

UPPSALA UNIVERSITY



BAYESIAN STATISTICS AND DATA ANALYSIS

Assignment 6

General information

- The recommended tool in this course is R (with the IDE R-Studio). You can download R [here](#) and R-Studio [here](#). There are tons of tutorials, videos and introductions to R and R-Studio online. You can find some initial hints from [RStudio Education pages](#).
- When working with R, we recommend writing the report using R markdown and the provided [R markdown template](#). The remplate includes the formatting instructions and how to include code and figures.
- Instead of R markdown, you can use other software to make the PDF report, but the the same instructions for formatting should be used. These instructions are available also in [the PDF produced from the R markdown template](#).
- We supply a Google Colab notebook that can also be used for the assignments. We have included installation of all nessecary R packages and hence this can be an alternative to using your own local computer. You can find the notebook [here](#). You can also open the notebook in Colab [here](#).
- Report all results in a single and *anonymous* pdf.
- The course has its own R package **bsda** with data and functionality to simplify coding. To install the package just run the following (upgrade="never" skips question about updating other packages):

```
1. install.packages("remotes")
2. remotes::install_github("MansMeg/BSDA",
  subdir = "rpackage", upgrade="never")
```

- Many of the exercises can be checked automatically using the R package **markmyassignment**. Information on how to install and use the package can be found [here](#). There is no need to include **markmyassignment** results in the report.
- Common questions and answers regarding installation and technical problems can be found in [Frequently Asked Questions \(FAQ\)](#).
- Deadlines and information on how to turn in the assignments can be found in Studium.
- You are allowed to discuss assignments with your friends, but it is not allowed to copy solutions directly from other students or from internet. Try to solve the actual assignment problems with your own code and explanations. Do not share your answers publicly. Do not copy answers from the internet or from previous years. We compare the answers with urkund. All suspected plagiarism will be reported and investigated.
- If you have any suggestions or improvements to the course material, please post in the course chat feedback channel, create an issue, or submit a pull request to the public repository [here](#)

Information on this assignment

This assignment is related to Chapters 10 and 11.

Reading instructions: Chapter 10 and 11 in BDA3, see reading instructions.

Reporting accuracy: For posterior statistics of interest, only report digits for which the Monte Carlo standard error (MCSE) is zero. *Example:* If you estimate $E(\mu) = 1.234$ with $\text{MCSE}(E(\mu)) = 0.01$, you should report $E(\mu) = 1.2$.

Installing and using stan: To install Stan on your laptop, <https://github.com/stan-dev/rstan/wiki/RStan-Getting-Started>. If you encounter problems, see additional answers in the [FAQ](#). Recently there have been reports of installation problems with Windows and R 4.0 (see Stan discourse for more).

Installing and using CmdStanR: If you want to use Stan in R on local computer, it can be easier to install CmdStanR interface mc-stan.org/cmdstanr/.

Additional useful packages are `loo`, `bayesplot` and `shinystan` (but you don't need these in this assignment). For Python users, `PyStan`, `CmdStanPy`, and `Arviz` packages are useful.

Stan manual can be found at <https://mc-stan.org/users/documentation/>. From this website, you can also find a lot of other useful material about Stan.

1. Generalized linear model: Bioassay with Stan

Replicate the computations for the bioassay example of section 3.7 (BDA3) using Stan.

1. Write down the model for the bioassay data in Stan syntax. For instructions in reporting your implementation, you can refer to parts 2 c) - g) in Assignment 5. More information on the bioassay data can be found in Section 3.7 of the course book. To get access to data, use the following code:

```
library(bsda)
data("bioassay")
```

Use the Gaussian prior

$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix} \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0), \quad \text{where} \quad \boldsymbol{\mu}_0 = \begin{bmatrix} 0 \\ 10 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}_0 = \begin{bmatrix} 2^2 & 12 \\ 12 & 10^2 \end{bmatrix}.$$

Hint! You will need Stan functions `multi_normal` and `binomial_logit` for implementing the prior and observation model, respectively. In Stan code, it is easiest to declare a variable (say `theta`) which is a two-element vector so that the first value denotes α and latter one β . This is because the `multi_normal` function that you need for implementing the prior requires a vector as an input.

2. Use \hat{R} for convergence analysis. You can either use Eq. (11.4) in BDA3 or the later version that can be found [here](#). You should specify which \hat{R} you used. In R the best choice is to use function `Rhat` from package `rstan` (see `?rstan::Rhat`). To check \hat{R} and other diagnostics, you can also call `monitor(fit)`, where `fit` is the fit object returned by Stan's sampling function. Report the \hat{R} values both for α and β and discuss the convergence of the chains. **Briefly explain in your own words how to interpret the obtained \hat{R} values.**
3. Plot the draws for α and β (scatter plot) and include this plot in your report. You can compare the results to Figure 3.3b in BDA3 to verify that your code gives sensible results. Notice though that the results in Figure 3.3b are generated from posterior with a uniform prior, so even when your algorithm works perfectly, the results will look slightly different (although fairly similar).