Bayesian Statistical Methods

Homework #1

January 26, 2022

Akshay Umashankar* **Baris Bilir** †

*akshayumashankar38@gmail.com, au3692.
†baris.bilir@austin.utexas.edu, bb39649.

Problem 1.

Observe that

$$
\begin{aligned}
[\sigma^2|\mathbf{y},\mu] &\propto \prod_{i=1}^{n}[y_i|\sigma,\mu][\sigma,\mu] \\
&= \prod_{i=1}^{n}[y_i|\sigma,\mu][\sigma][\mu] \\
&\propto \prod_{i=1}^{n}[y_i|\sigma,\mu][\sigma] \\
&\propto \left(\sigma^2\right)^{-n/2}\exp\Big\{-\frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\Big\}\left(\sigma^2\right)^{-(v/2+1)}\exp\big\{-1/2\sigma^2\big\},
\end{aligned}
$$

where the first equality holds since μ and σ are independent, second proportional follows from the fact that σ^2 do not appear in $[\mu]$. By arranging the terms above, we get

$$
[\sigma^2|\mathbf{y},\mu] = \left(\sigma^2\right)^{-\left((v+n)/2+1\right)}\exp\Big\{-\frac{0.5+\sum_{i=1}^{n}(y_i-\mu)^2}{\sigma^2}\Big\}. \tag{1}
$$

Recall, the pdf of Inverse Gamma distribution with shape parameter r and rate parameter q is

$$
[\theta|q,r] = \frac{1}{r^q\Gamma(q)}\theta^{-(q+1)}\exp\Big\{-\frac{1}{r\theta}\Big\}.
$$

Notice that in (1), the right hand side is similar to the pdf of inverse gamma distribution. Summing up, the full conditional posterior for σ^2 is

$$
\boxed{[\sigma^2|\mathbf{y},\mu] \sim \mathrm{IG}(q,r) \quad \text{with} \quad q=(v+n)/2 \quad \text{and} \quad r = \text{‘}\left(0.5+\sum_{i=1}^{n}(y_i-\mu)^2\right)^{-1}.} \tag{2}
$$

It is noteworthy that Normal-Normal-Inverseχ^2 is a special case of Normal-Normal-Inverse Gamma model since Inverseχ^2 is a special case of Inverse Gamma distribution.

□

Problem 2.

To construct an MCMC algorithm with Gibbs updates for the Normal-Normal-Inverseχ^2 model, we need full conditional posteriors for μ and σ^2. The full conditional posterior for σ^2 is derived in Problem 1. With similar arguments as in Problem 1, the full conditional posterior for μ can be derived as follows:

$$
[\mu|\mathbf{y},\sigma^2] \propto \prod_{i=1}^{n}[y_i|\sigma,\mu][\mu]
$$

$$
\begin{aligned}
&= \Big(\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big\{ -\frac{1}{2\sigma^2}(y_i-\mu)^2 \Big\} \Big) \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\Big\{ -\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2 \Big\} \\
&\propto \exp\Big\{ -\frac{1}{2\sigma^2}\Big(\sum_{i}^{n}(y_i-\mu)^2 \Big) \Big\} \exp\Big\{ -\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2 \Big\} \\
&= \exp\Big\{ -\frac{1}{2}\Big(\frac{\sum_i^n (y_i^2) - \sum_{i=1}^n (2y_i\mu) + n\mu^2}{\sigma^2} + \frac{\mu^2 - 2\mu\mu_0 + \mu_0^2}{\sigma_0^2} \Big) \Big\} \\
&\propto \exp\Big\{ -\frac{1}{2}\Big(\frac{-2\mu\sum_{i=1}^n y_i + n\mu^2}{\sigma^2} + \frac{\mu^2 - 2\mu\mu_0}{\sigma_0^2} \Big) \Big\} \\
&= \exp\Big\{ -\frac{1}{2}\Big(\frac{-2\mu\sigma_0^2\sum_{i=1}^n y_i + n\sigma_0^2\mu^2 + \mu^2\sigma^2 - 2\mu\mu_0\sigma^2}{\sigma^2\sigma_0^2} \Big) \Big\} \\
&= \exp\Big\{ -\frac{1}{2}\Big(\frac{(n\sigma_0^2+\sigma^2)\mu^2 - 2(\sigma_0^2\sum_{i=1}^n y_i + \mu_0\sigma^2)\mu}{\sigma^2\sigma_0^2} \Big) \Big\} \\
&= \exp\Big\{ -\frac{1}{2}\Big(\frac{\mu^2 - \frac{2(\sigma_0^2\sum_{i=1}^n y_i + \mu_0\sigma^2)}{n\sigma_0^2+\sigma^2}\mu}{\frac{\sigma^2\sigma_0^2}{(n\sigma_0^2+\sigma^2)}} \Big) \Big\} \\
&\propto \exp\Big\{ -\frac{\Big(\mu - \frac{\sigma_0^2\sum_{i=1}^n y_i + \mu_0\sigma^2}{n\sigma_0^2+\sigma^2}\Big)^2}{2\frac{\sigma^2\sigma_0^2}{n\sigma_0^2+\sigma^2}} \Big\},
\end{aligned}
$$

