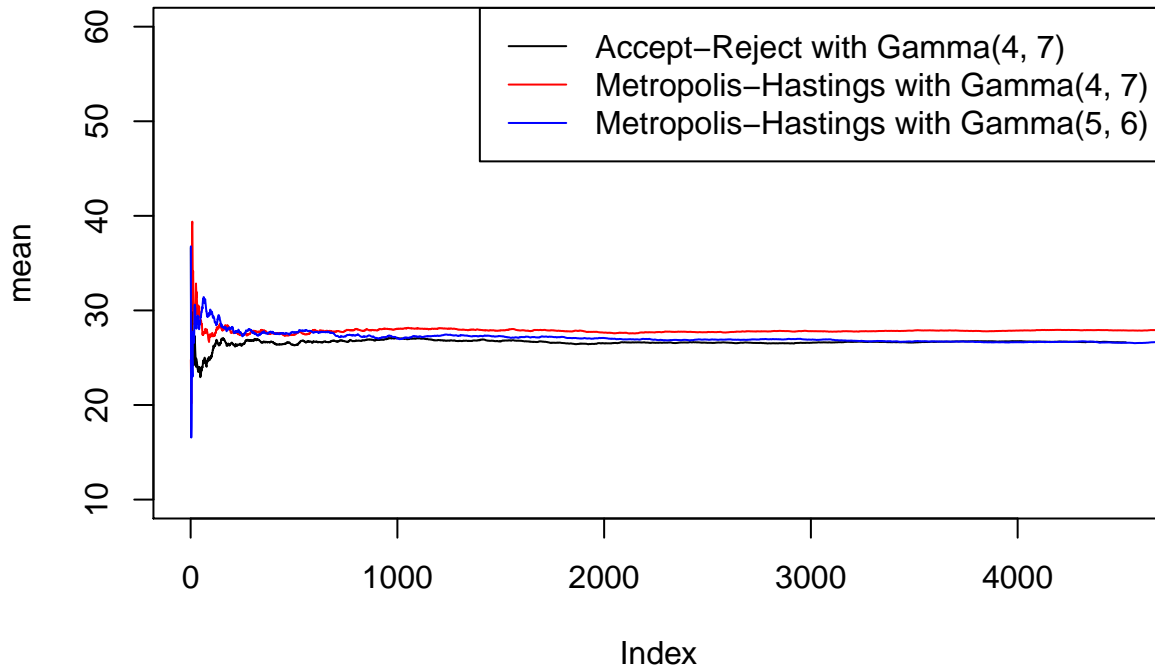**a: Accept–Rej with Ga(4.7) pro** **b: Met–Hast with Ga(4,7) pro** **c: Met–Hast with Ga(5,6) pro**

```
mean(X2)
```

```
## [1] 26.59489
```

```
plot(cumsum(x)/1:length(x), type = "l", ylim = c(10, 60), ylab = "mean")
lines(cumsum(X)/1:length(X), col="red")
lines(cumsum(X2)/1:length(X2), col="blue")
legend("topright", legend=c("Accept-Reject with Gamma(4, 7)", "Metropolis-Hastings with Gamma(4, 7)",
```

The mean of a Gamma(4.3, 6.2) random variable using (a) Accept-Reject with a Gamma(4, 7) candidate is 26.61735. (b) Metropolis-Hastings with a Gamma(4, 7) candidate is 27.8449. (c) Metropolis-Hastings with a Gamma(5, 6) candidate is 26.59489.

