

Advanced Macroeconometrics - Assignment 2

Lucas Unterweger, Katharina König, Miravet Jaime

2023-04-21

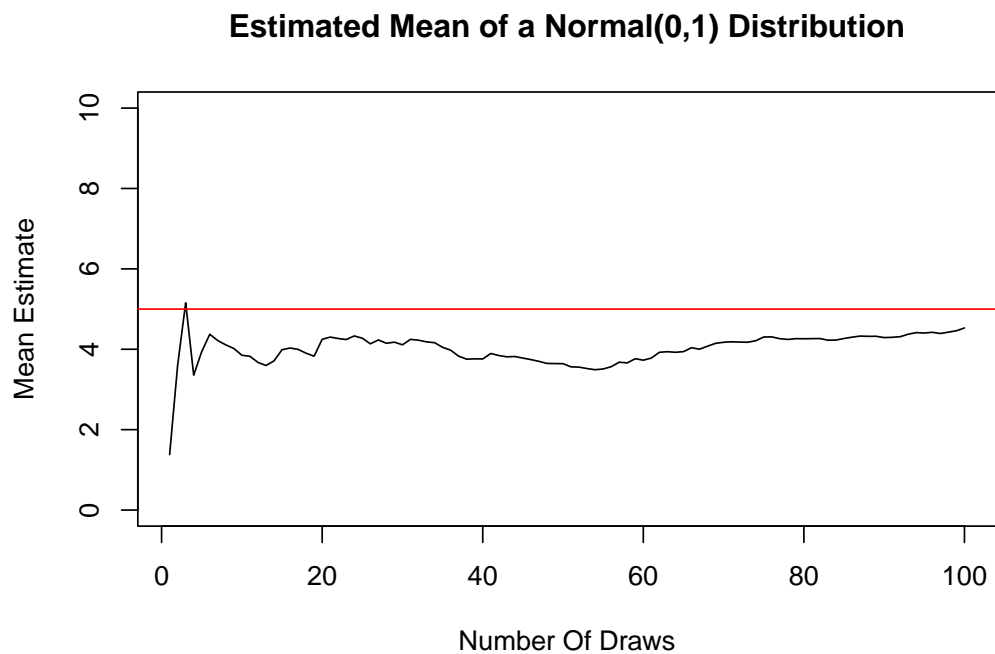
Preliminary

As always, the code which has been used to generate the plots etc. can be found in our GitHub repository [here](#).

Exercise 1

Question 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.

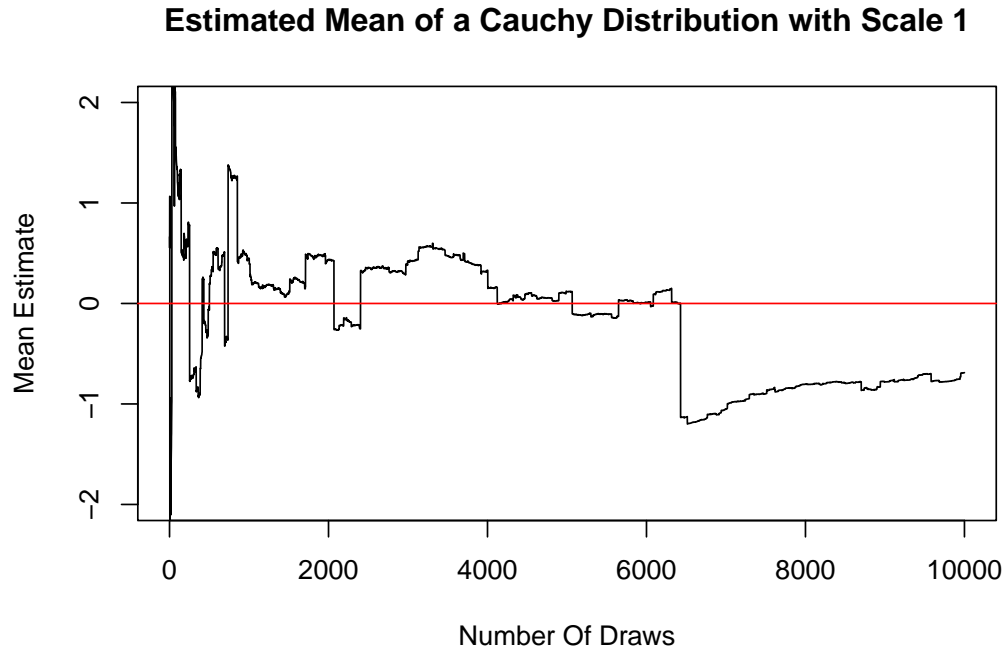


Here it can clearly be seen that with an increasing number of draws, the estimated mean

converges to the true value of the mean ($\mu = 5$).

