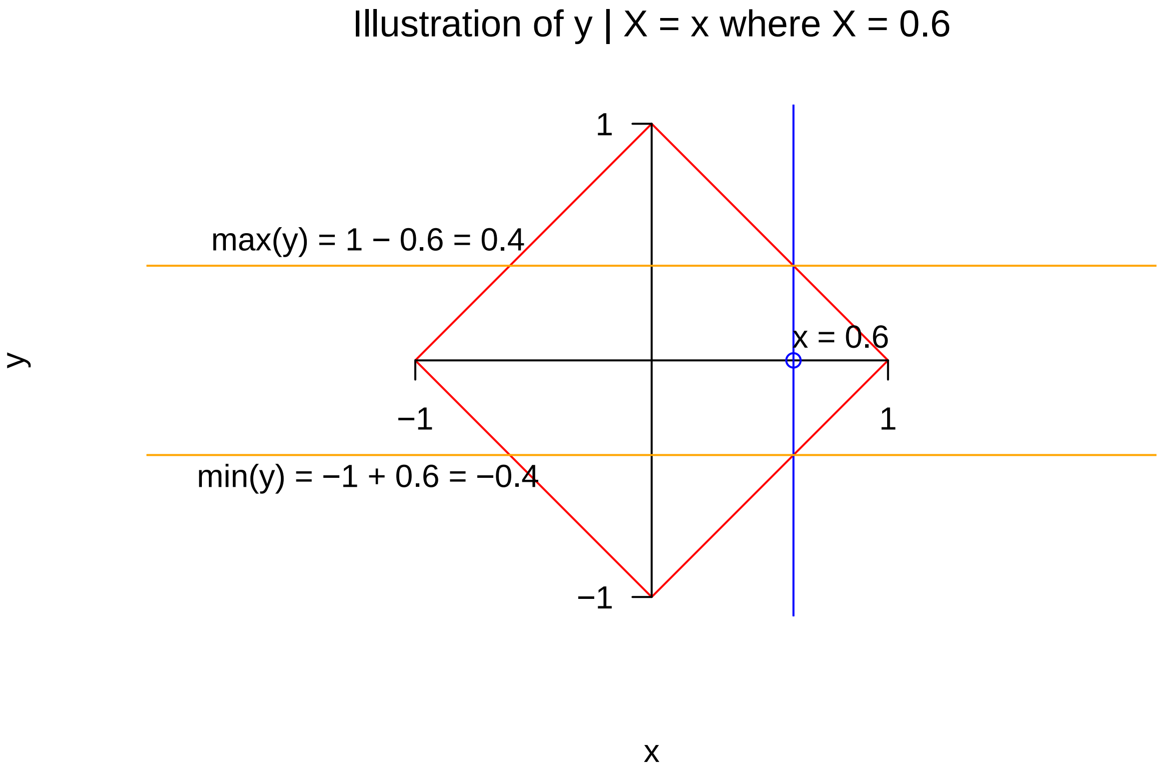
STAT 5443 Midterm

Daniel Kiser

March 28, 2018

Problem 1

One possible method for generating samples uniformly from the given diamond-shaped area is to use a Gibbs sampler. To implement the Gibbs sampler, we must first find the conditional probabilities $x|y$ and $y|x$.



If, for example, $x = 0.6$, it is clear from the above diagram that y can only take values less than 0.4 or greater than -0.4, since any values of y outside of that range would result in the point (x, y) being outside of the designated area. Thus, any point y within the range (-0.4, 0.4) is equally likely with probability density function

$$\frac{1}{b-a} = \frac{1}{0.4-(-0.4)} = \frac{1}{0.8},$$

and any point outside the range (-0.4, 0.4) has a probability of zero. Thus, $y|X = 0.6 \sim Unif(-0.4, 0.4)$. We can generalize and say that for any value of x,

$$y|X = x \sim Unif(-1 + |x|, 1 - |x|)$$

Since the area we are attempting to sample from is symmetric, we can use similar reasoning to find the conditional distribution of x given y:

