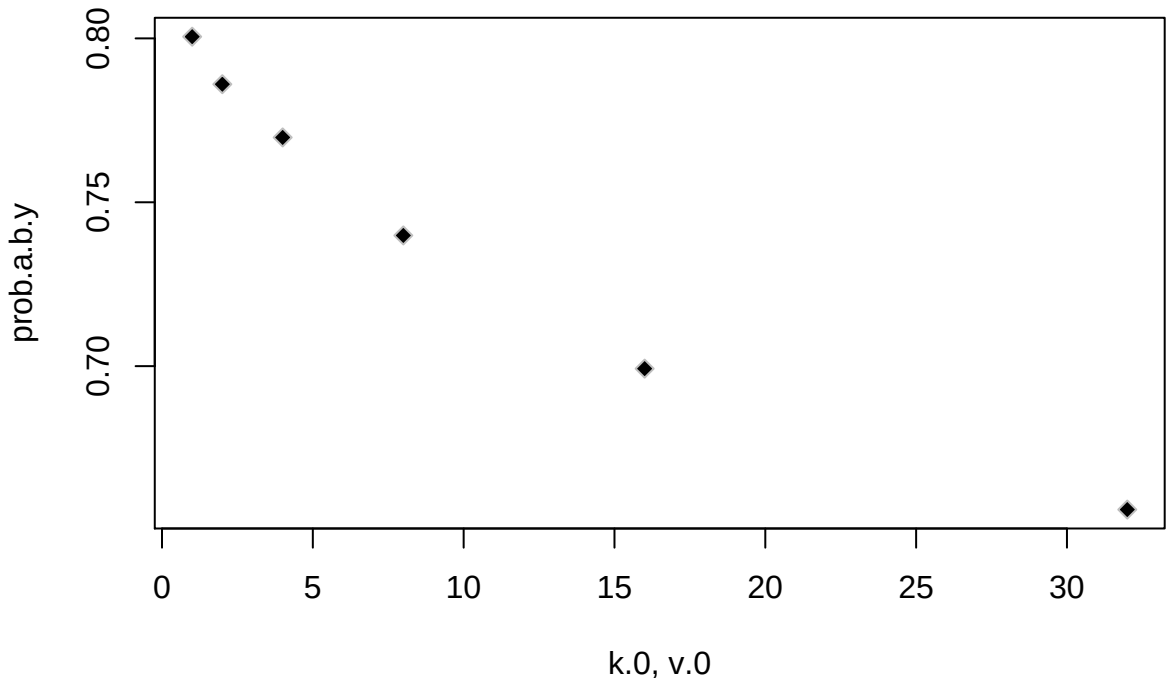
Homework 5

Thomas Fleming

October 14, 2016

5.2

```
set.seed(1)

n.a <- n.b <- 16

ybar.a <- 75.2
sd.y.a <- 7.3

ybar.b <- 77.5
sd.y.b <- 8.1

m.0 <- 75
sigma2.0 <- 100

k.0 <- v.0 <- 2^seq(0, 5, by = 1)

k.n.a <- k.0 + n.a
k.n.b <- k.0 + n.b
v.n.a <- v.0 + n.a
v.n.b <- v.0 + n.b

m.n.a <- (k.0*m.0 + n.a*ybar.a)/k.n.a
m.n.b <- (k.0*m.0 + n.b*ybar.b)/k.n.b

sigma2.n.a <- (v.0*sigma2.0 + (n.a - 1)*(sd.y.a)^2 + ((k.0*n.a)/k.n.a)*(ybar.a - m.0)^2)/v.n.a
sigma2.n.b <- (v.0*sigma2.0 + (n.b - 1)*(sd.y.b)^2 + ((k.0*n.b)/k.n.b)*(ybar.b - m.0)^2)/v.n.b

prob.a.b.y <- numeric(length(k.0))

for (i in 1:length(prob.a.b.y)) {
  sigma2.n.a.mc <- 1/rgamma(50000, (v.n.a[i])/2, ((sigma2.n.a[i])*(v.n.a[i]))/2)
  theta.n.a.mc <- rnorm(50000, m.n.a[i], sqrt(sigma2.n.a.mc/(k.n.a[i])))
  sigma2.n.b.mc <- 1/rgamma(50000, (v.n.b[i])/2, ((sigma2.n.b[i])*(v.n.b[i]))/2)
  theta.n.b.mc <- rnorm(50000, m.n.b[i], sqrt(sigma2.n.a.mc/(k.n.b[i])))
  prob.a.b.y[i] <- mean(theta.n.a.mc < theta.n.b.mc)
}
plot(k.0, prob.a.b.y, type = "p", pch = 23, col = "gray", bg = "black", xlab = "k.0, v.0")
```



