

Advanced Macroeconometrics – Assignment 2

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

May 10, 2023

Contents

Exercise 1	2
Exercise 1.1 - Convergence of Means of n Normal Draws	2
Exercise 2.2 - Convergence of Means of n Draws From a Cauchy Distribution	2
Exercise 2	4
Exercise 2.1 - Conjugate Prior to Binomial distributions	4
Exercise 2.2 - Point Estimates for θ	4
Exercise 2.3 - Impact of different prior information on point estimates	4
Exercise 2.4	6
Exercise 3	8
Exercise 3.1 - Simulating Data with $k = 1$ and $\sigma = 1$	8
Exercise 3.2 - Latent values of σ is known	8
Exercise 3.3 - An interesting regression	8
Exercise 3.4 - Simulating data with known σ and a Normal prior STILL OPEN	9
Exercise 4	10
Exercise 4.1 - Estimating μ	10
Exercise 4.2 - Estimating σ^2	11
Exercise 4.3 - A prior for η	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

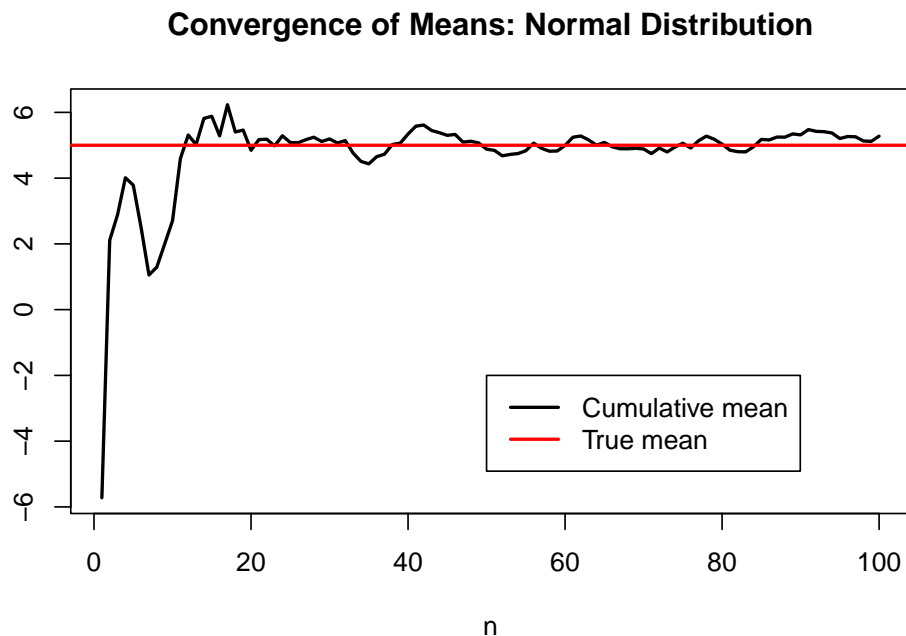
We draw 100 times from a $\mathcal{N}(5, 9)$ distribution and plot the cumulative mean as well as the expected value.

```
set.seed(2345)

n <- 100
normal_data <- data.frame(n = 1:n, x = rnorm(n, mean = 5, sd = 9), mean = NA)

for (i in 1:n) {
  normal_data$mean[i] <- mean(normal_data$x[1:i])
}

plot(x = normal_data$n, y = normal_data$mean, type = "l", lwd = 2, main = "Convergence of
Means: Normal Distribution",
      xlab = "n", ylab = "")
abline(h = 5, col = "red", lwd = 2)
legend(50, -2, c("Cumulative mean", "True mean"), lwd = c(2, 2), col = c("black",
"red"))
```



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 2.2 - 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.

```
set.seed(2345)

n <- 10000

cauchy_data <- data.frame(n = 1:n, x = rnorm(n, mean = 0, sd = 1)/rnorm(n, mean = 0,
sd = 1), mean = NA)

for (i in 1:n) {
```