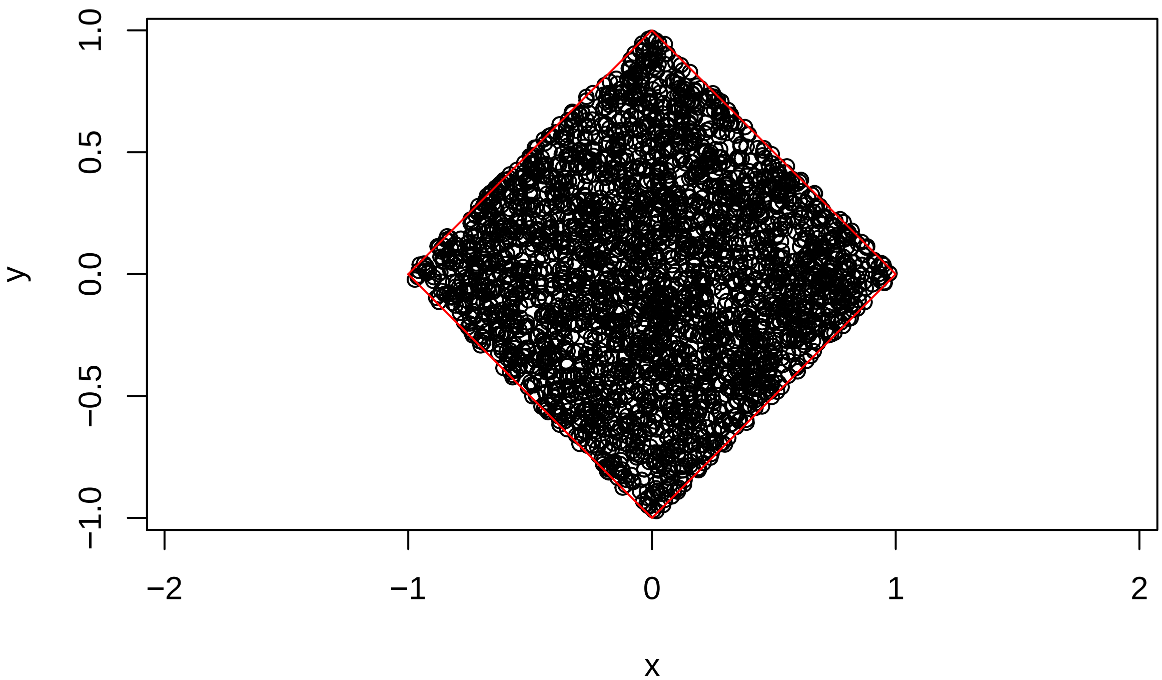
$$x|Y = y \sim Unif(-1 + |y|, 1 - |y|)$$

Below is our implementation of the Gibbs sampler in R:

```r
gibbs <- function(n, x, y) {
  for(i in 2:n+100) {
    x[i] <- runif(1, min = -1 + abs(y[i-1]), max = 1 - abs(y[i-1]))
    y[i] <- runif(1, min = -1 + abs(x[i]), max = 1 - abs(x[i]))
  }

  # discard burn-in
  return(list(x[100:n+100], y[100:n+100]))
}
```

We discard the first 100 samples at the burn-in samples. To run the Gibbs sampler, we set our initial values of x and y to 0 and and the number of repetitions (n) to 3000.

Our results are plotted below. All the points fall within the boundaries of the given area.



Problem 2

Part 1) First we find the full conditional distribution of θ given x and σ^2:

$$p(\theta|\mathbf{x}, \sigma^2) \propto f(\mathbf{x}|\theta, \sigma^2)p(\sigma^2|\theta)p(\theta)$$

We can drop the term $p(\sigma^2|\theta)$ since it does not depend on θ and behaves like a constant. So we have:

$$p(\theta|\mathbf{x}, \sigma^2) \propto f(\mathbf{x}|\theta, \sigma^2)p(\theta)$$

$$\propto \prod_{i=1}^{n} \left[(2\pi)^{-1/2}(\sigma^2)^{-1/2} exp\left\{-\frac{1}{2\sigma^2}(x_i-\theta)^2\right\}\right]\left[(2\pi)^{-1/2}(\tau^2)^{-1/2} exp\left\{-\frac{1}{2\tau^2}(\theta-\theta_0)^2\right\}\right]$$

$$\propto \left[(2\pi)^{-n/2}(\sigma^2)^{-n/2} exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2\right\}\right]\left[(2\pi)^{-1/2}(\tau^2)^{-1/2} exp\left\{-\frac{1}{2\tau^2}(\theta-\theta_0)^2\right\}\right]$$

$$\propto (2\pi)^{-n/2}(\sigma^2)^{-n/2}(2\pi)^{-1/2}(\tau^2)^{-1/2} exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2\right\} exp\left\{-\frac{1}{2\tau^2}(\theta-\theta_0)^2\right\}$$

$$\propto (2\pi)^{-n/2}(\sigma^2)^{-n/2}(2\pi)^{-1/2}(\tau^2)^{-1/2}exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2-\frac{1}{2\tau^2}(\theta-\theta_0)^2\right\}$$

We can drop the terms outside of the exponent, since they also behave like constants. Then we have:

$$p(\theta|\mathbf{x},\sigma^2)\propto exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2-\frac{1}{2\tau^2}(\theta-\theta_0)^2\right\}$$

$$\propto exp\left\{-\frac{1}{2}\left[\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2+\frac{1}{\tau^2}(\theta-\theta_0)^2\right]\right\}$$