Question 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.



As commonly known, the expected value/mean of a Cauchy distribution does not exist. This can be seen by the plot above as the average of the drawn Cauchy variables does not converge to the 0 one would expect it to. Even if we keep increasing the number of draws, it will not converge.

Exercise 2

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

Question 2.1

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

The class of conjugate priors for this problem, where the likelihood follows a binomial distribution, is given by Beta distributions $\text{Beta}(\alpha, \beta)$ with $\alpha, \beta > 0$. The pdf of a Beta-Distribution with parameters α and β is given by

$$p(x|\alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\text{Beta}(\alpha, \beta)}$$

with $\text{Beta}(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$, where $\Gamma(\cdot)$ is the Gamma function. Furthermore we know that $\forall i \in [n] : y_i \sim \text{Binom}(m, \theta)$ with PDF $p(y_i|m, \theta) = \binom{m}{y_i} \theta^{y_i} (1-\theta)^{m-y_i}$.

We first derive the likelihood function which is given by

$$p(\mathbf{y}|\theta) = \prod_{i=1}^n p(y_i|\theta) = \left[\prod_{i=0}^n \binom{m}{y_i} \right] \theta^{\sum_{i=1}^n y_i} (1-\theta)^{\sum_{i=1}^n (m-y_i)}$$

with $m = 20$. Using a general Beta distribution as the conjugate prior, the posterior distribution is - according to Bayes' theorem - proportional to

$$p(\theta|\mathbf{y}) \propto p(\mathbf{y}|\theta) \times p(\theta)$$

Thus we can compute

$$p(\theta|\mathbf{y}) \propto \left[\prod_{i=0}^n \binom{20}{y_i} \right] \theta^{\sum_{i=1}^n y_i} (1-\theta)^{\sum_{i=1}^n (20-y_i)} \times \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{\text{Beta}(\alpha, \beta)}$$

where we can drop constant values that do not depend on θ , yielding

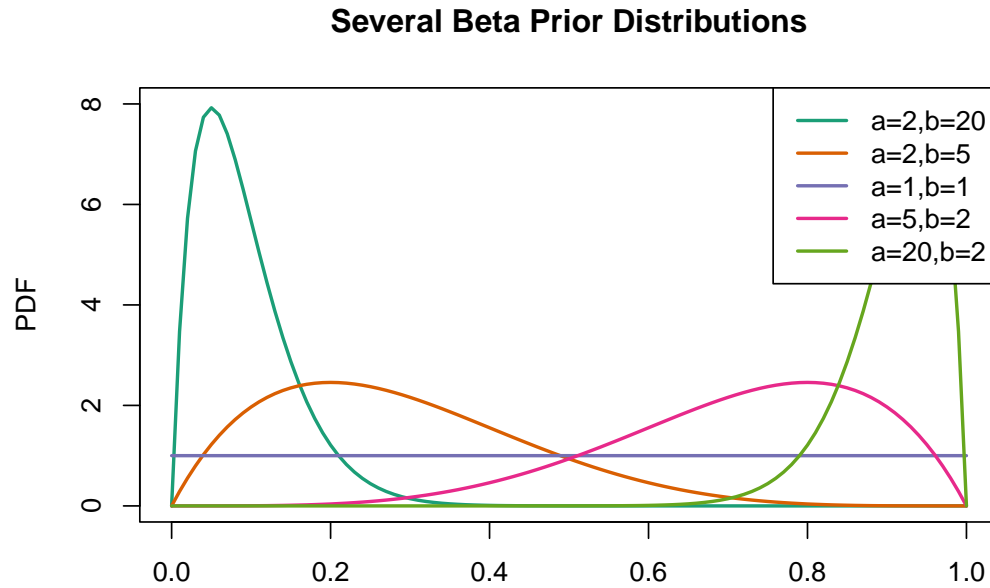
$$\propto \theta^{((\sum_{i=1}^n y_i) + \alpha) - 1} (1-\theta)^{(\sum_{i=1}^n (20-y_i) + \beta) - 1}$$

Which again is just a Beta distribution with parameters $\alpha' = \sum_{i=1}^n y_i + \alpha$ and $\beta' = \sum_{i=1}^n n - \sum_{i=1}^n y_i + \beta$.

