Cesium levels in Fish

Greg Johnson

Suppose the cesium levels of 20 fish are normally distributed, with a non-conjugate Normal-Inverse-Gamma prior. We are going to build a Bayesian model to produce inferences about the mean and variance of cesium levels.

$$Y_1, ..., Y_{20} \sim N(\mu, \sigma^2)$$

$$\mu \sim N(m_\mu, s_\mu^2)$$

$$\sigma^2 \sim \Gamma^{-1}(2, b_\sigma(scale))$$

```
#our data
y = c(52, 45, 106, 118, 87, 43, 101, 65, 87, 91, 58, 64, 86, 37, 98, 76, 52, 69, 72, 99 )
```

Set Hyperparameters

Say that nearby levels of cesium range from 40 to 120. We can use this information to set the hyperparameters. First, we note the relationship between the expectation of the data and the prior mean:

$$E(Y) = E(E(Y|\mu, \sigma^2)) = E(\mu) = m_\mu$$

So if we think of the middle of our range, 80, as the data mean, we can justify setting the prior mean as: $m_\mu = 80$.

Next we look at the variance:

$$Var(Y) = Var(E(Y|\mu, \sigma^2)) + E(Var(Y|\mu, \sigma^2)) = Var(\mu) + E(\sigma^2) = s_\mu^2 + b_\sigma/(2-1)$$

And if we think of the the mean plus and minus two standard deviations is a good approximation of the range, we get:

$$Var(Y) \approx (20)^2 = 400 = s_\mu^2 + b_\sigma$$

Since it's an underdetermined system, we arbitrarily choose $s_\mu^2 = 300$; $b_\sigma = 100$.

This gives us the hyperparameters:

$$\mu \sim N(80, 300)$$
$$\sigma^2 \sim \Gamma^{-1}(2, 100)$$

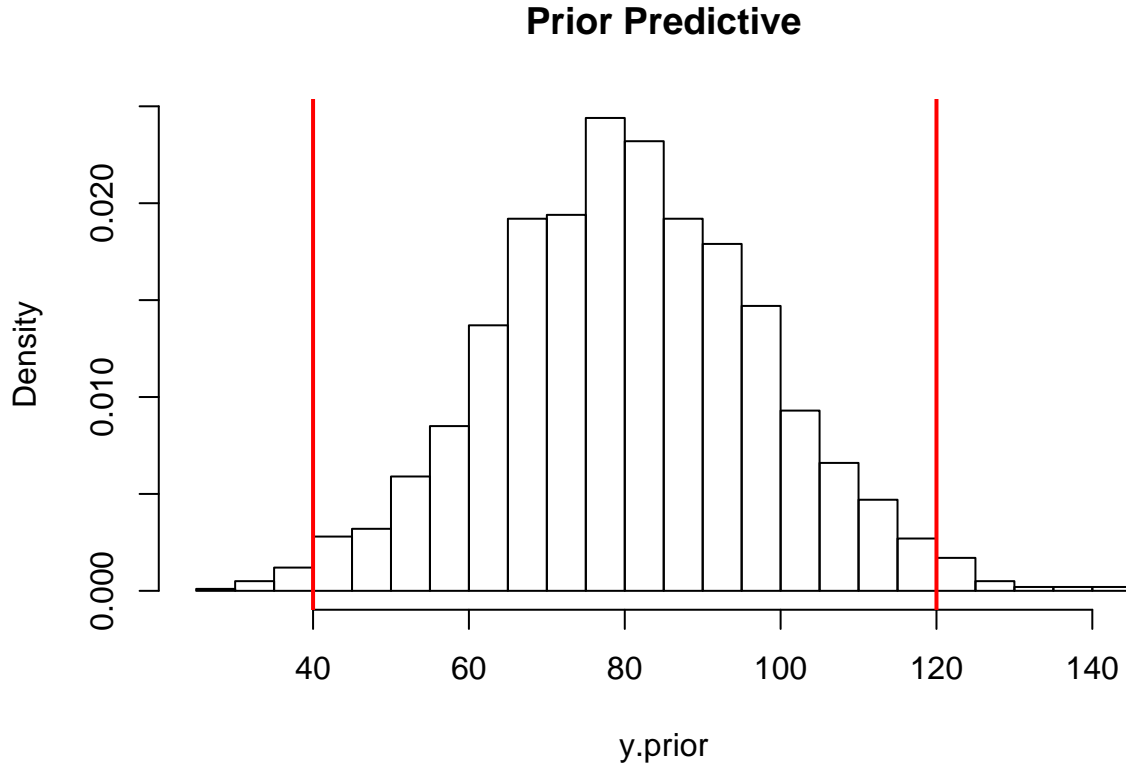**Plot Prior Predictive Distribution**

To test our hyperparameters we plot the prior predictive distribution.

```
sig2.prior = rinvgamma(n=2000,shape=2,scale=100)
mu.prior = rnorm(n=2000,mean=80,sd=sqrt(300))
y.prior = rnorm(n=2000,mean=mu.prior,sd=sqrt(sig2.prior))

hist(y.prior,freq=FALSE,main="Prior Predictive",breaks=20)
abline(v=40,col=2,lwd=2)
abline(v=120,col=2,lwd=2)
```

Prior Predictive

Monte Carlo and Grid Sampling

Our next step is to use Monte Carlo to obtain samples from the joint posterior (using marginal $p(\sigma^2|\mathbf{y})$ and conditional posteriors $p(\mu|\sigma^2, \mathbf{y})$); plot.

Determine Marginal Posterior of σ^2

We can obtain $p(\sigma^2|\mathbf{y})$ by noting the relationship:

$$p(\sigma^2|\mathbf{y}) = \int_{\forall \mu} p(\mu, \sigma^2|\mathbf{y})d\mu = \frac{1}{p(\mathbf{y})} \int_{\forall \mu} p(\mu, \sigma^2, \mathbf{y})d\mu$$

$$p(\sigma^2|\mathbf{y}) \propto \int_{\forall \mu} p(\mu, \sigma^2, \mathbf{y})d\mu$$

First we expand the log-quadratic terms in $p(\mu, \sigma^2, \mathbf{y})$:

$$\sum_i (y_i - \mu)^2 = \sum_i y_i^2 - 2\sum_i y_i\mu + n\mu^2$$

$$(\mu - m_\mu)^2 = \mu^2 - 2m_\mu\mu + m_\mu^2$$

2

Then we group the terms involving μ:

$$p(\sigma^2|\mathbf{y}) \propto C \int_{\forall \mu} \exp\left\{ -\frac{1}{2}S^2(\mu^2 - 2m_\mu\mu) - \frac{1}{2}T^2(\mu^2 - 2\bar{y}\mu) \right\}$$

$$C = (s_\mu^2)^{-1/2}(\sigma^2)^{-(n+6)/2} \exp\left\{ -\frac{1}{\sigma^2}b_\sigma - \frac{1}{2\sigma^2}\sum y_i^2 - \frac{1}{2s_\mu^2}m_\mu^2 \right\}$$

$$S^2 = \frac{1}{s_\mu^2}; \quad T^2 = \frac{1}{\sigma^2/n}$$

It should be evident that with some algebraic manipulation, we can get a normal kernel out of the log-quadratic terms.

