

# Advanced Macroeconometrics - Assignment 2

Group I

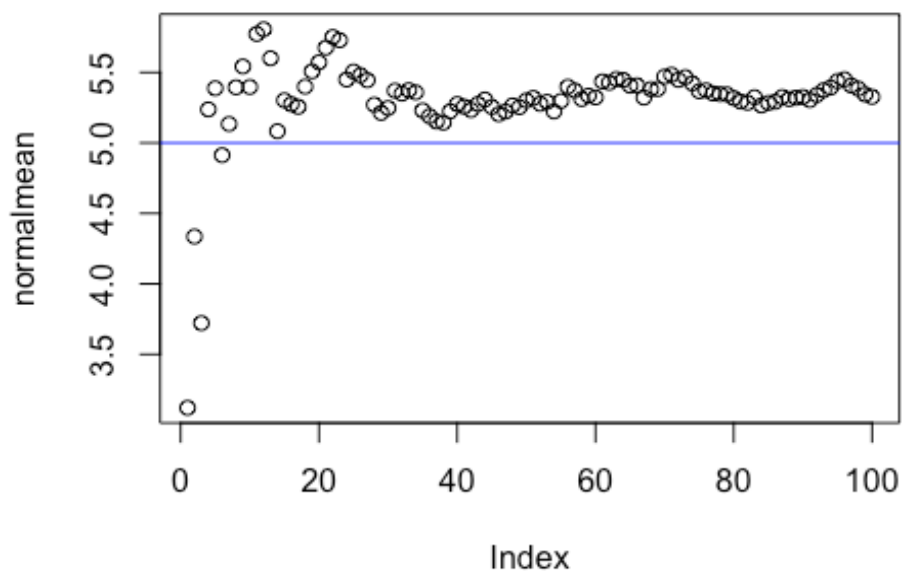
2023-05-10

## Exercise 1

1.1 Simulate  $n = 100$  draws from a Normal,  $N(5, 9)$ , distribution (using `rnorm`). Estimate the mean with the first 1,  $\dots$ ,  $n$  (very  $n$ ) draws. Discuss and visualise convergence of the estimates.

We simulate 100 draws from a  $N(5,9)$ . Then for each entry we calculate the mean up to that point and plot this data.

```
set.seed(1)
normaldraw <- rnorm(100, 5, 3)
normalmean <- c(rep(0,100))
for (i in 1:100) {
  normalmean[i] <- sum(normaldraw[1:i])/i
}
plot( normalmean)
abline(h = 5, col = "blue")
```

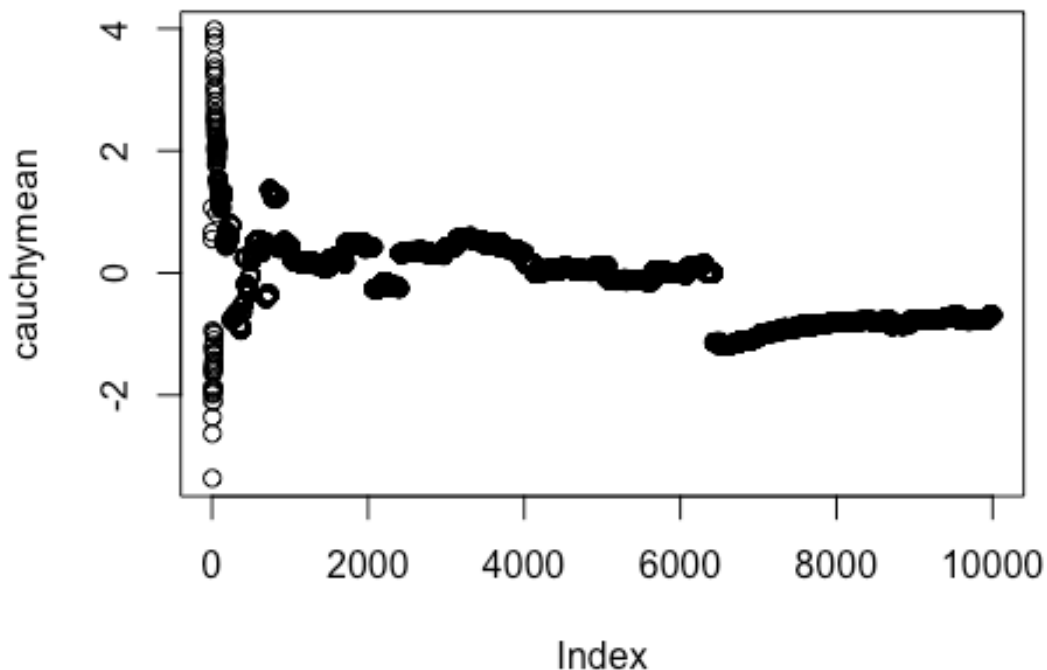


As you can see in the plot, the mean tends to converge to the mean of the normal distribution, i.e., 5.

1.2 Simulate  $n = 10000$  draws from a Cauchy distribution with scale one by drawing from  $N(0,1)/N(0,1)$ . Estimate the mean with the first  $1, \dots, n$  draws. Discuss and visualise convergence of the estimates.

We simulate 10.000 draws from a Cauchy distribution with scale one by drawing from a  $N(0,1)/N(0,1)$ . Then for each entry we calculate the mean up to that point and plot this data.

```
set.seed(1234)
norm1draw <- rnorm(10000, 0, 1)
norm2draw <- rnorm(10000, 0, 1)
cauchydraw <- c(rep(0,10000))
for (i in 1:10000) {
  cauchydraw[i] <- norm1draw[i]/norm2draw[i]
}
cauchymeans <- c(rep(0,10000))
for (i in 1:10000) {
  cauchymeans[i] <- sum(cauchydraw[1:i])/i
}
plot(cauchymeans)
```



The Cauchy-distribution has no mean, this can be seen in the plot as there is no convergence to one value but sudden jumps happening on numerous occasions.

## Exercise 2

You have observations on daily Alles Gurgelt tests in your office —  $\mathbf{y} = (y_1, \dots, y_n)$  — and want to learn about the prevalence. There are 20 colleagues who test everyday. Assume that the data is independent and identically distributed.