The summation $\sum_{i=1}^{n}(x_i-\theta)^2$ can be expanded into $\sum_{i=1}^{n}(x_i-\bar{x})^2+n(\bar{x}-\theta)^2$. Thus, we get:

$$\propto exp\left\{-\frac{1}{2}\left[\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\bar{x})^2+\frac{n}{\sigma^2}(\bar{x}-\theta)^2+\frac{1}{\tau^2}(\theta-\theta_0)^2\right]\right\}$$

Since the summation $\sum_{i=1}^{n}(x_i-\bar{x})^2$ no longer includes θ, we can drop the summation and sum the remaining terms in the exponent:

$$\propto exp\left\{-\frac{1}{2}\left[\frac{n}{\sigma^2}(\bar{x}-\theta)^2+\frac{1}{\tau^2}(\theta-\theta_0)^2\right]\right\}\propto exp\left\{-\frac{1}{2}\left[\frac{n\tau^2(\bar{x}-\theta)^2+\sigma^2(\theta-\theta_0)^2}{\sigma^2\tau^2}\right]\right\}$$

$$\propto exp\left\{-\frac{1}{2\sigma^2\tau^2}\left[n\tau^2(\bar{x}-\theta)^2+\sigma^2(\theta-\theta_0)^2\right]\right\}$$

Expanding the squared terms of the exponent, we get:

$$\propto exp\left\{-\frac{1}{2\sigma^2\tau^2}\left[n\tau^2(\bar{x}^2-2\theta\bar{x}+\theta^2)+\sigma^2(\theta^2-2\theta\theta_0+\theta_0^2)\right]\right\}$$

$$\propto exp\left\{-\frac{1}{2\sigma^2\tau^2}\left[n\tau^2\bar{x}^2-2n\tau^2\theta\bar{x}+n\tau^2\theta^2+\sigma^2\theta^2-2\sigma^2\theta\theta_0+\sigma^2\theta_0^2\right]\right\}$$

We group together terms that include θ and factor out θ and θ^2, and we drop the terms in the exponent that do not include θ:

$$\propto exp\left\{-\frac{1}{2\sigma^2\tau^2}\left[n\tau^2\theta^2+\sigma^2\theta^2-2n\tau^2\theta\bar{x}-2\sigma^2\theta\theta_0+n\tau^2\bar{x}^2+\sigma^2\theta_0^2\right]\right\}$$

$$\propto exp\left\{-\frac{1}{2\sigma^2\tau^2}\left[\theta^2(\sigma^2+n\tau^2)-2\theta(n\tau^2\bar{x}+\sigma^2\theta_0)+n\tau^2\bar{x}^2+\sigma^2\theta_0^2\right]\right\}$$

$$\propto exp\left\{-\frac{1}{2\sigma^2\tau^2}\left[\theta^2(\sigma^2+n\tau^2)-2\theta(n\tau^2\bar{x}+\sigma^2\theta_0)\right]\right\}$$

We both divide and multiply the terms in the exponent by $(\sigma^2+n\tau^2)$ to obtain:

$$\propto exp\left\{-\frac{\sigma^2+n\tau^2}{2\sigma^2\tau^2}\left[\frac{\theta^2(\sigma^2+n\tau^2)-2\theta(n\tau^2\bar{x}+\sigma^2\theta_0)}{\sigma^2+n\tau^2}\right]\right\}$$

$$\propto exp\left\{-\frac{\sigma^2+n\tau^2}{2\sigma^2\tau^2}\left[\theta^2-2\theta\left(\frac{n\tau^2\bar{x}+\sigma^2\theta_0}{\sigma^2+n\tau^2}\right)\right]\right\}$$

We can add the term $\left(\frac{n\tau^2\bar{x}+\sigma^2\theta_0}{\sigma^2+n\tau^2}\right)^2$ to the exponent without changing the proportionality, since the term does not depend on θ:

$$\propto exp\left\{-\frac{\sigma^2+n\tau^2}{2\sigma^2\tau^2}\left[\theta^2-2\theta\left(\frac{n\tau^2\bar{x}+\sigma^2\theta_0}{\sigma^2+n\tau^2}\right)+\left(\frac{n\tau^2\bar{x}+\sigma^2\theta_0}{\sigma^2+n\tau^2}\right)^2\right]\right\}$$

Finally, we can write the equation in the form that is typically used for a Normal distribution:

$$\propto exp\left\{-\frac{\sigma^2+n\tau^2}{2\sigma^2\tau^2}\left[\theta-\frac{n\tau^2\bar{x}+\sigma^2\theta_0}{\sigma^2+n\tau^2}\right]^2\right\}$$

$$\propto exp\left\{-\frac{1}{2}\frac{1}{\frac{\sigma^2\tau^2}{\sigma^2+n\tau^2}}\left(\theta-\frac{n\tau^2}{\sigma^2+n\tau^2}\bar{x}-\frac{\sigma^2}{\sigma^2+n\tau^2}\theta_0\right)^2\right\}$$