where we complete the square to get the last proportional. Then,

$$
\boxed{[\mu|\mathbf{y},\sigma^2] \sim \mathcal{N}(a,b) \quad \text{with} \quad a = \frac{\sigma_0^2\sum_{i=1}^n y_i + \mu_0\sigma^2}{n\sigma_0^2+\sigma^2} \quad \text{and} \quad b = \frac{\sigma^2\sigma_0^2}{n\sigma_0^2+\sigma^2}.} \tag{3}
$$

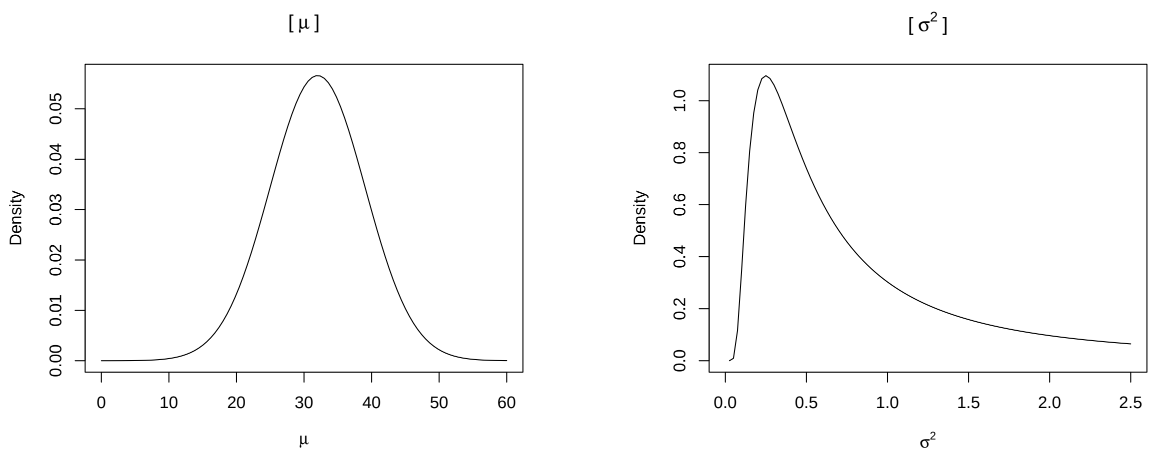

Figure 1: Prior distributions for μ (left) and σ^2 (right). $[\mu] \sim \mathcal{N}(\mu_0, \sigma_0^2)$, $[\sigma^2] \sim \text{Inverse}\chi^2(v)$, where μ_0, σ_0^2 and v are hyperparameters. In our implementation, we set $\mu_0 = 32$, $\sigma_0^2 = 49$ and $v = 2.02$.

