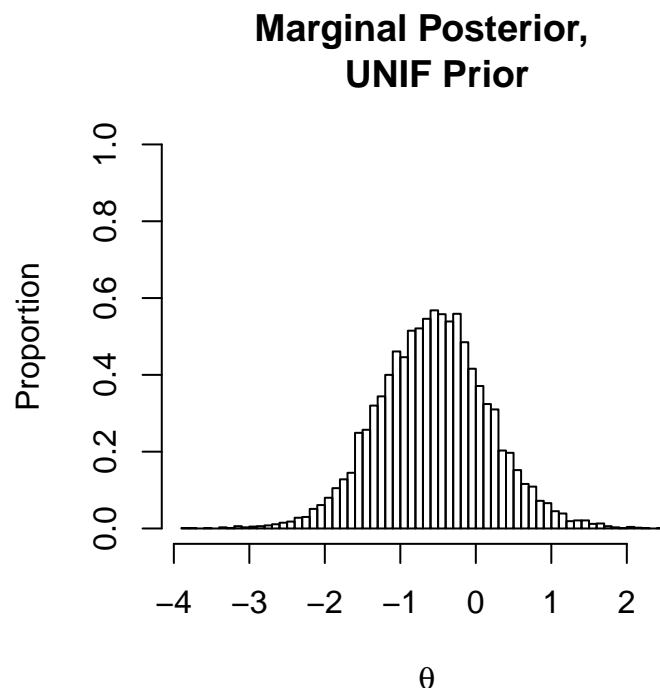


- (15 points) For this question we are going to take a deeper look at the impact of our prior distributions and the number of observed data points. Generate fifteen observations from a standard normal distribution and use independent priors on the parameters from the normal distribution:  $p(\theta) \sim N(\mu_o, \tau_o^2)$  and  $p(\sigma_o^2) \sim IG(\nu_o/2, \nu_o^2/2)$ . This will require running the a Gibbs sampler to estimate the posterior distribution.

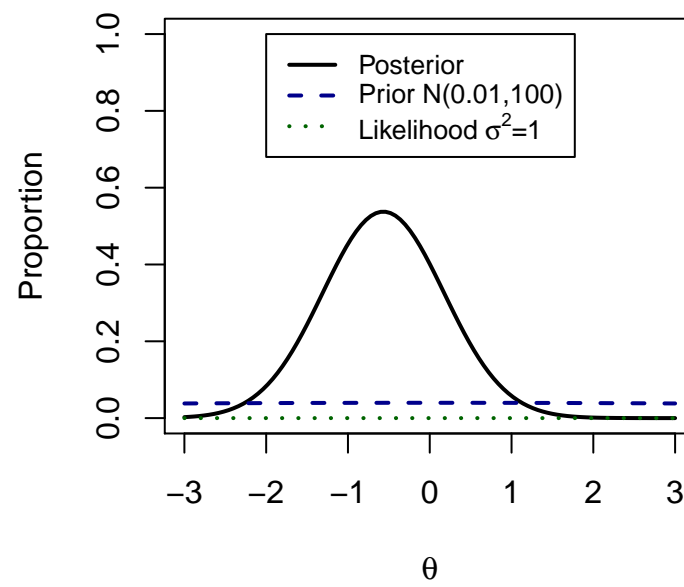
First plot the marginal posterior distributions for  $\theta$  under the flat prior  $p(\theta, \sigma^2) \propto 1$ : Note that this is proportional to the likelihood function.

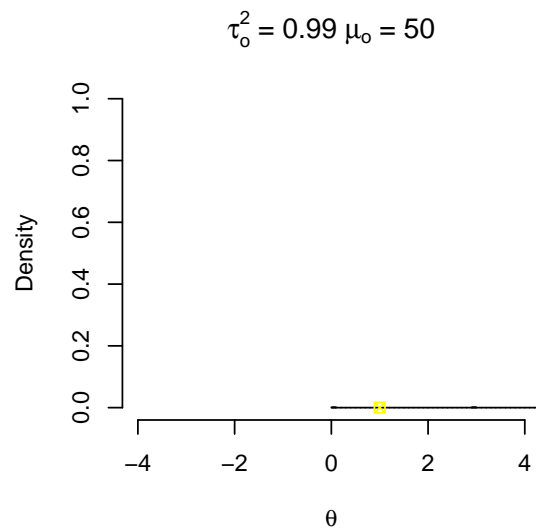
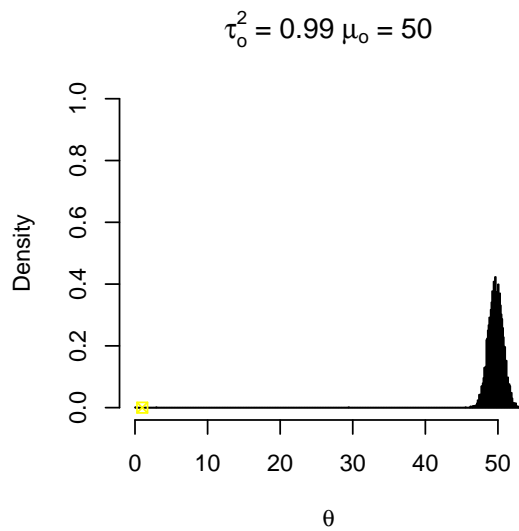
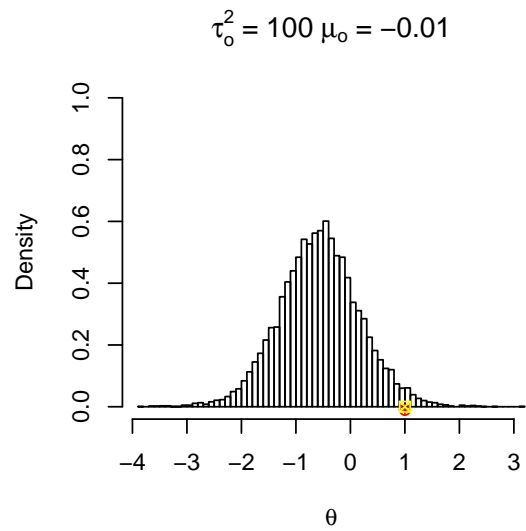
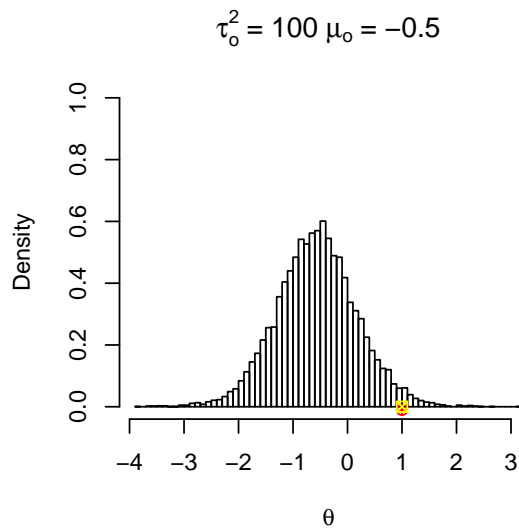
Under the flat bivariate uniform prior on  $p(\theta, \sigma^2)$ , the marginal posterior on  $\theta$  takes the form:

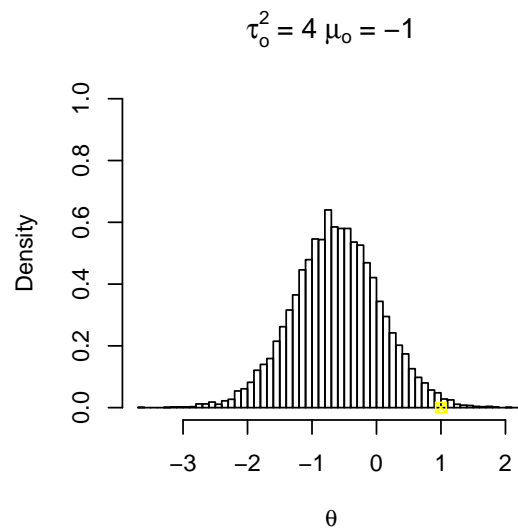
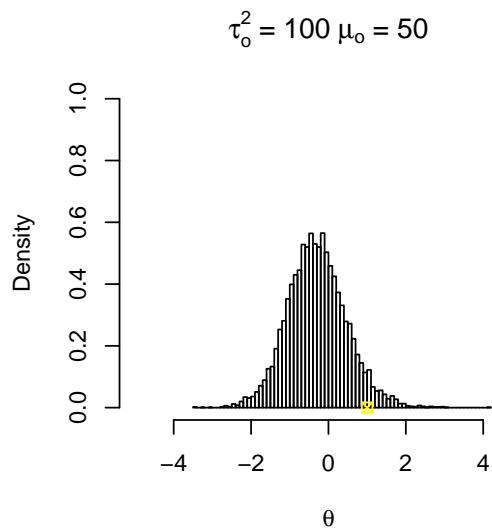
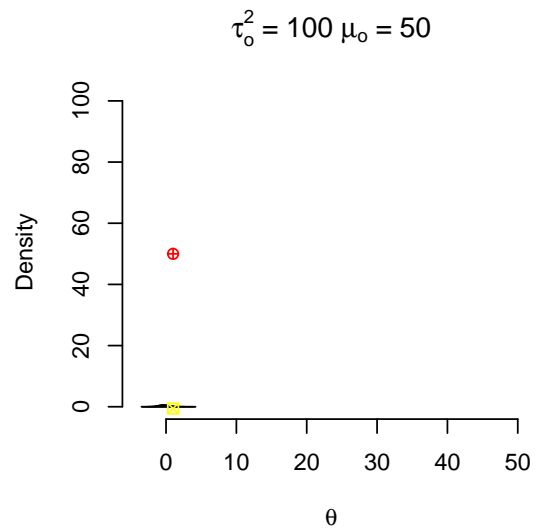
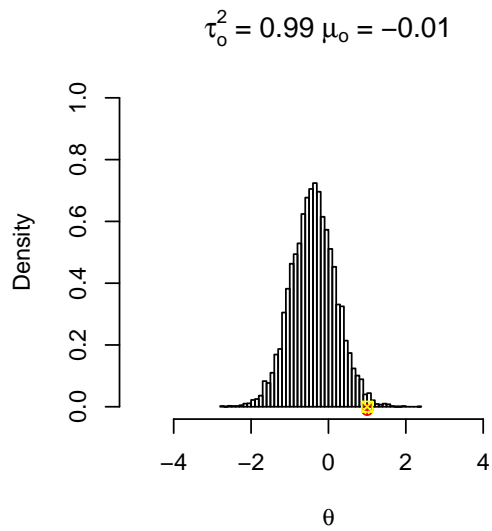
$(2\pi)^{-\frac{(n+1)}{2}} * \Sigma(y_i - \theta)^2$  Using the observed sample of 15  $y_i$  and  $n = 15$  from the standard normal distribution, the marginal posterior of  $\theta$  is shown below.  $\sigma^2$  is fixed to be 1.

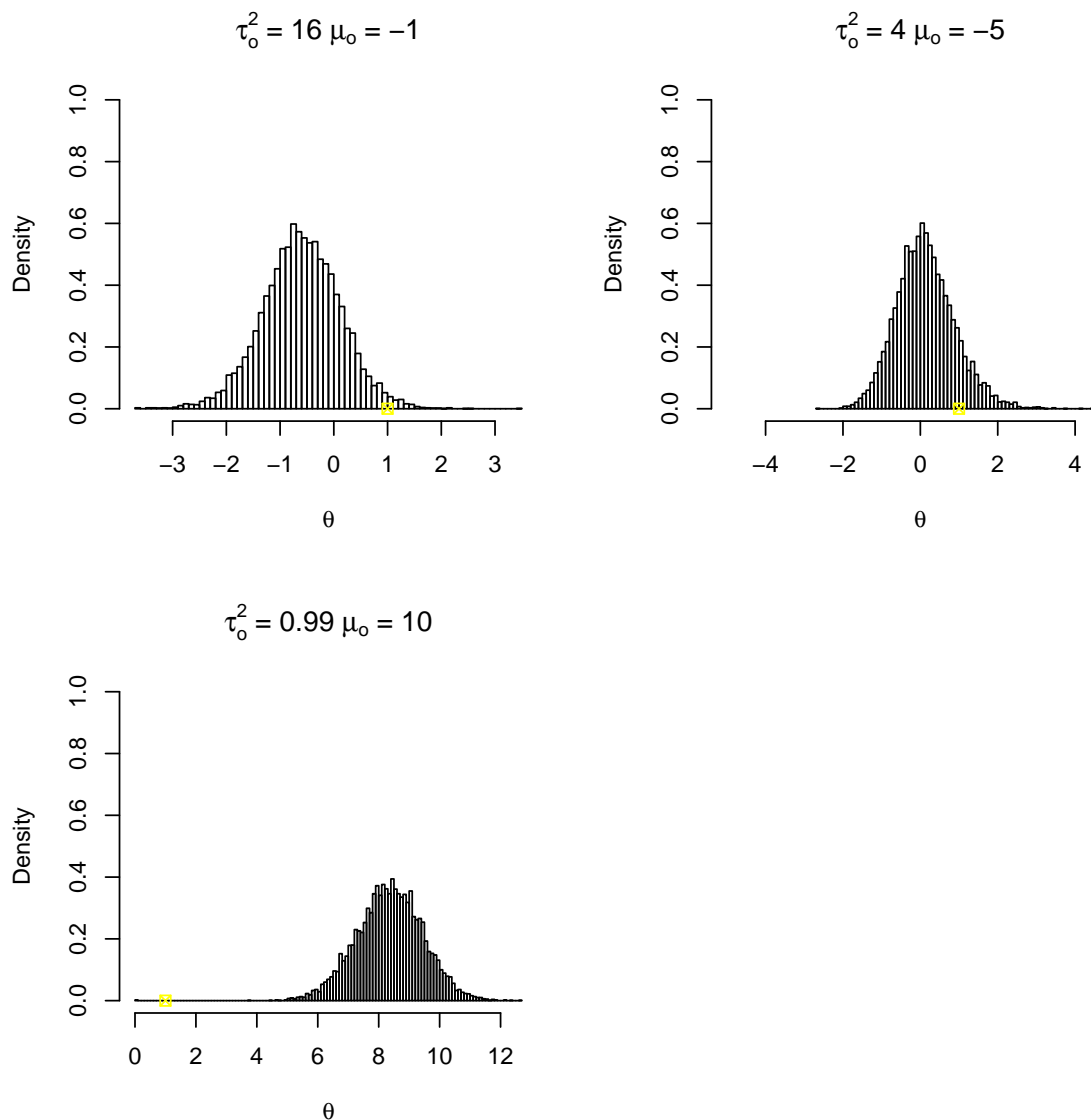


Next, rerun the Gibbs sampler using at least 5 different values for  $\mu_o, \tau_o^2$ , where some of the values should be close to the truth and some should be very different. The same principles apply for  $\sigma_o^2$ , but fix  $p(\sigma_o^2) \propto \frac{1}{\sigma^2}$  so that we can see the difference for  $\theta$ . Make sure to look at a few situations where  $\tau_o^2$  is very small and very large. Explain what you have learned from this question and create a figure that contains the prior, likelihood, and posterior on a single graph.









