Bayesian data analysis – exercise 8

This assignment is related to Chapter 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.

Model assessment: LOO-CV for factory data with Stan (6p)

Use leave-one-out cross validation (LOO-CV) to assess the predictive performance of the pooled, separate and hierarchical Gaussian models for the factory dataset (see the second exercise in Assignment 7). Use Stan for fitting the models, and the provided PSIS-LOO (Pareto smoothed importance sampling LOO) code for computing the approximate LOO-CV given the posterior samples provided by Stan. Your results should include:

- PSIS-LOO values, the effective number of parameters p_{eff}, and the k-values for each of the three models
- an assessment of how reliable the PSIS-LOO estimates are for the three models based on the k-values
- an assessment of whether there are differences between the models, and if so, which model should be selected according to PSIS-LOO

Hints and further advice:

- Fit the models with Stan as instructed in Assignment 6.
- In R, you can use the function `loo` in package `loo`. You can install the package with command `install.packages(loo)`. The Python and Matlab implementations are available at `https://github.com/avehtari/PSIS`. For both languages, the function that you will need is `psisloo`.
- In order to use the `loo` or `psisloo` functions, you need to compute the log-likelihood values of each observation for every posterior draw (i.e. an S-by-n matrix, where S is the number of posterior draws and $n = 30$ is the total number of observations). This can be done in the `generated quantities` block in the Stan code; for a demonstration, see the Gaussian linear model in the R/Python/Matlab Stan examples (file `lin.stan` in R and Python demos).
- It will be convenient to visualize the k-values for each model, so that you can easily see how many of these values fall in the range $k > 0.7$ to assess the reliability of the PSIS-LOO estimate for each model. You can read more about the theoretical guarantees for the accuracy of the estimate depending on k from the original article (see the link below), but regarding this assignment, it suffices to understand that if all the k-values are $k \lesssim 0.7$, the

PSIS-LOO estimate can be considered to be reliable, otherwise there is a concern that it may be biased (too optimistic, overestimating the predictive accuracy of the model).

- The estimated effective number of parameters in the model can be computed from equation (7.15) in the book, where $\text{lppd}_{\text{loo-cv}}$ is the PSIS-LOO value (sum of the LOO log densities) and lppd is given by equation (7.5) in the book.
- PSIS-LOO is a recently developed method for approximating the exact LOO and is thus not in BDA3. For more information, see the lecture slides and the original paper `https://arxiv.org/pdf/1507.04544`.