Now, we are ready to present the Gibbs Sampler for the Normal-Normal-Inverseχ^2 model using the full conditional posteriors in (2) and (3). Note that in the following μ_0, σ_0^2 and v are hyperparameters that should be specified by the user. The computation of the posterior distribution isn't effected by the selection of hyperparameters if the MCMC algorithm run long enough. In our implementation, we set μ_0 to $\overline{\mu}$, σ_0^2 to s^2 and v to $\frac{1}{s^2} + 2$ (since mean of Inverse$\chi^2(v)$ is $1/(v-2)$ for $v > 2$), where $\overline{\mu}$ is the sample mean and s^2 is the sample variance. Prior distributions are depicted in Figure 1. The MCMC algorithm that we implemented is as follows:

1. Set hyperparameters $\mu_0 = 32$, $\sigma_0^2 = 49$ and $v = 2.2$.

2. Set ${\sigma^2}^{(0)} = 1$.

3. Set $k = 1$.

4. Sample μ from its full conditional distribution that depends on the latest value for σ^2.

$$\mu^{(k)} \sim \mathcal{N}(a, b),$$

   where

$$a = \frac{\sigma_0^2 \sum_{i=1}^n y_i + \mu_0 \sigma^2}{n\sigma_0^2 + \sigma^2} \quad \text{and} \quad b = \frac{\sigma^2 \sigma_0^2}{n\sigma_0^2 + \sigma^2}.$$

5. Sample σ^2 from its full conditional distribution that depends on the latest value for μ.

$${\sigma^2}^{(k)} \sim \mathrm{IG}(q, r),$$

   where

$$q = (v+n)/2 \quad \text{and} \quad r = `\left(0.5 + \sum_{i=1}^n (y_i - \mu)^2\right)^{-1}.$$

6. Set $k = k + 1$. If $k = 1{,}000{,}000$ stop, otherwise go to step 4.

It is noteworthy that there is no problem with using the data at hand while setting hyperparameter values. However, data should not be used while specifying prior or the proposal prior in the MCMC algorithm. Finally, we use $k = 1{,}000{,}000$ as a threshold for termination of our algorithm. If this number is sufficient enough for another model where Gibbs sampler is used, the user can increase the threshold for the termination condition.

In the Normal-Normal-Inverseχ^2 model, we have two parameters μ and σ^2 that we want to do inference on. The prior for mu is a Normal distribution with mean μ_0 and variance σ_0^2 and the prior for σ^2 is Inverseχ^2

distribution with v degrees of freedom. As indicated before, μ_0, σ^2 and v are hyperparameters of the model. By running the MCMC algorithm presented above, we can sample from the marginal posterior distributions of μ and σ^2, that is we can sample from $[\mu|\mathbf{y}, \sigma^2]$ and $[\sigma^2|\mathbf{y}, \mu]$.