2.1 What is the class of conjugate priors for this problem? Derive the posterior distribution  $p(\theta|\mathbf{y})$ .

We understood the prevalence as the probability that one test is positive on a given day. The number of positive Alles Gurgelt tests out of the 20 colleagues can be modeled using a binomial distribution with probability of success  $\theta$ , which represents the prevalence. Assuming a conjugate prior for  $\theta$ , the posterior distribution will also be a binomial distribution, and the conjugate prior for the binomial distribution is the beta distribution. Therefore, the class of conjugate priors for this problem is the beta distribution.

Let  $\theta \sim \text{Beta}(\alpha, \beta)$  be the prior distribution for  $\theta$ . The likelihood function for the observed data  $y$  is given by:

$$p(y | \theta) = B(y + 1, n - y + 1)\theta^y(1 - \theta)^{(n-y)}$$

where  $B$  is the beta function and  $n$  is the total number of colleagues tested.

Then, by applying Bayes' theorem, the posterior distribution of  $\theta$  given the data  $y$  is:

$$\begin{aligned} p(y | \theta) &\propto p(y | \theta)p(\theta) \\ &\propto B(y + 1, n - y + 1)\theta^y(1 - \theta)^{(n-y)}\theta^{(\alpha-1)}(1 - \theta)^{(\beta-1)} \\ &\propto \theta^{(y+\alpha-1)}(1 - \theta)^{(n-y+\beta-1)} \end{aligned}$$

The last expression is proportional to the probability density function of a beta distribution with parameters  $y + \alpha$  and  $n - y + \beta$ . Therefore, the posterior distribution of  $\theta$  given the data  $y$  is:

$$\theta | y \sim \text{Beta}(y + \alpha, n - y + \beta)$$

This is the desired posterior distribution for  $\theta$ , which represents our updated knowledge about the prevalence of Alles Gurgelt tests after observing the data  $y$ .

2.2 Assume you have observations for thirty days ( $n = 30$ ) with a total of ten positive test ( $y = 10$ ). Determine and briefly explain several point estimators of  $\theta$ .

As we have already derived the posterior distribution of  $\theta$  given the data in part 1 of this question, which is:

$$\theta | y \sim \text{Beta}(y + \alpha, n - y + \beta)$$

We can obtain the point estimators of  $\theta$  by computing the mean or mode of the posterior distribution.

$$E(\theta | y) = \frac{(y + \alpha)}{(n + \alpha + \beta)} = \frac{(10 + 2)}{(30 + 2 + 2)} = 0.4$$

$$\text{Mode}(\theta | y) = \frac{(y + \alpha - 1)}{(n + \alpha + \beta - 2)} = \frac{(10 + 2 - 1)}{(30 + 2 + 2 - 2)} = 0.364$$

The Bayesian estimator takes into account our prior knowledge about the prevalence of positive tests, as well as the information provided by the data. In this case, we assumed a prior distribution with Hyperparameters  $\alpha = 2$  and  $\beta = 2$ , in order to have a weakly informative prior that is centered at 0.5. The posterior distribution shifts our knowledge towards the data, but the effect of the prior can still be seen in the Bayesian estimator.

### 2.3 Discuss sources of prior information for this problem and compare the impact of different priors on your point estimates.

Prior information can be obtained from various sources such as previous studies, expert opinions, historical data, and theoretical considerations. For the problem of estimating the prevalence of a disease based on test results, we can consider data published by national authorities and statistic institutes such as Statistik Austria.

The impact of different priors on the point estimates can be significant. If the prior is informative, it can significantly affect the posterior distribution and hence the point estimate. For example, if the prior distribution is centered around a particular value, the posterior distribution is likely to be more peaked around that value. On the other hand, if the prior is uninformative, the posterior distribution is likely to be more influenced by the data.

To illustrate the impact of different priors, we can consider the example we discussed earlier. Let's assume that we have the same data ( $y=10$ ,  $n=30$ ) but we use different prior distributions. Specifically, we consider the following three prior distributions:

```
# An informative prior based on a previous study that estimated the prevalence to be 0.2:
```

```
alpha1 <- 4
beta1 <- 16
```

```
# An uninformative prior based on a Beta distribution with parameters (1, 1):
```

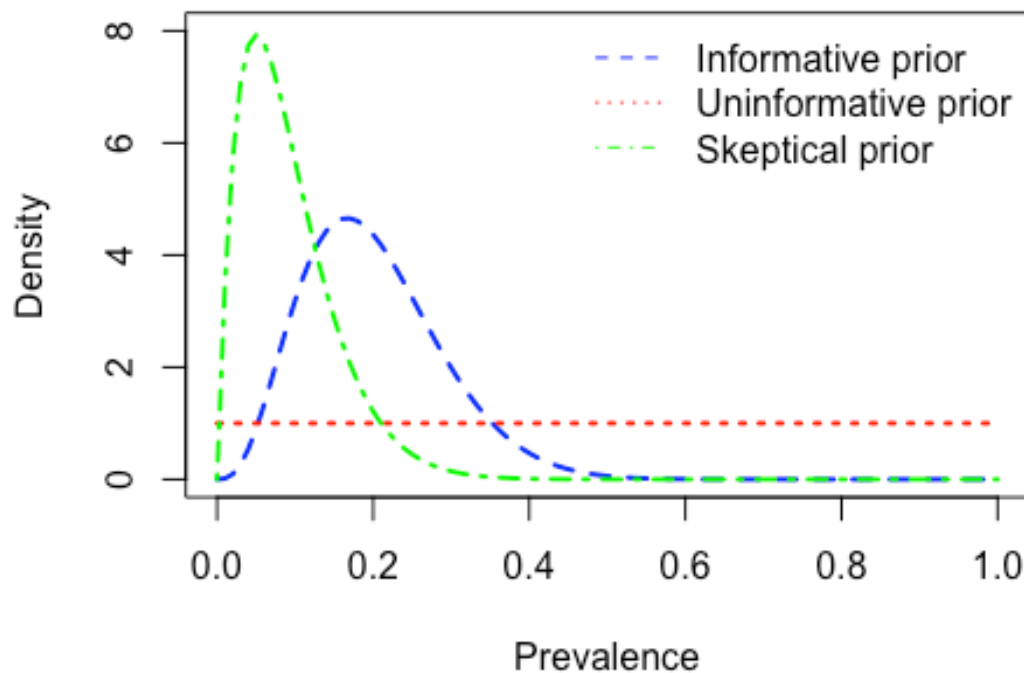
```
alpha2 <- 1
beta2 <- 1
```

```
# A skeptical prior based on a Beta distribution with parameters (2, 20):
```

```
alpha3 <- 2
beta3 <- 20
```

```
# The three priors are plotted below:
```

```
curve(dbeta(x, alpha1, beta1), from=0, to=1, lty=2, lwd=2, col="blue", ylab="Density", xlab="Prevalence", ylim = c(0,8))
curve(dbeta(x, alpha2, beta2), from=0, to=1, lty=3, lwd=2, col="red", add=TRUE)
curve(dbeta(x, alpha3, beta3), from=0, to=1, lty=4, lwd=2, col="green", add=TRUE)
legend("topright", legend=c("Informative prior", "Uninformative prior", "Skeptical prior"), lty=c(2, 3, 4), col=c("blue", "red", "green"), bty="n")
```