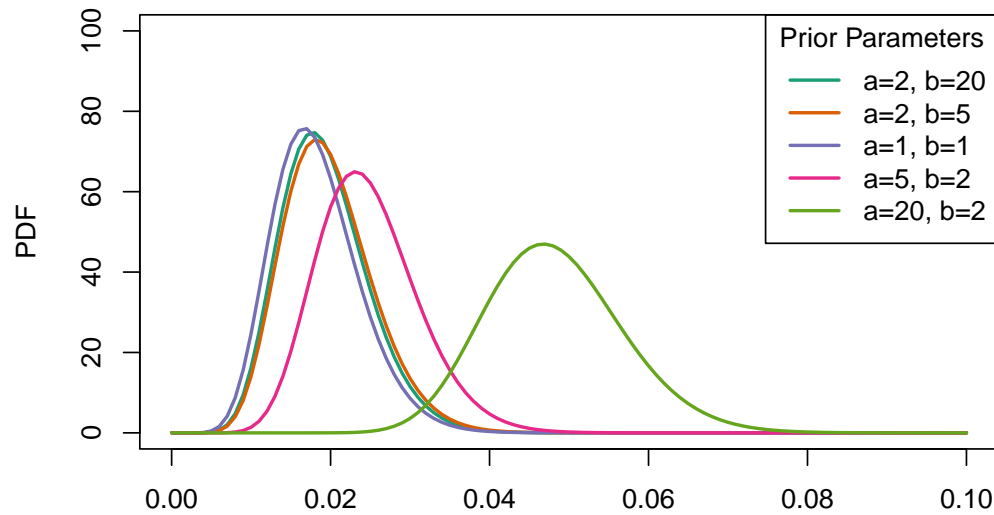
Question 2.2

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

We are now given observations for $n = 30$ with a total of ten positive tests. For computing the point estimates, we need a suitable prior distribution. For educational reasons and also for the follow up question in 2.3, we will test the following Beta-distributions:



Here, a and b refer to the parameters α and β of a Beta distribution. Using the given prior distributions and the given information on the data, we can compute the posterior distributions which are given by



We will focus on the point estimates which capture measures of central tendency, i.e. the mean, the mode and the median. For a random variable X which follows a Beta-Distribution, those are given by

$$Median \approx \frac{\alpha - 1/3}{\alpha + \beta - 2/3} \text{ if } \alpha, \beta > 1 \quad Mean = \frac{\alpha}{\alpha + \beta} \quad Mode = \frac{\alpha - 1}{\alpha + \beta + 2}$$

Given the chosen priors, the posterior point estimates are given by :

##	Alpha	Beta	Mean	Mode	Median
## 1	12	610	0.01929	0.01763	0.01878
## 2	12	595	0.01977	0.01806	0.01924
## 3	11	591	0.01827	0.01656	0.01774
## 4	15	592	0.02471	0.02299	0.02419
## 5	30	592	0.04823	0.04647	0.04775

We can see that the expected values are all in the neighborhood of 0.02, which means that the mean point estimate is roughly around 2%. Median and mode follow similar patterns which can be seen from the fact that in the posterior plot above, the distribution look very symmetric. The only outliers come from the (20, 2) prior, which will be covered in the next question.

Question 2.3

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

A credible source of prior information could be the official government infection rates. Of course, there several additional information one could incorporate into the prior decision. For example, the vaccination rates in the office are possibly higher than in the overall population. (could imply lower infection rates) Also, regional differences can be taken into account. Another source of information could be the COVID history of the office. How many people have already been infected? How high is the amount of regular contact between colleagues?

After setting up prior distributions for different parameter values of α and β , we computed the respective posterior distributions. For the point estimates of θ , we observed the effects of α and β on the distribution mean, median and mode. Both priors which expect the prevalence to be a rate below 50% and also the flat prior are not influencing the posterior very much as data provides enough information to result in similar posterior curves with mean roughly between 0.018 and 0.019. The Beta(5,2) prior starts to distort the posterior as we used a prior which expected the infection rate to be higher then fifty percent, which is definitely not present in the data. The worst prior - the Beta(20,2) has the biggest influence on the posterior as the prior does not coincide with the data at all.

Question 2.4

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

In case that the assumption of i.i.d would not hold, then it could happen that the results of day i depend on day $i - 1$. In this specific example, we know that the Binomial distribution is the probability distribution of the number of successes in a sequence of independent experiments, where each experiment is a Bernoulli experiment. Were this assumption violated, e.g. the different y_i depended on each other, we could not accurately retrieve the likelihood, which could lead to biased estimates.