$$p(\sigma^2|\mathbf{y}) \propto C \int_{\forall \mu} \exp\left\{ -\frac{1}{2}\left[(S^2 + T^2)\mu^2 - 2(S^2 m_\mu + T^2\bar{y})\mu\right] \right\}$$

$$\propto C \int_{\forall \mu} \exp\left\{ -\frac{1}{2}(S^2 + T^2)\left[\mu^2 - 2\frac{(S^2 m_\mu + T^2\bar{y})}{(S^2 + T^2)}\mu\right] \right\}$$

After completing the square and adding in a variance term $(S^2 + T^2)^{-1}$:

$$\propto C(S^2 + T^2)^{-1/2} \exp\left\{ \frac{1}{2}\frac{(S^2 m_\mu + T^2\bar{y})^2}{(S^2 + T^2)} \right\} \int_{\forall \mu} (S^2 + T^2)^{1/2} \exp\left\{ -\frac{1}{2}(S^2 + T^2)\left[\mu - \frac{(S^2 m_\mu + T^2\bar{y})}{(S^2 + T^2)}\right]^2 \right\}$$

The normal kernel $N\left(\frac{(S^2 m_\mu + T^2\bar{y})}{(S^2 + T^2)}, (S^2 + T^2)^{-1} \right)$ integrates out to a constant.

This leaves us with the following closed form of the marginal posterior:

$$p(\sigma^2|\mathbf{y}) \propto (\sigma^2)^{-(n+6)/2} \exp\left\{ -\frac{1}{\sigma^2}b_\sigma - \frac{1}{2\sigma^2}\sum y_i^2 - \frac{1}{2s_\mu^2}m_\mu^2 \right\}(S^2 + T^2)^{-1/2} \exp\left\{ \frac{1}{2}\frac{(S^2 m_\mu + T^2\bar{y})^2}{(S^2 + T^2)} \right\}$$

$$p(\sigma^2|\mathbf{y}) \propto (\sigma^2)^{-(n+6)/2} \exp\left\{ -\frac{1}{\sigma^2}b_\sigma - \frac{1}{2\sigma^2}\sum y_i^2 \right\}\left(\frac{1}{s_\mu^2} + \frac{1}{\sigma^2/n} \right)^{-1/2} \exp\left\{ \frac{1}{2}\frac{(m_\mu/s_\mu^2 + n\bar{y}/\sigma^2)^2}{(1/s_\mu^2 + n/\sigma^2)} \right\}$$

Determine Conditional Posterior for μ

Now we can easily obtain the corresponding conditional posterior for μ.

$$p(\mu|\sigma^2, \mathbf{y}) = \frac{p(\mu, \sigma^2, \mathbf{y})}{p(\sigma^2, \mathbf{y})} = \frac{1}{p(\mathbf{y})p(\sigma^2|\mathbf{y})} \cdot p(\mu, \sigma^2, \mathbf{y})$$

$$p(\mu|\sigma^2, \mathbf{y}) \propto p(\mu, \sigma^2, \mathbf{y})$$

$$p(\mu|\sigma^2, \mathbf{y}) \propto \exp\left\{ -\frac{1}{s_\mu^2}(\mu - m_\mu)^2 - \frac{1}{2\sigma^2}\sum_i (y_i - \mu)^2 \right\}$$

Which as we saw earlier, is the kernel for a normal distribution. Therefore, given σ^2 and the data, μ is normal:

$$N\left(\frac{(S^2 m_\mu + T^2\bar{y})}{(S^2 + T^2)}, (S^2 + T^2)^{-1} \right)$$

3

Draw from Joint Posterior using Grid Sampling

$$p(\mu, \sigma^2|\mathbf{y}) \propto p(\mu)p(\sigma^2)p(\mathbf{y}|\mu, \sigma^2)$$

$$p(\mu, \sigma^2|\mathbf{y}) \propto \exp\left\{-\frac{1}{2s_\mu^2}(\mu - m_\mu)^2\right\}(\sigma^2)^{-(n+6)/2}\exp\left\{-\frac{1}{2\sigma^2}\left[2b_\sigma + \sum_i(y_i - \mu)^2\right]\right\}$$

```r
# set hyperparameters
bsig = 100
m_mu = 80
s2_mu = 300

post = function(mu, sig2) {
    # data
    n = 20
    sum_y2 = 123638
    sum_y = 1506

    exp(-1/(2 * s2_mu) * (mu - m_mu)^2) * sig2^(-(n + 6)/2) * exp(-1/(2 * sig2) *
        (2 * bsig + sum_y2 - 2 * mu * sum_y + n * mu^2))
}

mu_grid = seq(40, 120, length.out = 300)
sig2_grid = seq(100, 1200, length.out = 300)
post_grid = matrix(0, nrow = length(mu_grid), ncol = length(sig2_grid))

for (i in 1:length(mu_grid)) {
    for (j in 1:length(sig2_grid)) {
        post_grid[i, j] = post(mu_grid[i], sig2_grid[j])
    }
}
C_joint = sum(post_grid)
image(mu_grid, sig2_grid, post_grid/C_joint, col = topo.colors(20), xlab = expression(mu),
    ylab = expression(sigma^2), main = "Monte Carlo Joint Posterior Density")
contour(mu_grid, sig2_grid, post_grid/C_joint, add = TRUE)
```

4

Monte Carlo Joint Posterior Density

Plot Marginal Posterior of μ

$$p(\mu|\mathbf{y}) \propto \int_{\forall \sigma^2} p(\mu, \sigma^2, \mathbf{y}) d\sigma^2$$

We proceed by isolating terms of σ^2 :

$$C \cdot (\sigma^2)^{-(n+6)/2} \exp\left\{ -\frac{1}{2\sigma^2} A \right\}$$

$$C = (s_\mu^2)^{-1/2} \exp\left(-\frac{1}{2s_\mu^2} (\mu - m_\mu)^2 \right)$$

$$A = 2b_\sigma + \sum (y_i - \mu)^2$$

Now it should be apparent that we are one substitution away from a Gamma kernel:

$$p(\mu|\mathbf{y}) \propto C \int_{\forall \sigma^2} (\sigma^2)^{-(n+6)/2} \exp\left\{ -\frac{1}{2\sigma^2} A \right\} d\sigma^2$$

$$z = \left(\frac{A}{2} \right) (\sigma^2)^{-1}$$

$$\sigma^2 = \left(\frac{2}{A} \right) z^{-1}$$

$$\left(\frac{A}{2}\right)z^{-2}dz = d\sigma^2$$

$$p(\mu|\mathbf{y}) \propto C \cdot \left(\frac{A}{2}\right)^{-(n+4)/2} \int_{\forall z} z^{(n+2)/2} \exp(-z)dz$$

Revealing the Gamma kernel: $Gamma(n/2 + 2, -1)$ which will integrate to a constant.

