

# Advanced Macroeconometrics – Assignment 2

Max Heinze ([h11742049@s.wu.ac.at](mailto:h11742049@s.wu.ac.at))      Gabriel Konecny ([h11775903@s.wu.ac.at](mailto:h11775903@s.wu.ac.at))  
Patrick Schüssele ([h11712262@s.wu.ac.at](mailto:h11712262@s.wu.ac.at))

May 10, 2023

## Contents

<b>Exercise 1</b>	<b>2</b>
Exercise 1.1 - Convergence of Means of $n$ Normal Draws . . . . .	2
Exercise 1.2 - Convergence of Means of $n$ Draws From a Cauchy Distribution . . . . .	2
<b>Exercise 2</b>	<b>3</b>
Exercise 2.1 - The conjugate prior . . . . .	3
Exercise 2.2 - Point estimates of $\theta$ . . . . .	3
Exercise 2.3 - Prior information and different priors . . . . .	4
Exercise 2.4 - Improving the model . . . . .	5
<b>Exercise 3</b>	<b>6</b>
Exercise 3.1 - Simulating Data with $k = 1$ and $\sigma = 1$ . . . . .	6
Exercise 3.2 - Latent values if $\sigma$ is known . . . . .	7
Exercise 3.3 - An interesting regression . . . . .	7
Exercise 3.4 - Simulating data with known $\sigma$ and a Normal prior . . . . .	8
<b>Exercise 4</b>	<b>10</b>
Exercise 4.1 - Estimating $\mu$ . . . . .	10
Exercise 4.2 - Estimating $\sigma^2$ . . . . .	11
Exercise 4.3 - A prior for $\eta$ . . . . .	12

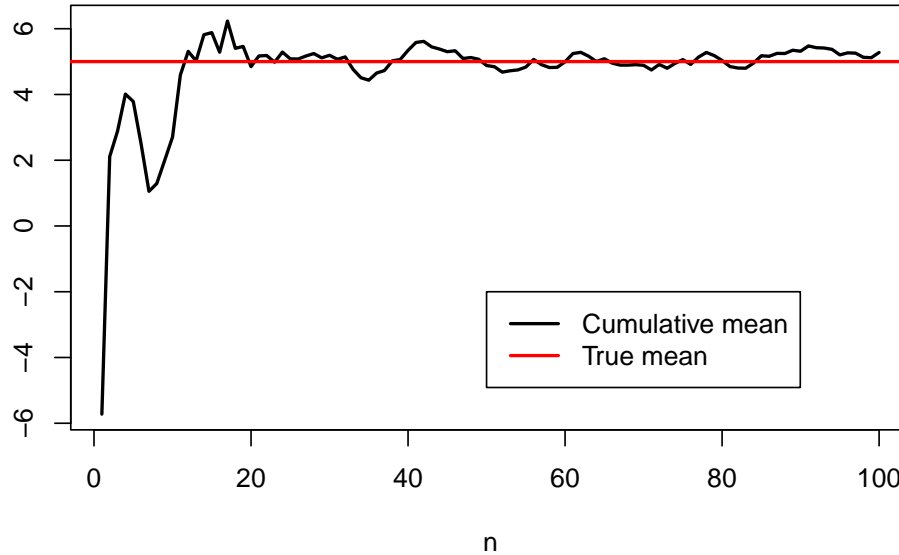
*The executable code that was used in compiling the assignment is available on GitHub at  
<https://github.com/maxmheinze/macrometrics>.*

## Exercise 1

### Exercise 1.1 - Convergence of Means of $n$ Normal Draws

We draw 100 times from a  $\mathcal{N}(5, 9)$  distribution and plot the cumulative mean as well as the expected value.

#### Convergence of Means: Normal Distribution



We can easily see that the cumulative mean, that is,

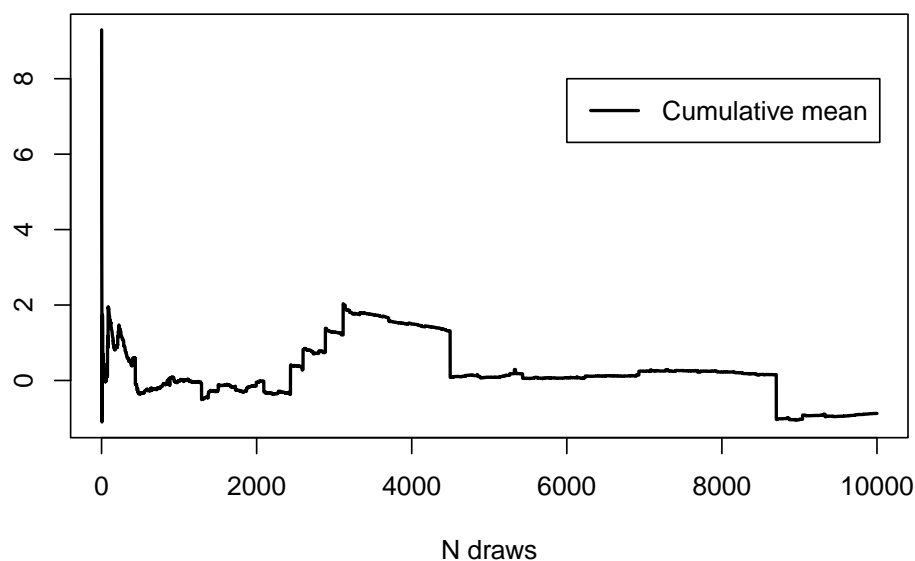
$$\frac{1}{n} \sum_{i=1}^n x_i, \quad n = 1, \dots, 100,$$

converges to the true mean, the expected value of the distribution,  $E(N(5, 9)) = 5$ .