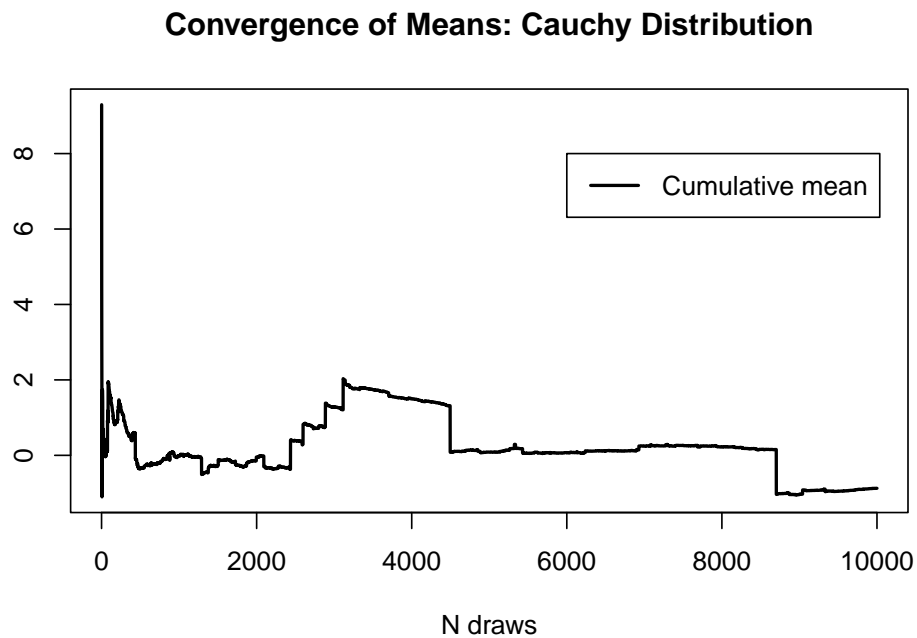
```

cauchy_data$mean[i] <- mean(cauchy_data$x[1:i])
}

plot(x = cauchy_data$n, y = cauchy_data$mean, type = "l", lwd = 2, main = "Convergence of
Means: Cauchy Distribution",
     xlab = "N draws", ylab = "")

legend(6000, 8, "Cumulative mean", lwd = 2, col = "black")

```



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 - Conjugate Prior to Binomial distributions

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.¹

The class of conjugate priors for a Binomial distribution are Beta(α, β) 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 $y_i \sim \text{Binom}(20, \theta)$. The likelihood is then the product of n such iid variables.

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

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

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: $S_n = \sum_i y_i$ 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 conjugate.

Exercise 2.2 - Point Estimates for θ

30 observations for each of the 20 individuals result in a total of $N=600$ observations. In total, $\{S_n = 10\}$ 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 α_0 by $\alpha_1 = \alpha_0 + S_n$, and β_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 again.

Exercise 2.3 - Impact of different prior information on point estimates

To see, how these point estimators change with different values for α_0 , and β_0 , see the table underneath. Also, this table gives first indications of how different beta priors affect the posterior.

¹Bernoulli distribution is just a special case of a Binomial distribution with $n=1$.

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	0.50000	0.50000	any in (0,1)	11.0	591.0	0.01827	0.01774	0.01667
0.1	0.1	0.50000	0.50000	0.5	10.1	590.1	0.01683	0.01629	0.01521
0.1	5.0	0.01961	-0.05263	-0.29032	10.1	595.0	0.01669	0.01616	0.01509
5.0	0.1	0.98039	1.05263	1.29032	15.0	590.1	0.02479	0.02427	0.02321