$$p(\mu|\mathbf{y}) \propto \left(2b_\sigma + \sum(y_i - \mu)^2\right)^{-(n+4)/2} \times \exp\left(-\frac{1}{2s_\mu^2}(\mu - m_\mu)^2\right)$$
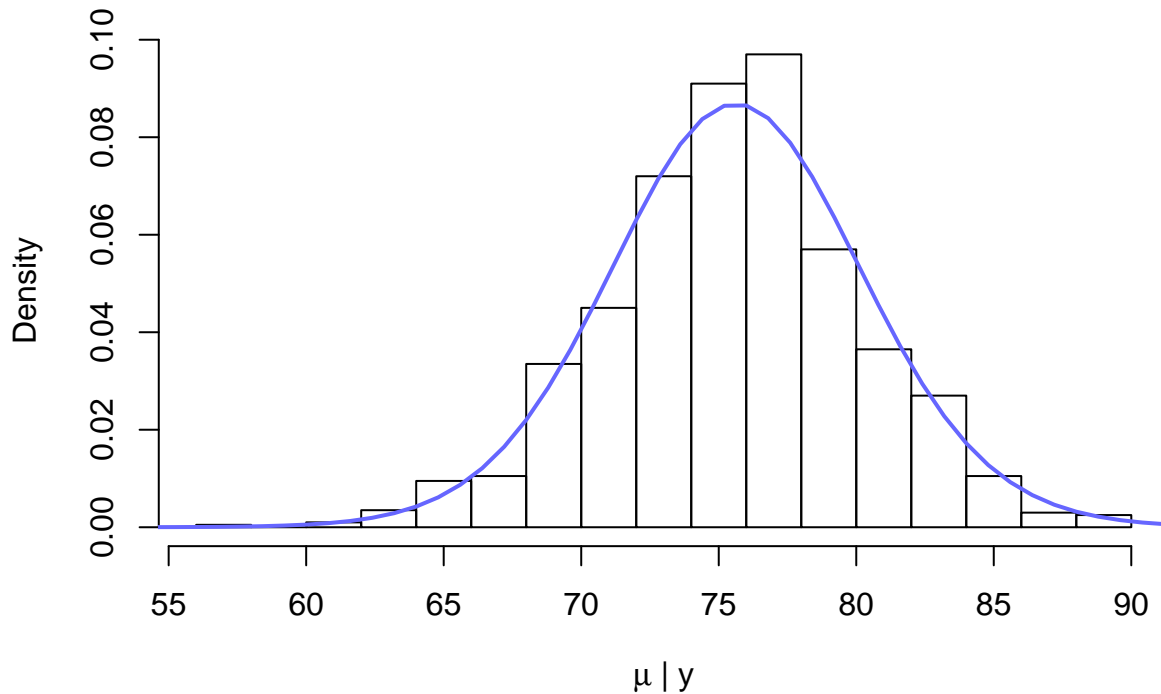
```r
# set hyperparameters
bsig = 100
m_mu = 80
s2_mu = 300

mu_margin = function(mu) {
    # data
    n = 20
    sum_y2 = 123638
    sum_y = 1506
    (2 * bsig + sum_y2 - 2 * mu * sum_y + n * mu^2)^(-(n + 4)/2) * exp(-1/(2 *
        s2_mu) * (mu - m_mu)^2)
}

# grid approach
mu_grid = seq(40, 120, length.out = 300)
post_grid = numeric()
for (i in 1:length(mu_grid)) {
    post_grid[i] = mu_margin(mu_grid[i])
}

mu_margin_MC = sample(mu_grid, 1000, replace = TRUE, prob = post_grid/sum(post_grid))

hist(mu_margin_MC, breaks = 20, main = "Histogram of Monte Carlo Draws of Marginal Mean",
    freq = FALSE, xlab = expression(paste(mu, " | ", y)))
curve(mu_margin(x) * 1.5e+47, from = 40, to = 120, add = TRUE, lwd = 2, col = rgb(0.4,
    0.4, 1, 1))
```

Histogram of Monte Carlo Draws of Marginal Mean

Plot Marginal Posterior of σ^2

$$p(\sigma^2|\mathbf{y}) \propto (\sigma^2)^{-(n+6)/2} \exp\left\{-\frac{1}{\sigma^2}b_\sigma - \frac{1}{2\sigma^2}\sum y_i^2\right\}\left(\frac{1}{s_\mu^2} + \frac{1}{\sigma^2/n}\right)^{-1/2} \exp\left\{\frac{1}{2}\frac{(m_\mu/s_\mu^2 + n\bar{y}/\sigma^2)^2}{(1/s_\mu^2 + n/\sigma^2)}\right\}$$

```r
sig2_margin = function(sig2) {
    # data
    n = 20
    sum_y2 = 123638
    sum_y = 1506

    (sig2)^(-(n + 6)/2) * exp(-1/sig2 * bsig - 1/(2 * sig2) *
        sum_y2) * (1/s2_mu + 1/(sig2/n))^(-1/2) * exp(1/2 * (m_mu/s2_mu +
        sum_y/sig2)^2/(1/s2_mu + n/sig2))
}

# grid approach
sig2_grid = seq(100, 1200, length.out = 300)
post_grid2 = numeric()
for (i in 1:length(sig2_grid)) {
    post_grid2[i] = sig2_margin(sig2_grid[i])
}

sig2_margin_MC = sample(sig2_grid, 1000, replace = TRUE, prob = post_grid2/sum(post_grid2))

hist(sig2_margin_MC, breaks = 20, main = "Histogram of Monte Carlo Draws of Marginal Variance",
    freq = FALSE, xlab = expression(paste(sigma^2, " | ", y)))
```
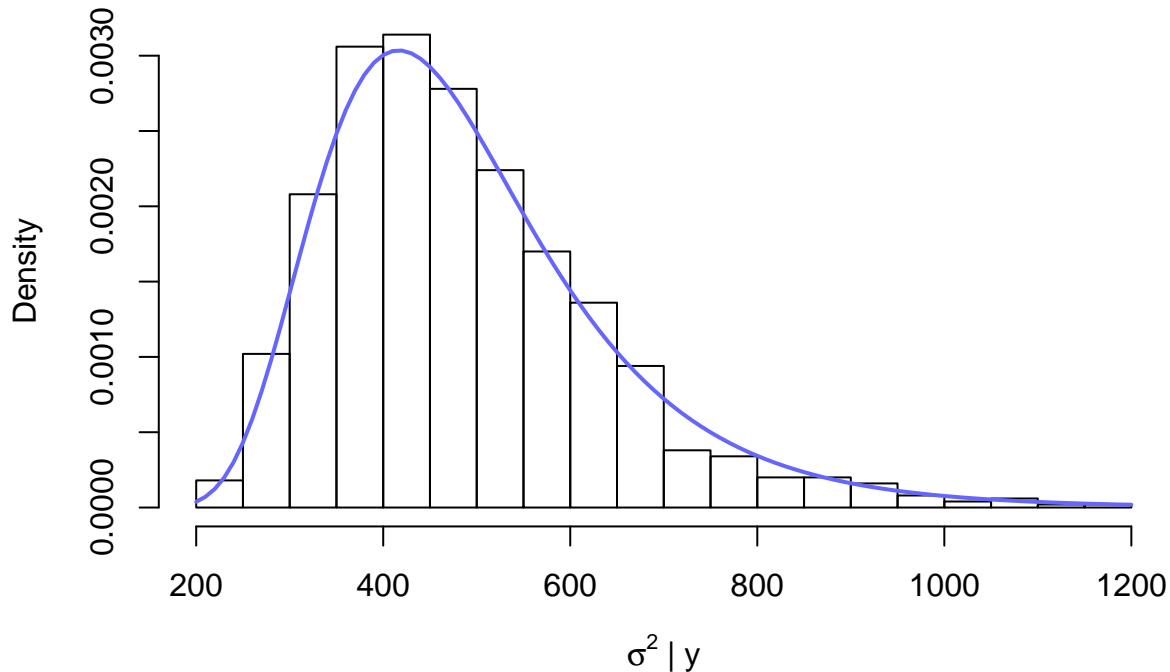
```
curve(sig2_margin(x) * 5.2e+31, from = 200, to = 1200, add = TRUE,
    lwd = 2, col = rgb(0.4, 0.4, 1, 1))
```