The axes may change. In each plot, the red represents the initializations of  $\mu_o$  and  $\tau_o$  and the yellow represents the true mean and variance of the 15 simulated points from the standard normal distribution where how close the two initializations are to the truth varied. Note that  $\tau_o^2$  is not  $\sigma^2$  since  $\tau_o^2 = \frac{\sigma^2}{\kappa_o}$ , and neither are not plotted.

If  $\tau_o^2$  is somewhat far from the truth, the marginal posterior of  $\theta$  is generally unbiased. However, if not, then the marginal posterior of  $\theta$  is centered near the initialized  $\mu_o$ . The spread appears to be robust for all initializations of  $\tau_o^2$ .

2. (10 points) Assume you are faced with a modeling scenario for a multivariate normal response where you do not have a good sense of the covariance structure. Select and justify a prior distribution for  $\Sigma$  of  $(\Sigma^{-1})$ .

Generally the gamma distribution is used as a prior on the precision parameter in the univariate normal

case. The Wishart distribution is a multivariate form of the gamma distribution, and so makes sense for the prior on the multivariate covariance matrix,  $\Sigma$ .

The prior on  $\Sigma$  requires parameters,  $\nu_o$  and  $\Psi_o$ . Let  $p$  = the number of groups in the multivariate normal distribution. Choose  $\nu_o = p+2$  to make the prior less informative as the  $E[\Sigma] = \frac{1}{\nu_o - p - 1} * S_o = S_o$ . Note  $\nu_o$  can be thought of as the prior sample size, and we need at least one observation per group, so setting  $\nu_o$  this ways would mean we have very little information, nearly one observation, in each group.

$S_o$  is the prior sums of squares. In class, choosing the off-diagonal elements to be 0 was less restrictive than assuming non-zero covariance terms between observations. Setting the diagonal elements to 100 is fairly nonrestrictive as it allows for enough variation to make the prior have more spread and therefore be less informative.

3. (10 points) Verify that for a multivariate normal distribution, the variance and expectation can be extracted from the kernel,  $\exp[\frac{-1}{2}(\tilde{\theta}^T A \tilde{\theta} - \tilde{\theta}^T B + \dots)]$ , where the variance is  $A^{-1}$  and the expectation is  $A^{-1}B$ .

The kernel of a multivariate distribution of  $\theta|A, B$  is given in the problem description. Consider  $\exp[\phi] = \exp[\frac{-1}{2}(\tilde{\theta}^T A \tilde{\theta} - \tilde{\theta}^T B + \dots)] \ni$

$$\phi = -\frac{1}{2}\tilde{\theta}^T A \tilde{\theta} - \tilde{\theta}^T B + \dots$$

$$\propto -\frac{1}{2}(A^{-1})^{-1}(\tilde{\theta}^T \tilde{\theta} - \tilde{\theta}^T A^{-1}B - c\tilde{\theta}) \text{ from p.58 in the notes}$$

$$\propto -\frac{1}{2}(A^{-1})^{-1}(\tilde{\theta}^T \tilde{\theta} - \tilde{\theta}^T A^{-1}B - \tilde{\theta}^T c) \text{ c is symmetric}$$

$$\propto -\frac{1}{2}(A^{-1})^{-1}(\tilde{\theta}^T \tilde{\theta} - 2\tilde{\theta}^T A^{-1}B)$$

$$\propto -\frac{1}{2}(A^{-1})^{-1}(\tilde{\theta}^T \tilde{\theta} - 2\tilde{\theta}^T A^{-1}B + (A^{-1}B)^T(A^{-1}B)) + \frac{1}{2}(A^{-1})^{-1}(A^{-1}B)^T(A^{-1}B)$$

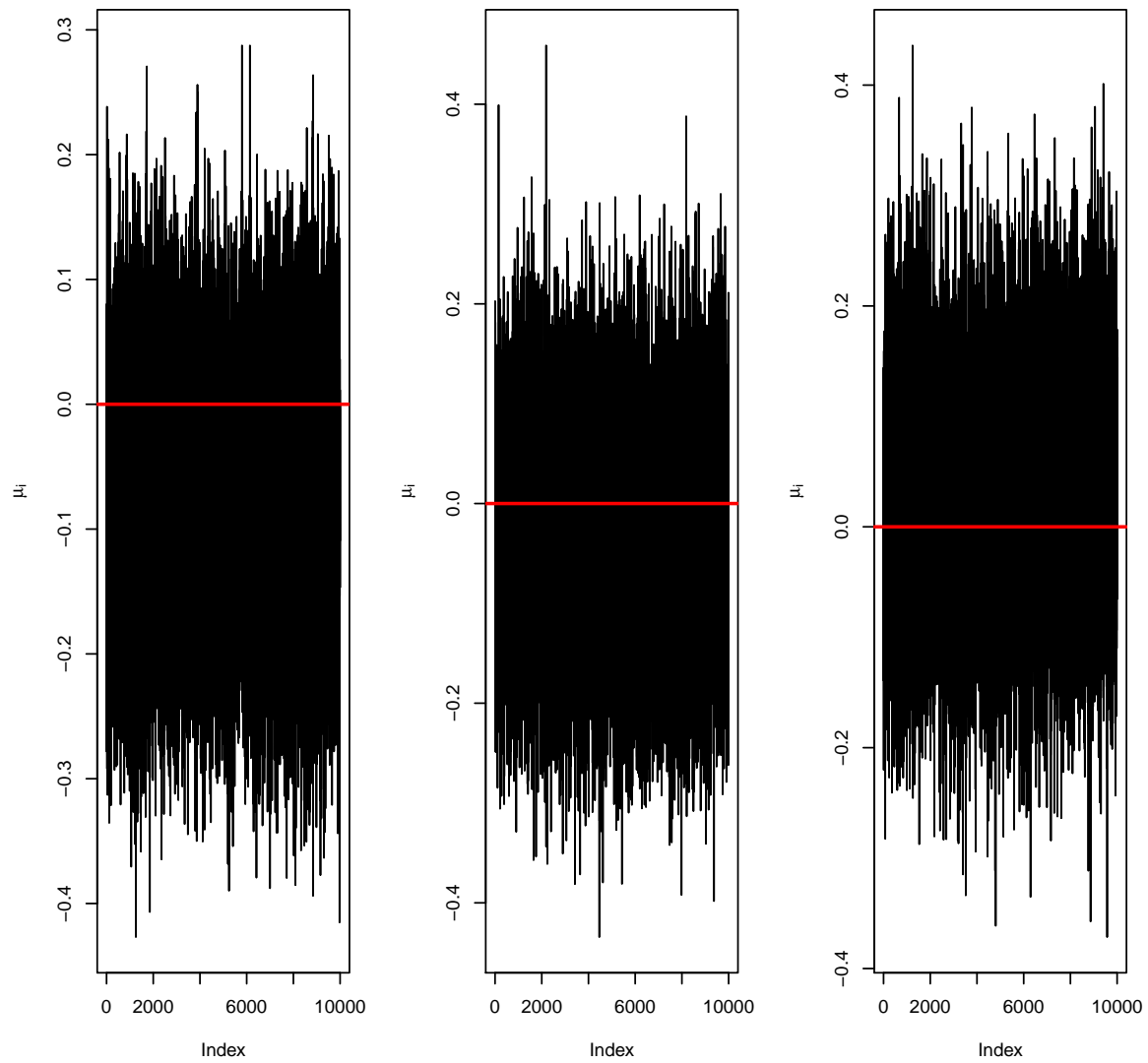
$$\propto -\frac{1}{2}(A^{-1})^{-1}(\tilde{\theta}^T - (A^{-1}B)^T)(\tilde{\theta} - (A^{-1}B))$$