The assumption of independent and identically distributed data may not be realistic in this case, since the observations are drawn from the same office and represent the results of a test for a contagious disease. This implies that it is quite likely that the probability of a positive test depends on whether some other colleges have been infected or not. Also, if I get a positive COVID result today from an actual infection, I have a higher probability to receive a positive test tomorrow as well. To improve the model, we could use a model that allows for dependence or serial correlation in the data, like an ARMA model.

Exercise 3

We start by writing the function.

```
simObs <- function(n,k,alpha,beta,sigma){
  if(length(beta)==k){

    # Generate errors from a (0,sigma)-normal distribution
    errors <- rnorm(n,0,sigma)

    # Generate k covariates from a uniform distribution in
    # the intervall [-10,10]
    x <- vector()
    for (i in seq(k)) {
      x <- cbind(x,runif(n, min=-10,max=10))
    }

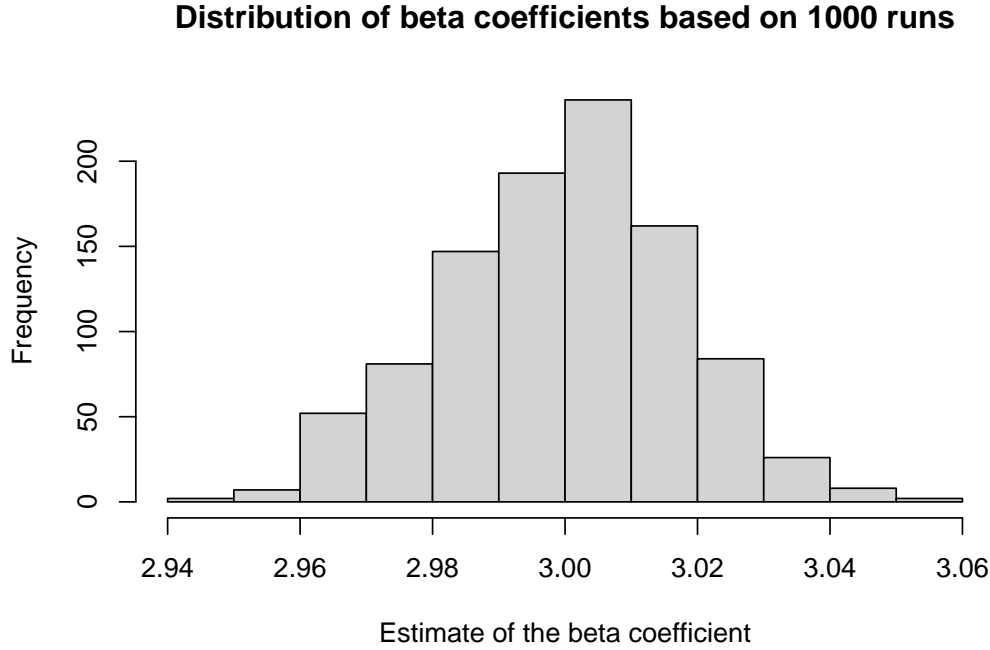
    # Generate the dependent variables based on the given
    # independent variables and errors
    y <- alpha + x %*% beta + errors

    # Return data frame with the data
    output <- data.frame(Y = y, X = x)
    return(output)
  }
  else{
    print("Number of variables K and length beta vector does not match!")
    break
  }
}
```

Question 3.1

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

With our function, we generate data with $k=1$, $\sigma=1$, $\alpha=5$ and $\beta=3$.



We can see that the estimates are somewhat normally distributed with mean 3 (the true value). Which makes sense as the expected value of the estimated beta coefficient is - in the case of unbiasedness - equal to the true value. The variance is given by 3.1496696×10^{-4} , which would decrease in case we'd increase the sample size of the simulation. (here set at $n = 100$)