Histogram of Monte Carlo Draws of Marginal Variance

Finally, we simulate from the posterior predictive distribution and plot the results.

Simulate from Posterior Predictive Distribution

```
# produce discrete approximate to joint posterior
post = function(mu, sig2) {
    # data
    n = 20
    sum_y2 = 123638
    sum_y = 1506

    exp(-1/(2 * s2_mu) * (mu - m_mu)^2) * sig2^(-(n + 6)/2) * exp(-1/(2 * sig2) *
        (2 * bsig + sum_y2 - 2 * mu * sum_y + n * mu^2))
}

mu_grid = seq(40, 120, length.out = 300)
sig2_grid = seq(100, 1200, length.out = 300)
mu_post = numeric()
post_grid = matrix(0, nrow = length(mu_grid), ncol = length(sig2_grid))

for (i in 1:length(mu_grid)) {
    for (j in 1:length(sig2_grid)) {
        post_grid[i, j] = post(mu_grid[i], sig2_grid[j])
    }
    mu_post[i] = sum(post_grid[i, ])
}
```

```
C_joint = sum(post_grid)
C_mu = sum(mu_post)

### sample from discrete joint posterior sample mu marginal
B = 1000
place = sample.int(length(mu_grid), B, prob = mu_post, replace = TRUE)
mu_post_sample = mu_grid[place]

sig2_post_sample = place2 = numeric()
for (b in 1:B) {
    place2[b] = sample.int(length(sig2_grid), 1, prob = post_grid[place[b],
        ], replace = TRUE)
    sig2_post_sample[b] = sig2_grid[place2[b]]
}

mu_post_sample = mu_post_sample + runif(B, -0.27/2, 0.27/2)
sig2_post_sample = sig2_post_sample + runif(B, -3.7/2, 3.7/2)

# sample posterior predictive draws
ytilde = rnorm(1000, mu_post_sample, sqrt(sig2_post_sample))
```

Plot Posterior Predictive Distribution

```
hist(ytilde, main = "Histogram of Posterior Predictive Monte Carlo Draws", freq = FALSE,
    xlab = expression(tilde(y)))
```


Histogram of Posterior Predictive Monte Carlo Draws
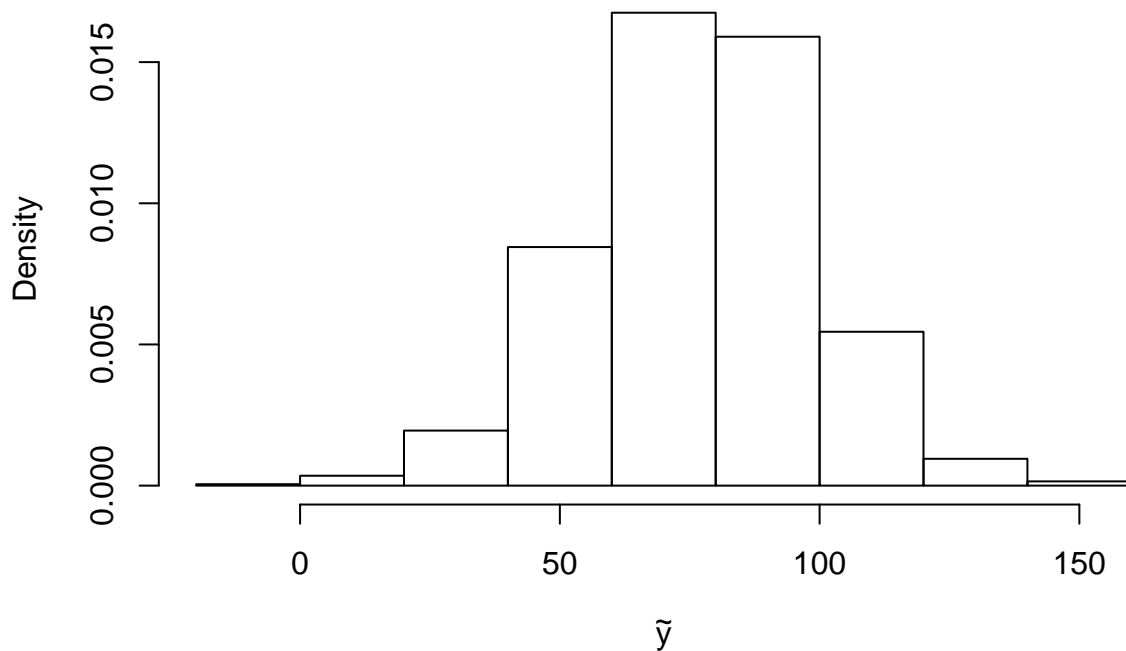
Estimate Probability of Cesium Levels over 100

What is the probability of a fish having Cesium levels over 100?

```
sum(ytilde>100)/length(ytilde)
```

```
## [1] 0.131
```

Gibbs Sampling

We need to derive the conditional posteriors necessary to perform Gibbs Sampler.

Determine Conditional Posterior for σ^2

We're only missing $p(\sigma^2|\mu, \mathbf{y})$ which we can obtain through:

$$p(\sigma^2|\mu, \mathbf{y}) \propto \frac{p(\sigma^2, \mu, \mathbf{y})}{p(\mu, \mathbf{y})} = \frac{1}{p(\mathbf{y})p(\mu|\mathbf{y})} p(\sigma^2, \mu, \mathbf{y})$$

$$p(\sigma^2|\mu, \mathbf{y}) \propto (\sigma^2)^{-(n+6)/2} \exp\left\{ -\frac{1}{2\sigma^2}\left[2b_\sigma + \sum_i (y_i - \mu)^2\right]\right\}$$

Which we found out earlier with the proper substitution:

$$\sigma^2 = \left(\frac{2}{A}\right) z^{-1}$$

$$A = 2b_\sigma + \sum_i (y_i - \mu)^2$$

has the gamma distribution $Gamma(n/2 + 2, -1)$.