From the above equation, it is evident that we have a Normal distribution with mean

$$\frac{n\tau^2}{\sigma^2+n\tau^2}\bar{x}+\frac{\sigma^2}{\sigma^2+n\tau^2}\theta_0$$

and variance

$$\frac{\sigma^2\tau^2}{\sigma^2+n\tau^2}$$

Thus, we can write:

$$\theta|\mathbf{x},\sigma^2 \sim N\left(\frac{n\tau^2}{\sigma^2+n\tau^2}\bar{x}+\frac{\sigma^2}{\sigma^2+n\tau^2}\theta_0, \frac{\sigma^2\tau^2}{\sigma^2+n\tau^2}\right)$$

Next, we find the full conditional distribution of σ^2 given θ and x:

$$p(\sigma^2|\mathbf{x},\theta) \propto f(\mathbf{x}|\theta,\sigma^2)p(\theta|\sigma^2)p(\sigma^2)$$

We can drop the term $p(\theta|\sigma^2)$ since it behaves like a constant. Thus, we have:

$$p(\sigma^2|\mathbf{x},\theta) \propto f(\mathbf{x}|\theta,\sigma^2)p(\sigma^2)$$

$$\propto \prod_{i=1}^{n}\left[(2\pi)^{-1/2}(\sigma^2)^{-1/2}exp\left\{-\frac{1}{2\sigma^2}(x_i-\theta)^2\right\}\right]\frac{b^a\left(\frac{1}{\sigma^2}\right)^{a+1}e^{-b/\sigma^2}}{\Gamma(a)}$$

$$\propto (2\pi)^{-n/2}(\sigma^2)^{-n/2}exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2\right\}\frac{b^a\left(\frac{1}{\sigma^2}\right)^{a+1}e^{-b/\sigma^2}}{\Gamma(a)}$$

$$\propto \frac{(2\pi)^{-n/2}b^a}{\Gamma(a)}(\sigma^2)^{-n/2-a-1}exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2-\frac{b}{\sigma^2}\right\}$$

The constant term $\frac{(2\pi)^{-n/2}b^a}{\Gamma(a)}$ can be dropped, leaving

$$\propto (\sigma^2)^{-n/2-a-1}exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2-\frac{b}{\sigma^2}\right\}$$

$$\propto (\sigma^2)^{-n/2-a-1}exp\left\{-\frac{\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)^2+b}{\sigma^2}\right\}$$

The above equation matches the form of the Inverse Gamma pdf

$$x^{-\tilde{a}-1}exp{-\tilde{b}/x}$$

if we let

$$\tilde{a}=\frac{n}{2}+a \ , \qquad \tilde{b}=\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)^2+b$$

Thus, $\sigma^2|\mathbf{x}, \theta \sim IG(\tilde{a}, \tilde{b})$ or

$$\sigma^2|\mathbf{x}, \theta \sim IG\left(\frac{n}{2} + a, \frac{1}{2}\sum_{i=1}^{n}(x_i - \theta)^2 + b\right)$$

Part 2) We set the hyperparameters as follows:

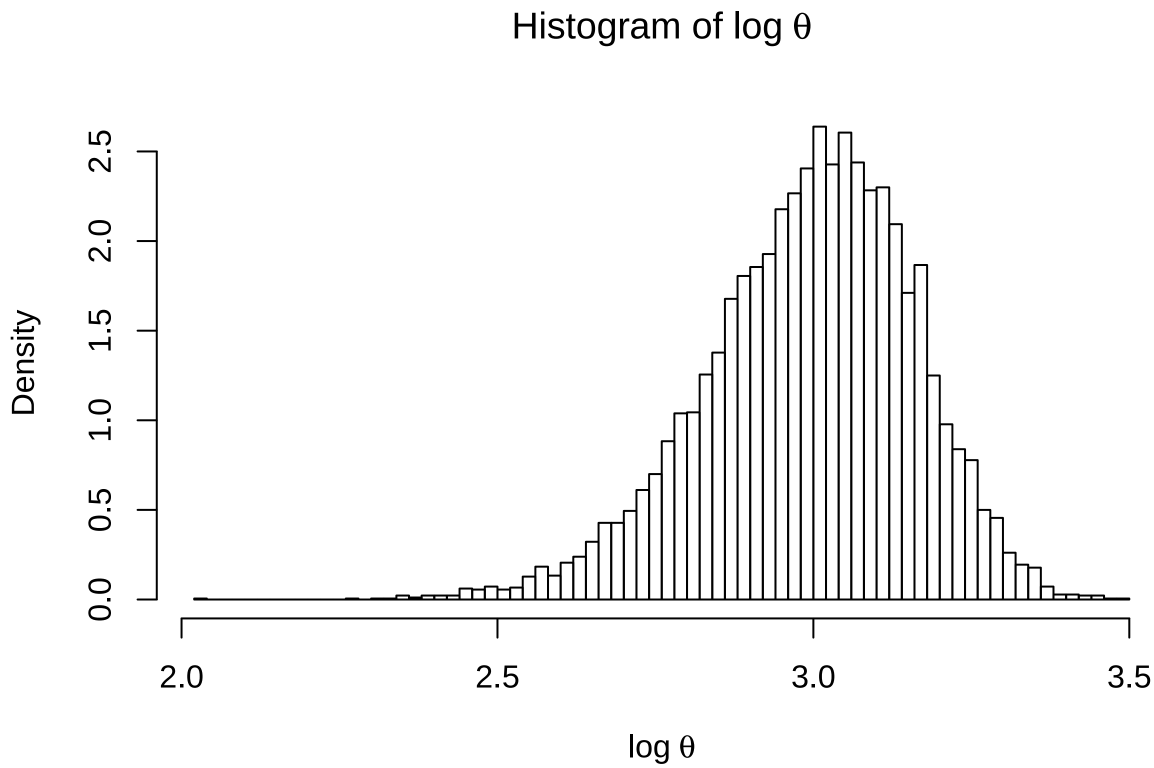
$$a = b = 3, \tau^2 = 10, \theta_0 = 20$$