### Exercise 1.2 - Convergence of Means of $n$ Draws From a Cauchy Distribution

Next, we draw 10,000 times from a Cauchy distribution as specified and plot the cumulative mean.

#### Convergence of Means: Cauchy Distribution



This time, we do not observe convergence of the cumulative mean. Rather, we observe repeated random jumps in the cumulative mean. This is due to the Cauchy distribution not having an expected value that it could converge to.

## Exercise 2

### Exercise 2.1 - The conjugate prior

A Covid test can be represented by a Bernoulli trial, where our random variable takes over value = 1 if test is *positive* and value = 0 if *negative*. As we conduct the tests for 20 colleagues, we have a sequence of Bernoulli trials, where the random variable now is the sum of all single Bernoulli trials. This random variable now follows a Binomial distribution. That is, our likelihood follows a Binomial distribution.<sup>1</sup>

The class of conjugate priors for a Binomial distribution are **Beta**( $\alpha$ ,  $\beta$ ) distributions where  $\alpha, \beta \in \mathbb{R}_+$ .

Notice that we have vector of observations  $y = (y_1, \dots, y_n)$  for  $n$  days, where  $y_i$  contains 20 observations, one for each colleague. Thus  $y_i \sim \text{Binom}(20, \theta)$  Since the  $n$  observations across time are iid the likelihood originates from  $\text{Binom}(20n, \theta)$

Thus the distribution of  $y_i$ , the likelihood function and our prior are respectively:

$$\begin{aligned} p(y_i | \theta) &= \binom{20}{k_i} \theta^{k_i} (1 - \theta)^{20 - k_i} \\ p(y | \theta) &= \binom{N}{S_n} \theta^{S_n} (1 - \theta)^{N - S_n} \propto \theta^{S_n} (1 - \theta)^{N - S_n} \\ p(\theta) &= p(\theta | \alpha, \beta) = \frac{1}{B(\alpha, \beta)} \theta^{\alpha - 1} (1 - \theta)^{\beta - 1} \propto \theta^{\alpha - 1} (1 - \theta)^{\beta - 1} \end{aligned}$$

where  $k_i = \sum_j y_{ij}$ ,  $S_n = \sum_i \sum_j y_{ij}$  and  $N = 20n$ . We could simplify the sum and rewrite it as:  $S_n = \sum_i^N y_i$  where  $y_i$  now denotes individual test results for all days.

Then the posterior can be derived using Bayes formula:

$$\begin{aligned} p(\theta | y) &\propto p(y | \theta) \cdot p(\theta) \\ &\propto \theta^{S_n} \cdot (1 - \theta)^{N - S_n} \cdot \theta^{\alpha - 1} \cdot (1 - \theta)^{\beta - 1} \\ &= \theta^{\alpha + S_n - 1} \cdot (1 - \theta)^{\beta + N - S_n - 1} \\ \implies p(\theta | y) &\sim \text{Beta}(\alpha + S_n, \beta + N - S_n) \end{aligned}$$

Our posterior is proportional to a Beta distribution. Thus the Beta prior is indeed a conjugate prior.

### Exercise 2.2 - Point estimates of $\theta$

30 observations for each of the 20 individuals are in total 600 observations. In total, 10 of the 600 observations/tests are positive. Then, given the fact that our prior as well as our posterior follows a beta distribution, we can look up the formulae of point estimators such as *mean*, *median*, *mode*, etc. for a beta distribution  $\text{Beta}(\alpha, \beta)$ . Applied to the case at hand, we obtain for the prior the following point estimators:

- Mean:  $E[\theta] = \frac{\alpha_0}{\alpha_0 + \beta_0}$
- Median
  - for  $\alpha_0, \beta_0 > 1$ :  $\frac{\alpha_0 - \frac{1}{3}}{\alpha_0 + \beta_0 - \frac{2}{3}}$
  - for  $\alpha, \beta = 1$ : Median = Mean.
- Mode:
  - for  $\alpha_0, \beta_0 > 1$ :  $\frac{\alpha_0 - 1}{\alpha_0 + \beta_0 - 2}$
  - for  $\alpha_0, \beta_0 = 1$ : Any value  $x \in (0, 1)$  is mode.

Note, that for posterior point estimators, we simply replace  $\alpha_0$  by  $\alpha_1 = \alpha_0 + S_n$ , and  $\beta_0$  by  $\beta_1 = \beta_0 + N - S_n$ , where  $S_n = 10$  is the number of successes; i.e. positive test results, and  $N = 600$  is the number of observations.

To see how these point estimators change with different values for  $\alpha_0$ , and  $\beta_0$ , see the table underneath. This gives a first indication of how different beta priors affect the posterior.

