

Bayesian data analysis – assignment 2

This exercise is related to Chapters 1 and 2.

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.

You may find an additional discussion about choosing priors by Andrew Gelman useful <http://andrewgelman.com/2017/10/04/worry-rigged-priors/>.

Data file used in this exercise can be found in https://github.com/avehtari/BDA_R_demos/blob/master/data/algae.txt

Inference for binomial proportion (Computer)

Algae status is monitored in 274 sites at Finnish lakes and rivers. The observations for the 2008 algae status at each site are presented in file `algae.txt` ('0': no algae, '1': algae present). Let π be the probability of a monitoring site having detectable blue-green algae levels.

Use a binomial model for observations and a $\text{Beta}(2, 10)$ prior for π in Bayesian inference. Formulate Bayesian model likelihood $p(y|\pi)$, prior $p(\pi)$, and the resulting posterior $p(\pi|y)$. Here it is not necessary to derive the posterior distribution as it has already been done in the book. Also it is not necessary to write out the distributions; label-parameter format suffices. Although recommended, plotting is not required in this exercise. Use your model to answer the following questions:

- What can you say about the value of the unknown π according to the observations and your prior knowledge? Summarize your results using point and interval estimates (figure and numerical values).
- What is the probability that the proportion of monitoring sites with detectable algae levels π is smaller than $\pi_0 = 0.2$ that is known from historical records?
- What assumptions are required in order to use this kind of a model with this type of data?
- Make prior sensitivity analysis by testing a couple of different reasonable priors. Summarize the results by one or two sentences.

Hint: With a conjugate prior, a closed form posterior is Beta form (see equations in the book). Useful functions: `dbeta`, `pbeta`, `qbeta` in R and `pdf`, `cdf` and `ppf` from `scipy.stats.beta` in Python.