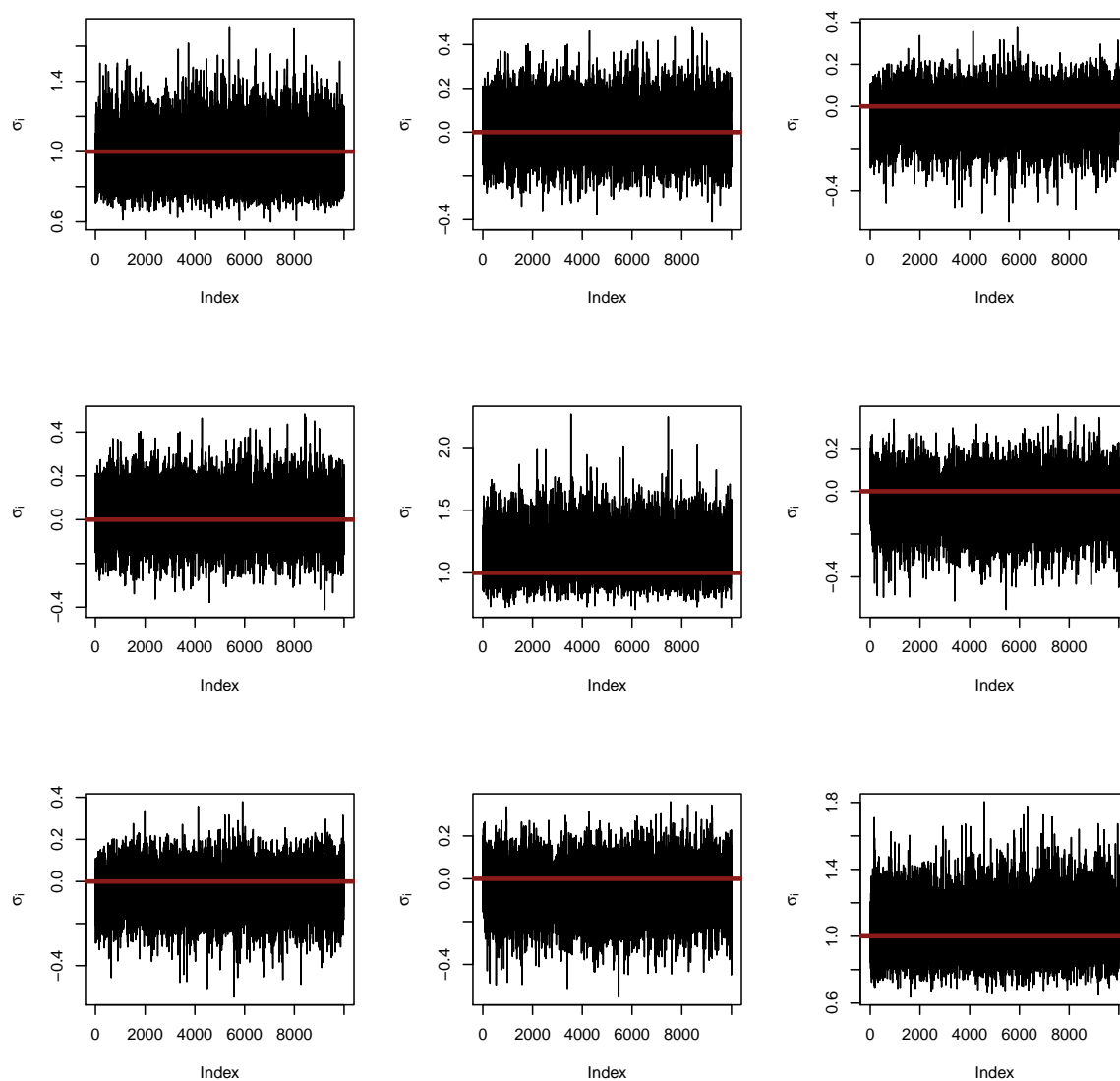
$$\propto -\frac{1}{2}[(\tilde{\theta} - (A^{-1}B))^T](A^{-1})^{-1}(\tilde{\theta} - (A^{-1}B))$$

$$\rightarrow E[\tilde{\theta}] = (A^{-1}B); \text{Var}[\tilde{\theta}] = A^{-1}$$

4. (20 points) Simulate data from a multivariate normal distribution, with  $p = 3$ . Choose a mean vector  $\mu$  and covariance matrix  $\sigma$ . Run a Gibbs sampler and verify that your sampler returns the true values. Comment on the results.

The true expected value in all groups is 0 and the true correlation matrix is the identity matrix. The prior means on  $\tilde{\theta}$  were all 0 and the prior covariance matrix was fairly uninformative as  $\text{diag}(100)$ . The prior sample size for  $\Sigma$  must be at least  $p$  = the number of groups = 3. Making the prior sample size small makes the prior less informative, let the prior sample size be 5 for all groups. Set the prior residual sums of squares matrix to be  $\text{diag}(10)$  which is uninformative relative to the prior means.





The trace plots for the posterior on  $\tilde{\theta}$  show unbiased convergence to the true mean, 0, quite fast which was to be expected because both the true means for the random draws and the prior means were set to 0. By making the posteriors fairly uninformative, we see unbiased convergence of the posterior on  $\tilde{\sigma}^2$  to the prior correlation matrix (the identity) quickly as well which is seen in the diagonal plots being centered at 1 and the off diagonal trace plots being centered at 0.

#### 5. Modeling heavy tailed distributions.

$$(a) \ y \sim N(\theta, \sigma^2 * \gamma)$$

$$\sigma^2 \sim \text{IG}(a, b)$$

$$\text{Find } p(y | \theta, \gamma) = \int p(y | \theta, \gamma, \sigma^2) p(\sigma^2) d\sigma^2$$



$$\begin{aligned}
& \propto \int (\sigma^2 \gamma \pi)^{-1/2} \exp\left[\frac{-1}{2\sigma^2 \gamma} (y - \theta)^2\right] \frac{\left(\frac{1}{\sigma^2}\right)^{a-1} \exp\left[\frac{-1}{\sigma^2} b\right] b^a}{\Gamma(a)} d\sigma^2 \\
& \propto (\pi \gamma)^{-1/2} * \frac{b^a}{\Gamma(a)} \int \left(\frac{1}{\sigma^2}\right)^{a+\frac{1}{2}-1} \exp\left[\frac{-1}{\sigma^2} \left(\frac{(y-\theta)^2}{2\gamma} + b\right)\right] d\sigma^2 \\
& \propto (\pi \gamma)^{-1/2} \Gamma\left(a + \frac{1}{2}\right) * \frac{b^a}{\Gamma(a) * \left[\frac{(y-\theta)^2}{2\gamma} + b\right]^{a+\frac{1}{2}}} * 1.
\end{aligned}$$

This is because we get out a  $\text{GAM}(a + \frac{1}{2}, \frac{(y-\theta)^2}{2\gamma} + b)$