Table 1: Point estimators for various alpha and beta.

a_0	b_0	prior_mean	prior_med	prior_mod	a_1	b_1	post_mean	post_med	post_mod
1.0	1.0	0.5000	0.5000	any in (0,1)	11.0	591.0	0.0183	0.0177	0.0167

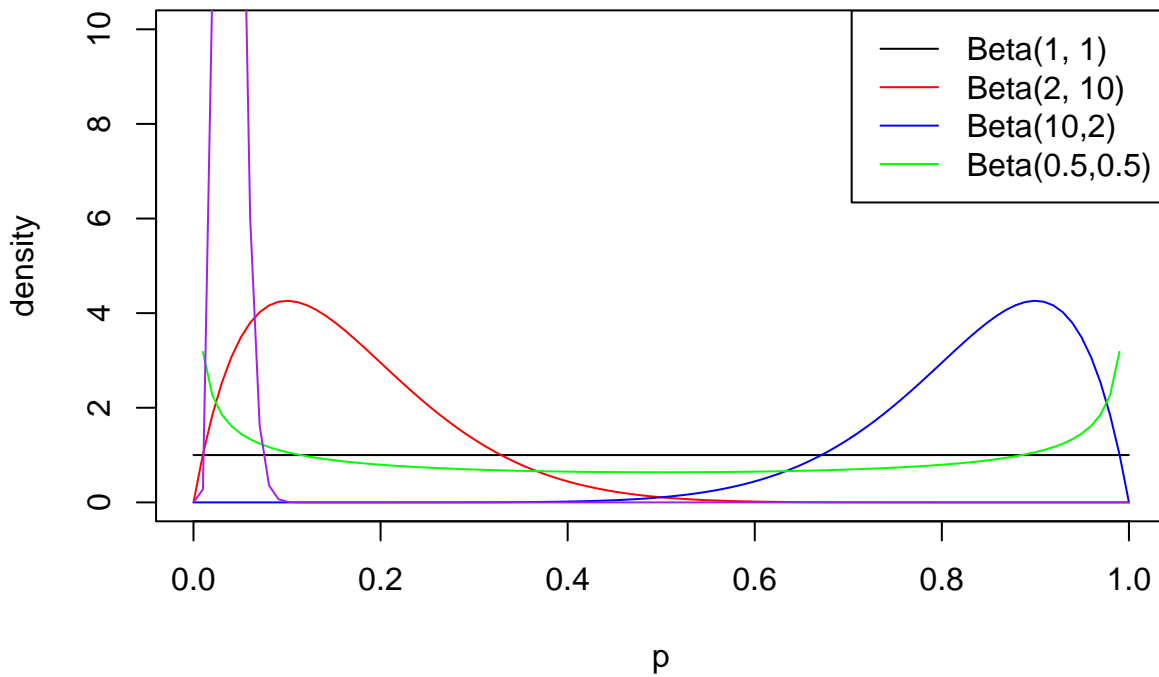
<sup>1</sup>Bernoulli distribution is just a special case of a Binomial distribution with  $n=1$ .

a_0	b_0	prior_mean	prior_med	prior_mod	a_1	b_1	post_mean	post_med	post_mod
2.0	10.0	0.1667	0.1471	0.1	12.0	600.0	0.0196	0.0191	0.0180
10.0	2.0	0.8333	0.8529	0.9	20.0	592.0	0.0327	0.0322	0.0311
0.5	0.5	0.5000	0.5000	0.5	10.5	590.5	0.0175	0.0169	0.0159
1.0	2.0	0.3333	0.2857	0	11.0	592.0	0.0182	0.0177	0.0166
2.0	1.0	0.6667	0.7143	1	12.0	591.0	0.0199	0.0194	0.0183
10.0	250.0	0.0385	0.0373	0.0349	20.0	840.0	0.0233	0.0229	0.0221

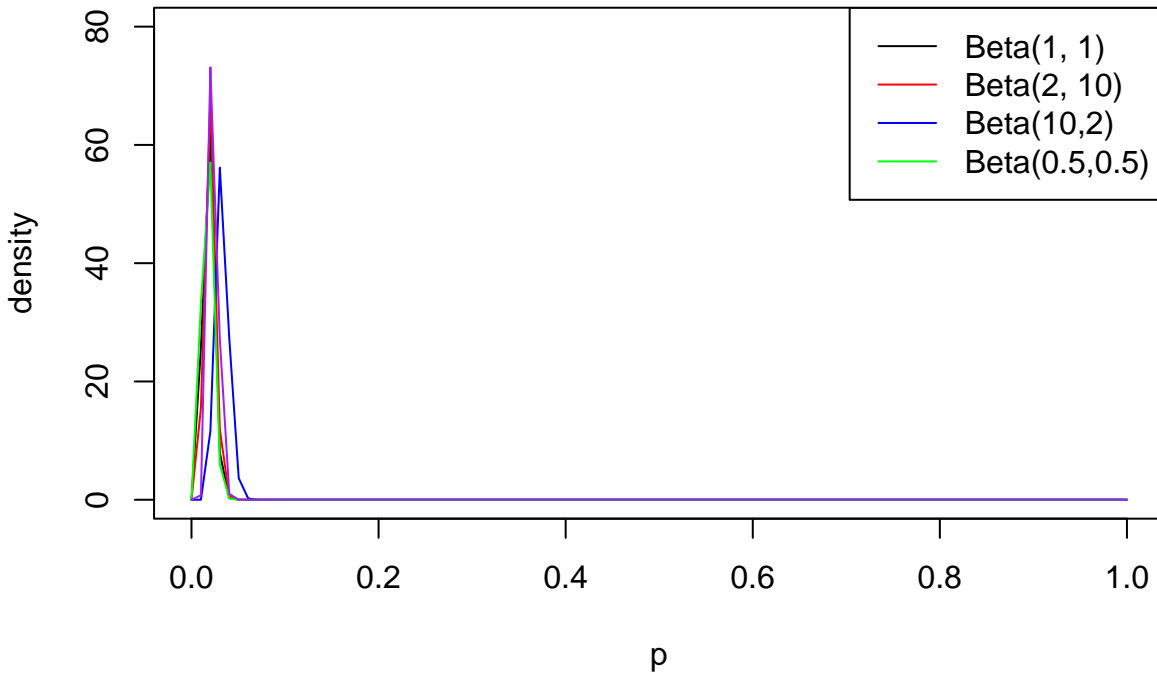
### Exercise 2.3 - Prior information and different priors

To visualise the impact different beta priors have on the posterior distribution, see the corresponding density plots that make use of selected  $\alpha_0$  and  $\beta_0$  shape parameters used in the table above.