## 2.4 Discuss the assumption of independent and identically distributed data. How could you (conceptually) improve the model with this in mind?

The assumption of i.i.d. data is a common assumption in many statistical models, including the Bayesian model for estimating the prevalence of a disease based on test results. This assumption implies that each observation is independent of the others and that they all come from the same underlying distribution. In the context of this problem, it means that the probability of a positive test result for each day is independent of the results of the others.

While the i.i.d. assumption simplifies the modeling process and allows for the use of standard statistical techniques, it may not always be a reasonable assumption in practice. In the context of disease prevalence estimation, there are several reasons why the i.i.d. assumption may be violated:

\*Clustering: The colleagues may be clustered in some way (e.g., they work in the same department or are located in the same office), which may affect the probability of a positive test result. In this case, the i.i.d. assumption is not appropriate, and alternative modeling techniques (e.g., hierarchical modeling) may be required.

\*Time dependence: The probability of a positive test result may change over time due to changes in the prevalence of the disease or changes in testing protocols. In this case, the i.i.d. assumption is not appropriate, and time-series models may be more appropriate.

Conceptually, one way to improve the model while relaxing the i.i.d. assumption is to consider a more flexible model that allows for dependence between the observations. One possible approach is to use a hierarchical model, which allows for clustering of the observations and allows the

probability of a positive test result to vary across clusters. Another approach is to use a time-series model, which allows for time dependence in the probability of a positive test result.

### Exercise 3

Write an R function to simulate  $n$  observations from the model  $\mathbf{y} = \alpha + \mathbf{X}\beta + \mathbf{e}$ . Draw the  $k$  independent variables from distributions of your choice, and the error from a Normal with mean zero and standard deviation  $\sigma$ . The function should have arguments to set  $n$ ,  $k$ ,  $\alpha$ ,  $\beta$ , and  $\sigma$ ; it should return a list with the simulated data,  $\mathbf{y}$  and  $\mathbf{X}$ .

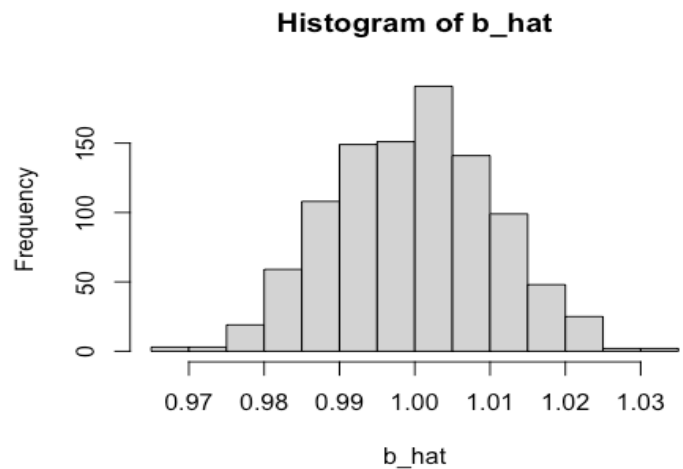
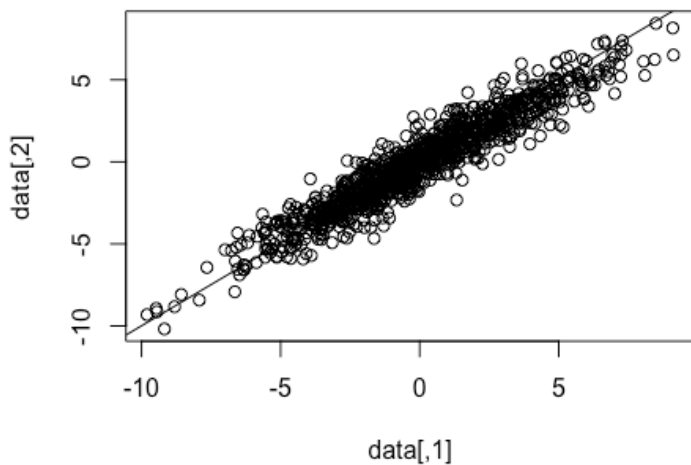
```
#Exercise 3
simulatedata <- function(n,k,alpha,beta,sigma) { #beta can be a vector
  X <- matrix(rep(c(rep(0,n)),k),byrow =F,ncol=k) #k must correspond to the number of entries in beta
  for(i in 1:k) {
    x <- rnorm(n,0,3)
    X[,i] <- x
  }
  e <- c(rnorm(n,0,sigma))
  a <- c((rep(alpha,n)))
  b <- as.vector(beta)
  y = a + X%*%b + e
  yX <- matrix(c(y,X),byrow=F, ncol=k+1)
  return(yX)
}
```

3.1 Simulate data with  $k = 1$  and  $\sigma = 1$ . Plot the regressor  $\mathbf{x}$  and regressand  $\mathbf{y}$  in a scatterplot; add a LS regression line. Repeat this 1,000 times and store  $\hat{\beta}_{LS}$  every time. Then create a histogram of the LS estimates — what do you see?

```
#Exercise 3a
#Simulating data
data <- simulatedata(1000,1,0,1,1) #n=1000, k=1, alpha=0, beta=1, sigma=1

#Plotting data and ls regression line
plot(data)
abline(lm(data[,1] ~ data[,2]))

#Repeating 1,000 times, storing b_hat and plotting
b_hat <- c(rep(0,1000))
for(i in 1:1000) {
  data <- simulatedata(1000,1,0,1,1)
  b_hat[i]<-lm(data[,1] ~ data[,2])$coefficients[2]
}
hist(b_hat)
```



It seems as if the estimates of beta are normally distributed around the true value of beta.