Question 3.2

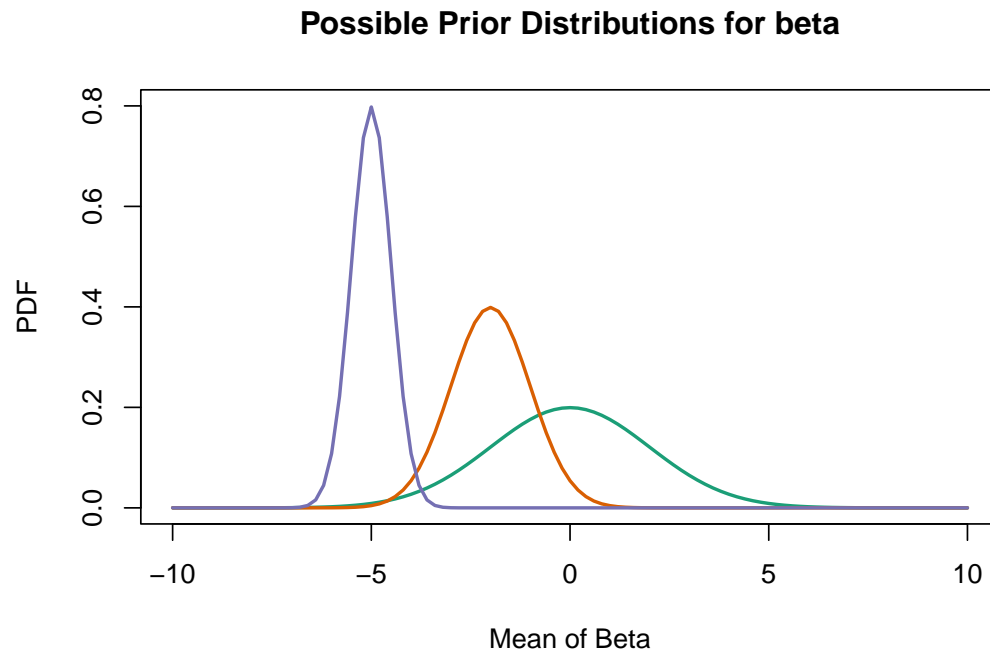
The remaining latent variable is just the parameter vector β .

Question 3.3

A potentially interesting regression would be the relationship between the rise of populism - let's use the change in votes for populist parties in percentage points - and the economic development of the respective regions in e.g. GDP change compared to 25 years ago. More formally,

$$Y_{VoteChange} = \alpha + \beta X_{EconomicGrowth} + u, \quad u \sim N(0, \sigma^2)$$

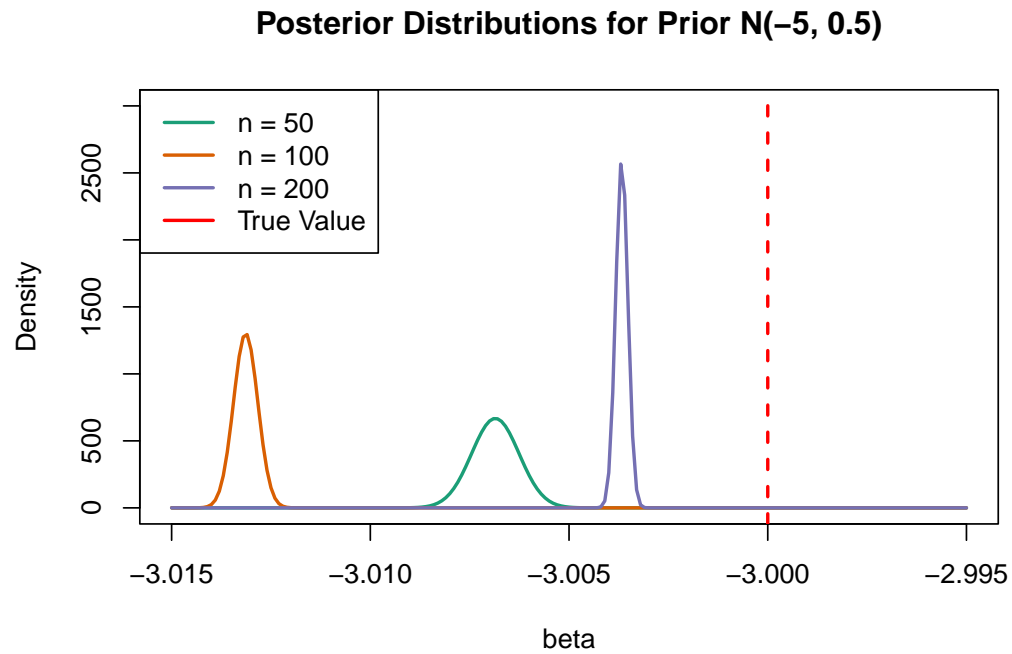
In recent years, various studies have found evidence for the hypothesis that economic decline causes rise in populist support. Hence, one could assume that the relationship is negative. Similarly, for simplicity we assume the prior to be normally distributed. Some prior distribution could look like:



Here the prior center around 0 would describe the guess that there is no relationship at all but with a lot uncertainty. The other two priors would suggest - in a stringer and weaker version - a negative relationship between economic growth and populist party votes.