#### Priors



## Posteriors



For  $\alpha_0 = \beta_0 = 1$  the prior is simply a uniform(1,1) distribution. This uninformative prior gives us a posterior distribution which is narrowly centered around approx. 0.02. As we see in the graphs above, a change in the prior parameters changes the pattern for the posterior only slightly in absolute terms. For arbitrary values  $\alpha_0 = \beta_0 > 1$  prior has the property that mean = median = mode<sup>2</sup>, but this does not hold for the posterior as  $\alpha_1$  and  $\beta_1$  are non-proportional linear combinations of  $\alpha_0$  and  $\beta_0$ .

For an increase in  $\alpha_0$  holding  $\beta_0$  constant we see an increase in posterior point estimates. If we increase  $\beta_0$  holding  $\alpha_0$  constant, estimates decrease.

As for sources of prior information, information for this particular problem can stem from previous research or available data. For one of our Beta-priors we tried to match the mean to the most recent value of COVID-19 test rates, which seem to be around 3.7%. After trying out few different parameters, this led us to use Beta(10,250), which has mean close to 3.7%. For more serious work, one would solve the equation for mean, but also try to match at least the variance of prior to the data as well.

### Exercise 2.4 - Improving the model

First, the assumption of *identical distribution* might not be fulfilled in real life, because the situation outside of our trial can influence the probability of testing positive on a particular day. Ceteris paribus, on a day with high Covid incidence, the probability that one of our employees will get in contact with an infected person and contracts Covid is higher. Also the life-style of the employees can influence their probability of testing positive. The people who tend to participate more in social activities might be more likely to test positive as compared to people who mostly isolate at home. Also someone who already had covid might be partially immune thus effecting the probability. Second, if one of our employees in the office tests positive on a particular day, then this employee will probably test positive on the second day as well. The probability that other colleagues test positive on the second day increases, too. Hence the *independence* assumption is not fulfilled either.

*How could you (conceptually) improve the model with this in mind?*

We could take into account whether person tested has symptoms and whether there were previous contacts with positively tested colleagues. Based on this we would change the prior for employees who have higher risk of having COVID by shifting their prior to the right side. Since the prior has relatively low effect we would need to change it drastically.

<sup>2</sup>for  $\alpha_0 = \beta_0 = 1$ , mean = median = 0.5, and mode is any value in interval (0,1).