Let  $a = \frac{1}{2}$  and  $b = 1$ :

$$\begin{aligned}
& \propto (\pi \gamma)^{-1/2} \Gamma\left(\frac{1}{2} + \frac{1}{2}\right) * \frac{1}{\Gamma\left(\frac{1}{2}\right) * \left[\frac{(y-\theta)^2}{2\gamma} + 1\right]^1} \\
& \propto (\gamma)^{-1/2} * \frac{1}{\pi} * \frac{1}{\left[\frac{(y-\theta)^2}{2\gamma} + 1\right]}
\end{aligned}$$

The above is not in the correct form. Now let  $a = \frac{1}{2}$  and  $b = \frac{1}{2}$ :

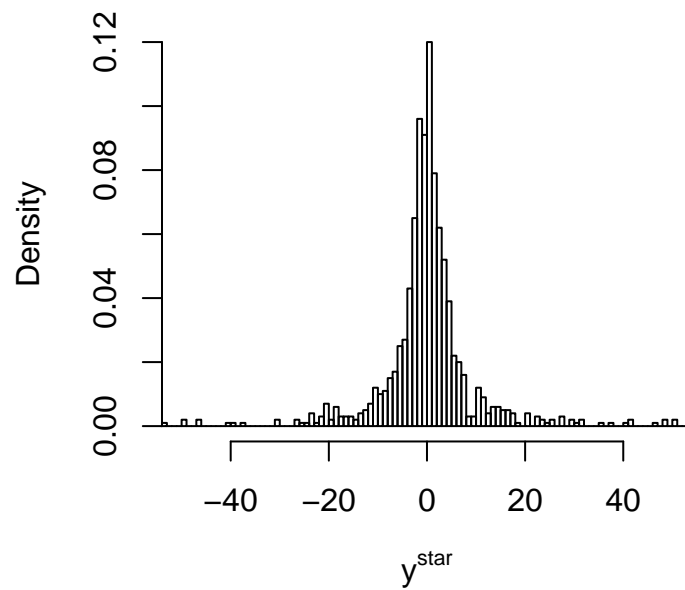
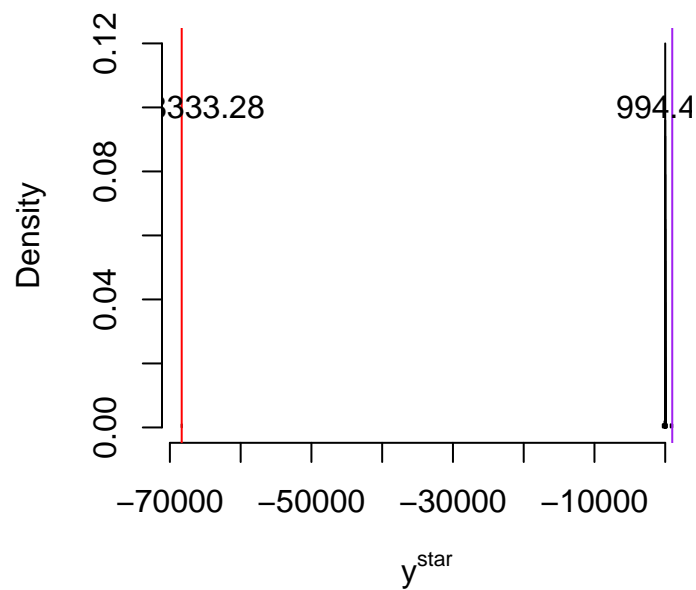
$$\begin{aligned}
& \propto (\pi \gamma)^{-1/2} \Gamma\left(\frac{1}{2} + \frac{1}{2}\right) * \frac{1}{\Gamma\left(\frac{1}{2}\right) * \left[\frac{(y-\theta)^2}{2\gamma} + \frac{1}{2}\right]^1 2^{\frac{1}{2}}} \\
& \propto 2^* (\gamma)^{-1/2} * \frac{1}{\pi} * \frac{1}{\left[\frac{(y-\theta)^2}{\gamma} + 1\right] 2^{\frac{1}{2}}} \\
& \propto (\gamma)^{-1/2} * \frac{1}{\pi} * \frac{1}{\left[\frac{(y-\theta)^2}{\gamma} + 1\right]}
\end{aligned}$$

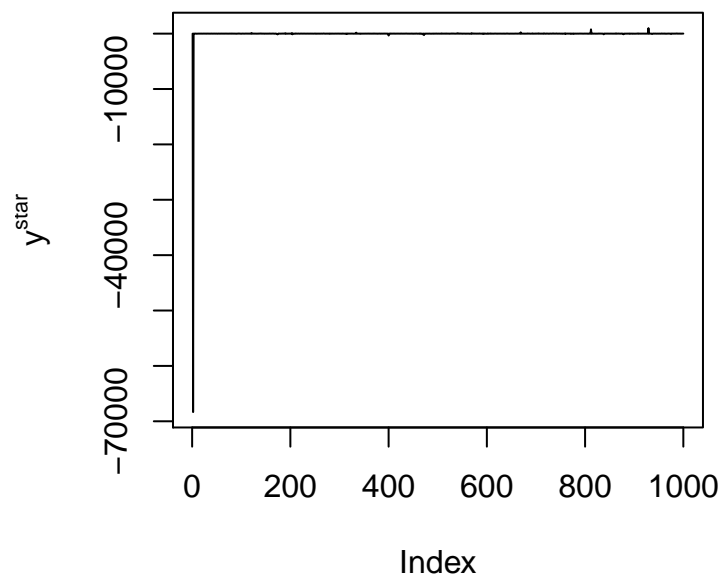
This shows  $p(y|\theta, \gamma) \propto \text{Cauchy}(\theta, \text{variance} = \gamma)$ .

- (b) The location is  $\theta$  and the scale is actually the standard deviation, so  $\gamma^{\frac{1}{2}}$  which holds when  $a = \frac{1}{2}$  and  $b = \frac{1}{2}$ .
- (c) *Generate a posterior predictive distribution for  $y^*$ .*

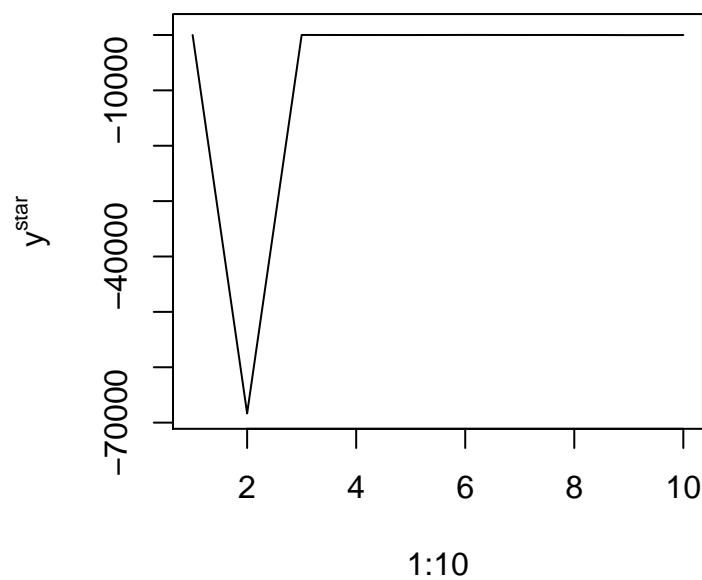
I assumed we were supposed to fix  $\gamma$  and  $\theta$ , so that  $p(\lambda) = p(\sigma^2)$  and  $p(y|\lambda^*) = p(y|\gamma, \theta, \sigma^2)$  where  $\gamma$  and  $\theta$  are somewhat arbitrarily chosen to be fixed at 2 and 0, respectively.