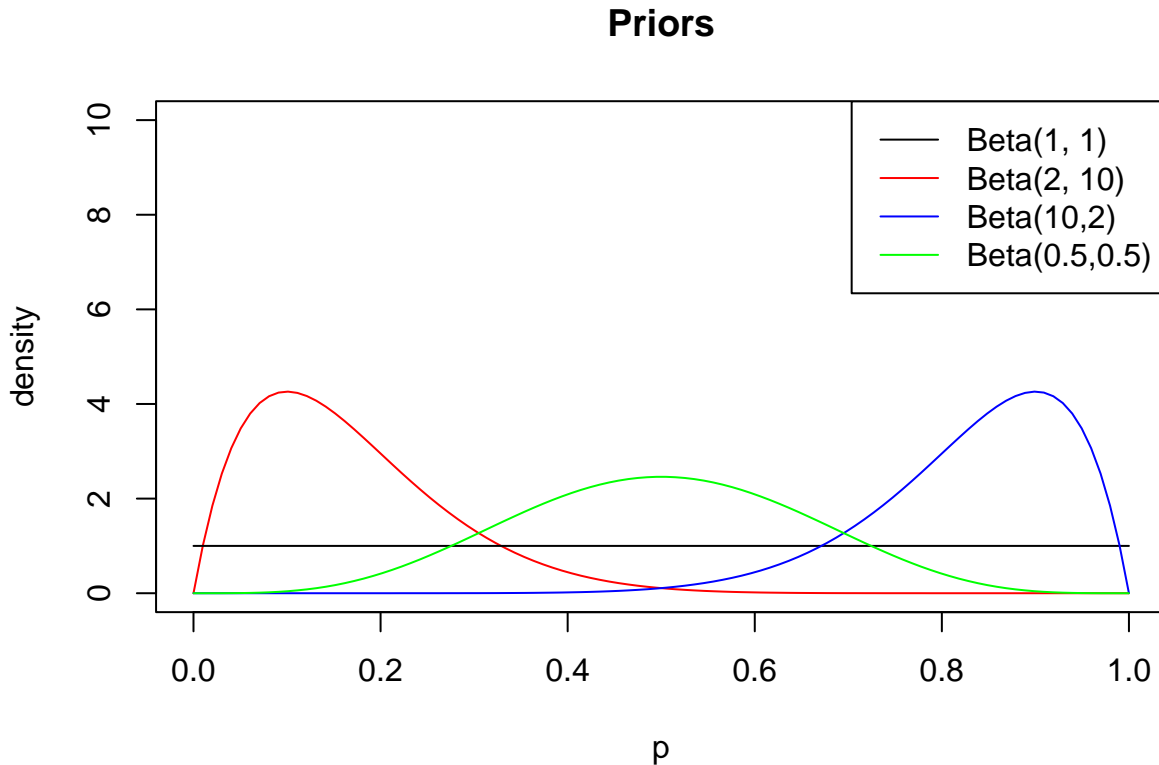
To visualise the impact different beta priors have on the posterior distribution, see the corresponding density plots that make use of the same α_0 and β_0 shape parameters used in the table above. Note that we adjusted the x-axis limit of the posterior plots, to 0.1 rather than 1 as we wanted to precisely see what impact different beta priors have on the posterior. Of course, we only did that because there was not much going on in terms of posterior density weights on the range from 0.1 to 1, and thus, including the whole range from 0 to 1 would not allow us to clearly see the different impacts.

```
# define range
p = seq(0, 1, length = 100)

# plot several Beta distributions
plot(p, dbeta(p, 2, 10), ylab = "density", type = "l", col = "red", main = "Priors",
      ylim = c(0, 10))
a_0 = c(1, 0.1, 0.1, 5)
b_0 = c(1, 0.1, 5, 0.1)

lines(p, dbeta(p, 1, 1), col = "black")
lines(p, dbeta(p, 10, 2), col = "blue")
lines(p, dbeta(p, 5, 5), col = "green")

# add legend
legend("topright", c("Beta(1, 1)", "Beta(2, 10)", "Beta(10,2)", "Beta(0.5,0.5)"),
      lty = c(1, 1, 1), col = c("black", "red", "blue", "green"))
```

