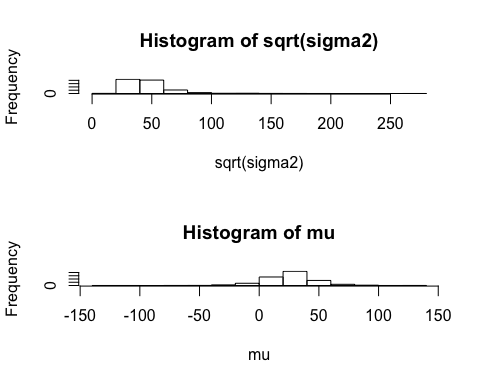
Homework2.R

EmilyErnst

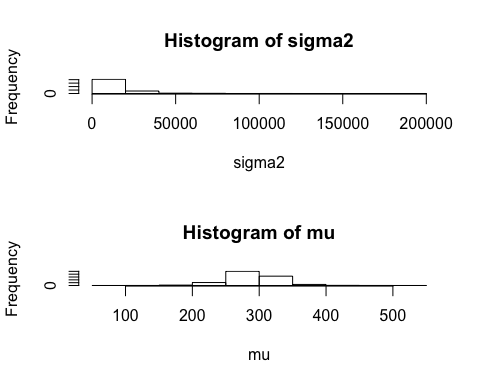
Tue Feb 28 00:34:27 2017

# Homework 2   
# Emily Ernst  
  
#Problem 1: Refer to Lab 4. Re-analyze Newcomb’s speed of light experiment, but this time  
#using a conjugate prior for μ,σ2. For σ2, select a prior scale parameter that is in agreement with the sample variance and for the prior degrees of freedom, choose a value that, relative to the sample size, does not reflect much confidence in the prior scale. For μ, choose whatever values of the prior parameters seem reasonable to you, and justify your choices.  
  
newcomb\_data = c(28, 26, 33, 24, 34, -44, 27, 16, 40, -2,  
 29, 22, 24, 21, 25, 30, 23, 29, 31, 19,  
 24, 20, 36, 32, 36, 28, 25, 21, 28, 29,  
 37, 25, 28, 26, 30, 32, 36, 26, 30, 22,  
 36, 23, 27, 27, 28, 27, 31, 27, 26, 33,  
 26, 32, 32, 24, 39, 28, 24, 25, 32, 25,  
 29, 27, 28, 29, 16, 23)  
  
y = newcomb\_data  
  
#degrees of freedom = 5 because less degrees of freedom decreases confidence in the data

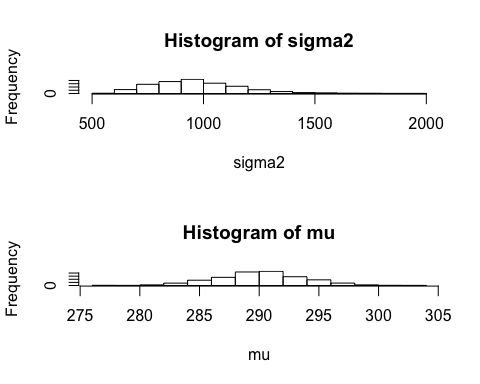
#sigma2 and mu = based on data variance and mean because I don't have any more info than that   
  
S=sum((y-mean(y))^2)  
n=length(y)  
sigma2=S/rchisq(1000, 5)  
par(mfrow=c(2,1))  
hist(sqrt(sigma2))  
  
k = 5  
  
mu = rnorm(1000, mean = mean(y), sd = sqrt(sigma2)/sqrt(k))  
hist(mu)



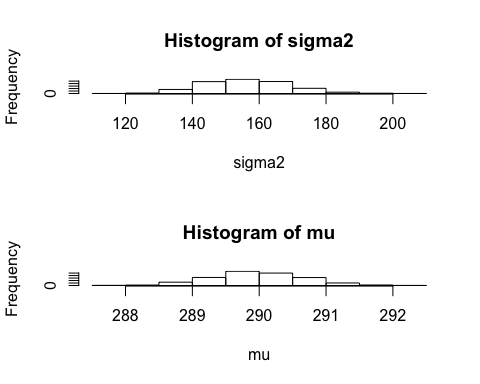
#Problem 2: Albert (2011) describes a dataset that includes marathon completion times (in minutes) for men aged 20 to 29 years who participate in the New York marathon. The dataset has times for 20 runners and you can access the data from R by issuing the commands:  
  
library(LearnBayes)  
data(marathontimes)  
attach(marathontimes)  
  
#Let y1, y2, ..., y20 denote the completion times for the 20 runners and assume that yi ∼ N(μ,σ2), where both μ and σ2 are unknown. Use a conjugate prior for μ,σ2 and select values for the prior parameters which you think are good choices. Justify those choices. Carry out a test of sensitivity to the prior by setting ν0, κ0 at three different values reflecting low, medium and high confidence in your prior guesses for μ,σ2.  
  
x = marathontimes$time  
  
#mean and variance are based on the data again because I have no other data to compare to  
  