In parts (a) and (b) we showed that if  $a = \frac{1}{2}$  and  $b = \frac{1}{2}$ , then  $p(y|\theta, \gamma)$  is  $\text{Cauchy}(\theta, \gamma)$ . Use these values for the prior on  $\sigma^2$  and generate  $y^*$  from a  $N(0, 2\sigma^2)$  distribution where each  $\sigma^2$  is drawn independently.

histogram of  $y^{\text{star}}$  zoomed inhistogram of  $y^{\text{star}}$  zoomed out



### Zoomed In First 10 Random Draws



Zooming in we see the majority of the mass of  $p(y^*|\theta, \gamma)$  forms a symmetric distribution about 0. There was an outlier in the draws for the precision term, making the histogram of all observations not helpful. Zooming in,  $p(y^*|\theta, \gamma)$ , is still quite dispersed.

- (d) In the context of modeling over-dispersed data,  $\gamma$  can be set to a fractional number, making thicker tails in the normal distribution. The distribution is still a bell-shaped curve, but with more observations in the tails. The result that the posterior predictive distribution is Cauchy is useful because it is easy to sample from.

In terms of modeling over-dispersed data, the result is useful because we only need to set values of  $\theta$  and  $\gamma$ , the mean and over-dispersion parameters and the prior on  $\sigma^2$  is always the same.

6. I will choose to read *Bayesian Geostatistical Design* by Diggle and Lophaven.

For the project, what will we need to be included in the summary of our work?

## R Code

```
1. #draw randomly from stand norm 15 observations

set.seed(53205)
y15 <- rnorm(15,0,1)

#simulate the joint posterior
num.sims <- 10000
Phi <- matrix(0,nrow=num.sims,ncol=2)
Phi[1,1] <- 0 # initialize theta
Phi[1,2] <- 1 # initialize (1/sigma^2)
mu.0 <- 0
tausq.0 <- 100
nu.0 <- 1
sigmaq.0 <- 100
num.obs <- length(y15)

mean.y <- mean(y15)

for (i in 2:num.sims){
  # sample theta from full conditional
  mu.n <- (mu.0 / tausq.0 + num.obs * mean.y * Phi[(i-1),2]) / (1 / tausq.0 + num.obs * Phi[(i-1),2] )
  tausq.n <- 1 / (1/tausq.0 + num.obs * Phi[(i-1),2])
  Phi[i,1] <- rnorm(1,mu.n,sqrt(tausq.n))

  # sample (1/sigma^2) from full conditional
  nu.n <- nu.0 + num.obs
  sigmaq.n.theta <- 1/nu.n*(nu.0*sigmaq.0 + sum((y15 - Phi[i,1])^2))
  Phi[i,2] <- rgamma(1,nu.n/2,nu.n*sigmaq.n.theta/2)
}

## first figure out what the "truth" is for my simulations
## although not really necessary, truth = 0,1
y15.true.mean <- mean(y15)
y15.true.var <- var(y15)

par(mfrow = c(1,1))
hist(Phi[,1], freq = FALSE, breaks = "fd", xlab = expression(theta), ylab = "Proportion", main = "Marginal Posterior, \n UNIF Prior", ylim=c(0,1))

## generating from bivariate uniform
## since theta and sigma^2 are orthogonal, generate two observations randomly from unif(0,1)

## rerun for several different values of mu.o and tau.o

## 1 close mean, far var

#simulate the joint posterior
num.sims <- 10000
Phi <- matrix(0,nrow=num.sims,ncol=2)
Phi[1,1] <- 0 # initialize theta
Phi[1,2] <- 1 # initialize (1/sigma^2)
mu.0 <- -0.01
tausq.0 <- 100
nu.0 <- 1
sigmaq.0 <- 100
num.obs <- length(y15)

mean.y <- mean(y15)

for (i in 2:num.sims){
  # sample theta from full conditional
  mu.n <- (mu.0 / tausq.0 + num.obs * mean.y * Phi[(i-1),2]) / (1 / tausq.0 + num.obs * Phi[(i-1),2] )
  tausq.n <- 1 / (1/tausq.0 + num.obs * Phi[(i-1),2])
  Phi[i,1] <- rnorm(1,mu.n,sqrt(tausq.n))

  # sample (1/sigma^2) from full conditional
  nu.n <- nu.0 + num.obs
```

```

sigmasq.n.theta <- 1/nu.n*(nu.0*sigmasq.0 + sum((y15 - Phi[i,1])^2))
Phi[i,2] <- rgamma(1,nu.n/2,nu.n*sigmasq.n.theta/2)
}

par(mfrow=c(1,1))
#plot prior, posterior, and likelihood
post.x <- function(x){2*pi*((-length(y15))/2)*exp(-(sum(y15)^2) + 2*sum(y15)*x - x^2)/2)}

curve(dnorm(x, mean = mean(Phi[,1]), sd = sqrt(var(Phi[,1]))), ylim = c(0,1), xlim = c(-3,3),lwd = 2, xlab = expression(theta), ylab = "Proportion", lty = 1)
curve(dnorm(x, mean=0, sd=10),
      col="darkblue", lwd=2, add=TRUE, yaxt="n", lty = 2)
curve(post.x, col="darkgreen", lwd=2, add=TRUE, yaxt="n", lty = 3)
legend(-2,1, legend = c("Posterior", "Prior N(0.01,100)", expression(paste("Likelihood ", sigma^2, "=1"))),lty=c(1,2,3),lwd = c(2,2,2),col=c("black",
"darkblue", "darkgreen"))

par(mfrow=c(2,2))
hist(Phi[,1], xlab = expression(theta), freq = FALSE, breaks = "fd", main = expression(paste(tau[o]^2, " = 100", mu[o], " = -0.5")), ylim = c(0,1))
points(x = mu.0, col = "red", pch = 10)
points(x = 0, col = "yellow", pch = 7)

hist(Phi[,1], xlab = expression(theta), freq = FALSE, breaks = "fd", main = expression(paste(tau[o]^2, " = 100", mu[o], " = -0.01")), ylim = c(0,1))
points(x = mu.0, col = "red", pch = 10)
points(x = 0, col = "yellow", pch = 7)

#####
# 2) far from mean, close to var
#####

Phi <- matrix(0,nrow=num.sims,ncol=2)
Phi[1,1] <- 0 # initialize theta
Phi[1,2] <- 1 # initialize (1/sigmasq)
mu.0 <- 50
tausq.0 <- 0.99
nu.0 <- 1
sigmasq.0 <- 100
num.obs <- length(y15)

mean.y <- mean(y15)

for (i in 2:num.sims){
  # sample theta from full conditional
  mu.n <- (mu.0 / tausq.0 + num.obs * mean.y * Phi[(i-1),2]) / (1 / tausq.0 + num.obs * Phi[(i-1),2] )
  tausq.n <- 1 / (1/tausq.0 + num.obs * Phi[(i-1),2])
  Phi[i,1] <- rnorm(1,mu.n,sqrt(tausq.n))

  # sample (1/sigma.sq) from full conditional
  nu.n <- nu.0 + num.obs
  sigmasq.n.theta <- 1/nu.n*(nu.0*sigmasq.0 + sum((y15 - Phi[i,1])^2))
  Phi[i,2] <- rgamma(1,nu.n/2,nu.n*sigmasq.n.theta/2)
}

hist(Phi[,1], xlab = expression(theta), freq = FALSE, breaks = "fd", main = expression(paste(tau[o]^2, " = 0.99", mu[o], " = 50")), ylim = c(0,1))
points(x = mu.0, col = "red", pch = 10)
points(x = 0, col = "yellow", pch = 7)

hist(Phi[,1], xlab = expression(theta), freq = FALSE, breaks = "fd", main = expression(paste(tau[o]^2, " = 0.99", mu[o], " = 50")), ylim = c(0,1), xlim = c(-4,4))
points(x = mu.0, col = "red", pch = 10)
points(x = 0, col = "yellow", pch = 7)

#####
#3) close to mean, close to var
#####
Phi <- matrix(0,nrow=num.sims,ncol=2)
Phi[1,1] <- 0 # initialize theta
Phi[1,2] <- 1 # initialize (1/sigmasq)
mu.0 <- -0.01
tausq.0 <- 0.99
nu.0 <- 1
sigmasq.0 <- 100
num.obs <- length(y15)

mean.y <- mean(y15)

for (i in 2:num.sims){
  # sample theta from full conditional
  mu.n <- (mu.0 / tausq.0 + num.obs * mean.y * Phi[(i-1),2]) / (1 / tausq.0 + num.obs * Phi[(i-1),2] )
  tausq.n <- 1 / (1/tausq.0 + num.obs * Phi[(i-1),2])
  Phi[i,1] <- rnorm(1,mu.n,sqrt(tausq.n))

  # sample (1/sigma.sq) from full conditional
  nu.n <- nu.0 + num.obs
  sigmasq.n.theta <- 1/nu.n*(nu.0*sigmasq.0 + sum((y15 - Phi[i,1])^2))
  Phi[i,2] <- rgamma(1,nu.n/2,nu.n*sigmasq.n.theta/2)
}

hist(Phi[,1], xlab = expression(theta), freq = FALSE, breaks = "fd", main = expression(paste(tau[o]^2, " = 0.99", mu[o], " = -0.01")), ylim = c(0,1), xlim = c(-5,5))
points(x = mu.0, col = "red", pch = 10)
points(x = 0, col = "yellow", pch = 7)

#####
# 4 far from mean, far from var
#####

Phi <- matrix(0,nrow=num.sims,ncol=2)
Phi[1,1] <- 0 # initialize theta
Phi[1,2] <- 1 # initialize (1/sigmasq)
mu.0 <- 50