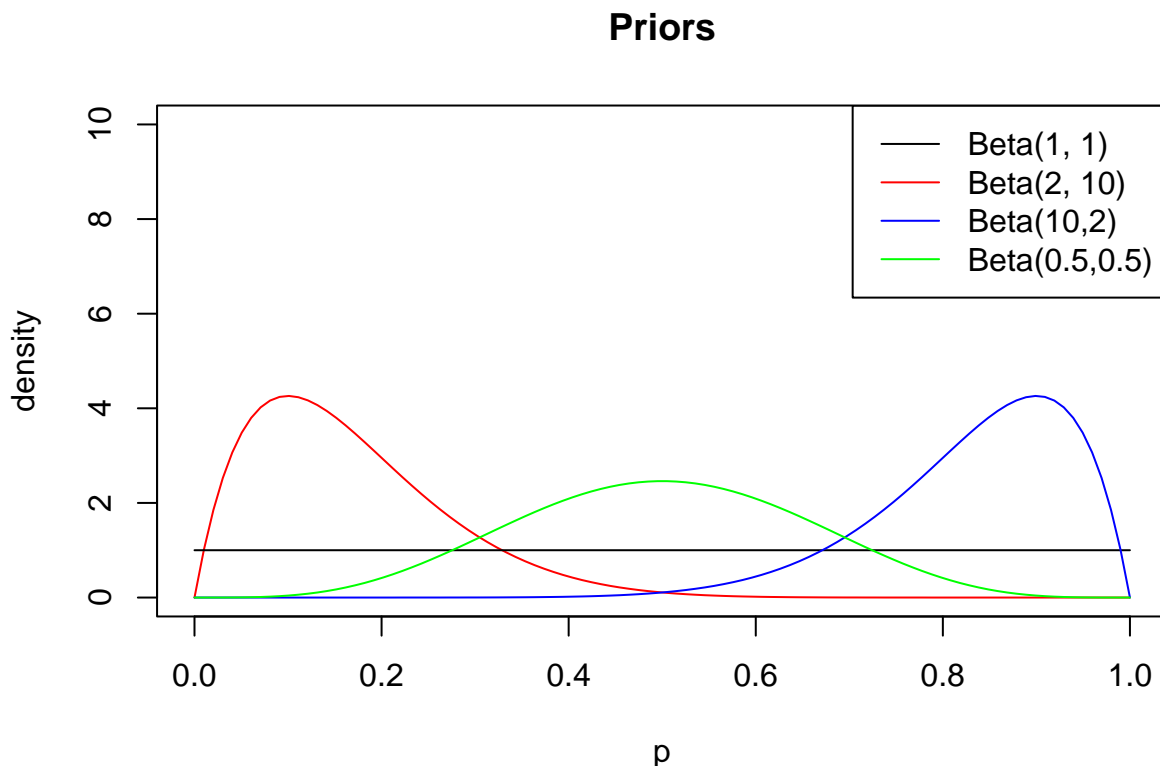


```
# define range
p = seq(0, 1, length = 100)
```

```
# plot several Beta distributions
plot(p, dbeta(p, 2, 10), ylab = "density", type = "l", col = "red", main = "Priors",
     ylim = c(0, 10))
a_0 = c(1, 0.1, 0.1, 5)
b_0 = c(1, 0.1, 5, 0.1)

lines(p, dbeta(p, 1, 1), col = "black")
lines(p, dbeta(p, 10, 2), col = "blue")
lines(p, dbeta(p, 5, 5), col = "green")

# add legend
legend("topright", c("Beta(1, 1)", "Beta(2, 10)", "Beta(10,2)", "Beta(0.5,0.5)"),
      lty = c(1, 1, 1), col = c("black", "red", "blue", "green"))
```



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 = mode², but this does not hold for the posterior as α_1 and β_1 are non-proportional linear combinations of α_0 and β_0 .

For an increase in α_0 holding β_0 constant we see an increase in all posterior point estimates. If we increase β_0 holding α_0 constant, estimates decrease.

As to sources of prior information, information for this particular problem can stem from previous research. In a similar example in class, we had the probability of having the disease given by a value of 1 in 10,000. If this is what the literature agrees upon, we could take this value into account. However, as we are rather interested in a probability distribution rather than a single estimate, we could consult the literature again and see if this value of 1 in 10,000 is simply the mean of the prior distribution. If that is so, then we know from Exercise 2.1 that the class of conjugate priors for this problem follows a Beta distribution. So, taking this into account, we could agree on values for α_0 and β_0 that give approximately a mean value of 1 in 10,000. As to additional sources of prior information, technically, we could include any type of additional prior information such as information on symptoms (coughing, sore throat, fever, etc.) that a subject might have, or locations where the subject stayed at (crowded places, inside areas, etc.), or the simple fact whether or not a subject was wearing a mask.

Exercise 2.4

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

2.4a - The iid assumption for data

Conducting n Covid-19 tests can be seen as conducting a sequence of n Bernoulli trials, where in each round we either obtain a positive (= value of 1) or a negative (= value of 0) test result. The sum of these Bernoulli trials follows a Binomial distribution, where the binomial random variable is defined as the number of “successes”. Here, we define a “success” as obtaining a positive test result. Now, for a random variable to follow a Binomial distribution, it is necessary that the probability of “success” is

- a) the same in every Bernoulli trial, i.e., it is identically distributed; and
- b) is not affected by any previous (or future) outcome, i.e., the outcome in any Bernoulli trial is independent to any other outcome of any other Bernoulli trial in the sequence.

If these two conditions do not hold, then our random variable does not follow a Binomial distribution. Hence, they are necessary.

However, in real life, things might be different, and these two conditions are likely to fail as is explained hereinafter. 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. Second, if one of our employees in the office tests positive on a particular day, then this employee will almost certainly 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. Therefore, the scope of inference based on a posterior derived using this likelihood function is limited.