Question 3.4

For the purpose of this question, we simulate data with $\alpha = 1$ and $\beta = -3$, which roughly captures the previously given relationship. We will use the prior distribution with mean -5 and standard deviation 0.5 , hence a prior that is $N(-5, 0.5)$. The posterior distribution is then given by



Here, the dotted red line indicates the true value of the simulated data. It can be clearly seen that the posterior density's variance decreases with an increasing number of simulations. In other words, we are getting more precise with an increasing number of data points. Also, the posterior density with the highest number of simulations is closest to the true value of β , which makes sense as we are having more data that supports the true value of 3 and the prior is becoming less influential.

Exercise 4

Question 4.1

We have data $y \sim N(\mu, 1)$ and want to estimate μ , hence a posterior distribution $p(\mu|y)$, which we can compute using Bayes' Theorem

$$p(\mu|y) \propto p(y|\mu) \times p(\mu)$$

We start by computing the likelihood $p(y|\mu, 1)$ which is

$$\begin{aligned} p(y|\mu, 1) &= \prod_{i=1}^N p(y_i|\mu, 1) = (2\pi)^{-N/2} \cdot \exp \left\{ -\frac{1}{2} \sum_{i=1}^N (y_i - \mu)^2 \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \sum_{i=1}^N (y_i - \mu)^2 \right\} \propto \exp \left\{ -\frac{N(\mu - \bar{y})^2}{2} \right\} \end{aligned}$$

The prior is given by $\mu \sim N(\mu_0, \sigma_0^2)$. which means that after dropping constants we gain

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

Using Bayes's theorem we can compute the posterior density

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

with $\sigma^* = (N + \sigma_0^{-2})^{-1}$ and $\mu^* = \frac{\bar{y}N + \sigma_0^{-2}\mu_0}{N + \sigma_0^{-2}}$, this yields

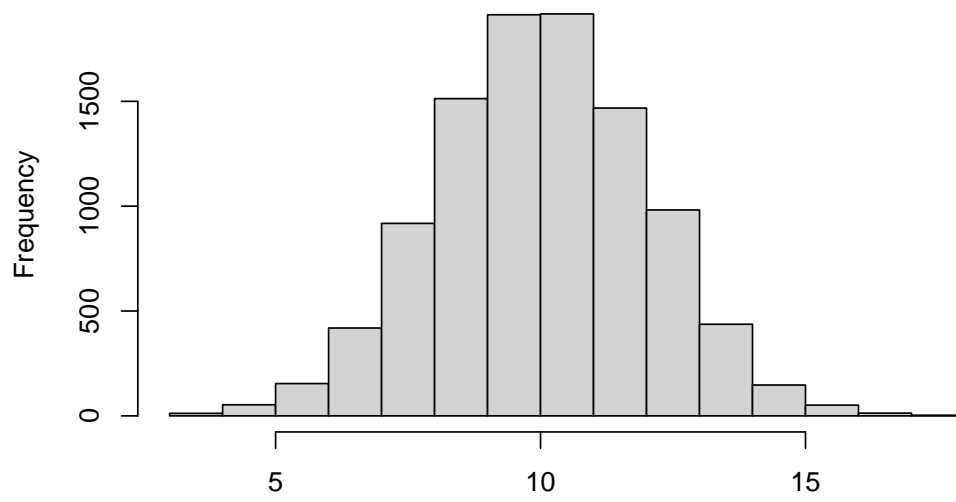
$$\exp \left\{ -\frac{1}{2\sigma^*} (\mu^2 - \mu^* 2\mu) \right\} \propto \exp \left\{ -\frac{1}{2\sigma^*} (\mu^2 - 2\mu^* \mu + (\mu^*)^2) \right\} = \exp \left\{ -\frac{1}{2\sigma^*} (\mu - \mu^*)^2 \right\}$$