Since the marginal posteriors of theta.a and theta.b are not very sensitive to changes in k and v, it would not be too difficult to convince people with different prior opinions that theta.a < theta.b is within a certain range (roughly .65 to .8). As k and v increase, the probability of theta.a < theta.b goes down because the prior data is given more weight in the posterior, due to a larger prior sample size. Thus, those favoring a more objective approach to statistics would have more confidence that theta.a < theta.b, while those with more subjective views would be less confident, but would still believe that theta.a < theta.b given the data.

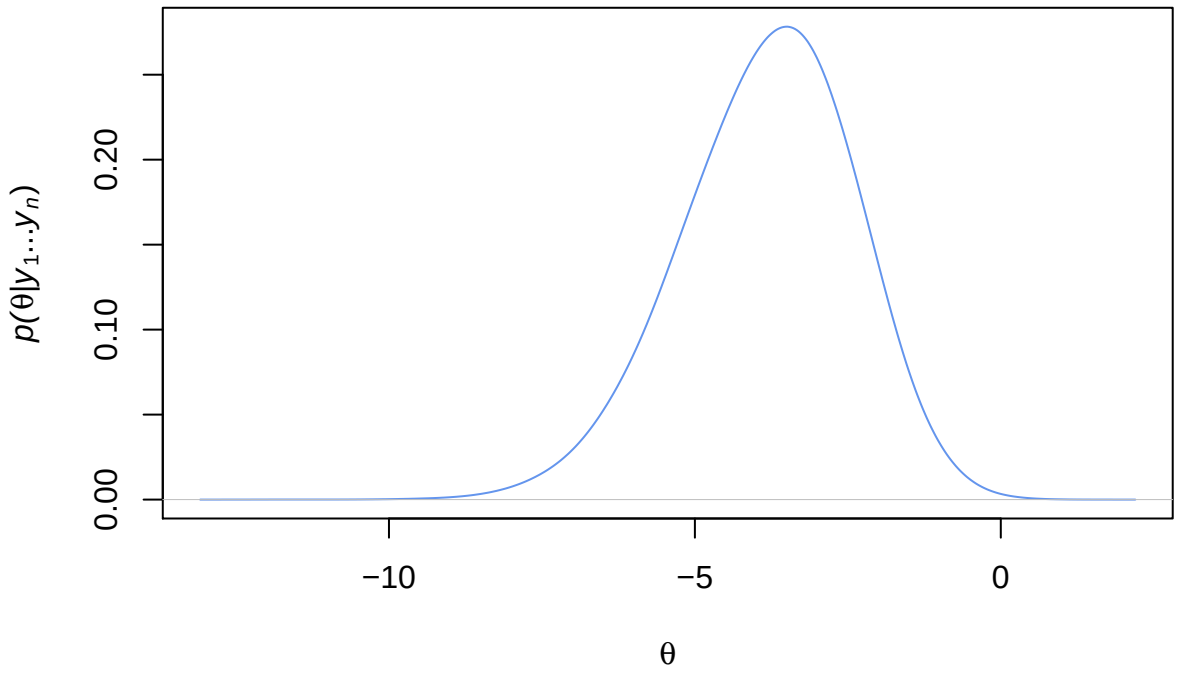
5.3

```
theta <- seq(1.5, 2, by = 1/length(k.0))
marg.theta <- rt(50000, v.n.a, (theta - m.n.a)/(sigma2.0/sqrt(k.n.a)))
```

```
## Warning in theta - m.n.a: longer object length is not a multiple of shorter
## object length
```