2.4b - Improving the model

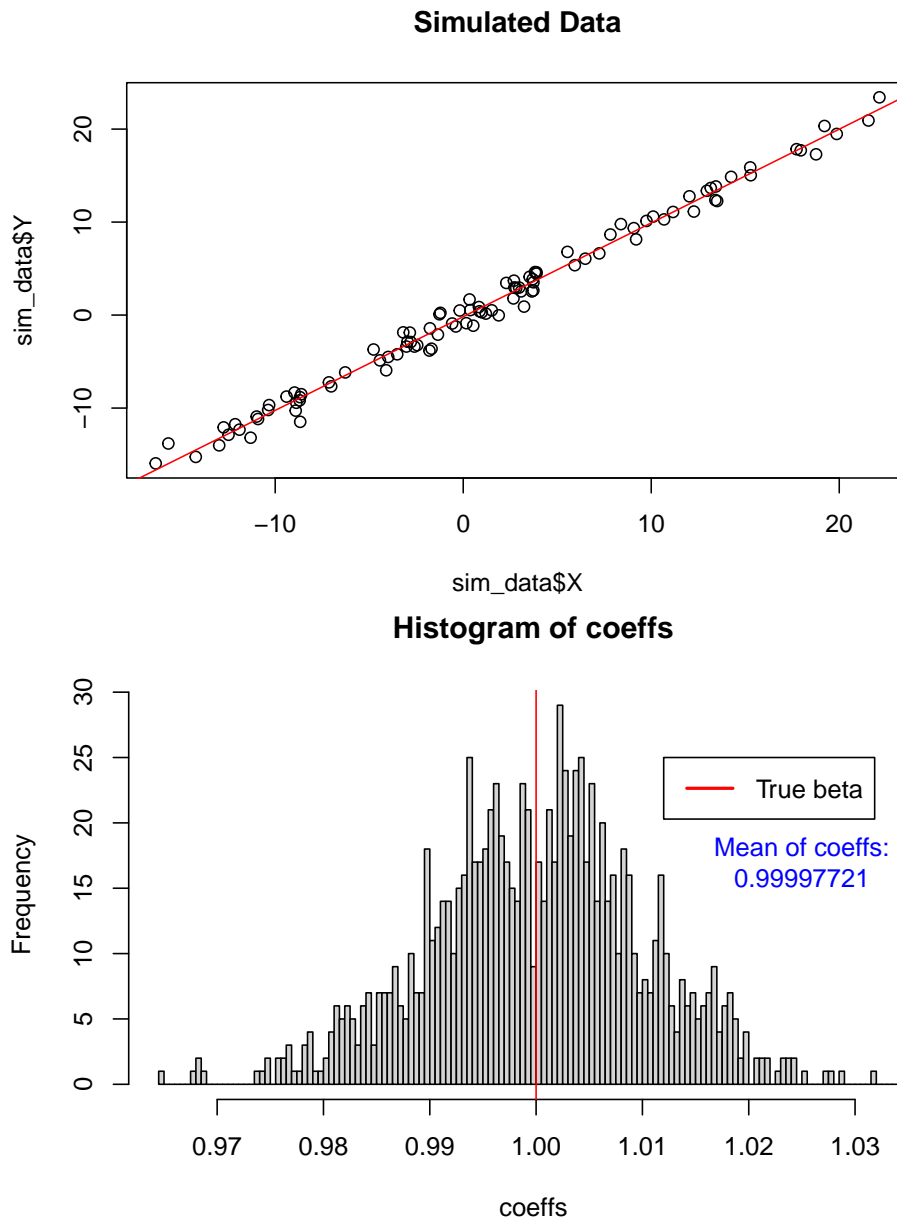
In class, we used probabilities rather than probability distributions for the Covid-19 test example. If we make use of a Bayesian point estimator of θ , e.g. the mean of the prior distribution, we can basically replicate the example from class with own values depending on our estimates from Exercise 2.3, for instance. However, rather than just picking a single point, we could also directly do it with a probability distribution. The procedure is analogous, and in the end, we could summarise the computed posterior distribution by means of a single point estimate – again the mean of the distribution, for instance.

As to re-testing, we can run several Covid-19 tests on a single individual, and update our prior information with the posterior obtained in the previous test. This procedure is called Bayesian updating. Depending on the size of the probability of a *false positive* and the *marginal probability* of having the disease, a test result obtained might be more or less informative. In class, we had a high probability of a *false positive* and a low *marginal probability* of having the disease, which led to the case that a positive test result was more informative than a negative one.