Perform Gibbs Sampling

To perform Gibbs we follow these steps:

1. Initialize $(\sigma^2)^{(0)}$.
2. For b in $1, ..., B$.

i) Sample from:

$$\mu^{(b)} \sim N\left(\frac{m_\mu/s_\mu^2 + n\bar{y}/(\sigma^2)^{(b-1)}}{1/s_\mu^2 + n/(\sigma^2)^{(b-1)}}, \left(1/s_\mu^2 + n/(\sigma^2)^{(b-1)}\right)^{-1}\right)$$

ii) Sample from:

$$z^{(b)} \sim Gamma(n/2 + 2, -1)$$

iii) Transform the z into σ^2 :

$$(\sigma^2)^{(b)} = \left(\frac{2}{2b_\sigma + \sum_i (y_i - \mu^{(b)})^2}\right)\left(z^{(b)}\right)^{-1}$$

```r
#b - number of iterations
Gibbs = function(B,sig2_start){

  #our data
  y = c(52,45,106,118,87,43,101,65,87,91,58,64,86,37,98,76,52,69,72,99)
  n = length(y)

  #set hyperparameters
  bsig = 100
  m_mu = 80
  s2_mu = 300


  #initialize vectors to collect posterior draws from gibbs
  mu_draws = numeric()
  sig2_draws = numeric()

  #start gibbs
  sig2_old = sig2_start
  for(b in 1:B){


    #sample posterior mu conditional on sig2
    condmu_mean = (m_mu/s2_mu+sum(y)/sig2_old)/(1/s2_mu+n/sig2_old)
    condmu_var = (1/s2_mu+n/sig2_old)^(-1)

    mu = rnorm(1,condmu_mean,sqrt(condmu_var))

    #sample posterior sig2 conditional on mu

    z = rgamma(1,n/2+2,1)
    sig2 = (2*bsig+sum((y-mu)^2))/2*z^{-1}

    #save
    mu_draws[b] = mu
    sig2_draws[b] = sig2

    #update
    sig2_old = sig2
  }
  return(list(mu_draws=mu_draws,sig2_draws=sig2_draws))
}

GSpostdraws = Gibbs(B=2000,sig2_start=200)
```
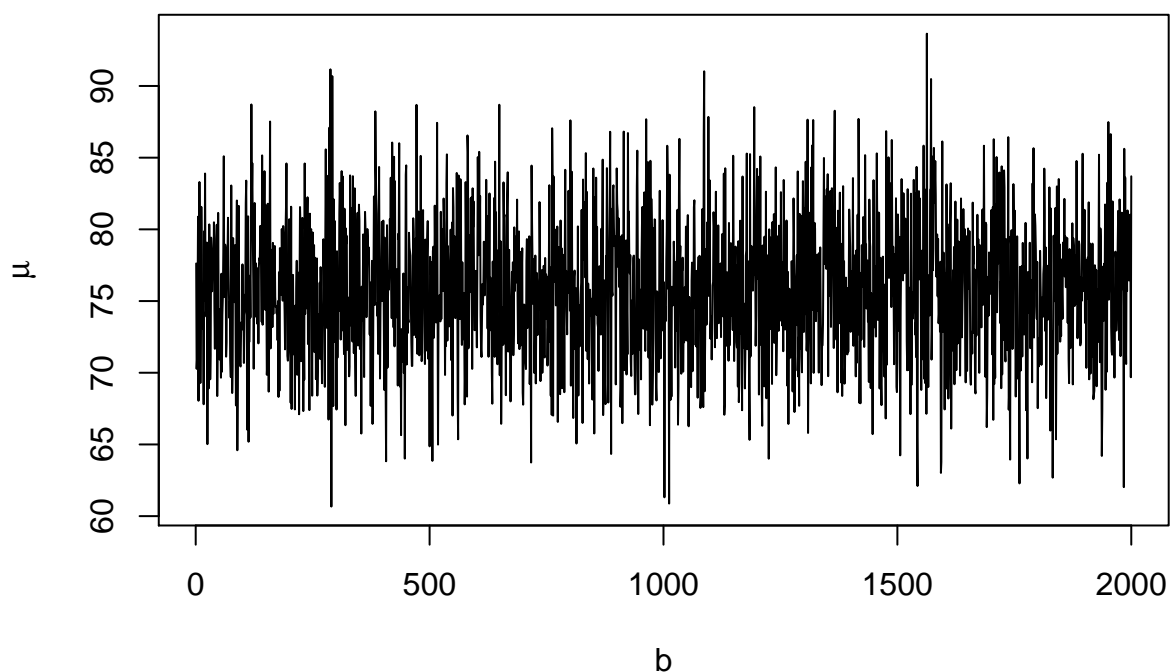
Gibbs Sampling Convergence Diagnostics

```r
# trace plots
plot(GSpostdraws[["mu_draws"]], type = "l", xlab = c("b"), ylab = expression(mu),
    main = expression(paste("Trace plot of ", mu)))
```
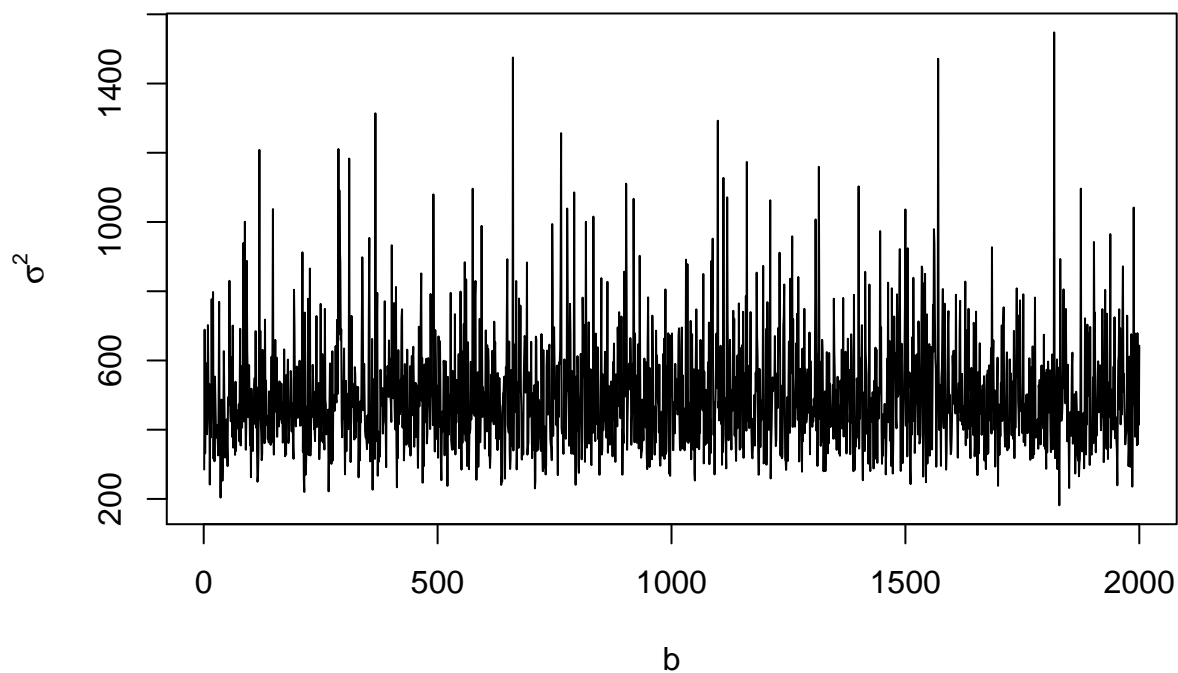
Trace plot of μ


```
plot(GSpostdraws[["sig2_draws"]], type = "l", xlab = c("b"), ylab = expression(sigma^2),
    main = expression(paste("Trace plot of ", sigma^2)))
```