3.2 Assume you know that  $\sigma^2 = 1$ . What are the latent values of the model?

If we know sigma, the only latent parameter is beta.

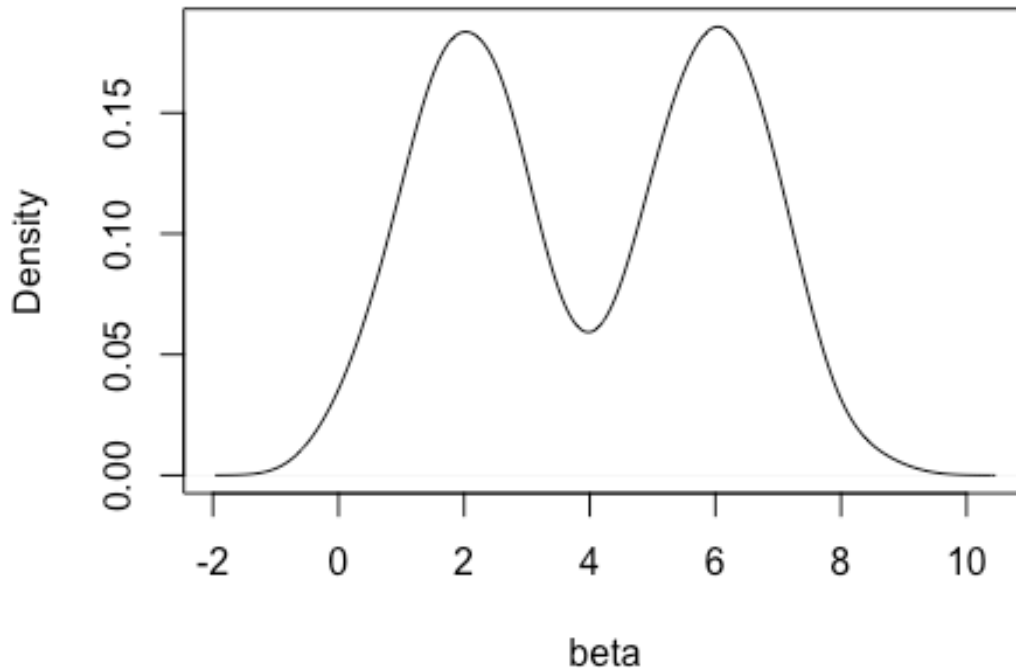
3.3 Come up with a potentially interesting regression you want to run. Explain and draw ways you expect a single coefficient of interest,  $\beta_j$ , to look like a priori.

An interesting regression would be to regress the attendance of university lectures on a Dummy-Variable indicating whether lecture attendance is mandatory or not. One could imagine that some people would attend lectures anyway, for them the corresponding effect of mandatory attendance on actual attendance would be small. Other people might never attend a lecture without mandatory attendance, for them the effect of mandatory attendance would be large. Assuming that most people are in one of these two extreme categories, the a priori distribution of the coefficient  $\beta_j$  would have two local maxima and could look like this:

```
# Generate random data from two normal distributions
x1 <- rnorm(1000, mean = 2, sd = 1)
x2 <- rnorm(1000, mean = 6, sd = 1)

# Combine the data
x <- c(x1, x2)
xdens <- density(x)
# Plot the density
plot(xdens, main = "a priori distribution of beta", xlab="beta")
```

### a priori distribution of beta



3.4 Simulate data with  $k = 1$  and  $\sigma = 1$  — you can assume you know  $\sigma$ . Set a Normal prior,  $N(\mu_0, \sigma_0)$ , for  $\beta$  — decide on parameters  $\mu_0$  and  $\sigma_0$  for this prior. Compute and plot the posterior density for simulated data with increasing  $n$  (e.g.  $n \in \{50, 100, 200\}$ ).

First we decide on the Normal prior for beta:

```
mu_0 <- 5 #Parameters for Normal prior for beta
sigma_0 <- 1
```