As to extending the model, we could include any type of additional prior information such as taking into account symptoms (coughing, sore throat, fever, etc.) or stays at places which are crowded, inside, etc., and whether or not a mask was being worn, as was discussed in Exercise 2.3.

Exercise 3

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



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 β 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 of σ is known

If σ is known, α and β 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.

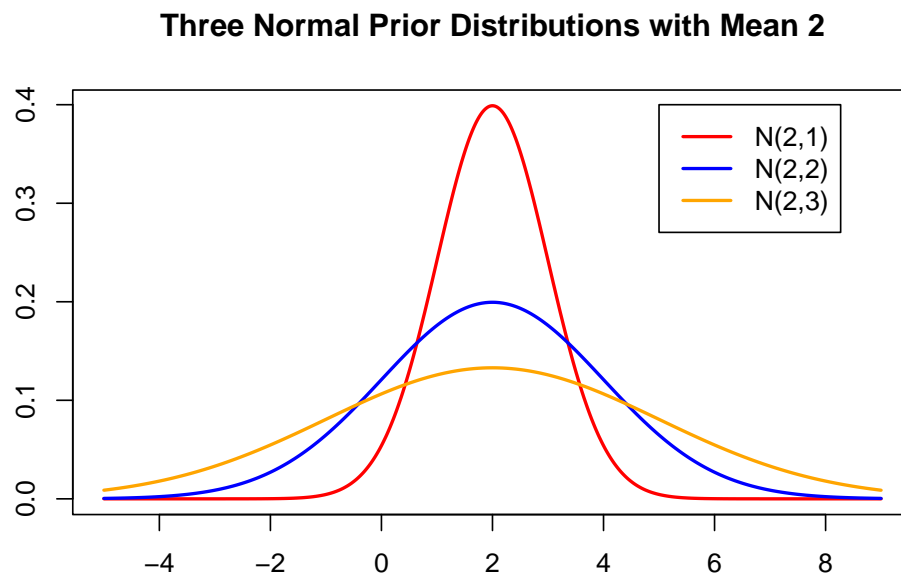
How can we come up with a prior distribution for β_1 , the parameter associated with study time?

- 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:

```
plot(x = seq(-5, 9, length = 1000), y = dnorm(seq(-5, 9, length = 1000), 2, 1), type = "l",
     lwd = 2, col = "red", main = "Three Normal Prior Distributions with Mean 2",
     xlab = "", ylab = "")
lines(x = seq(-5, 9, length = 1000), y = dnorm(seq(-5, 9, length = 1000), 2, 2),
      type = "l", lwd = 2, col = "blue")
lines(x = seq(-5, 9, length = 1000), y = dnorm(seq(-5, 9, length = 1000), 2, 3),
      type = "l", lwd = 2, col = "orange")

legend(5, 0.4, c("N(2,1)", "N(2,2)", "N(2,3)"), lwd = c(2, 2, 2), col = c("red",
"blue", "orange"))
```



Exercise 3.4 - Simulating data with known σ and a Normal prior STILL OPEN

Exercise 4

Exercise 4.1 - Estimating μ

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.