Trace plot of σ²

The trace plots indicate that our initial values were within the target distribution and there's no need to burn the early part of the chain since convergence was essentially instantaneous. Next we want to check

autocorrelation in the chain in case thinning is needed.

```r
acf(GSpostdraws[["mu_draws"]],main=expression(paste("Autocorrelation Plot of ",mu)))
```

Autocorrelation Plot of μ
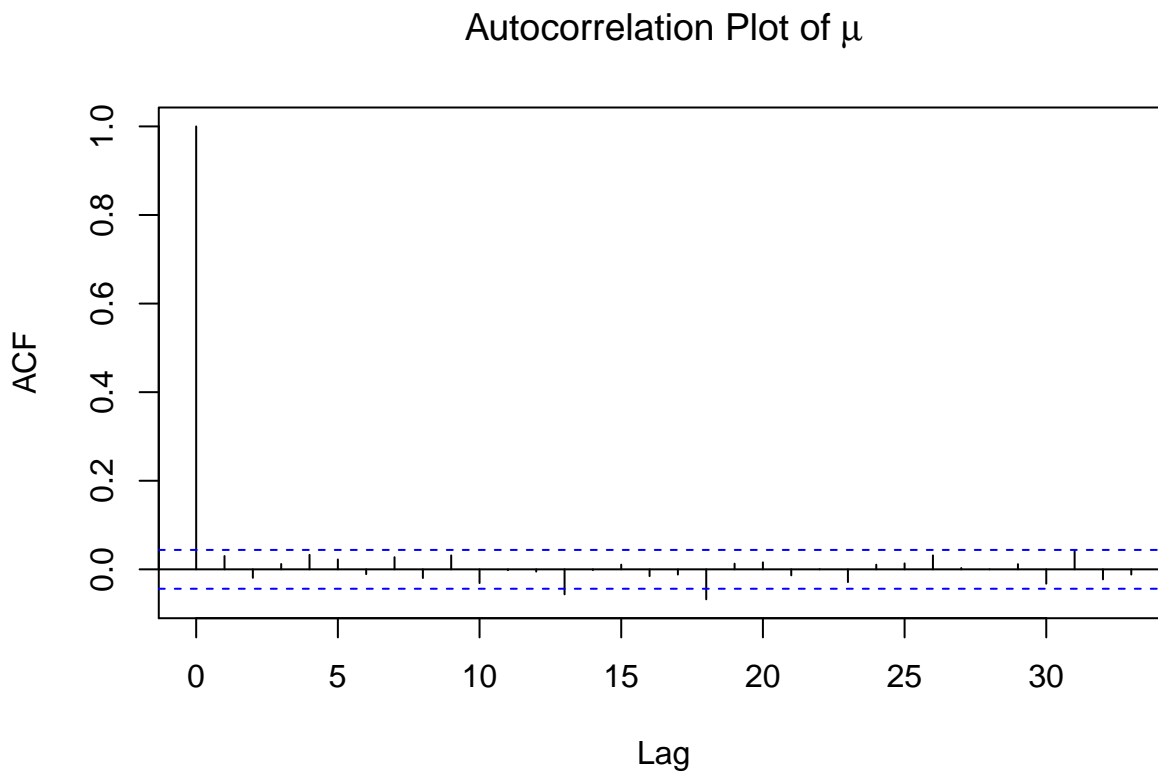

```r
acf(GSpostdraws[["sig2_draws"]],main=expression(paste("Autocorrelation Plot of ",sigma^2)))
```

## Autocorrelation Plot of σ^2
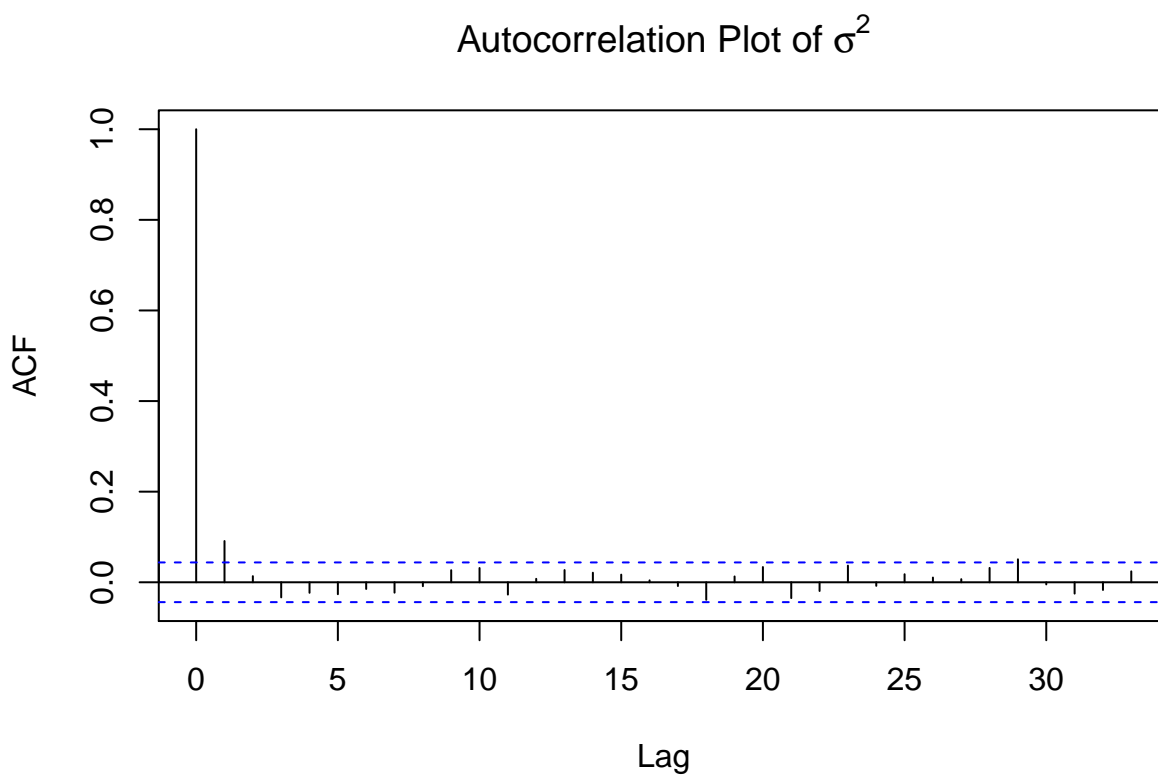
μ and σ^2 shows no significant autocorrelation.

Simulate from Posterior Predictive Distribution

```
ytildeGS = rnorm(2000, GSpostdraws[["mu_draws"]], sqrt(GSpostdraws[["sig2_draws"]]))
hist(ytildeGS, main = "Histogram of Posterior Predictive Gibbs Sampling Draws",
    freq = FALSE, xlab = expression(tilde(y)))
```