Problem 7.20

(a) Verify the likelihood function

$logit(p_{ij}) = a + bx_i + cz_{ij}$, $i = 1, ..., k$, $j = 1, ..., n_i$ Solving for p_{ij}, we get,

$$p_{ij} = \frac{exp(a + bx_i + cz_{ij})}{1 + exp(a + bx_i + cz_{ij})}$$

The likelihood function is,

$$\prod_{i=1}^{k} \prod_{j=1}^{n_i} (p_{ij})^{y_{ij}} (1 - p_{ij})^{1-y_{ij}}$$

$$= \prod_{i=1}^{k} \prod_{j=1}^{n_i} (\frac{exp(a + bx_i + cz_{ij})}{1 + exp(a + bx_i + cz_{ij})})^{y_{ij}} (1 - \frac{exp(a + bx_i + cz_{ij})}{1 + exp(a + bx_i + cz_{ij})})^{1-y_{ij}}$$

$$= \prod_{i=1}^{k} \prod_{j=1}^{n_i} (\frac{exp(a + bx_i + cz_{ij})}{1 + exp(a + bx_i + cz_{ij})})^{y_{ij}} (\frac{1}{1 + exp(a + bx_i + cz_{ij})})^{1-y_{ij}}$$

(b) Run a standard GLM on these data9 and get the estimated mean and variance of a, b, and c.

```
LogisticData<-read.table("~/Downloads/Stats 202C/HW2/LogisticData.txt", header = T)

Logisticmodel<-glm(LogisticData$erodd~ LogisticData$np + LogisticData$metq, family = "binomial")
summary(Logisticmodel)

##
## Call:
```

3

```
## glm(formula = LogisticData$erodd ~ LogisticData$np + LogisticData$metq,
##     family = "binomial")
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.8057  -0.7120  -0.6269  -0.5400   1.9988
##
## Coefficients:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)        -1.97391    0.22113  -8.926  < 2e-16 ***
## LogisticData$np     0.16221    0.07974   2.034  0.04193 *
## LogisticData$metq   0.28438    0.09274   3.066  0.00217 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1229.2  on 1179  degrees of freedom
## Residual deviance: 1215.9  on 1177  degrees of freedom
## AIC: 1221.9
##
## Number of Fisher Scoring iterations: 4
```

The mean and variance of a are -1.97391, 0.0489 respectively. The mean and variance of b are 0.16221, 0.00636 respectively. The mean and variance of c are 0.28438, 0.0086 respectively.

