Report submissions are accepted in the following formats: One report file in pdf format denoted **name_lab3.pdf**. *Also submit an email with your R-files, with a file named* **name_lab3.R** *that can be used to run your analysis, remember to submit* **all** *of the files you have created, and also that the codes are possible to run. Email the files to* **jonas.wallin@stat.lu.se**.

Discussion between groups is permitted (and encouraged), as long as your answers and code reflects your own work.

Deadline: Friday 8/12 at 23.59

Bullets, indicates mandatory exercises, whereas star indicates voluntarily.

Tutorial

Go through and make sure you understand the tutorial about Rstan on the homepage. tutorial about Rstan on the homepage.

Fitting Bayesian models in Rstan

- Do exercise 10H3 in the book. You should use rstan to solve the exercise, that is you not allowed to use map2stan or map for fitting the model .

  [7.5p]

- In this exercise you are supposed to join data from several different sources together to preform an Bayesian analysis. Here we study the relation between mass shootings and legislation in the US. The file mass_shootings_2016.csv contains reported mass shooting in the US 2016, the data is originally from GunViloence. From the site statefirearmlaws we have information about protective laws by state in the file lawdata.csv. The file state_pol contains data on what the stated voted in the presidential election. Finally in the library **usmap** contains state populations in the data **statepop**.

  - Structure the data so it can be analyzed:
    - ⋆ Go through the tutorial on the homepage. Modify the mass shooting data so you have number deaths by state.
    - ⋆ For the law data remove everything but the year 2016 data and you will need the columns state and lawtotal.
    - ⋆ Merge all dataset into one **data.frame** or **list**.
  - Fit a Poisson model for number of people killed by state. As covariates use number of protection laws in the state, how the state voted in the presidential election and number of people living in the state. Think if some covariates needs to be transformed given the link function you use. Which effects are significant (for chosen level)? Does the posterior sample mix well?
  - Fit a Poisson model for number of protection laws by state. Choose some covariates you feel are sensible. Does number people killed by mass shooting have an effect on number of protection laws?

  [7.5p]

Markov Chain Monte Carlo

*) In this exercise you are going to build your own MCMC algorithm, more specifically a Metropolis hasting random walk. The file poiss.txt contains 20 observation from a Poisson random variable. Set up a simple Bayesian model with a, sensible, prior of your choice for the mean parameter.

- Write a Metropolis-Hastings random walk (MHRW) algorithm to samples from the posterior distribution of your model given the data. Check if the samples from the algorithm are reasonable.
- Test the MHRW uses one parameter, σ, which be tuned to improve the mixing of the MCMC algorithm. Examine various values of σ see how it effects the mixing of the MCMC algorithm. Visualize with a figure the relation the σ parameter and the autocorrelation between the consecutive samples in your MCMC algorithm.

[5p]

Tobit regression

*) In this exercise you are supposed to build a non-standard regression model. In 1978 C. Fair developed a model for extramarital affairs using tobit regression. In the article comes from magazine survey (which is very unreliable so no real conclusions can be made from the data). The data is one the homepage in the file Affair_data.txt The data is $y_i = \frac{q_i}{v_i}$, where q_i is the number of times the persons (only females) had sexual relation outside marriage, and v_i is the number of years in the marriage. For covariates we have the age of the persons, number of years married, happiness (scale 1-5) and intercept. The original article is also available at the live@lund homepage.

- First fit a regular regression model for the data, using suitable priors for the variance and regression factors. Report the posterior distributions of the covariates, which effects are significant?

After this you are supposed to fit a Tobit regression which means that you truncate all data above and below a certain level. That is you have your unobserved regression

$$y_i^* \sim N(\beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2 + x_{i3}\beta_3 + x_{i4}\beta_4, \sigma^2),$$

however you only observe

$$y_i = \begin{cases} y_i & \text{if } y_i^* \in (a, b), \\ b & \text{if } y_i^* \geq b, \\ a & \text{if } y_i^* \leq a. \end{cases}$$

- Read the section about Estimating Censored values in Stan manual section 12.3, on page 188. And use this to implement the Tobit regression for the affairs data, with $a = 0$ and $b = 5$.
- Compare the result to the regular model, how does it affect the output.

[5p]