```

```

tausq.0 <- 100
nu.0 <- 1
sigmasq.0 <- 100
num.obs <- length(y15)

mean.y <- mean(y15)

for (i in 2:num.sims){
  # sample theta from full conditional
  mu.n <- (mu.0 / tausq.0 + num.obs * mean.y * Phi[(i-1),2]) / (1 / tausq.0 + num.obs * Phi[(i-1),2])
  tausq.n <- 1 / (1/tausq.0 + num.obs * Phi[(i-1),2])
  Phi[i,1] <- rnorm(1,mu.n,sqrt(tausq.n))

  # sample (1/sigma.sq) from full conditional
  nu.n <- nu.0 + num.obs
  sigmasq.n.theta <- 1/nu.n*(nu.0*sigmasq.0 + sum((y15 - Phi[i,1])^2))
  Phi[i,2] <- rgamma(1,nu.n/2,nu.n*sigmasq.n.theta/2)
}

hist(Phi[,1], xlab = expression(theta), freq = FALSE, breaks = "fd", main =
  expression(paste(tau[o] ~2, " = 100 ", mu[o], " = 50")), ylim = c(0,101),
  xlim = c(-4,51))
points(x = mu.0, col = "red", pch = 10)
points(x = y15.true.mean, col = "yellow", pch = 7)

hist(Phi[,1], xlab = expression(theta), freq = FALSE, breaks = "fd", main =
  expression(paste(tau[o] ~2, " = 100 ", mu[o], " = 50")), ylim = c(0,1),
  xlim = c(-5,5))
points(x = mu.0, col = "red", pch = 10)
points(x = 0, col = "yellow", pch = 7)

#####
# 5) fairly close to mean, fairly close to var
#####

Phi <- matrix(0,nrow=num.sims,ncol=2)
Phi[1,1] <- 0 # initialize theta
Phi[1,2] <- 1 # initialize (1/sigmasq)
mu.0 <- -1
tausq.0 <- 4
nu.0 <- 1
sigmasq.0 <- 100
num.obs <- length(y15)

mean.y <- mean(y15)

for (i in 2:num.sims){
  # sample theta from full conditional
  mu.n <- (mu.0 / tausq.0 + num.obs * mean.y * Phi[(i-1),2]) / (1 / tausq.0 +
    num.obs * Phi[(i-1),2])
  tausq.n <- 1 / (1/tausq.0 + num.obs * Phi[(i-1),2])
  Phi[i,1] <- rnorm(1,mu.n,sqrt(tausq.n))

  # sample (1/sigma.sq) from full conditional
  nu.n <- nu.0 + num.obs
  sigmasq.n.theta <- 1/nu.n*(nu.0*sigmasq.0 + sum((y15 - Phi[i,1])^2))
  Phi[i,2] <- rgamma(1,nu.n/2,nu.n*sigmasq.n.theta/2)
}

hist(Phi[,1], xlab = expression(theta), freq = FALSE, breaks = "fd", main =
  expression(paste(tau[o] ~2, " = 4 ", mu[o], " = -1")), ylim = c(0,1))
points(x = mu.0, col = "red", pch = 10)
points(x = 0, col = "yellow", pch = 7)

#####
# 6) fairly close to mean, somewhat far from var
#####

Phi <- matrix(0,nrow=num.sims,ncol=2)
Phi[1,1] <- 0 # initialize theta
Phi[1,2] <- 1 # initialize (1/sigmasq)
mu.0 <- -1
tausq.0 <- 16
nu.0 <- 1
sigmasq.0 <- 100
num.obs <- length(y15)

mean.y <- mean(y15)

for (i in 2:num.sims){
  # sample theta from full conditional
  mu.n <- (mu.0 / tausq.0 + num.obs * mean.y * Phi[(i-1),2]) / (1 / tausq.0 +
    num.obs * Phi[(i-1),2])
  tausq.n <- 1 / (1/tausq.0 + num.obs * Phi[(i-1),2])
  Phi[i,1] <- rnorm(1,mu.n,sqrt(tausq.n))

  # sample (1/sigma.sq) from full conditional
  nu.n <- nu.0 + num.obs
  sigmasq.n.theta <- 1/nu.n*(nu.0*sigmasq.0 + sum((y15 - Phi[i,1])^2))
  Phi[i,2] <- rgamma(1,nu.n/2,nu.n*sigmasq.n.theta/2)
}

hist(Phi[,1], xlab = expression(theta), freq = FALSE, breaks = "fd", main =
  expression(paste(tau[o] ~2, " = 16 ", mu[o], " = -1")), ylim = c(0,1))
points(x = mu.0, col = "red", pch = 10)
points(x = 0, col = "yellow", pch = 7)