Histogram of Posterior Predictive Gibbs Sampling Draws
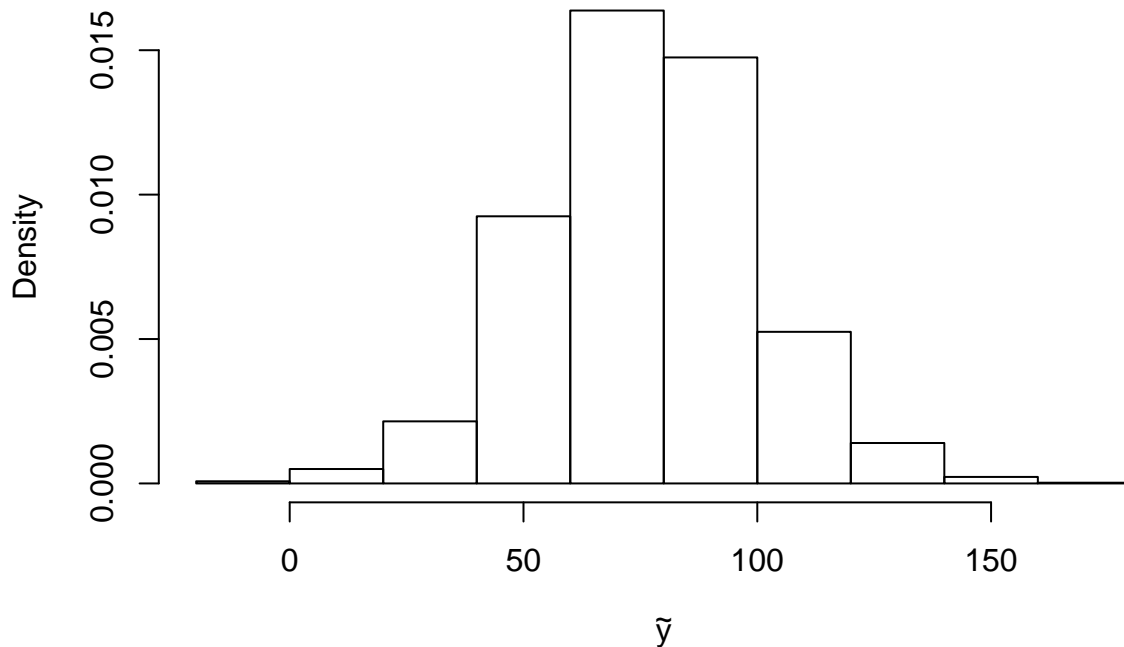
Compare Joint Posterior Draws of (μ, σ^2)

```
plot(GSpostdraws[["mu_draws"]][1:1000], GSpostdraws[["sig2_draws"]][1:1000],
    col = rgb(0.4, 0.4, 1, 1), pch = 20, xlab = expression(paste(mu, "|y")),
    ylab = expression(paste(sigma^2, "|y")), main = "Gibbs Sampling Posterior Draws")
points(mu_post_sample, sig2_post_sample, col = rgb(0, 0.4, 0.4, 0.5), pch = 20)
legend("topright", c("Gibbs", "MC"), col = c(rgb(0.4, 0.4, 1, 1), rgb(0, 0.4,
    0.4, 0.5)), pch = 20)
```

Gibbs Sampling Posterior Draws
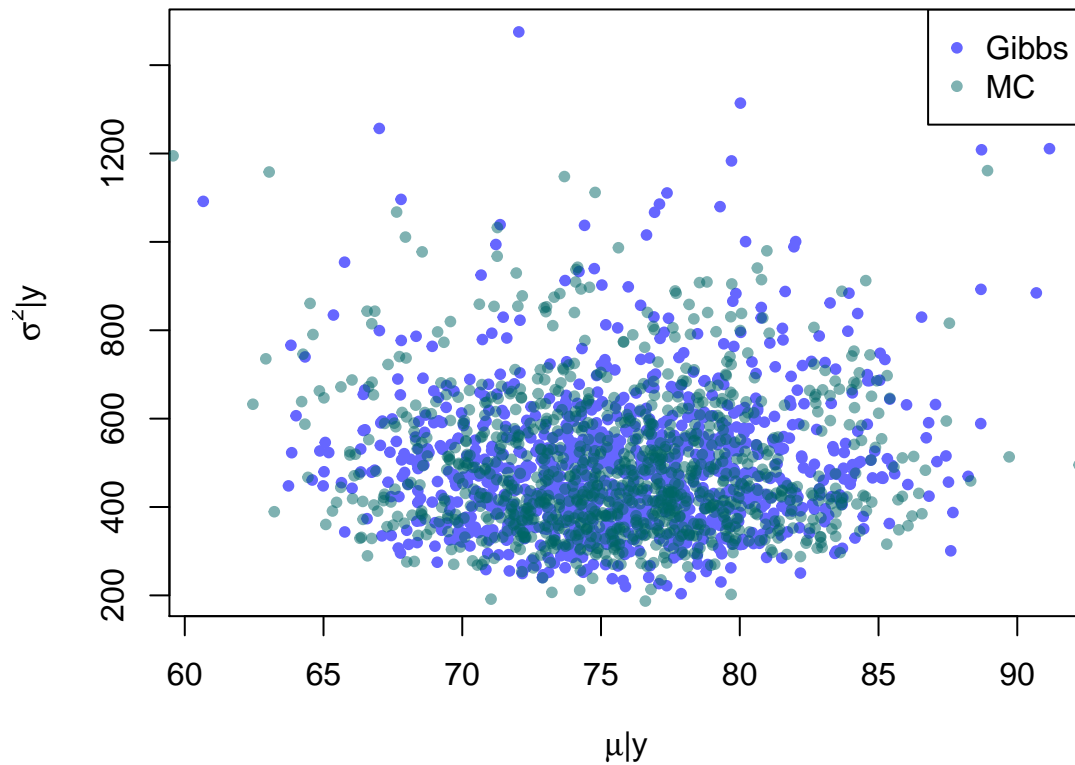
The draws both look like they're coming from the same target distribution: the joint posterior. They both look to be equally good approximate samples.