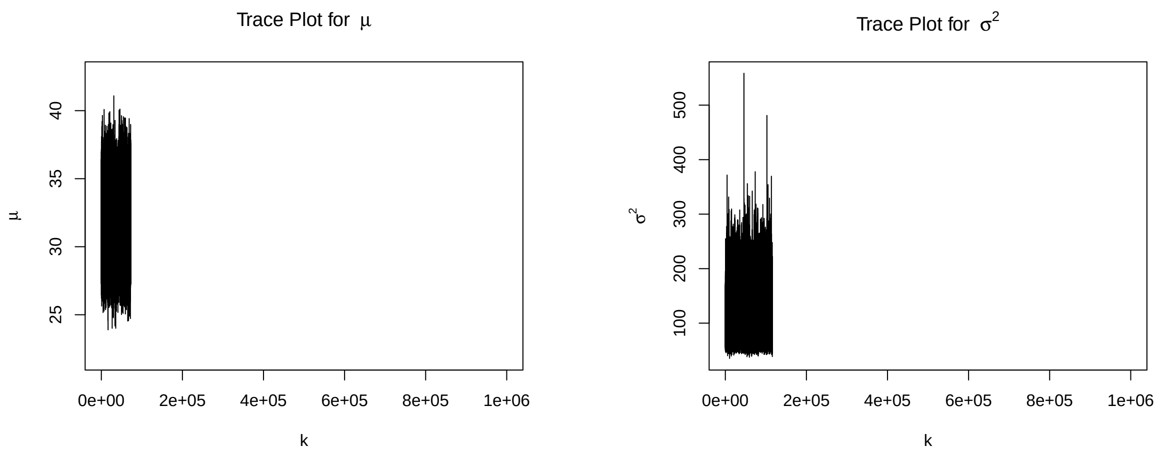

Figure 2: Prior distributions for μ (left) and σ^2 (right). $[\mu] \sim \mathcal{N}(\mu_0, \sigma_0^2)$, $[\sigma^2] \sim \text{Inverse}\chi^2(v)$, where μ_0, σ_0^2 and v are hyperparameters. In our implementation, we set $\mu_0 = 32$, $\sigma_0^2 = 49$ and $v = 2.02$.

Next, we present the computational results we get by running the MCMC algorithm. The trace plots for full conditional posteriors of μ and σ^2 are given in Figure 2. In our implementation, we didn't need burn-in (or warm-up period). The reason is that, as can be seen in the trace plots, the chains do not need time to reach stationary distribution. Furthermore, we don't see any flat regions in the trace plots, so we can conclude that there are no issues with mixing.

As a final remark, we check autocorrelation for mu and σ^2 in case we need thinning. As can be seen in Figure 3, we don't see any autocorrelation in the samples that we sampled from the posterior distribution. So, we didn't apply thinning to our posterior samples.

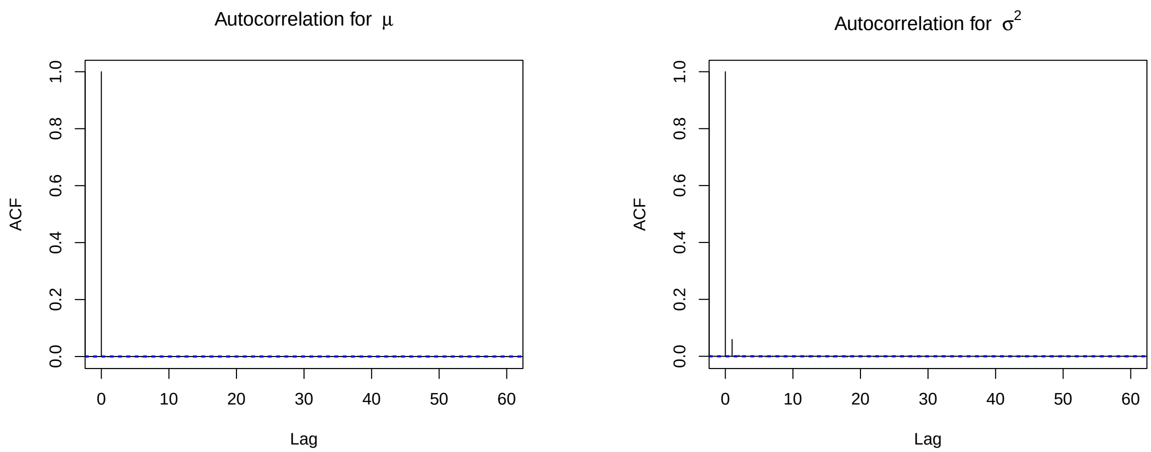

Figure 3: Autocorrelation plots for posterior samples of μ and σ^2.

□

Problem 3. The posterior distributions are illustrated in Figure 4. Posterior means, 95% equal-tailed credible interval for both μ and σ^2 are given in Table 3. Note that $95\% CI$ is the interval with end points corresponds to 2.5% and 97.5% quantiles of posterior samples.