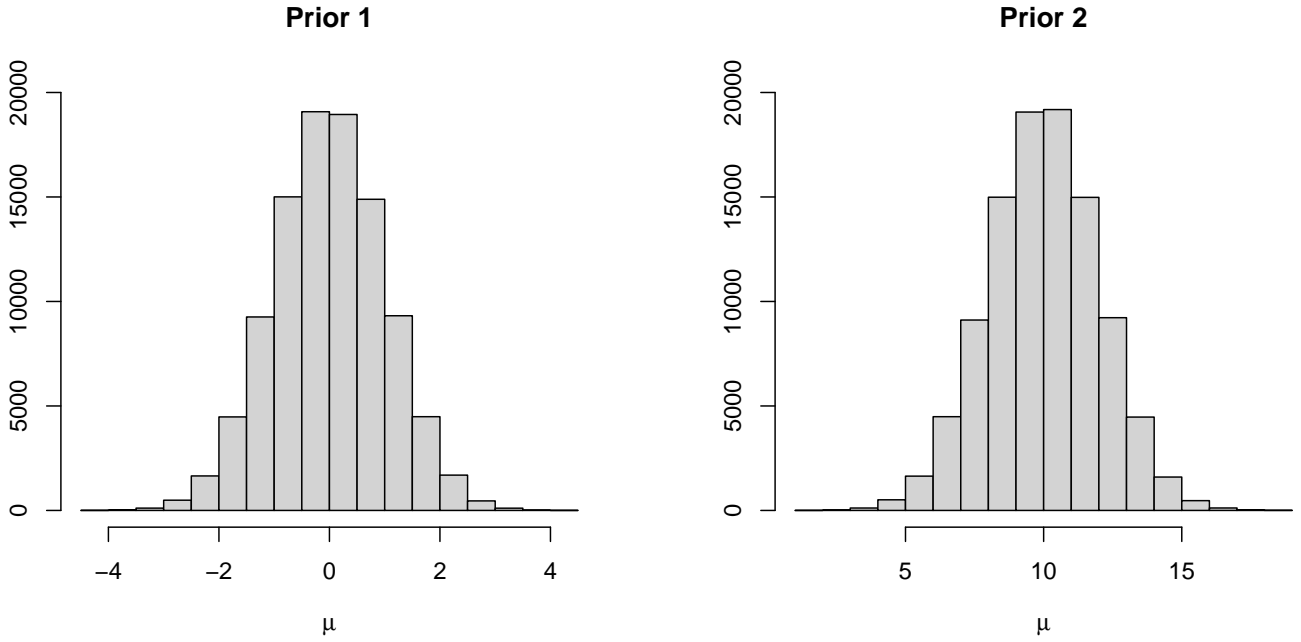
$$\begin{aligned}
 p(\mu|\mathbf{y}, 1) &\propto p(\mathbf{y}|\mu, 1) \times p(\mu|\mu_0, \sigma_0^2) \\
 &\propto \exp(-\frac{1}{2} \sum_{i=1}^n (y_i - \mu)^2) \times \exp(-\frac{1}{2} \frac{(\mu - \mu_0)^2}{\sigma_0^2}) \\
 &= \exp(-\frac{1}{2} (\sum_{i=1}^n (y_i - \mu)^2) + \frac{1}{\sigma_0^2} (\mu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

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 σ^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 $d = \sum_{i=1}^n (y_i - 5)^2/2$.

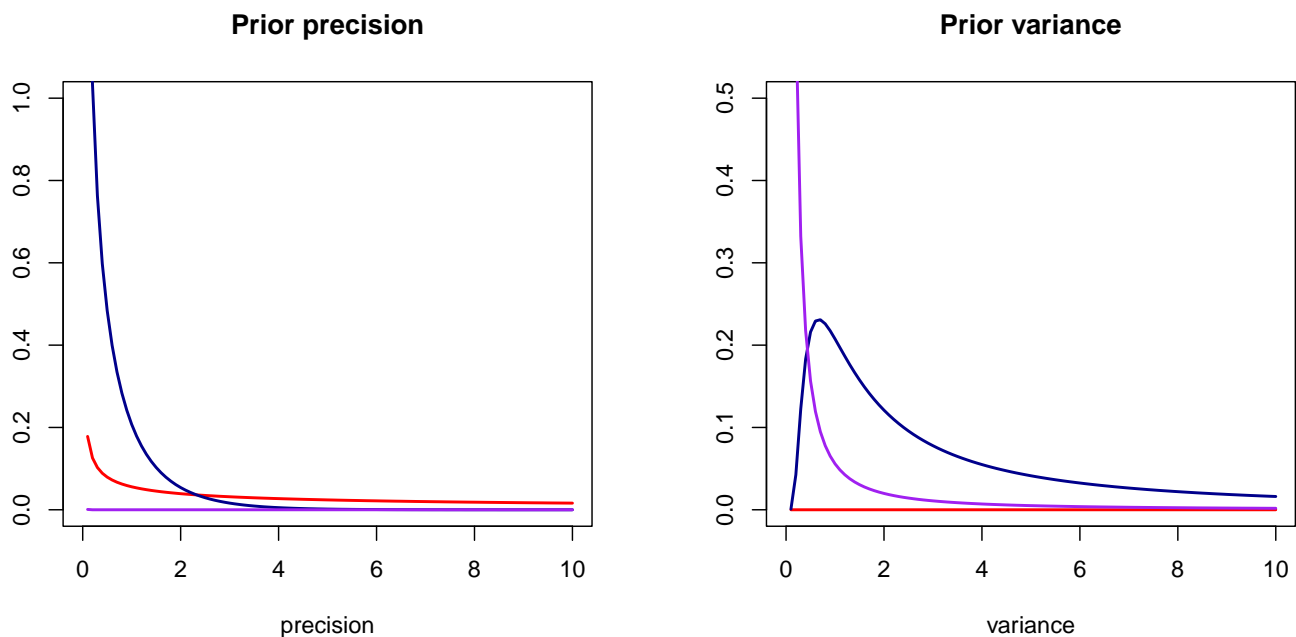
For the prior, we assume that the precision follows a $\sigma^{-2} \sim G(0.5, \eta)$. Therefore, we have $\sigma^2 \sim G^{-1}(0.5, \eta)$.³ This gives us the prior for σ^2 in the form of $p(\sigma^2|0.5, \eta) = \frac{\sqrt{\eta}}{\Gamma(0.5)} (\sigma^2)^{-1.5} \exp(-\frac{\eta}{\sigma^2})$.