(c) Use normal candidate densities with mean and variance at the GLM estimates in a Metropolis-Hastings algorithm that samples from the likelihood. Get histograms of the parameter values.

```
np = LogisticData$np
metq = LogisticData$metq
erodd = LogisticData$erodd

am = -1.97391 # coef (mean)
as = 0.22113 # std err
bm = 0.16221
bs = 0.07974
cm = 0.28438
cs = 0.09274

log_likelihood<- function(a, b, c){
  sum(erodd * (a + b*np + c*metq) - log(1 + exp(a + b*np + c*metq)))
}

get_dens <- function(a, b, c){
  dnorm(a, am, as, log = T) + dnorm(b, bm, bs, log = T)+ dnorm(c, cm, cs, log = T)
}

n_sim = 100000
a_hat <- rep(0, n_sim)
b_hat <- rep(0, n_sim)
c_hat <- rep(0, n_sim)
a_hat[1] <- rnorm(1, am, as)
```

```r
b_hat[1] <- rnorm(1, bm, bs)
c_hat[1] <- rnorm(1, cm, cs)

for (i in 2: n_sim){
  a = a_hat[i-1]
  b = b_hat[i-1]
  c = c_hat[i-1]

  a_tr <- rnorm(1, am, as)
  b_tr <- rnorm(1, bm, bs)
  c_tr <- rnorm(1, cm, cs)

  tr_logl = log_likelihood(a_tr, b_tr, c_tr)
  logl = log_likelihood(a, b, c)
  A = tr_logl - logl

  dens  = get_dens(a, b, c)
  tr_dens = get_dens(a_tr, b_tr, c_tr)
  B = dens - tr_dens

  P = exp(A + B)
  rho = runif(1) < min(P, 1)
  a_hat[i] = a_tr*rho + a_hat[i-1]*(1-rho)
  b_hat[i] = b_tr*rho + b_hat[i-1]*(1-rho)
  c_hat[i] = c_tr*rho + c_hat[i-1]*(1-rho)

}
```

```r
print( paste0("a_mean: ", round( mean(a_hat), 4)))
```

```
## [1] "a_mean: -1.9919"
```

```r
print( paste0("b_mean: ", round( mean(b_hat), 4)))
```

```
## [1] "b_mean: 0.1649"
```

```r
print( paste0("c_mean: ", round( mean(c_hat), 4)))
```

```
## [1] "c_mean: 0.2901"
```

```r
print( paste0("a_variance: ", round( var(a_hat), 4)))
```

```
## [1] "a_variance: 0.0547"
```

```r
print( paste0("b_variance: ", round( var(b_hat), 4)))
```

```
## [1] "b_variance: 0.0064"
```

```r
print( paste0("c_variance: ", round( var(c_hat), 4)))
```
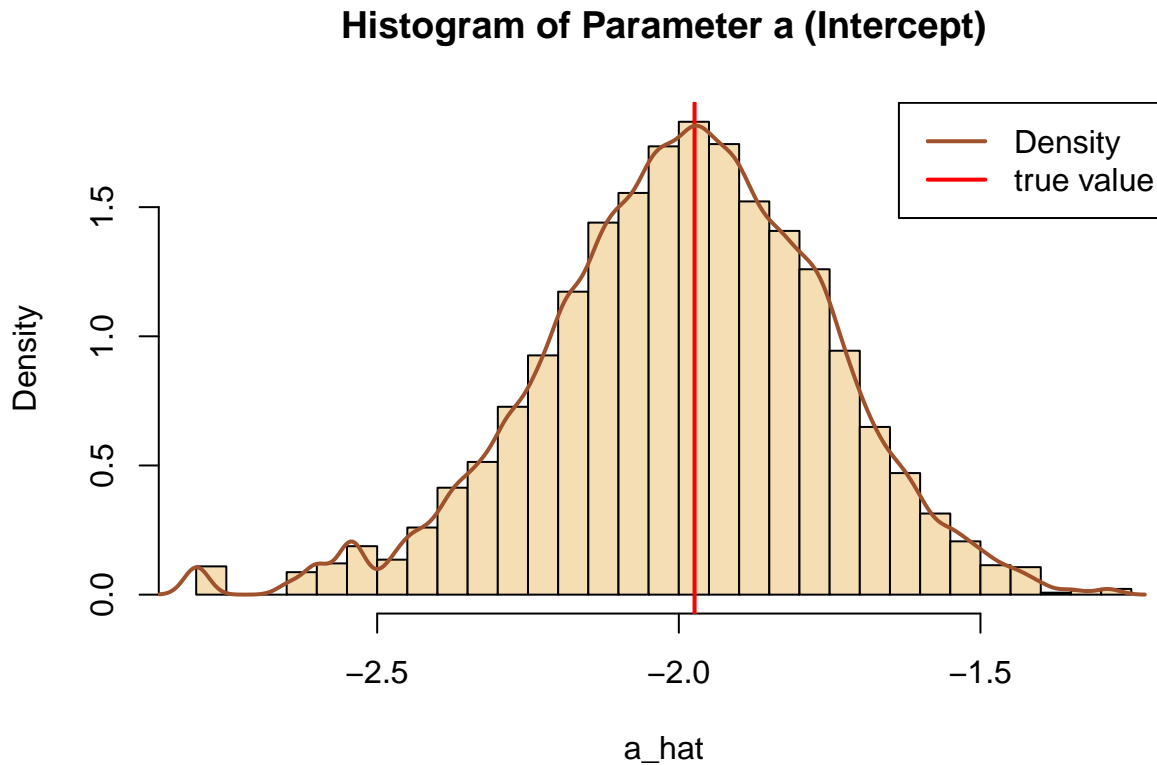
```
## [1] "c_variance: 0.0095"
```