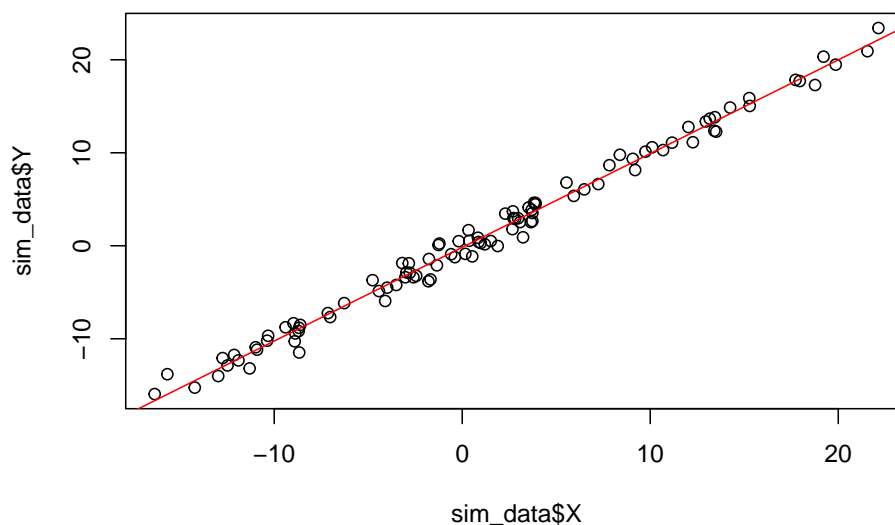
## Exercise 3

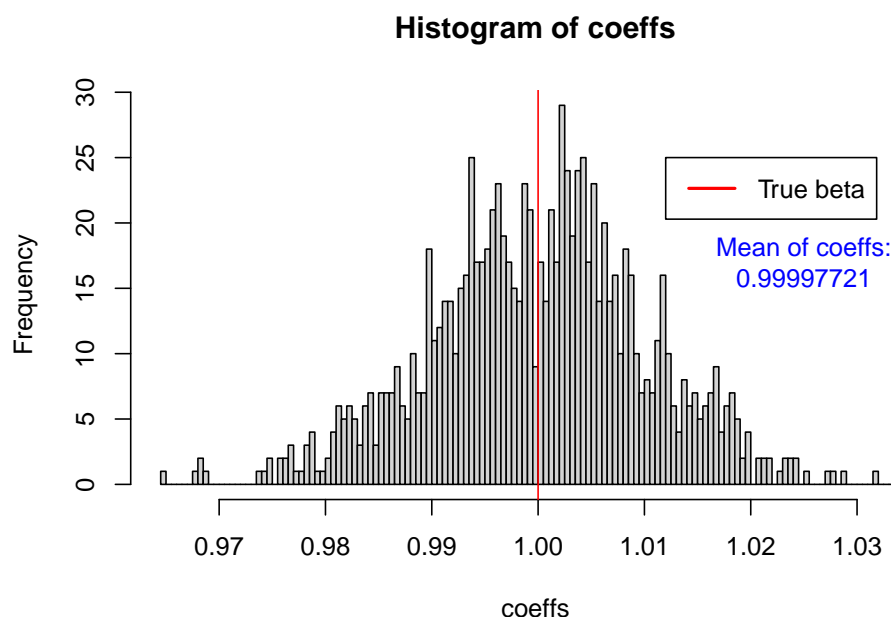
We write the following function to use in this exercise:

```
simulate_reg <- function(n, k = 1, alpha = 0, beta = 1, sigma = 1) {  
  
  # Remind the user to specify a vector as beta if they want to  
  if ((length(beta) == 1) * (k > 1) == 1) {  
    warning("You have specified k > 1 but only one value for beta.  
           This function then uses the same beta for all independent variables!")  
  }  
  
  # Create an empty matrix for the independent variables  
  X <- matrix(nrow = n, ncol = k)  
  
  # Fill them with draws from normal distributions. For fun we let the mean  
  # increase by 1 for every additional independent variable  
  for (i in 1:k) {  
    X[, i] <- rnorm(n, i, 10)  
  }  
  
  # Make beta a vector of length k if it is only a scalar  
  if ((length(beta) == 1) == TRUE) {  
    beta <- rep(beta, k)  
  }  
  
  # Compute y = a + Xb + e, e ~ N(0, sigma)  
  simul_dep <- rep(alpha, n) + X %*% beta + rnorm(n, 0, sigma)  
  
  # Create one output data frame  
  output_data <- data.frame(Y = simul_dep, X)  
  
  # Return it  
  return(output_data)  
}
```

### Exercise 3.1 - Simulating Data with $k = 1$ and $\sigma = 1$

**Simulated Data**





The true value  $\beta_0 = 1$ . We see that if we run this regression over and over and over again, we get a distribution of estimated  $\beta$ s that is centered around the true value, but with some variance. Indeed, the mean of our estimated betas equals 0.9999772, which is very close to  $\beta_0 = 1$ .

### Exercise 3.2 - Latent values if $\sigma$ is known

If  $\sigma$  is known,  $\alpha$  and  $\beta$  are the only latent parameters.

### Exercise 3.3 - An interesting regression

Let's say we are interested in the determinants of our Econometrics grades, and we propose the model

$$\text{points}_i = \alpha + \text{study time}\beta_1 + \text{talent}\beta_2 + \text{likeability}\beta_3 + u_i,$$

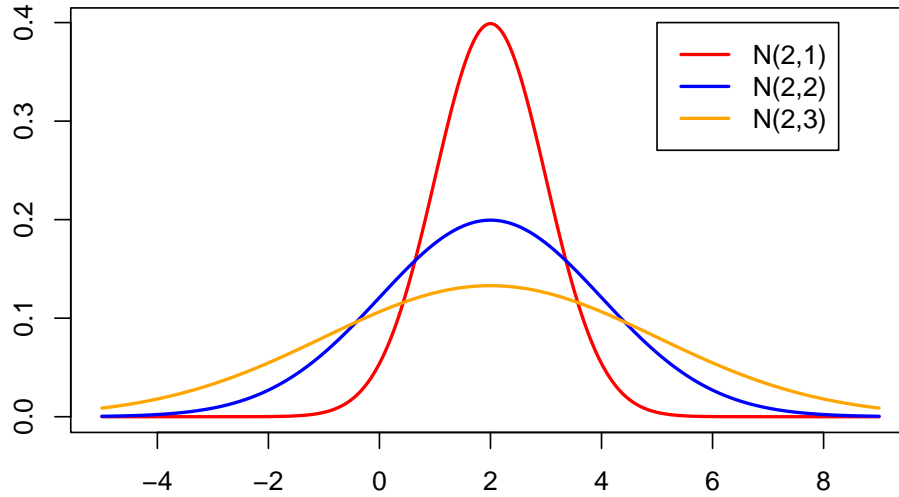
i.e. we regress the total course points (0–100) of student  $i$  on the time they devoted to studying, their inherent ability (assuming we can measure that), and how much they are deemed likable by the teachers of the course.

How can we come up with a prior distribution for  $\beta_1$ , the parameter associated with study time?

- It seems sensible to assume that the parameter associated with study time is **positive**, since it would be very weird if increasing the time devoted to studying for your Econometrics exam led to you getting a lower grade in your exam. If that were the case, nobody would be studying.
- We could even try to provide reasoning for a mean that we would want our prior distribution to be centered around. It would make sense if one hour of studying betters our result by, say, two points (and neither 0 nor 20). So we would use a prior distribution that is **centered at 2**.
- Then, we could assume for simplicity that the prior distribution is **normal**.

We could then go for one of the following priors:

### Three Normal Prior Distributions with Mean 2



### Exercise 3.4 - Simulating data with known $\sigma$ and a Normal prior

We simulate data with  $n \in \{50, 100, 200\}$ ,  $k = 1$ ,  $\alpha = 0$ ,  $\beta = 1$  and  $\sigma = 1$ . We set the prior density to be  $\mathcal{N}(0, 1)$ . Then, our posterior distribution will be given by