```

```
#####
# 7) somewhat far from mean, near var
#####

Phi <- matrix(0,nrow=num.sims,ncol=2)
Phi[1,1] <- 0 # initialize theta
Phi[1,2] <- 1 # initialize (1/sigma^2)
mu.0 <- 5
tausq.0 <- 4
nu.0 <- 1
sigmaq.0 <- 100
num.obs <- length(y15)

mean.y <- mean(y15)

for (i in 2:num.sims){
  # sample theta from full conditional
  mu.n <- (mu.0 / tausq.0 + num.obs * mean.y * Phi[(i-1),2]) / (1 / tausq.0 +
    num.obs * Phi[(i-1),2])
  tausq.n <- 1 / (1/tausq.0 + num.obs * Phi[(i-1),2])
  Phi[i,1] <- rnorm(1,mu.n,sqrt(tausq.n))

  # sample (1/sigma^2) from full conditional
  nu.n <- nu.0 + num.obs
  sigmaq.n.theta <- 1/nu.n*(nu.0*sigmaq.0 + sum((y15 - Phi[i,1])^2))
  Phi[i,2] <- rgamma(1,nu.n/2,nu.n*sigmaq.n.theta/2)
}

hist(Phi[,1], xlab = expression(theta), freq = FALSE, breaks = "fd", main =
  expression(paste(tau[o] ~2, " = 4 ", mu[o], " = -5")), ylim = c(0,1), xlim = c(-5,5))
points(x = mu.0, col = "red", pch = 10)
points(x = 0, col = "yellow", pch = 7)

#####
# 8) need to look at where mean breaks down
#####
Phi <- matrix(0,nrow=num.sims,ncol=2)
Phi[1,1] <- 0 # initialize theta
Phi[1,2] <- 1 # initialize (1/sigma^2)
mu.0 <- 10
tausq.0 <- 0.99
nu.0 <- 1
sigmaq.0 <- 100
num.obs <- length(y15)

mean.y <- mean(y15)

for (i in 2:num.sims){
  # sample theta from full conditional
  mu.n <- (mu.0 / tausq.0 + num.obs * mean.y * Phi[(i-1),2]) / (1 / tausq.0 +
    num.obs * Phi[(i-1),2])
  tausq.n <- 1 / (1/tausq.0 + num.obs * Phi[(i-1),2])
  Phi[i,1] <- rnorm(1,mu.n,sqrt(tausq.n))

  # sample (1/sigma^2) from full conditional
  nu.n <- nu.0 + num.obs
  sigmaq.n.theta <- 1/nu.n*(nu.0*sigmaq.0 + sum((y15 - Phi[i,1])^2))
  Phi[i,2] <- rgamma(1,nu.n/2,nu.n*sigmaq.n.theta/2)
}

hist(Phi[,1], xlab = expression(theta), freq = FALSE, breaks = "fd", main =
  expression(paste(tau[o] ~2, " = 0.99 ", mu[o], " = 10")), ylim = c(0,1))
points(x = mu.0, col = "red", pch = 10)
points(x = 0, col = "yellow", pch = 7)

par(mfrow = c(1,1))

4. set.seed(53205)
#####
###simulate multivariate normal data
#####
n=100
p=3
true.mu <- rep(0,p)

# create correlation matrix
rho.12 <- 0
rho.23 <- 0
rho.13 <- 0
R <- matrix(c(1,rho.12,rho.13,rho.12,1,
  rho.23,rho.13,rho.23,1),p,p) #corr = diag(1) -> cov = diag(sigma2)

y.vec <- rmnorm(n,true.mu,R) #used correlation matrix here in place of cov?
#colMeans(y.vec)
#cor(y.vec)

#####
# Run Gibbs Sampler
#####
##initialize
num.sims <- 10000
mu <- matrix(0,nrow=num.sims,ncol=p)
Sigma <- array(0,dim=c(p,p,num.sims))
Sigma[,1] <- diag(p) # initialize first to be identity
```

```

# note when p is large it makes sense to only save unique
# elements in Sigma

## priors
# theta
mu.0 <- rep(0,p) # all mu.0[i] = 0
Lambda.0 <- matrix(0,p,p)
diag(Lambda.0) <- 100 # choose uninformative prior covariance matrix
Lambda.0 <- Lambda.0 + 10000
Lambda.0.inv <- solve(Lambda.0)
# visualize a draws from prior
#rmnorm(5,mu.0,Lambda.0)

# Sigma
nu.0 <- p + 2
nu.n <- nu.0 + n
S.0 <- matrix(0,p,p)
diag(S.0) <- 10 # choose fairly uninformative prior on RSS matrix
# visualize a few draws from prior
#riwish(nu.0,solve(S.0)) # note nu.0 must be atleast p
#riwish(nu.0,solve(S.0))
#riwish(nu.0,solve(S.0))

for (i in 2:num.sims){
  #sample mu
  Sig.inv <- solve(Sigma[,,(i-1)])
  Lambda.n <- solve(n* Sig.inv + Lambda.0.inv )
  mu.n <- Lambda.n %*% (Sig.inv %*% colSums(y.vec) + Lambda.0.inv %*% mu.0)
  mu[i,] <- rmnorm(1,mu.n,Lambda.n) #why did he use correlation
  #matriz in place of covariance
  #matriz in drawing the random observations?

  # Sample Sigma
  mu.matrix <- matrix(mu[i,],n,p,byrow=T)
  S.theta <- t(y.vec - mu.matrix) %*% (y.vec - mu.matrix)
  Sigma[,i] <- riwish(nu.n,(S.theta + S.0)) # note the parameterization
}

# look at theta
#colMeans(mu)
#colMeans(y.vec)
par(mfrow=c(1,3))
for (i in 1:p){
  plot(mu[,i],type='l',ylab=expression(mu[i]))
  abline(h=true.mu[i],col='red',lwd=2)
}

# look at Sigma
#var(y.vec)
#matrix(c(mean(Sigma[1,]),mean(Sigma[1,2]),mean(Sigma[1,3]),
#          mean(Sigma[2,1]),mean(Sigma[2,2]),mean(Sigma[2,3]),
#          mean(Sigma[3,1]),mean(Sigma[3,2]),mean(Sigma[3,3])),p,p)
par(mfcol=c(3,3))
for(i in 1:p){
  for (j in 1:p){
    plot(Sigma[i,j],type='l',ylab=expression(sigma[i,j]))
    abline(h=R[i,j],col='firebrick4',lwd=3)
  }
}

```

5. # set a = 0.5 and b = 2 to make  $p(y/\sigma) = \text{cauchy}(\theta, \gamma)$

```

a <- 0.5
b <- 0.5

set.seed(53205)
precision <- replicate(1000, rgamma(1, shape = a, scale = b))

g <- 2
theta <- 0
ystar <- sapply(1/precision, function(t){rnorm(n = 1, mean = theta, sd = sqrt(g*t))})
# for some reason I am getting the precision to be values the var should be
par(mfrow=c(1,1))
hist(ystar, freq = FALSE, breaks = "fd", xlim = c(-50,50), main =
  expression(paste("histogram of ", y~star, "zoomed in")), xlab = expression(paste(y~star)))

hist(ystar, freq = FALSE, breaks = "fd", main = expression(paste("histogram of ", y~star, "zoomed out")), xlab = expression(paste(y~star)),
  xlim = c(min(ystar)-10, max(ystar) +10))
abline(v = max(ystar), col = "purple")
abline(v = min(ystar), col = "red")
text(x = max(ystar)-1000, y = 0.1, labels = round(max(ystar),2))
text(x = min(ystar)+1000, y = 0.1, labels = round(min(ystar),2))

par(mfrow=c(1,1))
plot(ystar,type='l', ylim = c(min(ystar)-10, max(ystar) +10),ylab=
  expression(y~star))

plot(1:10,ystar[1:10],type='l', ylim = c(min(ystar)-10, max(ystar) +10),ylab=expression(y~star), main = "Zoomed In First 10 Random Draws")

```