Bayesian data analysis – exercise 4

This exercise is related to Chapters 3 and 10.

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.

Generalized linear model: Bioassay with importance sampling

In the bioassay example (Chapter 3 in the book), replace the uniform prior density by a joint normal prior distribution on (α, β), with $\alpha \sim N(0, 2^2)$, $\beta \sim N(10, 10^2)$, and $\text{corr}(\alpha, \beta) = 0.5$.

Repeat parts of the computations and plots of Section 3.7 with this new prior distribution and using importance sampling and resampling.

- Calculate the posterior density in a grid of points around the prior (α: 0 ± 4, β: 10 ± 20) and plot a heatmap of the density.

- Sample draws of α and β from the prior distribution.

- Compute the importance ratios for each draw (target distribution is the posterior).

- Using the importance ratios, compute the effective sample size S_{eff} and report it. If S_{eff} is less than 1000, repeat the importance sampling with more draws.

- *BDA3 1st (2013) and 2nd (2014) printing have an error for $\tilde{w}(\theta^s)$ used in the effective sample size equation 10.4. The normalized weights equation should not have the multiplier S (the normalized weights should sum to one). Errata for the book http://www.stat.columbia.edu/~gelman/book/errata_bda3.txt. The later printings and slides have the correct equation.*

- Compute the posterior mean using importance sampling and report it.

- Use importance resampling to obtain a posterior sample of size 1000 of α and β.

- Using the posterior sample obtained via importance resampling:

  - Plot a scatterplot of the obtained posterior sample and compare that to the heatmap plotted earlier.

  - Report an estimate for $p(\beta > 0|x, n, y)$, that is, the probability that the drug is harmful.

  - Draw a histogram of the draws from the posterior distribution of the LD50 conditional on $\beta > 0$ (see Figure 3.4).

Hints

- See `demo3_6.R` / `demo3_6.ipynb` and `demo10_3.R` / `demo10_3.ipynb`

- Compute the posterior density and importance ratios as logarithms, and exponentiate at the end

- Scale the log-posterior by subtracting its maximum value before exponentiating

- Normalize the posterior and importance ratios to sum to 1

- Possibly useful R functions: `dmvnorm` and `rmvnorm` from the `mvtnorm` package

- Possibly useful Python functions: `scipy.stats.multivariate_normal, numpy.meshgrid, numpy.ravel`

- You may use the `bioassaylp.R` and `bioassaylp.py` from the `additional_files` folder to compute the log-density of the likelihood at any point (α, β)

- When computing the importance ratios, you will have to take logarithms of very small numbers, which may round to zero. If this happens, replace the small numbers with `1e-12` to prevent `log(0) = NaN` (not a number).