$$\mathcal{N}(\mu_n, \sigma_n^2),$$

where

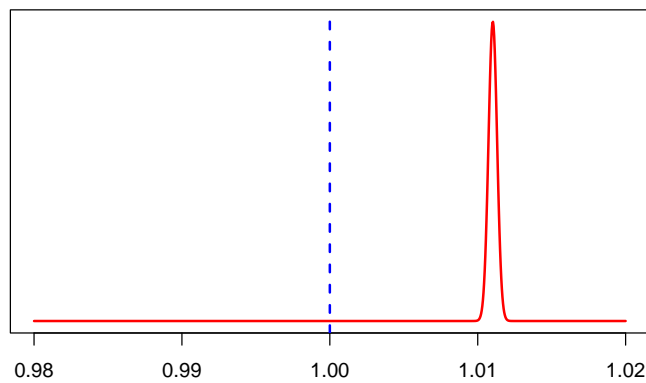
$$\sigma_n = (\sigma_0^{-1} + \mathbf{X}'\mathbf{X})^{-1}$$

and

$$\mu_n = \sigma_n(\sigma_0^{-1}\mu_0 + \mathbf{X}'\mathbf{y}).$$

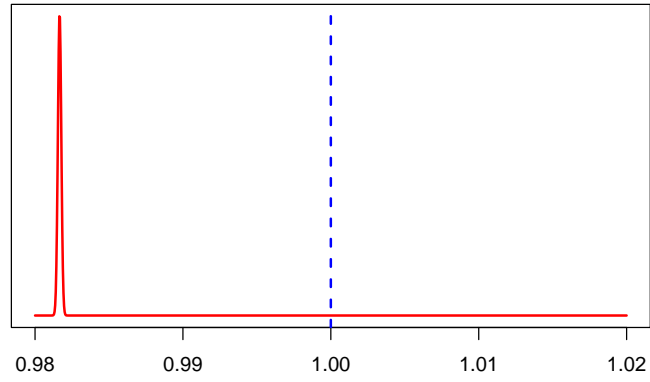
This yields the following posterior distributions:

**n = 50: Posterior  $\sim \mathcal{N}(1.011, 0.00029)$**

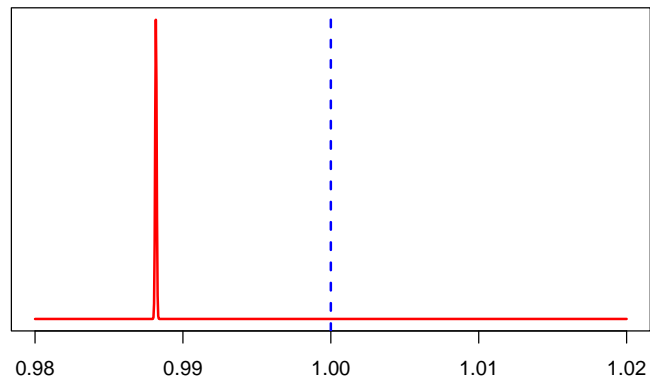




**n = 100: Posterior  $\sim N(0.9817, 0.00011)$**



**n = 200: Posterior  $\sim N(0.9882, 5e-05)$**



We can see that as we increase  $n$ , the priors' variance gets smaller and their mean moves closer to the true value.

## Exercise 4

### Exercise 4.1 - Estimating $\mu$

For the given normally distributed data, the probability density of each observation  $y_i$  looks like this:  $p(y_i|\mu, 1) = (2\pi)^{-\frac{1}{2}} \exp(-\frac{1}{2}(y_i - \mu)^2)$ . For iid data, we furthermore have the following likelihood:  $p(\mathbf{y}|\mu, 1) = \prod_{i=1}^n p(y_i|\mu, 1) = (2\pi)^{-\frac{n}{2}} \exp(-\frac{1}{2} \sum_{i=1}^n (y_i - \mu)^2)$ .

The normal pdf of the Normal prior is:  $p(\mu|\mu_0, \sigma_0^2) = (2\pi\sigma_0^2)^{-\frac{1}{2}} \exp(-\frac{1}{2} \frac{(\mu - \mu_0)^2}{\sigma_0^2})$ .

To get to our posterior, we utilize Bayes' theorem, dropping all constant factors in the process, as they don't change proportionality.

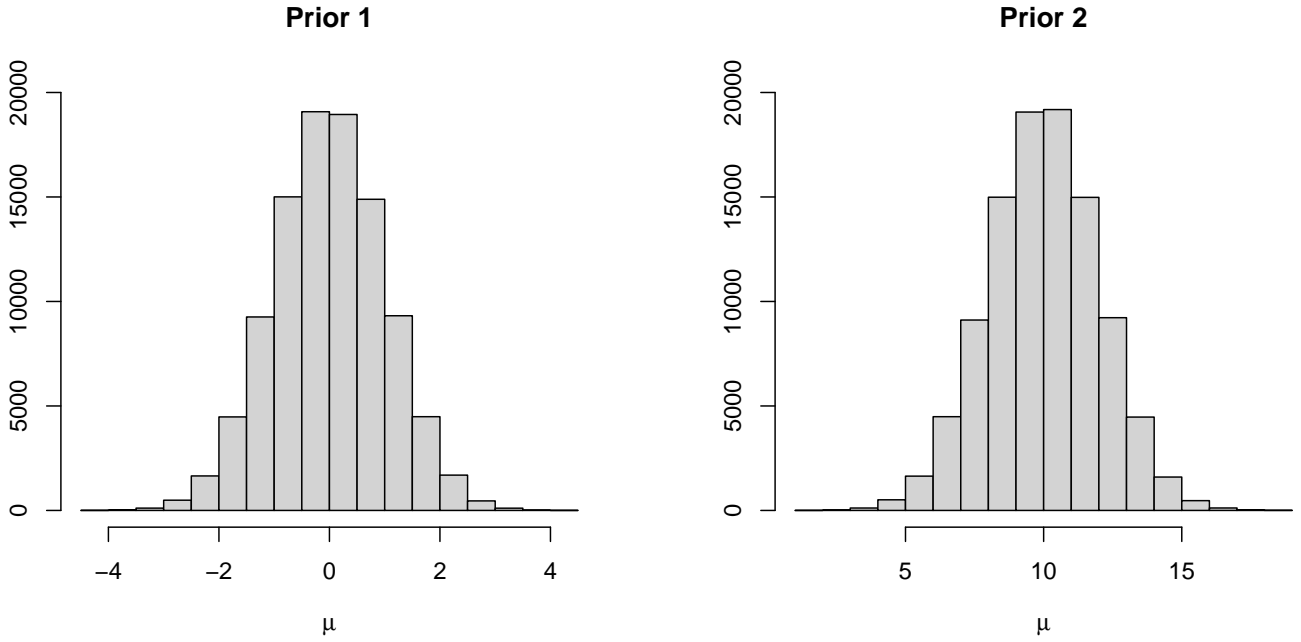
$$\begin{aligned}
 p(\mu|\mathbf{y}, 1) &\propto p(\mathbf{y}|\mu, 1) \times p(\mu|\mu_0, \sigma_0^2) \\
 &\propto \exp(-\frac{1}{2} \sum_{i=1}^n (y_i - \mu)^2) \times \exp(-\frac{1}{2} \frac{(\mu - \mu_0)^2}{\sigma_0^2}) \\
 &= \exp(-\frac{1}{2} (\sum_{i=1}^n (y_i - \mu)^2) + \frac{1}{\sigma_0^2} (\mu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