Then we are simulating data. After that we are calculating and plotting the posterior together with the prior. We do this for  $n=50$ ,  $n=100$  and  $n=200$ .

```
#n=50
set.seed(123)
data50 <- simulatedata(50,1,0,rnorm(1,mu_0,sigma_0^0.5),1) #Simulating data for
#or n=50 with beta some value from N(0,1)
X_50 <- data50[,2]
y_50 <- data50[,1]
sigma_50 <- 1/((1/sigma_0)+ (t(X_50)%*%X_50)) #applying formula for sigma_50
mu_50 <- sigma_50 *((1/sigma_0)*mu_0 + (t(X_50)%*%y_50))#applying formula for
mu_50

# Plotting prior and posterior density
# Generating prior and posterior density
```



```

prior_50 <- rnorm(10000, mean = mu_0, sd = sigma_0^0.5)
posterior_50 <- rnorm(10000, mean = mu_50, sd = sigma_50^0.5)

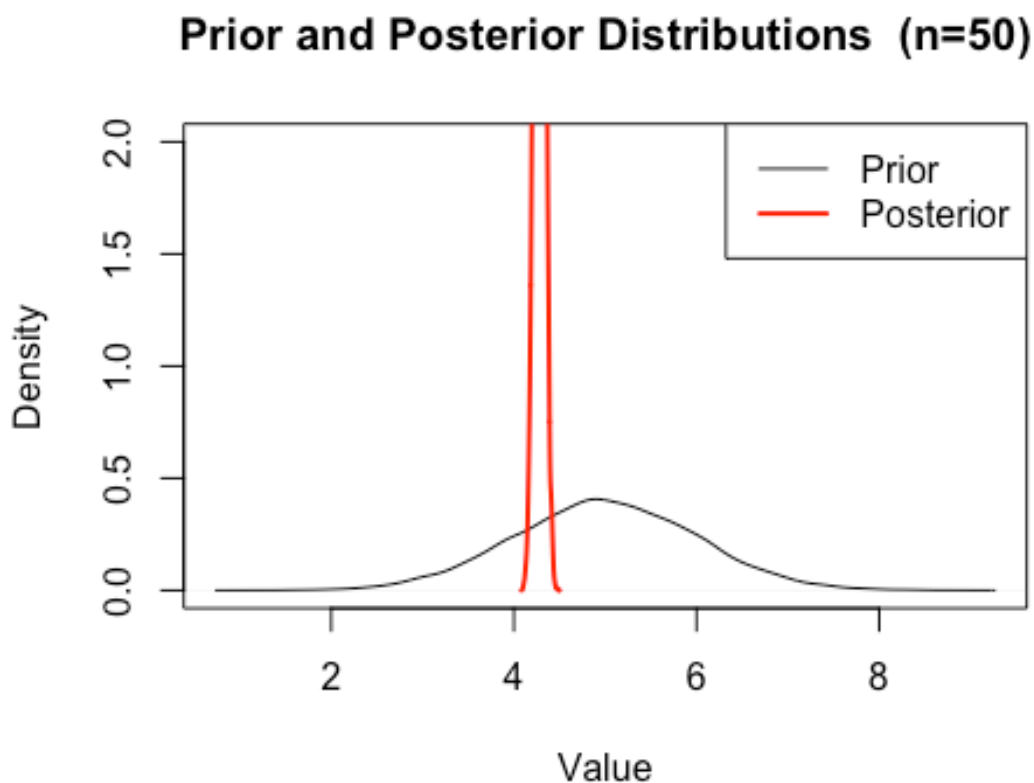
# Calculate the density estimates
prior_density_50 <- density(prior_50)
posterior_density_50 <- density(posterior_50)

# Plot the prior density
plot(prior_density_50, main="Prior and Posterior Distributions (n=50)", xlab="Value", ylab="Density", ylim=c(0,2))

# Plot the posterior density
lines(posterior_density_50$x, posterior_density_50$y, col="red", lwd=2)

# Add a Legend
legend("topright", legend=c("Prior", "Posterior"), lty=c(1,1), col=c("black", "red"), lwd=c(1,2))

```



```

#n=100
data100 <- simulatedata(100,1,0,rnorm(1,mu_0,sigma_0^0.5),1) #n=100
X_100 <- data100[,2]
y_100 <- data100[,1]
sigma_100 <- 1/((1/sigma_0)+ (t(X_100)%*%X_100)) #applying formula for sigma_

```

```

100
mu_100 <- sigma_100 * ((1/sigma_0)*mu_0 + (t(X_100)%*%y_100)) #applying formul
a for mu_100

#Plotting prior and posterior density
# Generating prior and posterior density
prior_100 <- rnorm(10000, mean = mu_0, sd = sigma_0^0.5)
posterior_100 <- rnorm(10000, mean = mu_100, sd = sigma_100^0.5)

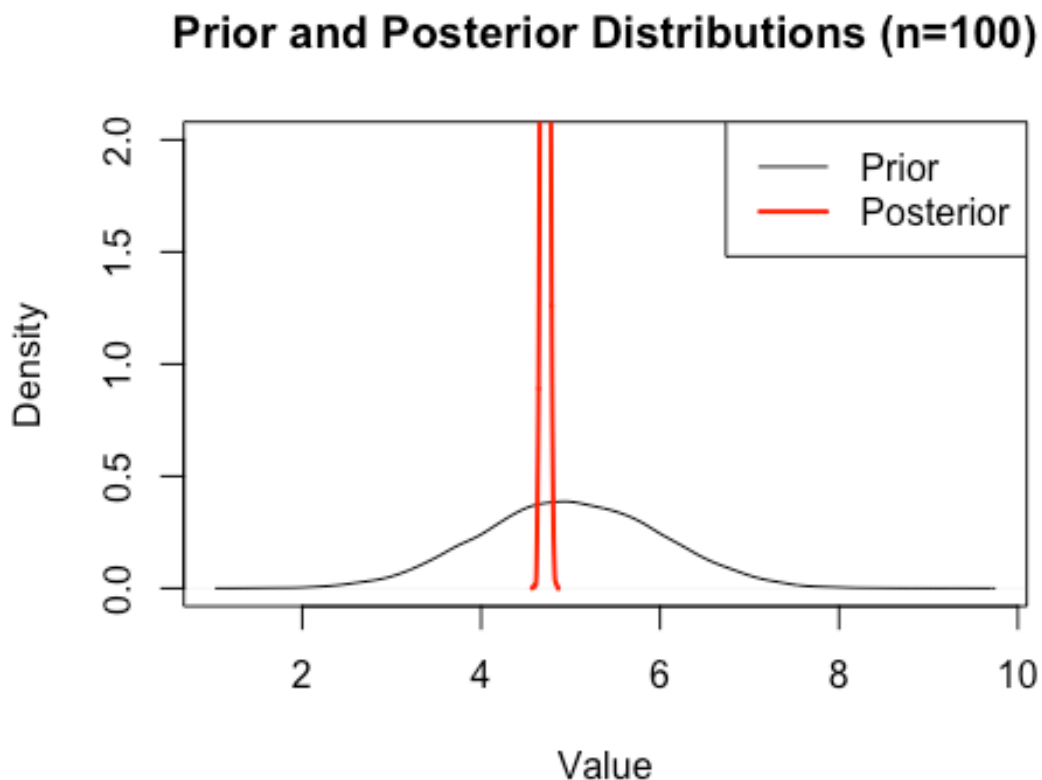
# Calculate the density estimates
prior_density_100 <- density(prior_100)
posterior_density_100 <- density(posterior_100)

# Plot the prior density
plot(prior_density_100, main="Prior and Posterior Distributions (n=100)", xla
b="Value", ylab="Density", ylim=c(0,2))

# Plot the posterior density
lines(posterior_density_100$x, posterior_density_100$y, col="red", lwd=2)

# Add a Legend
legend("topright", legend=c("Prior", "Posterior"), lty=c(1,1), col=c("black",
"red"), lwd=c(1,2))