#Low  
  
S=sum((x-mean(x))^2)  
n=length(x)  
sigma2=S/rchisq(1000, 5)  
  
k = 10  
  
mu = rnorm(1000, mean = mean(290), sd = sqrt(sigma2)/sqrt(k))  
  
par(mfrow=c(2,1))  
hist(sigma2)  
hist(mu)



#Medium  
  
S=sum((x-mean(x))^2)  
n=length(x)  
sigma2=S/rchisq(1000, 50)  
  
k = 70  
  
mu = rnorm(1000, mean = mean(290), sd = sqrt(sigma2)/sqrt(k))  
  
par(mfrow=c(2,1))  
hist(sigma2)  
hist(mu)

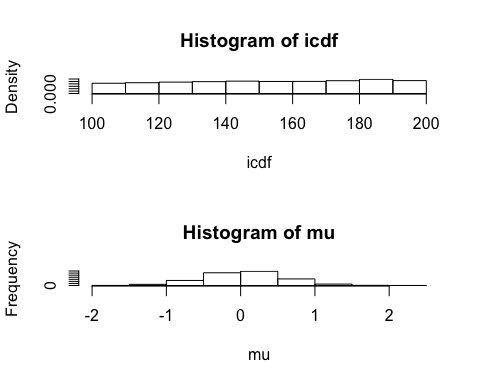


#High  
  
S=sum((x-mean(x))^2)  
n=length(x)  
sigma2=S/rchisq(1000, 300)  
  
k = 350  
  
mu = rnorm(1000, mean = mean(290), sd = sqrt(sigma2)/sqrt(k))  
  
par(mfrow=c(2,1))  
hist(sigma2)  
hist(mu)



#Problem 3: Revisit the football scores example, where now we consider non-conjugate prior distributions for μ, σ2 (see Notes, pp 27-33). We wish to obtain the posterior distributions for the two parameters μ,σ2, and we do so under two different priors.  
  
#3.a First choose a standard non-informative prior for μ, σ2, with p(μ, σ2) ∝ σ−2. Approximate the posterior distribution of μ, σ2 by first drawing values of σ2 from its marginal posterior distribution p(σ2|y) using the inverse CDF method (see Notes, p. 33) and then drawing values of μ from its conditional posterior distribution p(μ|σ2,y). Draw histograms of the posterior distributions of μ and of σ2.  
  
data("footballscores")  
attach(footballscores)  
  
z = favorite - underdog - spread  
  
par(mfrow=c(2,1))  
  
sig2 = seq(from=100, to=200, length=1000)  
  
nu0 = 0  
n = length(z)  
nun = nu0  
  
S=sum((z-mean(z))^2)  
n=length(z)  
sigma2n=S/rchisq(1000, n-1)  
  
p=seq(from=0, to=1, length=length(sig2))  
lp = p   
for(i in 1:length(lp)){  
 lp[i] = -(nun+1)\*log(sig2[i]) - sigma2n/sig2[i]  
}

p <- exp(lp - max(lp))  
p <- p/sum(p)  
  
cdf <- cumsum(p)  
  
u <- runif(5000)  
icdf <- rep(0,5000)  
for (i in 1:5000) {  
 icdf[i] <- sig2[1]  
 for (j in 1:1000) {  
 if (cdf[j] < u[i]) icdf[i] <- sig2[j]  
 }  
}  
  
hist(icdf,freq=F)  
  
mu = rnorm(1000, mean = mean(z), sd = sqrt(sigma2n)/sqrt(n))  
hist(mu)



#3.b Consider now a non-conjugate prior for μ, σ2 of the form p(μ, σ2) = p(μ)p(σ2) = N(μ|μ0, τ02)Inv − χ2(ν0, σ02) for some reasonable choices of μ0,τ02,ν0,σ02. Implement the inverse CDF method to draw values from the marginal posterior distribution of σ2, and given those draws, generate values of μ from its conditional posterior distribution, as in part [3.a].  
  
sig2 = seq(from=100, to=200, length=1000)  
  
nu0 = 600  
n = length(z)  
nun = nu0 + 0.5\*n  
  
sigma20 = 200  
sigma2n = sigma20 + 0.5\*(n-1)\*var(z) + 0.5\*n\*mean(z)\*mean(z)/(n)  
  
p=seq(from=0, to=1, length=length(sig2))  
lp = p   
for(i in 1:length(lp)){  
 lp[i] = -(nun+1)\*log(sig2[i]) - sigma2n/sig2[i]  
}  
  
p <- exp(lp - max(lp))  
p <- p/sum(p)  
  
cdf <- cumsum(p)  
  
u <- runif(5000)  
icdf <- rep(0,5000)  
for (i in 1:5000) {  
 icdf[i] <- sig2[1]  
 for (j in 1:1000) {  
 if (cdf[j] < u[i]) icdf[i] <- sig2[j]  
 }  
}  
  
hist(icdf,freq=F)  
  
tau0 = 250  
taun = 1/((1/tau0)+(n/sigma20))  
  
mu0 = 0  
mun = ((1/tau0)\*mu0) + ((n/sigma20)\*mean(z))/((1/taun)+(n/sigma2n))  
  
mu = rnorm(1000, mean = mean(mun), sd = sqrt(taun)/sqrt(n))  
hist(mu)