parameter a histogram and density estimate

```r
hist(a_hat, freq = F, main ='Histogram of Parameter a (Intercept)', col = "wheat", breaks = 50)
#curve(dnorm(x,mean(a_hat), sd(a_hat)),lwd=2,col="sienna",add=T)
 abline(v=am,lwd=2,col=2)
```
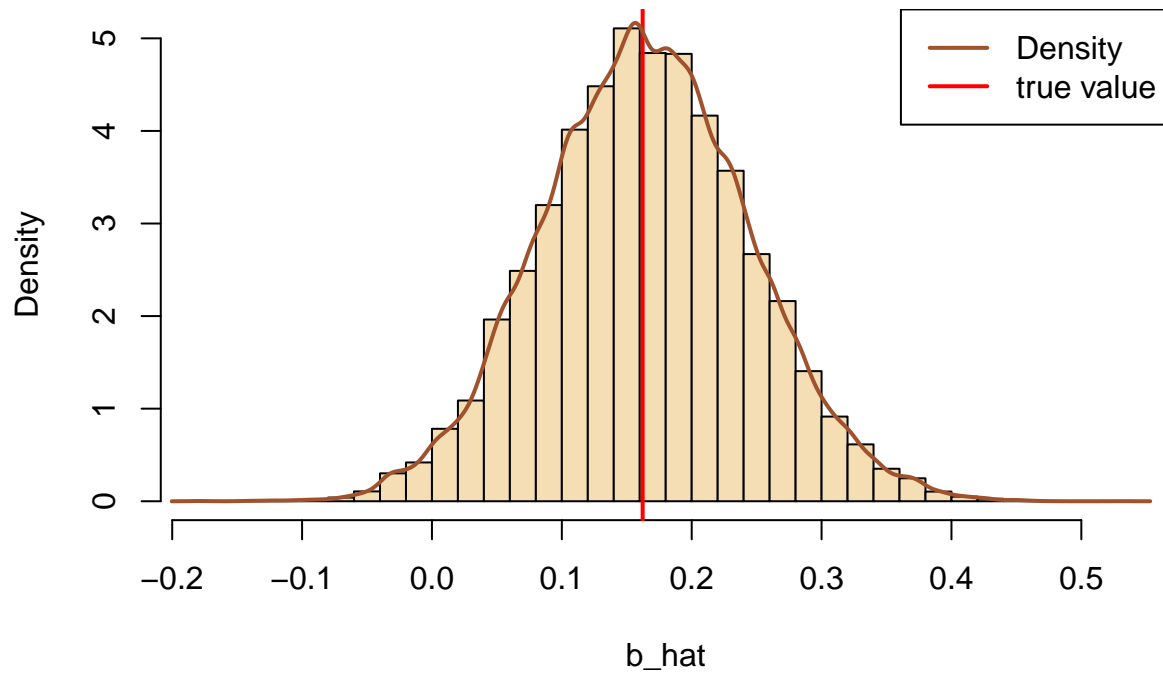
```
lines(density(a_hat), col="sienna", lwd=2)
legend("topright", legend = c("Density", "true value"), col=c("sienna", "red"), lty=1, lwd = 2)
```

Histogram of Parameter a (Intercept)

parameter a histogram and density estimate

```
hist(b_hat, freq = F, main ='Histogram of Parameter b (for HMO Type)', col="wheat", breaks = 50)
#curve(dnorm(x,mean(b_hat), sd(b_hat)),lwd=2,col="sienna",add=T)
 abline(v=bm,lwd=2,col=2)
lines(density(b_hat), col="sienna", lwd=2)
legend("topright", legend = c("Density", "true value"), col=c("sienna", "red"), lty=1, lwd = 2)
```

Histogram of Parameter b (for HMO Type)


```r
hist(c_hat, freq = F, main ='Histogram of Parameter c (for Health Status)', col="wheat", breaks = 50)
#curve(dnorm(x,mean(c_hat), sd(c_hat)),lwd=2,col="sienna",add=T)
 abline(v=cm,lwd=2,col=2)
lines(density(c_hat), col="sienna", lwd=2)
legend("topright", legend = c("Density", "true value"), col=c("sienna", "red"), lty=1, lwd = 2)
```

Histogram of Parameter c (for Health Status)