	μ	σ^2
Mean	32.00827	102.4641
95% CI	[28.43779, 35.57669]	[60.1834, 174.3460]

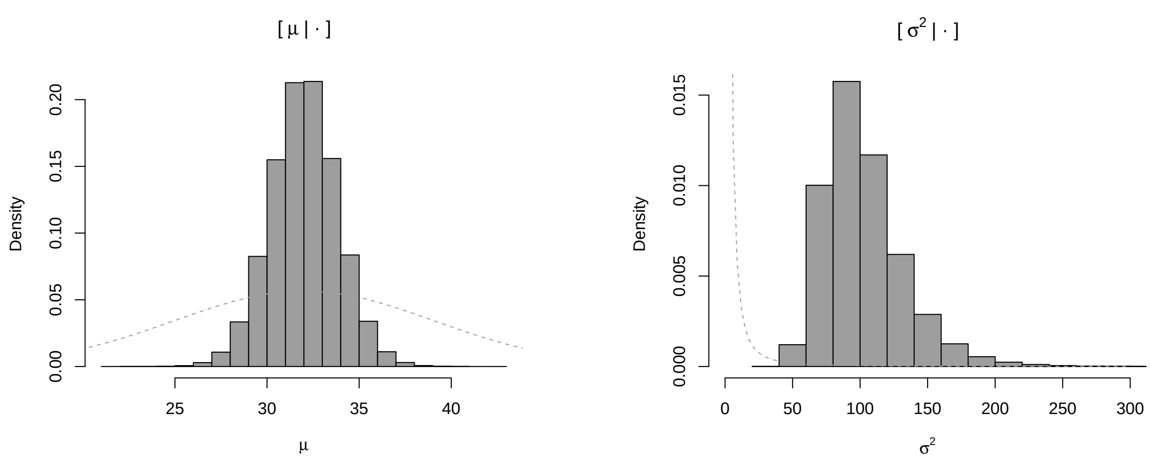

Figure 4: Posterior distributions for μ (left) and σ^2 (right). Dashed lines are the densities of the prior distributions.

□

Problem 4. We can compute $\mathbb{P}\{\mu > 30 \,|\, y\}$ using the full conditional posterior for μ that we computed in Problem 2. This probability can be estimated as the number of the samples of μ that are greater than 30 divided by the total number of samples. That is

$$\mathbb{P}\{\mu > 30 \,|\, y\} = \int\int \mathbb{1}_{\{\mu>30\}}[\mu, \sigma^2|y] d\mu d\sigma^2 \approx 0.86957.$$

□

Problem 5. The Normal-Normal-Invereχ^2 model assumes that mean follows a normal distribution and variance follows an Inverseχ^2 distribution for the BMI among young adults. In our opinion, believing that mean (prior) following a Normal distribution is a valid argument considering the fact that as sample size increases the sample mean converges to a normal distribution as a result of central limit theorem. Inverseχ^2 distribution is skewed to the right and for both the mean and the Inverseχ^2 trace plots do not have any issues. In our opinion this makes Inverseχ^2 a good assumption for prior distribution of σ^2. □

Appendix

The Appendix includes three R files:

- main.R: Running this script generates the results.
- mcmc_nnig.R: Implementation of MCMC algorithm. (Gibbs sampler) for Normal-Normal-Inverseχ^2 model.
- invchisquare.R: Density function for Inverseχ^2 distribution.

To replicate the results, all you need to do is to run main.R. mcmc_nnig.R and invchisquare.R are sourced in main.R.

main.R

```r
# set working directory
setwd("/Users/barisbilir/Desktop/Bayesian/hw1")

# source MCMC for Gibbs sampling
source("/Users/barisbilir/Desktop/Bayesian/hw1/mcmc_nnig.R")

# source invchisquare function
source("/Users/barisbilir/Desktop/Bayesian/hw1/invchisquare.R")

# read medical data
df.med_data <- read.csv("MedicalData.csv")

# subset young adults data
df.young_adult <- df.med_data[df.med_data$Age == "young adult",]

# get y vector (BMI of young adults)
y = df.young_adult$BMI

# Hyper parameters for mean
mu0 = mean(y)
sigma20 = var(y)