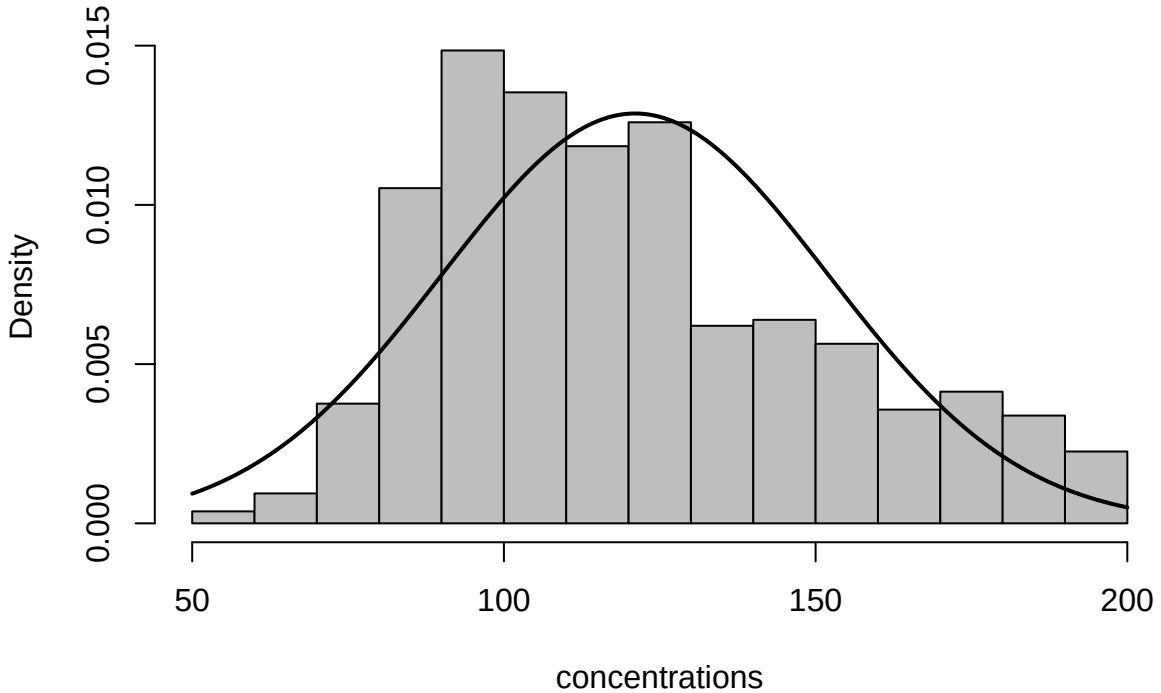
```
## Warning in rchisq(n, df)/df: longer object length is not a multiple of
## shorter object length
```

```
plot(density(marg.theta,adjust=3),main="", col = "cornflowerblue", xlab=expression(theta),ylab=expressi
     theta,"|",italic(y[1]),"...",italic(y[n]),")",sep="")))
```



6.2
a)

```
setwd("~/Downloads")
concentrations <- scan("glucose.dat")
hist(concentrations,main="", prob=TRUE,nclass=15,col="gray")
theta <- seq(50, 200, by = .01)
lines(theta, dnorm(theta, mean(concentrations), sd(concentrations)),lwd=2)
```