```



```

#n=200
data200 <- simulatedata(200,1,0,rnorm(1,mu_0,sigma_0^0.5),1) #n=200
X_200 <- data200[,2]
y_200 <- data200[,1]
sigma_200 <- 1/((1/sigma_0)+ (t(X_200)%*%X_200)) #applying formula for sigma_
100
mu_200 <- sigma_200 *((1/sigma_0)*mu_0 + (t(X_200)%*%y_200)) #applying formul
a for mu_100

#Plotting prior and posterior density
# Generating prior and posterior density
prior_200 <- rnorm(10000, mean = mu_0, sd = sigma_0^0.5)
posterior_200 <- rnorm(10000, mean = mu_200, sd = sigma_200^0.5)

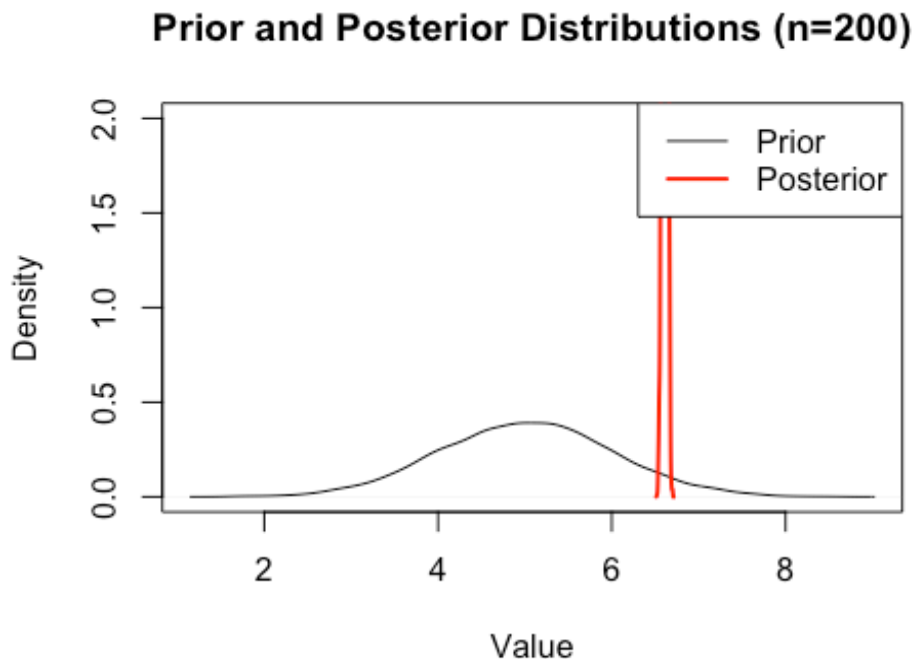
# Calculate the density estimates
prior_density_200 <- density(prior_200)
posterior_density_200 <- density(posterior_200)

# Plot the prior density
plot(prior_density_200, main="Prior and Posterior Distributions (n=200)", xla
b="Value", ylab="Density", ylim=c(0,2))

# Plot the posterior density
lines(posterior_density_200$x, posterior_density_200$y, col="red", lwd=2)

# Add a Legend
legend("topright", legend=c("Prior", "Posterior"), lty=c(1,1), col=c("black",
"red"), lwd=c(1,2))

```



From the plots we can see that, depending on the randomly drawn beta from the Normal prior distribution, the posterior distribution will concentrate around that value for beta. As expected, with increasing observations the distribution becomes narrower, i.e.,  $\sigma_n$  decreases in  $n$ .

## Exercise 4

4.1. Suppose you have data  $\mathbf{y} \sim N(\mu, 1)$ , and want to estimate  $\mu$ . Specify a Normal prior  $\mu \sim N(\mu_0, \sigma_0^2)$ . Derive the posterior  $p(\mu | \mathbf{y})$  by applying Bayes' theorem. Create histograms of two priors of your choice.

First, we specify a Normal prior for  $\mu \sim N(\mu_0, \sigma_0^2)$ . As our data,  $\mathbf{y} = (y_1, \dots, y_n)$ , is assumed to be i.i.d, the likelihood is  $p(\mathbf{y}|\mu) = p(y_1|\mu), \dots, p(y_n|\mu)$ . Hence, we get for the joint density function:

$$p(\mathbf{y}|\mu) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(y_i - \mu)^2}{2\sigma^2}\right)$$

By the decomposition of the variance we get:

$$p(\mathbf{y}|\mu) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(\mu - \bar{y})^2}{2\frac{\sigma^2}{n}}\right)$$

Moreover, we assume a Normal prior:

$$p(\mu) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(\frac{-(\mu - \mu_0)^2}{2\sigma_0^2}\right)$$

Now, we can derive the posterior using Bayes' theorem. We can drop all constant factors from the likelihood function and use the fact that  $\sigma^2 = 1$

$$\begin{aligned} p(\mu|\mathbf{y}) &\propto p(\mathbf{y}|\mu)p(\mu) \\ &\propto \exp\left(\frac{-(\mu - \bar{y})^2}{\frac{2}{n}} + \frac{-(\mu - \mu_0)^2}{2\sigma_0^2}\right) \\ &= \exp\left(-\frac{n}{2}(\mu^2 - 2\mu\bar{y} - \bar{y}^2) - \frac{1}{2\sigma_0^2}(\mu^2 - 2\mu\mu_0 + \mu_0^2)\right) \\ &= \exp\left(-\frac{1}{2}\mu^2\left(n + \frac{1}{\sigma_0^2}\right) + \mu\left(\bar{y}n + \frac{\mu_0}{\sigma_0^2}\right) - \frac{\mu_0^2}{2\sigma_0^2} - \frac{n\bar{y}^2}{2}\right) \end{aligned}$$

Now we can match the powers of  $\mu$ :

$$\frac{\mu^2}{2\sigma_n^2} = -\frac{1}{2}\mu^2\left(n + \frac{1}{\sigma_0^2}\right)$$