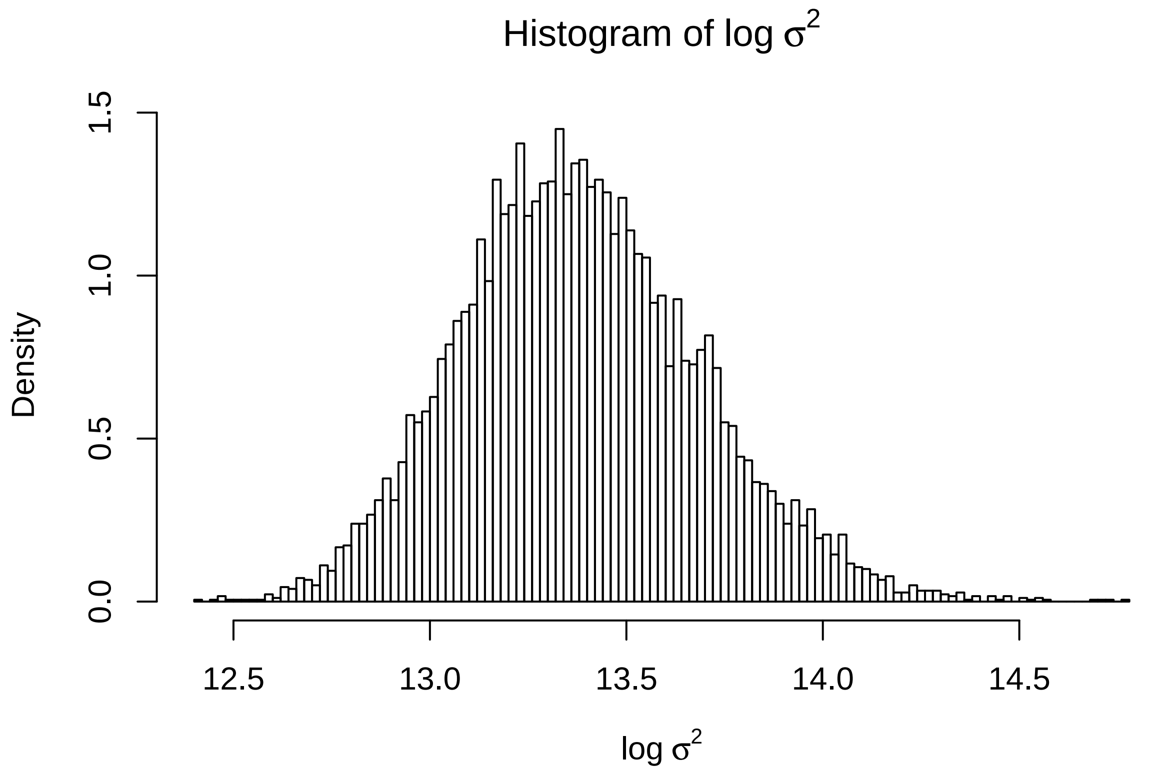
(θ_0 was set to 20 instead of 5 since setting θ_0 equal to 5 pulled the mean of the posterior distribution so low that negative values of θ were sometimes generated by the sampler. This resulted in non-real results for *log* θ). The implementation of the Gibbs sampler is shown below:

```r
library(invgamma)
gibbs <- function(M, data, theta, sigma2) {

  # summation function for calculating rate parameter
  summation <- function(data, theta) {
    sum <- 0
    for(j in 1:length(data)) {
      sum <- sum + (data[j] - theta)^2
    }
    return(sum)
  }
  # constants
  a <- 3
  b <- 3
  tau2 <- 10
  theta0 <- 20
  xbar <- mean(data)
  n <- length(data)

  for(i in 2:M+1000) {

    # sample theta conditioned on x and sigma2
    mean_theta <- ((sigma2[i-1]*theta0)/(sigma2[i-1] + n*tau2)) + ((n*tau2*xbar)/(sigma2[i-1] +
                                                                                      n*tau2))
    var_theta <- sigma2[i-1] * tau2 / (sigma2[i-1] + n*tau2)
    theta[i] <- rnorm(1, mean = mean_theta, sd = sqrt(var_theta))

    # sample sigma2 conditioned on x and theta
    shape <- n/2 + a
    rate <- (1/2) * summation(data, theta[i]) + b
    sigma2[i] <- rinvgamma(1, shape, rate)
  }

  # discard burn-in
  return(list(theta[1000:M+1000], sigma2[1000:M+1000]))
}
```