For simplicity, we defined  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ , and used  $\sum_{i=1}^n (y_i - \mu)^2 = \sum_{i=1}^n (y_i - \bar{y} + \bar{y} - \mu)^2 = \sum_{i=1}^n (y_i - \bar{y})^2 + n(\bar{y} - \mu)^2$ . We can now say  $\mu_n = (n + \frac{1}{\sigma_0^2})^{-1} (n\bar{y} + \frac{1}{\sigma_0^2} \mu_0)$  and  $\sigma_n^2 = (n + \frac{1}{\sigma_0^2})^{-1}$ .

After further simplification, our posterior yields:

$$\begin{aligned}
 p(\mu|\mathbf{y}, 1) &\propto \exp(-\frac{1}{2} (\sum_{i=1}^n (y_i - \bar{y})^2) + n\bar{y}^2 + \frac{1}{\sigma_0^2} \mu_0^2 - \frac{1}{\sigma_n^2} \mu_n^2 + \frac{1}{\sigma_n^2} (\mu^2 - 2\mu\mu_n + \mu_n^2)) \\
 &= \exp(-\frac{1}{2} (\sum_{i=1}^n (y_i - \bar{y})^2) + n\bar{y}^2 + \frac{1}{\sigma_0^2} \mu_0^2 - \frac{1}{\sigma_n^2} \mu_n^2 + \frac{1}{\sigma_n^2} (\mu - \mu_n)^2) \\
 &= \exp(-\frac{1}{2} (\sum_{i=1}^n (y_i - \bar{y})^2) + n\bar{y}^2 + \frac{1}{\sigma_0^2} \mu_0^2 - \frac{1}{\sigma_n^2} \mu_n^2) \times \exp(-\frac{1}{2\sigma_n^2} (\mu - \mu_n)^2) \\
 &\propto \exp(-\frac{1}{2\sigma_n^2} (\mu - \mu_n)^2)
 \end{aligned}$$

We now see that the posterior is proportional to a  $N(\mu_n, \sigma_n^2)$ . The following plot shows two histograms with different prior distributions - for Prior 1 we have  $\mu \sim N(0, 1)$  and for Prior 2 we use  $\mu \sim N(10, 4)$ .



### Exercise 4.2 - Estimating $\sigma^2$

For the given normally distributed data, the probability density of each observation  $y_i$  looks like this:  $p(y_i|5, \sigma^2) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp(-\frac{1}{2} \frac{(y_i-5)^2}{\sigma^2})$ . For iid data, the likelihood is given by:

$$\begin{aligned} p(\mathbf{y}|5, \sigma^2) &= \prod_{i=1}^n p(y_i|5, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - 5)^2) \\ &\propto (\sigma^2)^{-\frac{n}{2}+1-1} \exp(-\frac{\sum_{i=1}^n (y_i - 5)^2/2}{\sigma^2}) \\ &= (\sigma^2)^{c-1} \exp(-\frac{d}{\sigma^2}) \end{aligned}$$

This expression is proportional to an Inverse-Gamma distribution  $G^{-1}(c, d)$ , with  $c = -\frac{n}{2} + 1$  and  $d = \sum_{i=1}^n (y_i - 5)^2/2$ .

For the prior, we assume that the precision follows a  $\sigma^{-2} \sim G(0.5, \eta)$ . Therefore, we have  $\sigma^2 \sim G^{-1}(0.5, \eta)$ .<sup>3</sup> This gives us the prior for  $\sigma^2$  in the form of  $p(\sigma^2|0.5, \eta) = \frac{\sqrt{\eta}}{\Gamma(0.5)} (\sigma^2)^{-1.5} \exp(-\frac{\eta}{\sigma^2})$ .