Hence, the variance of the posterior distribution is:

$$\sigma_n^2 = \frac{\sigma_0^2}{1 + \sigma_0^2 n}$$

Same for  $\mu_n$ :

$$\frac{2\mu\mu_n}{2\sigma_n^2} = \mu \left( \bar{y}_n + \frac{\mu_0}{\sigma_0^2} \right)$$

Hence, the mean of the posterior distribution is:

$$\mu_n = \sigma_n^2 \left( \frac{\bar{y}_n \sigma_0^2 + \mu_0}{\sigma_0^2} \right)$$

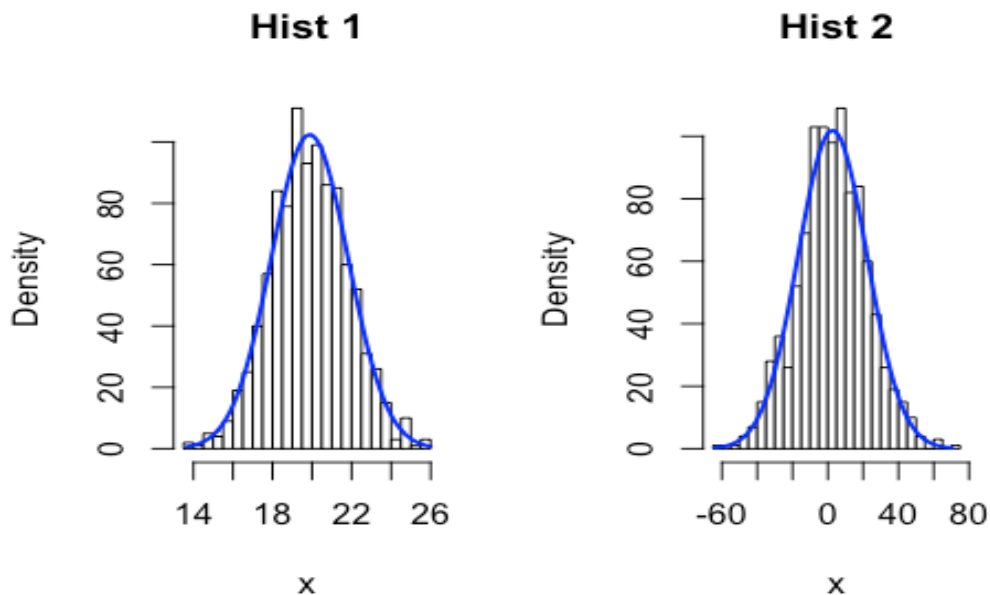
The prior distribution is:

$$\mu_n \sim N \left( \sigma_n^2 \left( \frac{\bar{y}_n \sigma_0^2 + \mu_0}{\sigma_0^2} \right), \frac{\sigma_0^2}{1 + \sigma_0^2 n} \right)$$

```
x1 = rnorm(1000, mean = 20, sd = 2)
x2 = rnorm(1000, mean = 2, sd = 20)

par(mfrow=c(1,2))
plotNormalHistogram(x1, col = "white",
                    border = "black", main = "Hist 1",
                    ylab = "Density", breaks = 20)

plotNormalHistogram(x2, col = "white",
                    border = "black", main = "Hist 2",
                    ylab = "Density", breaks = 20)
```



4.2 Suppose you have data  $\mathbf{y} \sim N(5, \sigma^2)$ , and want to estimate  $\sigma^2$ . Work with the precision,  $\sigma^{-2}$ , and specify a Gamma prior  $\sigma^{-2} \sim G(0.5, \eta)$  with single parameter  $\eta$ . Derive the posterior  $p(\sigma^2 | \mathbf{y})$  by applying Bayes' theorem. visualise the prior density for  $\eta \in \{0.01, 1, 100\}$ .

The likelihood function is given by:

$$\begin{aligned} p(\mathbf{y} | \mu, \sigma^2) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(y_i - \mu)^2}{2\sigma^2}\right) \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^{\frac{n}{2}} \exp\left(\frac{-(\sum_{i=1}^n (y_i - \mu)^2)}{2\sigma^2}\right) \end{aligned}$$

The Gamma prior is:

$$p(\sigma^{-2}) = \frac{\beta^\alpha}{\Gamma(\alpha)} (\sigma^{-2})^{\alpha-1} \exp(-\beta\sigma^{-2})$$

Now, we can derive the posterior using Bayes' theorem and dropping all constant factors:

$$\begin{aligned} p(\sigma^2 | \mathbf{y}) &\propto p(\mathbf{y} | \sigma^2) p(\sigma^2) \\ p(\sigma_n^{-2} | \mathbf{y}, \sigma^{-2}) &\propto p(\mathbf{y} | \sigma^{-2}) p(\sigma^{-2}) \\ p(\sigma_n^{-2} | \mathbf{y}, \sigma^{-2}) &\propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} \exp\left(-(\sigma^{-2}) \frac{-(\sum_{i=1}^n (y_i - \mu)^2)}{2}\right) * (\sigma^{-2})^{\alpha-1} \exp(-\beta\sigma^{-2}) \\ &= (\sigma^{-2})^{\alpha + \frac{n}{2} - 1} \exp\left(-\left(\beta + \frac{-(\sum_{i=1}^n (y_i - \mu)^2)}{2}\right) \sigma^{-2}\right) \\ &\propto G\left(\alpha + \frac{n}{2}, \beta + \frac{(\sum_{i=1}^n (y_i - \mu)^2)}{2}\right) \end{aligned}$$

where  $\alpha_n = \alpha + \frac{n}{2}$  and  $\beta_n = \frac{\sum_{i=1}^n (y_i - \mu)^2}{2}$