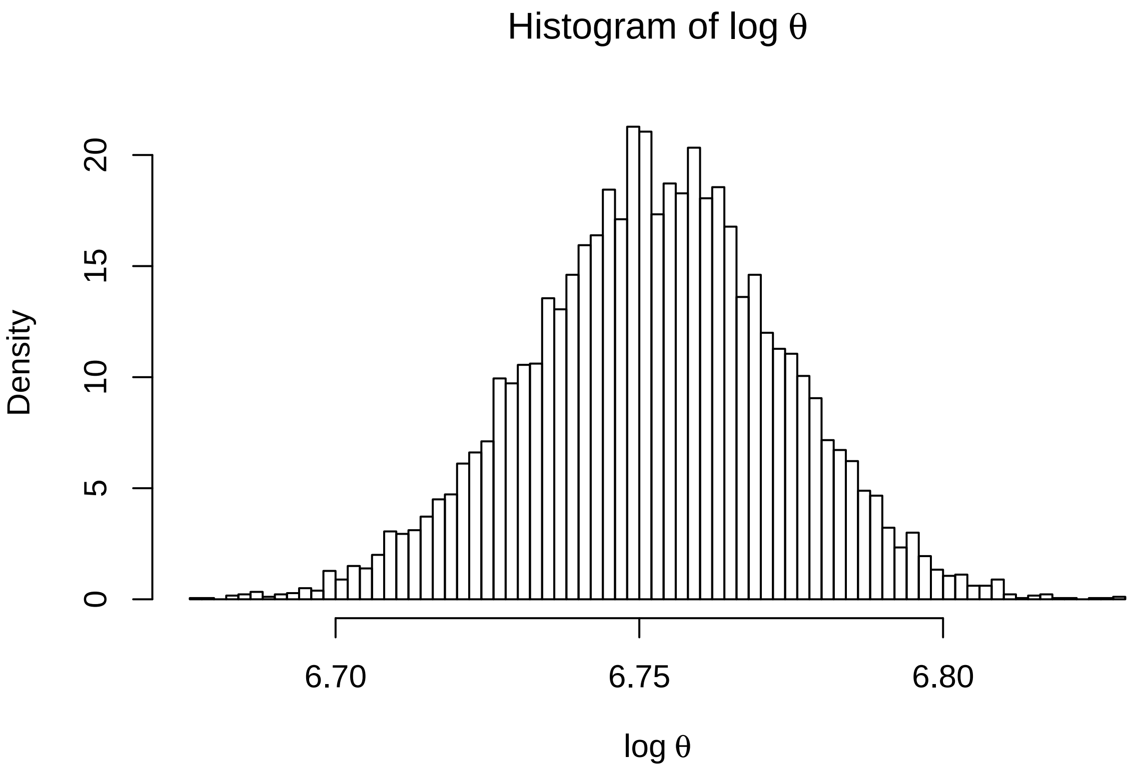
We then take 10000 samples of θ and σ^2 given x. We set our starting values for θ and σ^2 as the mean of x and the variance of x, respectively.

