Bayesian data analysis – exercise 9

This assignment is related to Chapter 9.

The maximum amount of points from this assignment is 3. 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.

Decision analysis for the factory data (3p)

This exercise is an example of decision analysis (DA). In broad context, this means optimizing over different decisions that lead to different outcomes that all have different utilities. In Bayesian context, this means using posterior distributions to make decisions.

In this exercise, you work as a data analyst in the company that owns the six machines that have produced the data in `factory.txt` (found from here: `https://github.com/avehtari/BDA_course_Aalto/blob/master/exercises/additional_files/factory.txt`). Your task is to decide whether or not to buy a new (7th) machine for the company. The decision should be based on our best knowledge about the machines.

The following is known about the production process:

- The given data contains quality measurements of single products from the six machines that are ordered from the same seller. (columns: different factories, rows: measurements)

- Customers pay 200 euros for each product.

  - If the quality of the product is below 85, the product cannot be sold

  - All the products that have sufficient quality are sold.

- Raw-materials, salary of the machine user and the usage cost of the machine for each product cost 106 euros in total.

  - Usage cost of the machine also involves all investment and repair costs divided by the number of products a machine can create. So there is no need to take the investment cost as a separate factor.

- The only thing the company owner cares about is money. Thus, as a utility function, use the profit of a new product from a machine.

Your task is the following:

- For each of the six machines, compute and report the expected utility of the products of that machine.

- Rank the machines based on the expected utilities (in other words order the machines from worst to best) and explain briefly what the utility values tell about the quality of these machines. E.g. Tell which machines are profitable and which are not (if any).

- Compute and report the expected utility of the products of a new (7th) machine.

- Based on your analysis, discuss briefly whether the company owner should buy a new (7th) machine.

- As usual, remember to include the source code (for both Stan and R/Python)!

Hints and further information:

- As noticed in the previous assignment, the hierarchical model fits best to the dataset, so use it to compute the utilities.

- Assume that the quality of the products of different machines have a common variance.

- Use uniform priors for hyper-parameters (population mean, population variance and common variance) as in previous assignment.

- Keep in mind that it is not necessary to compute the expected utility of the machines in Stan. Instead, you can take the posterior predictive samples given by Stan and use those in R/Python. This allows you to debug your Stan and R/Python codes separately and not having to always rerun the Stan sampler.

The assumptions for the hierarchical model are exactly the same as in the previous assignments! If you did things correctly then, solving the assignment only requires you to change the "generated quantities"-block to compute the correct predictive samples for products of all 7 (=6+1) machines.