This is now just a normal distribution with mean μ^* and variance σ^* , thus

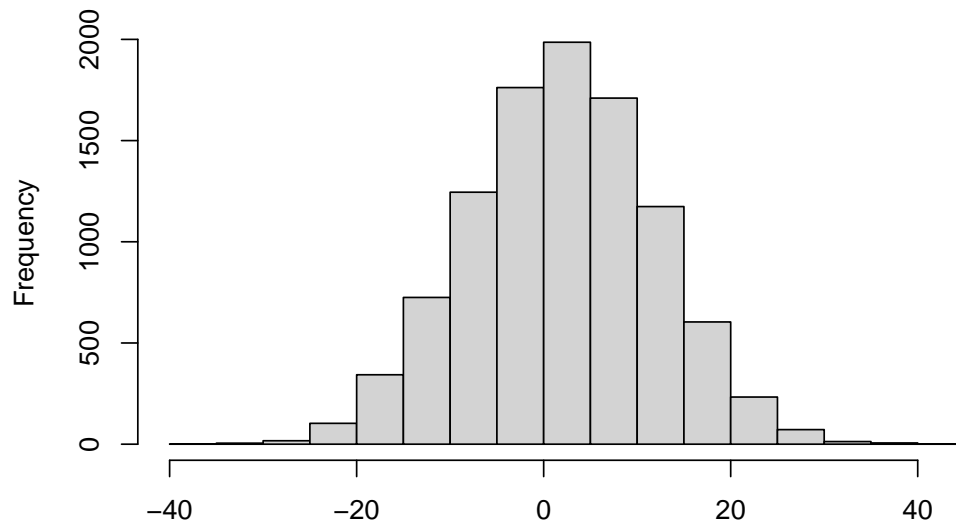
$$N(\mu^*, \sigma^*) = N\left(\frac{\bar{y}N + \sigma_0^{-2}\mu_0}{N + \sigma_0^{-2}}, (N + \sigma_0^{-2})^{-1}\right)$$

We will choose a $N(10, 2)$ and a $N(2, 10)$ prior and draw 10.000 times from each distribution to gain the histograms:

Histogramm of a $N(10,2)$ prior



Histogramm of a $N(2,10)$ prior



Question 4.2

We are given data $y \sim N(5, \sigma^2)$. We'll start by having a look at the likelihood which is given by

$$p(y|\mu, \sigma^2) = \prod_{i=1}^N p(y_i|5, \sigma^2) = (2\pi\sigma^2)^{-0.5} \exp\left\{-\frac{(y_1 - 5)^2}{2\sigma^2}\right\} \cdots (2\pi\sigma^2)^{-0.5} \exp\left\{-\frac{(y_N - 5)^2}{2\sigma^2}\right\}$$

Simplifying the expression yields

$$\propto (\sigma^2)^{-N/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^N (y_i - 5)^2\right\} = (\sigma^{-2})^{N/2} \exp\left\{-\frac{\sigma^{-2}}{2} \sum_{i=1}^N (y_i - 5)^2\right\}$$

The prior for σ^{-2} follows a Gamma distribution with parameters 0.5 and η . The pdf of such a distribution is given by

$$p(\sigma^{-2}|0.5, \eta) = \frac{\eta^{0.5}}{\Gamma(0.5)} (\sigma^{-2})^{0.5-1} \exp\{-\eta(\sigma^{-2})\} \propto (\sigma^{-2})^{-0.5} \exp\{-\eta(\sigma^{-2})\}$$

We can now use Bayes Theorem to compute the posterior density to gain

$$p(\sigma^{-2}|y) \propto p(y|\sigma^{-2}) \times p(\sigma^{-2})$$

Thus, by plugging in our previous results:

$$\begin{aligned} & (\sigma^{-2})^{N/2} \exp\left\{-\frac{\sigma^{-2}}{2} \sum_{i=1}^N (y_i - 5)^2\right\} \times (\sigma^{-2})^{-0.5} \exp\{-\eta(\sigma^{-2})\} = \\ & (\sigma^{-2})^{N/2-0.5} \exp\left\{-\frac{\sigma^{-2}}{2} \sum_{i=1}^N (y_i - 5)^2 - \eta(\sigma^{-2})\right\} \\ & = (\sigma^{-2})^{N/2+0.5-1} \exp\left\{-\sigma^{-2} \left(\frac{1}{2} \sum_{i=1}^N (y_i - 5)^2 - \eta\right)\right\} \end{aligned}$$