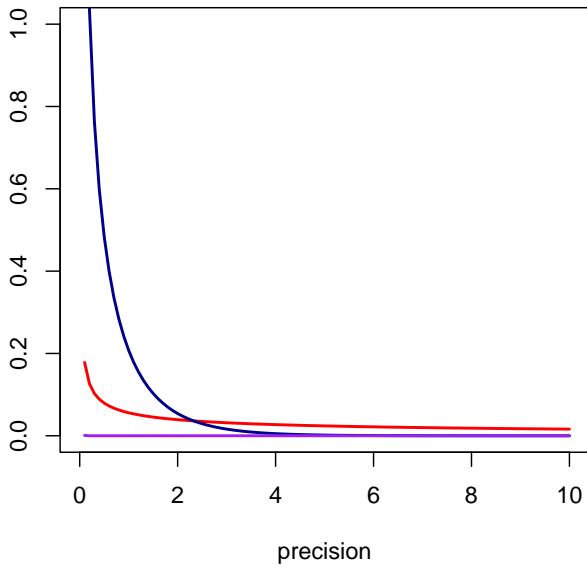
To get to our posterior, we again utilize Bayes' theorem and dropping constant terms:

$$\begin{aligned} p(\sigma^2|\mathbf{y}, 5) &\propto p(\mathbf{y}|5, \sigma^2) \times p(\sigma^2|0.5, \eta) \\ &\propto (\sigma^2)^{-\frac{n}{2}} \exp(-\frac{\sum_{i=1}^n (y_i - 5)^2/2}{\sigma^2}) \times (\sigma^2)^{-1.5} \exp(-\frac{\eta}{\sigma^2}) \\ &= (\sigma^2)^{-(\frac{n}{2}+0.5)-1} \exp(-\frac{1}{\sigma^2} (\eta + \sum_{i=1}^n (y_i - 5)^2/2)) \\ &= (\sigma^2)^{-c_n-1} \exp(-\frac{d_n}{\sigma^2}) \end{aligned}$$

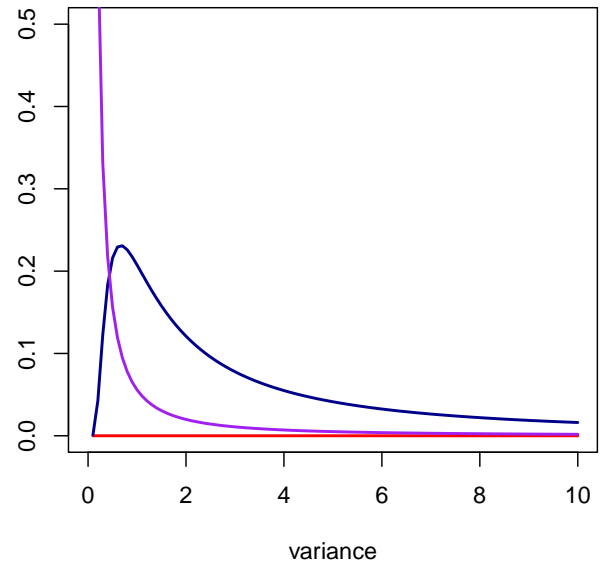
The posterior is thus proportional to an Inverse-Gamma distribution  $G^{-1}(c_n, d_n)$ , with  $c_n = \frac{n}{2} + 0.5$  and  $d_n = \eta + \sum_{i=1}^n (y_i - 5)^2/2$ . The following plot visualizes the prior density for the wanted values of  $\eta \in \{0.01, 1, 100\}$ . On the left, we show the plot with respect to the precision  $\sigma^{-2}$  and on the right with respect to variance  $\sigma^2$ .

<sup>3</sup>See derivation of Inverse-Gamma from Gamma on Wikipedia. For the Gamma distribution  $\beta$  denotes the *rate* parameter, whereas it denotes the *scale* parameter for the Inverse Gamma distribution!

**Prior precision**



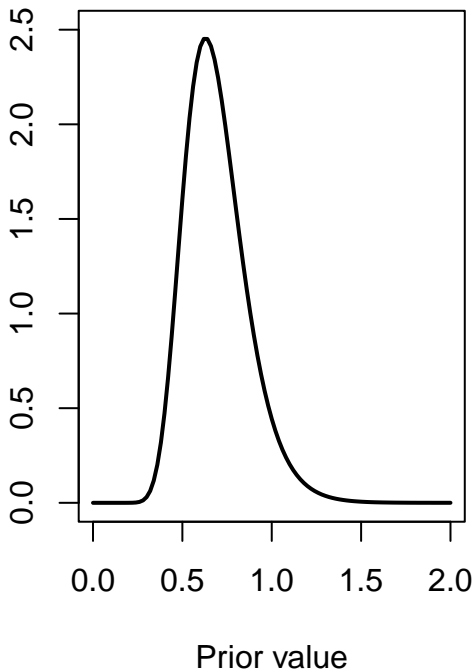
**Prior variance**



### Exercise 4.3 - A prior for $\eta$

A crucial consideration in prior choice is the one of allowed values - if we want to specify a prior for eg a variance, it makes no sense to use a Normal prior which allows for negative outcomes. A  $\eta$  is the scale parameter of an Inverse-Gamma distribution, it has to be positive, but there is no upper bound. This criteria is fulfilled by several distributions, such as the Gamma, the Inverse-Gamma or a Log-Normal. As it does seem plausible from our data, we argue that  $\sigma^2$  could be around 1, but we don't want to fully exclude larger values either; this could be achieved by drawing from a Inverse-Gamma with  $\alpha = 0.5$  and a scale parameter  $\eta$  between 0.5 and 1. To get those values for  $\eta$  often, we could draw  $\eta$  from a  $\text{Lognormal}(-0.4, 0.25)$ . The plot below shows the density of this Log-Normal on the left and the actual prior we get for  $\sigma^2$  on the right:

**The Lognormal density**



**The  $\sigma^2$  Prior**

