

## Bayesian data analysis – exercise 7

The maximum amount of points from this assignment is 6. In addition to the correctness of the answers, the overall quality and clearness of the report is evaluated.

Report all results to a single, **anonymous** \*.pdf -file and return it to peergrade.io. Include also source code to the report (either as an attachment or as a part of the answer). By anonymity it is meant that the report should not contain your name or student number.

### 1. Linear model: drowning data with Stan (3p)

The provided `drowning.txt` file contains the number of people drown per year in Finland 1980–2013. Using Stan, fit a linear model with Gaussian noise to these data (using time as the predictor and number of drownings as the target variable), and answer the following questions:

- i) What can you say about the trend in the number of people drown per year? Plot the histogram of the slope of the linear model.
- ii) What does the model predict for year 2019? Plot the histogram of the posterior predictive distribution for number of people drown at  $\tilde{x} = 2019$ .

Use uniform prior for the intercept, slope and noise variance.

Hints and further advice:

- See the linear model example for the Kilpisjärvi-temperature data in the example Stan codes. What you need to do is essentially just change the dataset.
- You need to implement the prediction for year 2019. This can be done using the `generated quantities` block in the Stan code, as in the demo example. Alternatively, you can perform the sampling from the predictive distribution outside Stan, given the posterior samples provided by Stan.

### 2. Hierarchical model: factory data with Stan (3p)

The provided `factory.txt` file contains quality control measurements from 6 machines in a factory (units of the measurements are irrelevant here). Quality control measurements are expensive and time-consuming, so only 5 measurements were done for each machine. In addition to the existing machines, we are interested in the quality of another machine (the seventh machine). Implement a separate, pooled and hierarchical Gaussian model described in Section 11.6 using Stan. Similarly as in the model description in the book, use the same measurement standard deviation  $\sigma$  for all the groups in the hierarchical model. In the separate model however, use separate measurement standard deviation  $\sigma_j$  for each group  $j$ . Using each of the three models – separate, pooled, and hierarchical – report (plot histogram and comment on the results):

- i) the posterior distribution of the mean of the quality measurements of the sixth machine

- ii) the predictive distribution for another quality measurement of the sixth machine
- iii) the posterior distribution of the mean of the quality measurements of the seventh machine.

Use Stan's default uniform prior for all the parameters.

Hints and further advice:

- In the data file, each column contains measurements for a single machine. In the pooled model all measurements are combined together and in the separate model each machine has its own model.
- See the example Stan-codes for the comparison of  $k$  groups with and without the hierarchical structure. What you need to do is change the dataset, implement the prediction for the future measurement of the sixth machine, and figure out the distribution for the mean of the quality measurements for the seventh machine in the hierarchical model.