As we know that  $\frac{1}{x} \sim G(\alpha, \beta)$ , we have that  $x \sim IG(\alpha, \beta)$ . Hence, the posterior distribution for  $\sigma^2$  is:

$$p(\sigma^2 | \mathbf{y}) \propto IG(\alpha_n, \beta_n)$$

Now, we can visualise the prior density.

```
## Visualise the prior density for  $\eta \in \{0.01, 1, 100\}$ .
# Define the range of sigma_2 values to plot
sigma2_values <- seq(0, 20, length.out = 1000)

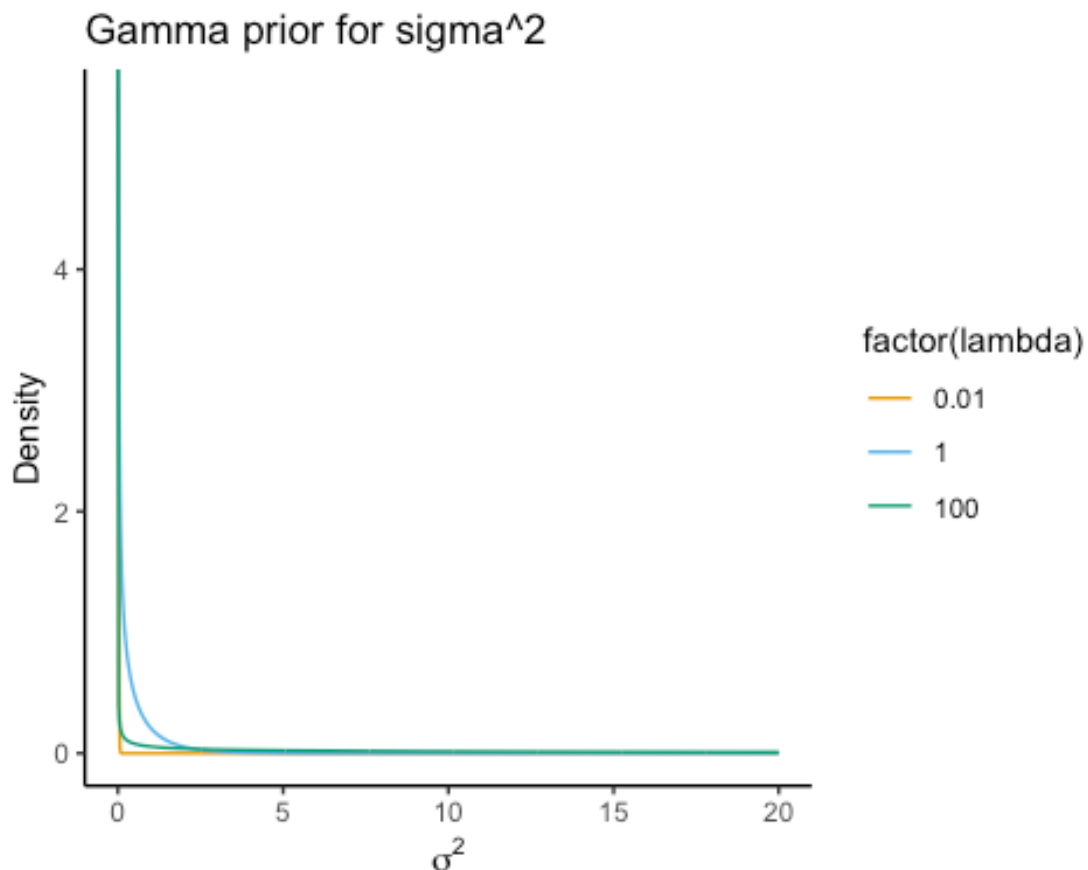
# Define the values of the hyperparameter Lambda
lambdas <- c(0.01, 1, 100)

# Create a data frame to store the sigma_2 and density values
```

```
df <- data.frame(sigma2 = rep(sigma2_values, length(lambdas)),
                 lambda = rep(lambdas, each = length(sigma2_values)))

# Add the density values to the data frame
df$density <- with(df, dgamma(sigma2, shape = 0.5, rate = 1/lambda))

# Plot the gamma prior for different values of Lambda
ggplot(df, aes(x = sigma2, y = density, color = factor(lambda))) +
  geom_line() +
  scale_x_continuous(breaks = seq(0, 20, by = 5)) +
  scale_color_manual(values = c("#E69F00", "#56B4E9", "#009E73")) +
  labs(x = expression(sigma^2), y = "Density",
       title = "Gamma prior for sigma^2") +
  theme_classic()
```



4.3 Suppose you are uncomfortable with choosing a value for  $\eta$ , and want to include this parameter in your model. Discuss a suitable prior distribution for  $\eta$ , and visualise the prior  $\sigma^2|\eta$  by first simulating draws from  $\eta$ , and then  $\sigma^2$  repeatedly.

A suitable prior distribution for the hyperparameter  $\eta$  in the gamma prior specification for the precision parameter,  $\sigma^2$ , would be the inverse gamma distribution. The inverse gamma distribution is a conjugate prior for the precision parameter  $\sigma^2$  when the likelihood is Gaussian, and is often used as a prior when the mean and variance are unknown.

To visualize the prior  $\sigma^2|\eta$ , we can simulate draws from  $\eta$  and  $\sigma^2$  repeatedly.

```
# Set up the data and prior parameters
y <- rnorm(100, mean = 5, sd = 1) # generate some data
n <- length(y) # sample size
mu_0 <- 0 # prior mean
sigma2_0 <- 1 # prior variance
lambda_0 <- 0.5 # prior shape parameter
nu_0 <- 2 * lambda_0 # prior degrees of freedom

# Sample values of nu and sigma2 from the inverse gamma prior
n_sims <- 1000 # number of simulation draws
nu_sims <- rinvgamma(n_sims, shape = nu_0, scale = lambda_0) # simulate from
inverse gamma
sigma2_sims <- 1/rgamma(n_sims, shape = nu_sims/2, scale = nu_sims/2) # trans
form to sigma^2

# Visualize the prior density
df_prior <- data.frame(sigma2 = sigma2_sims, nu = nu_sims)
ggplot(df_prior, aes(x = sigma2)) +
  geom_histogram(aes(y = ..density..), bins = 30, color = "black", fill = "wh
ite") +
  geom_density(fill = "blue", alpha = 0.3) +
  scale_x_continuous(limits = c(0, 20), breaks = seq(0, 20, by = 5)) +
  labs(x = expression(sigma^2), y = "Density",
       title = "Inverse gamma prior for sigma^2") +
  theme_classic()
```