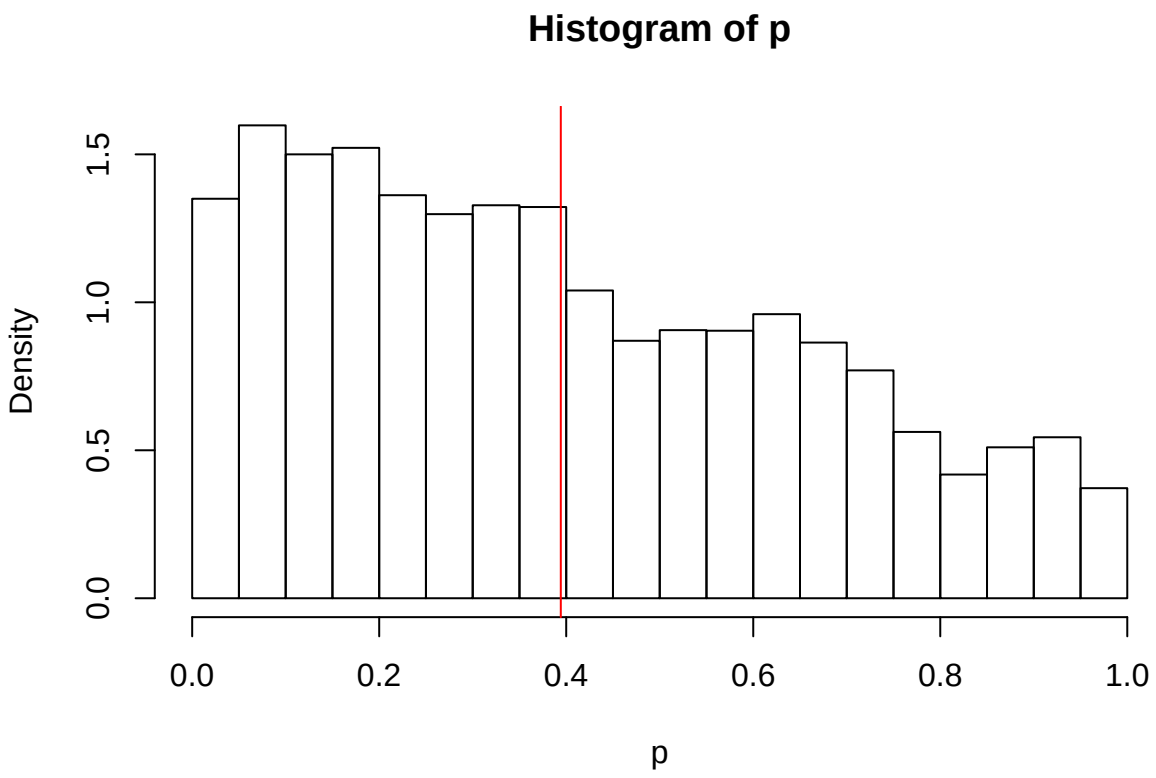


As shown in the plot, the histogram is slightly skewed to the right, with its peak being a little less than that of the normal distribution.

b) See handwritten work for derivations.

c)

```
a <- b <- 1

mu.0 <- 120
tau2.0 <- 200
sig2.0 <- 1000
kappa.0 <- 5
nu.0 <- 10

data <- concentrations
mean.data <- mean(concentrations)
var.data <- var(concentrations)
n.data <- length(data)
sum.data <- sum(concentrations)

I <- 10000
X <- matrix(NA, nrow=n.data, ncol=I)
p <- numeric(I)

#Random initialization to groups
X[,1] <- sample(1:2, n.data, prob=c(0.5, 0.5), replace= TRUE)
theta.1 <- mu.0
theta.2 <- mu.0
sigma2.1 <- 5
sigma2.2 <- 5
y.1 <- mu.0
y.2 <- mu.0
theta.1.full <- mu.0
theta.2.full <- mu.0
sigma2.1.full <- 5
sigma2.2.full <- 5

for(i in 2:I) {
#update p
n.1 <- sum(X[,i-1]==1)
n.2 <- sum(X[,i-1]==2)
p[i] <- rbeta(1, a + n.1, b + n.2)
#update X's
for(j in 1:n.data) {
  theta.1[j] <- theta.2[j] <- rnorm(1, mu.0, sqrt(tau2.0))
  sigma2.1[j] <- sigma2.2[j] <- 1/rgamma(1, v.0/2, (v.0*sig2.0)/2)
  w1 <- p[i]*dnorm(data[j], theta.1[j], sqrt(sigma2.1[j]))
  w2 <- (1 - p[i])*dnorm(data[j], theta.2[j], sqrt(sigma2.2[j]))
  w <- w1/(w1 + w2)
  X[j,i] <- sample(1:2, size = 1, prob=c(w, (1-w)))
}
}

#Marginal posterior for p
hist(p, prob=T)
abline(v=mean(p), col="red")
```




```
#apply(X,1,mean)
#update thetas and y's

#y.1[i] <- rnorm(1, theta.1, sqrt(sigma2.1))
#y.2[i] <- rnorm(1, theta.2, sqrt(sigma2.2))

#theta.1.full[i] <- rnorm(1, ((mu.0/tau2.0) + (y.1/sigma2.1)/(1/tau2.0) + (1/sigma2.1)), sqrt((1/tau2.0
#theta.2.full[i] <- rnorm(1, ((mu.0/tau2.0) + (y.2/sigma2.2)/(1/tau2.0) + (1/sigma2.2)), sqrt((1/tau2.0

#update sigma-squared
#sigma2.1.full[i] <- 1/gamma(1, (nu.0 + 1)/2, (nu.0*sig2.0 + (y.1 - mu.0)^2)/2)
#sigma2.2.full[i] <- 1/gamma(1, (nu.0 + 1)/2, (nu.0*sig2.0 + (y.2 - mu.0)^2)/2)
#}

#par(mfrow=c(1,2))
#acf(theta.1.full)
#acf(theta.2.full)

#library(coda)
#c(effectiveSize(mcmc(theta.1.full),
#effectiveSize(mcmc(theta.2.full))
```