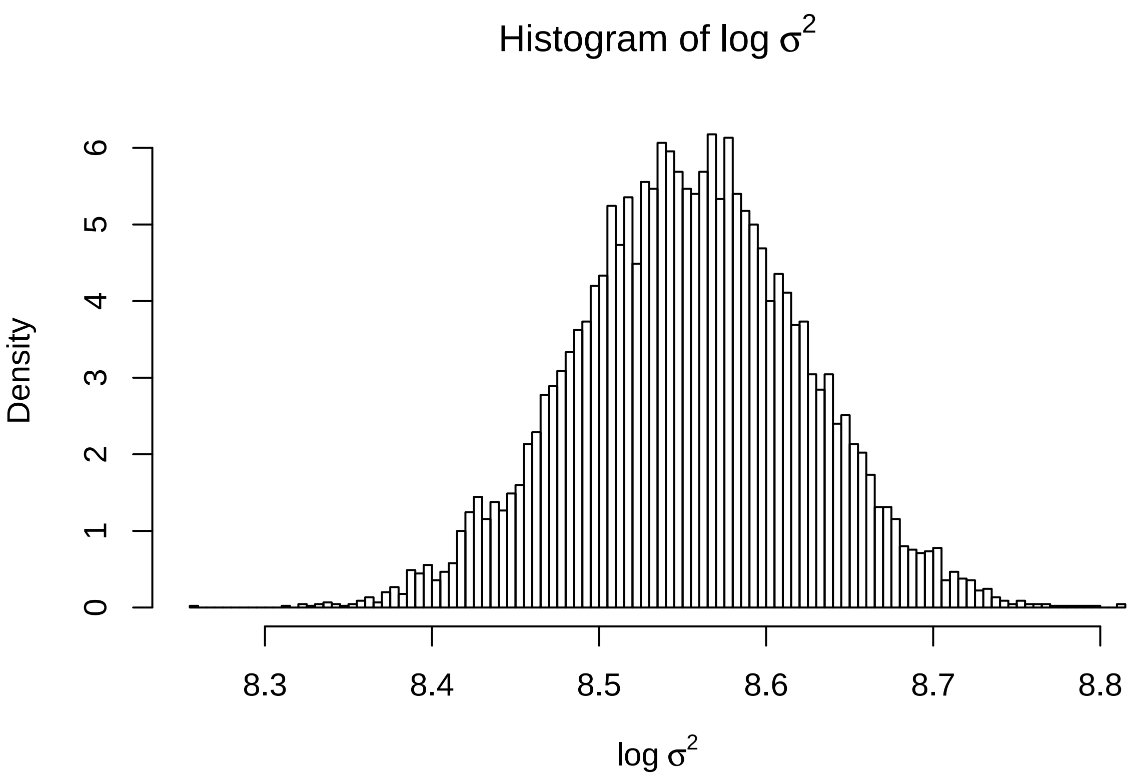
Part 3) We plot the histograms of both $log(\theta)$ and $log(\sigma^2)$ below, leaving out the first 1000 samples so that we only plot the values generated after the Gibbs sampler has stabilized and is no longer influenced by the initial values of θ and σ^2:





We note that the distributions of log θ and log σ^2 appear to be somewhat skewed, which seems to be a result of choosing values of the hyper-parameters of the prior distributions of θ and σ^2 that are much smaller than the typical values for X. If we increase τ^2 to 20000 and a to 200, we get histograms that look like the ones below, which appear much less skewed:





We report the 90% posterior probability intervals for our first samples of log θ and log σ^2 (the ones taken with small hyperparameters for the priors) below:

log θ:

```
##        5%      95%
## 2.700605 3.240775
```

log σ^2:

```
##        5%      95%
## 12.90928 13.92318
```

We also report the 90% posterior probabilility intervals for our second samples of log θ and log σ^2 (the ones taken with large hyperparameters):

log θ:

```
##        5%      95%
## 6.717421 6.786622
```

log σ^2:

```
##        5%      95%
## 8.438712 8.666029
```

Clearly, the hyperparameters of the priors can have a large effect on the posterior distribution, since the probability intervals do not even overlap.

Part 4) The fact that X is not symmetric and has large values means that it also has a very high variance. Thus, a prior distribution for the **mean** that has a very low variance will have a much larger influence on the posterior distribution than X—on the other hand, a prior distribution for the **variance** that has small hyper-parameters will have a much smaller influence on the posterior distribution than X. The choice of hyper-parameters ultimately depends on how much influence we desire for X to have on the final model, and how much influence we desire for the prior distributions to have on the final model. The trends mentioned above can be observed in the following plots:

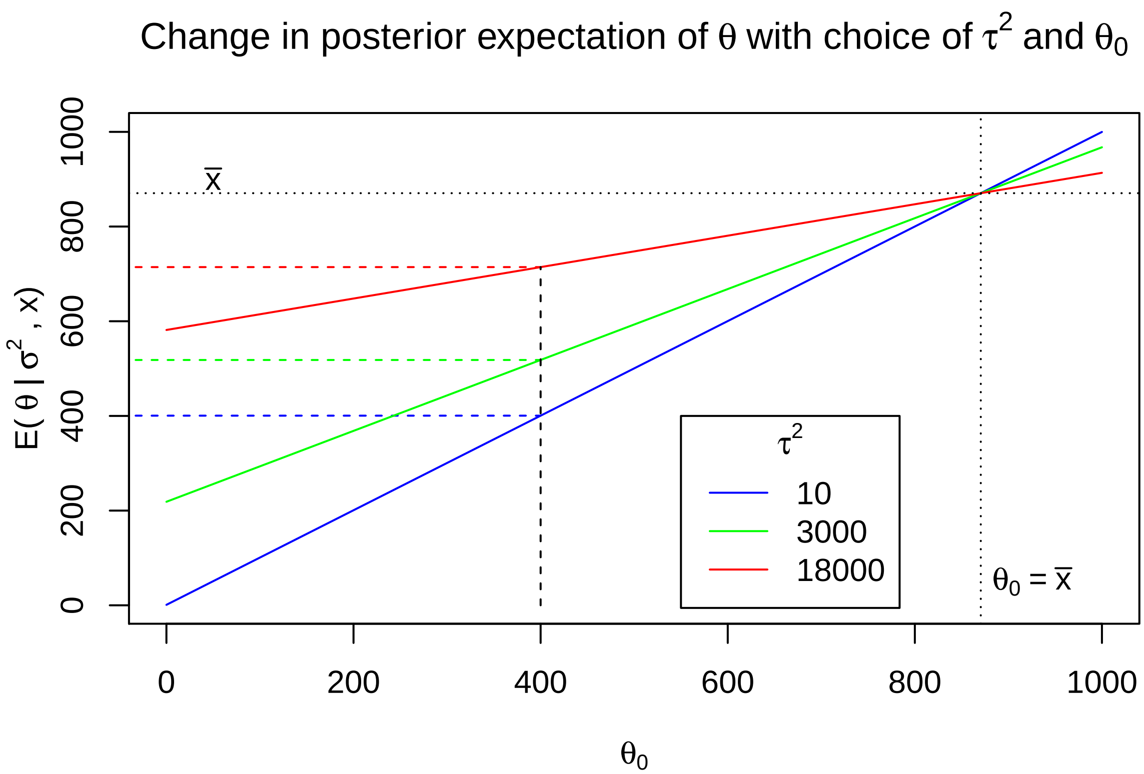


Figure 1: The expectation of the posterior distribution draws closer to the mean of the data as the variance of the prior distribution increases, as demonstrated when the expectation of the prior distribution equals 400.

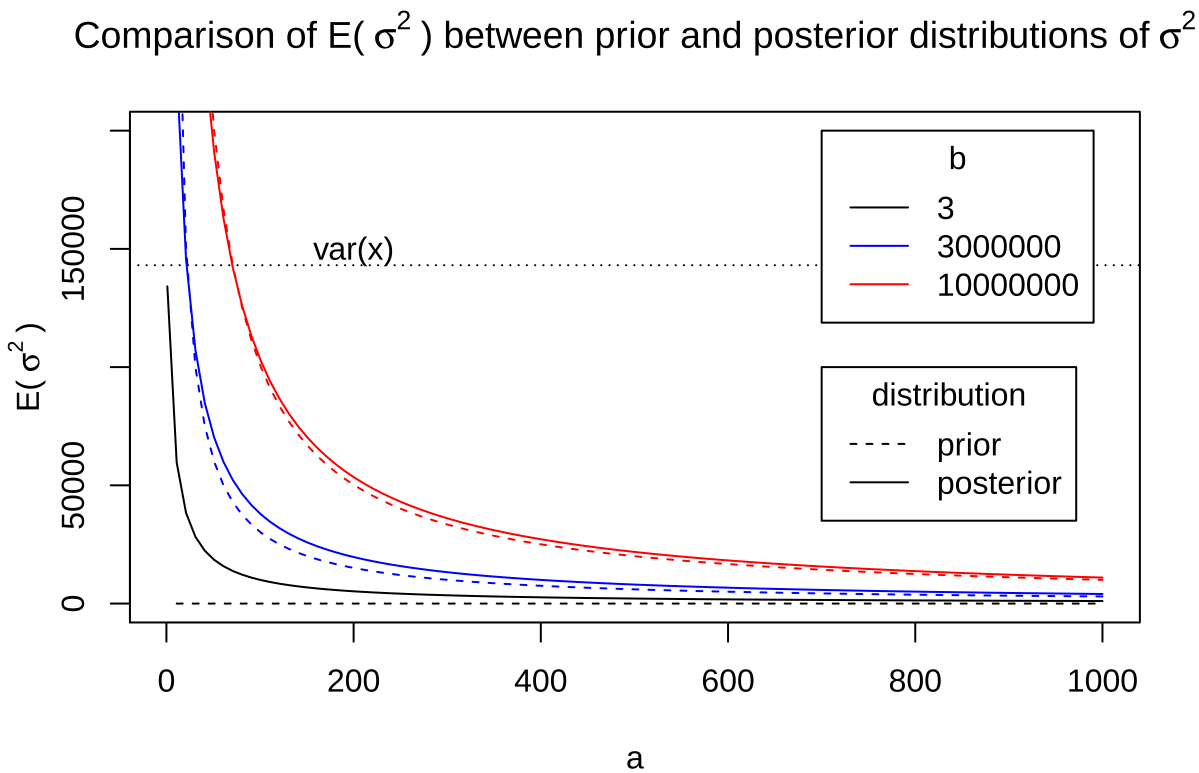


Figure 2: Using the expectation of each distribution as a benchmark, it is evident that a posterior distribution with small values of hyper-parameters a and b is more strongly influenced by x than by the prior distribution. As either a or b increases, the posterior distribution resembles the prior distribution more and more closely.