```

```r
# Hyper parameter for v
v = 1/var(y) + 2

# MCMC (Gibbs) for Normal-Normal-InverseChiSquare
mcmc.out <- mcmc.nnig(y, mu0, sigma20, v, sigma2.start=25, n.mcmc=1000000)

# trace plot
plot(mcmc.out$mu.save, type='l', xlab='k', ylab=bquote(mu),
     main =bquote("Trace Plot for " ~ mu ))

plot(mcmc.out$sigma2.save, type='l', xlab='k', ylab=bquote(sigma^2),
     main =bquote("Trace Plot for " ~ sigma^2 ))

# prior distributions
# density of the prior distribution of mu
curve(dnorm(x,mu0,sqrt(sigma20)),from=0, to=60, type='l',
      xlab=bquote(mu), ylab='Density', main =bquote("[" ~ mu ~ "]"))
curve(dinvchisquare(x,v), from=0, to=2.5, type='l', xlab=bquote(sigma^2),
      ylab='Density', main =bquote("[" ~ sigma^2 ~ "]"))

# posterior distributions
# marginal posterior distribution of mu
hist(mcmc.out$mu.save, col=8, probability=TRUE, xlab = bquote(mu),
     main =bquote("[" ~ mu ~ "| \U00B7 ]"))

# density of the prior distribution of mu
curve(dnorm(x,mu0,sqrt(sigma20)), from=15, to=50, col=8, lty=2, add=TRUE)
# marginal posterior distribution of sigma2
hist(mcmc.out$sigma2.save,xlim=c(0,300), col=8, probability=TRUE,
     xlab = bquote(sigma^2),
     main =bquote("[" ~ sigma^2 ~ "| \U00B7 ]"))
# density of the prior distribution of sigma2
curve(dinvchisquare(x,v), from=0, to=300, col=8, lty=2, add=TRUE)

# autocorrelation plots
```

```
acf(mcmc.out$mu.save, main=bquote("Autocorrelation for " ~ mu))
acf(mcmc.out$sigma2.save, main=bquote("Autocorrelation for " ~ sigma^2))

# 95% CI
CI_mean = quantile(mcmc.out$mu.save, probs = c(0.025, 0.975))
CI_var = quantile(mcmc.out$sigma2.save, probs = c(0.025, 0.975))

# calculate P(mu>30|y)
mu.posterior = mcmc.out$mu.save
prob = length(mu.posterior[mu.posterior>30])/length(mu.posterior)
print(paste('Probability of mean greater that 30 = ', prob))
```

mcmc_nnig.R

```
mcmc.nnig <- function(y, mu0, sigma20, v, sigma2.start, n.mcmc){

  # number of data points
  n <- length(y)
  # vector that holds mu updates
  mu.save <- rep(0, n.mcmc)
  # vector that holds sigma2 updates
  sigma2.save <- rep(0, n.mcmc)

  # Gibbs sampler

  # initialize mu
  sigma2.k <- sigma2.start
  # Gibbs updates for k=1 to the upper limit of MCMC iterations
  for (k in 1:n.mcmc){
    # parameters for the full conditional distribution of mu
    a_num <- sigma20*sum(y) + mu0*sigma2.k
    a_den <- n*sigma20 + sigma2.k
    a <- a_num/a_den

    b_num <- sigma2.k*sigma20
```

```
    b_den <- n*sigma20 + sigma2.k
    b <- b_num/b_den

    # sample mu from full conditional posterior of mu
    mu.k <- rnorm(1, mean=a, sd=sqrt(b))

    # save mu update
    mu.save[k] = mu.k

    # parameters for the full conditional distribution of sigma2
    q = (v + n)/2
    r_den <- 0.5 + sum((y-mu.k)^2)
    r <- 1/r_den

    # sample sigma2 from the full conditional distribution of sigma2
    sigma2.k <- 1/rgamma(1, shape=q ,scale = r)

    #save sigma update
    sigma2.save[k] <- sigma2.k
  }

  list(mu.save=mu.save, sigma2.save=sigma2.save, n.mcmc=n.mcmc)
}
```

invchisquare.R

```
dinvchisquare <- function(x,v){
 (2^(-v/2)/gamma(v/2))*x^(-v/2 - 1)*exp(-1/(2*x))
}
```