Compare Marginal Posterior of μ

```r
plot(density(mu_post_sample, adjust = 1.3), main = expression(paste("Marginal Posterior of ",
    mu)), lty = 1, xlim = c(50, 100))
lines(density(GSpostdraws[["mu_draws"]][1:1000], adjust = 1.3), col = 3, lty = 2)
legend("topleft", legend = c("1000 MC Samples", "Gibbs Sampling"), col = c(1,
    3), lty = c(1, 2))
```

Marginal Posterior of μ

N = 1000   Bandwidth = 1.397
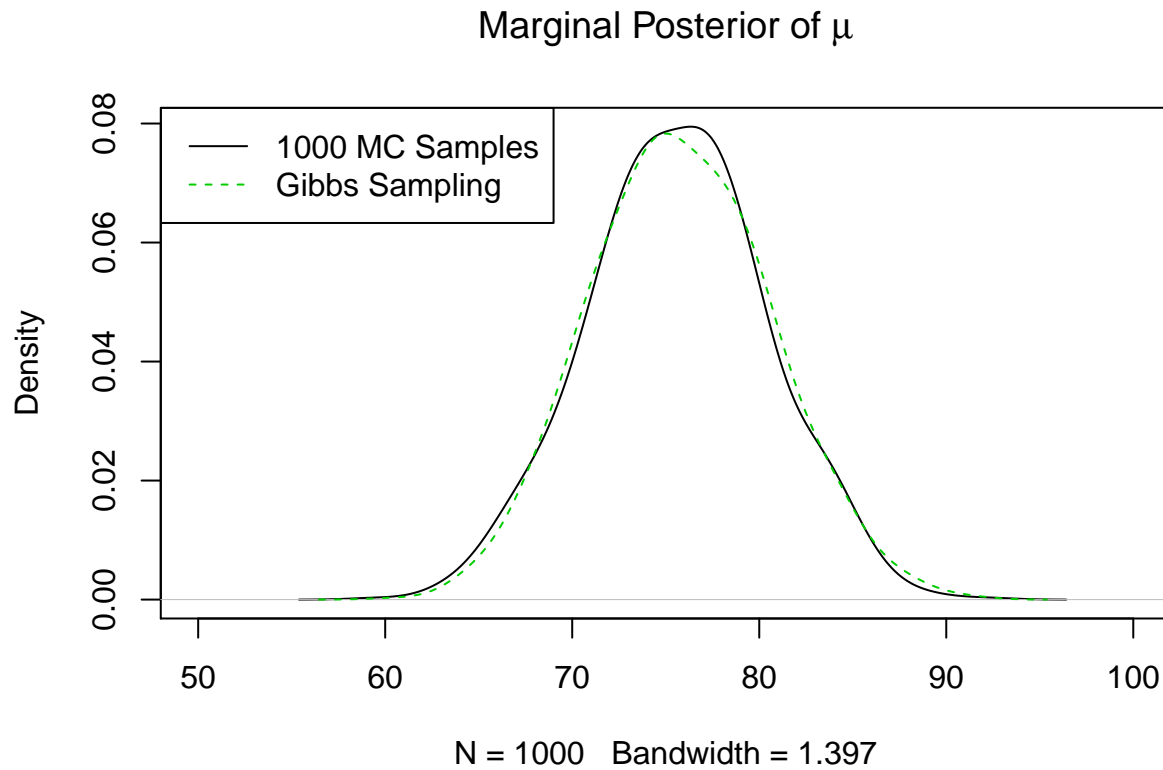
Looks like there isn't too much of a difference; obviously this also depends on the bandwidth.

Compare Marginal Posterior of σ^2

```r
plot(density(sig2_post_sample, adjust = 1.3), main = expression(paste("Marginal Posterior of ",
    sigma^2)), lty = 1)
lines(density(GSpostdraws[["sig2_draws"]][1:1000], adjust = 1.3), col = 3, lty = 2)
legend("topright", legend = c("1000 MC Samples", "Gibbs Sampling"), col = c(1,
    3), lty = c(1, 2))
```

Marginal Posterior of σ^2

N = 1000   Bandwidth = 45.58
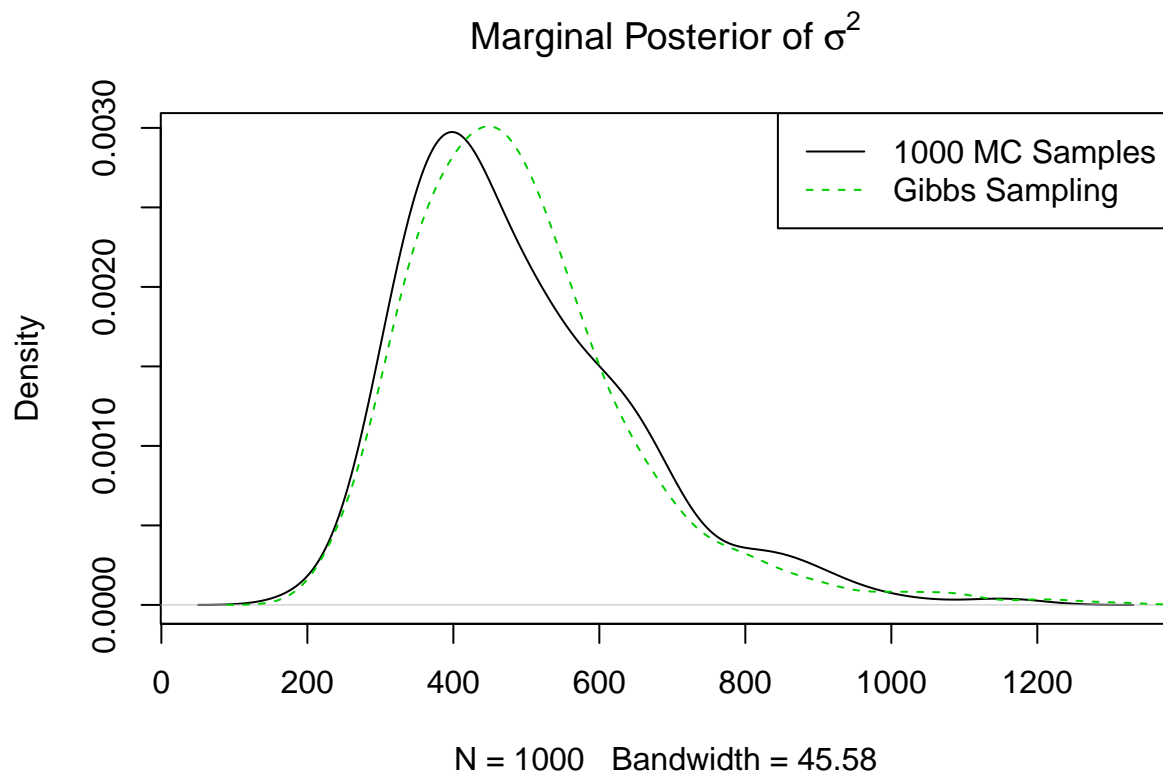
Again, the methods don't seem to differ much.