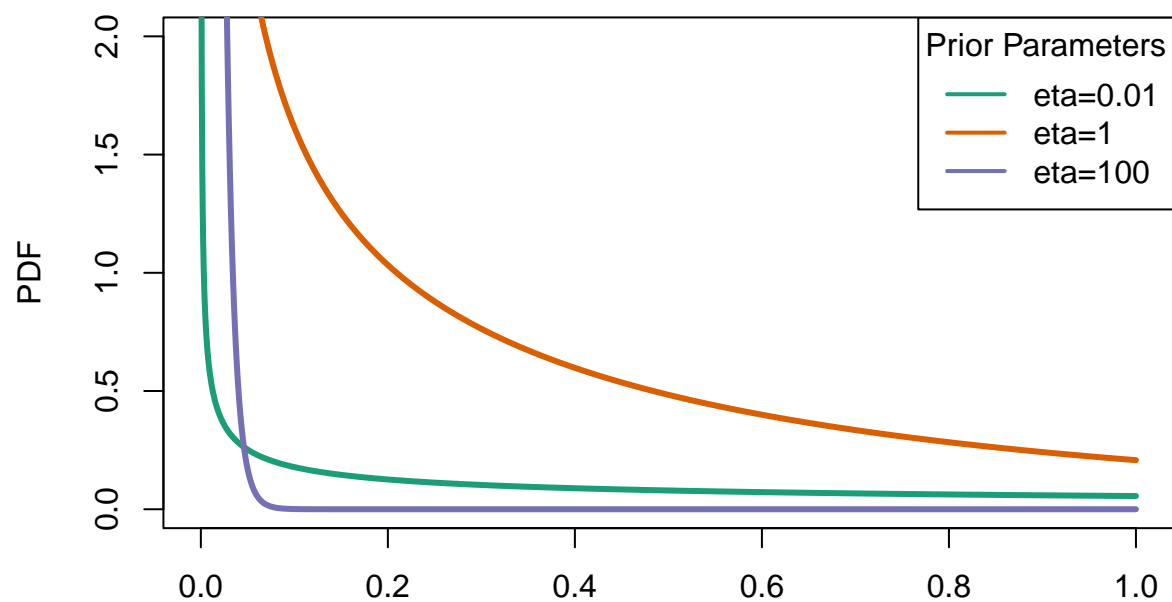
and this is just proportional to a Gamma distribution. Hence the posterior for the precision σ^{-2} is

$$\sigma^{-2}|y \sim G(N/2 + 0.5, \frac{1}{2} \sum_{i=1}^N (y_i - 5)^2 - \eta)$$

To gain the posterior for σ^2 , we can use the relationship between a Gamma and an inverted Gamma distribution:

$$\sigma^2|y \sim G^{-1}(N/2 + 0.5, \frac{1}{2} \sum_{i=1}^N (y_i - 5)^2 - \eta)$$

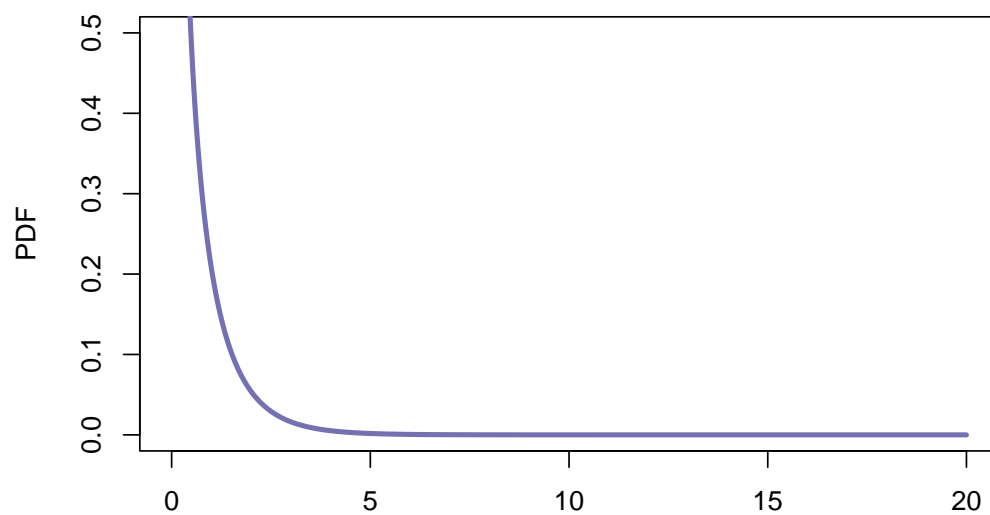
Here are the prior distributions for $\eta \in \{0.01, 1, 100\}$:



Question 4.3

First, it is important to note that η has to be strictly positive to be a parameter for the Gamma Distribution. Furthermore, we know that the first parameter, the shape, is fixed to 0.5. A conjugate prior for a Gamma with fixed shape is the Gamma distribution itself, which also satisfies the assumption that $\eta > 0$. Let's choose a Hyperprior with parameters 0.5 and 1. More formally, let $\eta \sim G(0.5, 1)$. This prior distribution looks like

Hyperprior Distribution as Gamma(0.5,1)



We can now simulate draws for η and then use these to draw the σ^2 . This yields the following Gamma-Gamma prior:

Prior Distribution as Gamma(0.5,Eta)