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 σ^{-2} and on the right with respect to variance σ^2 .

³See derivation of Inverse-Gamma from Gamma on Wikipedia. For the Gamma distribution β denotes the *rate* parameter, whereas it denotes the *scale* parameter for the Inverse Gamma distribution!

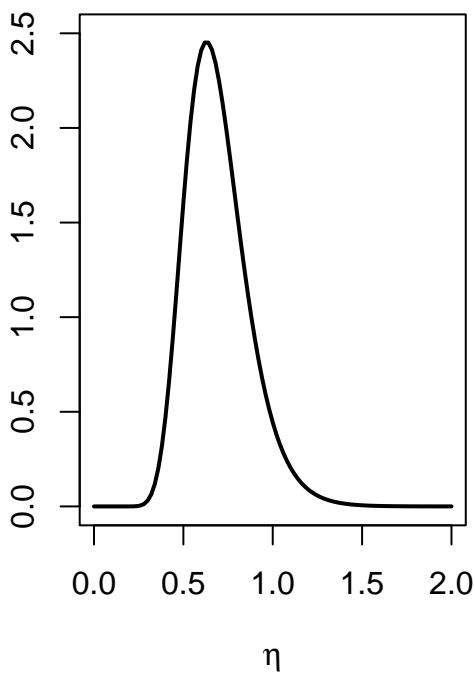


Exercise 4.3 - A prior for η

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 $\tilde{\eta}$ 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 & Inverse-Gamma. As we used those types before, we are going to use them as well for our prior distribution for $\tilde{\eta}$ is allowed to take. As the scale parameter of an Inverse-Gamma distribution has to be larger than 0, we need to find a hyperprior such that random draws result in $\eta \in (0, \infty)$. Several distributions such as the Gamma, Inverse-Gamma, or the Log-Normal distribution fulfill this criteria. Now from the type of data that we are working with, we may have some kind of an idea as to where σ^2 should be. Let's say that we are quite sure that σ^2 is most likely to be around 1, but we also want to allow for larger error terms. Then this would correspond to the density of an Inverse-Gamma function with $\alpha = 0.5$ (which is fixed by assumption) and a scale parameter η that is most of the time somewhere between 0.5 and 1. This can be achieved by drawing from $\eta \sim \text{Lognormal}(-0.4, 0.25)$ (see the density for this distribution on the left hand side of the plot below). The prior $\sigma^2|\eta$ is visualized on the right.

```
set.seed(2345)
eta.prior <- data.frame(eta = rlnorm(1000, meanlog = -0.4, sdlog = 0.25), sigma2 = NA)
for (i in 1:nrow(eta.prior)) {
  eta.prior$sigma2[i] <- invgamma::rinvgamma(1, shape = 0.5, scale = eta.prior$eta[i])
}
par(mfrow = c(1, 2))
plot(NA, NA, xlim = c(0, 2), ylim = c(0, 2.5), main = "Density: Lognormal(-0.4, 0.25)",
     xlab = expression(eta), ylab = NA)
curve(dlnorm(x, meanlog = -0.4, sdlog = 0.25), from = 0, to = 2, add = T, lwd = 2)
plot(density(eta.prior$sigma2), main = expression(paste("Prior: ", sigma^2, "|",
eta)), xlab = expression(sigma^2), ylab = NA)
```

Density: Lognormal(−0.4, 0.25)



Prior: $\sigma^2|\eta$

