STA623 - Bayesian Data Analysis - Assignment 1

26 August - 1 September 2023

Marc Henrion

Assigment

Please email your typed or scanned solutions before 23:59 on Monday 2 October 2023 to BOTH

- `mhenrion@mlw.mw`
- `biostat-unima@unima.ac.mw`

Please include `STA623 - Assignment 1` in the subject line.

While we used `R` and JAGS in the classroom, you can use any software package of your liking to fit models for this assignment. Please include your code, model output and graphs. Please comment any submitted code.

Notation

Please try to use the following notation where possible.

- X, Y, Z - random variables
- x, y, z - measured / observed values
- $\bar{X}$, $\bar{Y}, \bar{Z}$ - sample mean estimators for X, Y, Z
- $\bar{x}$, $\bar{y}, \bar{z}$ - sample mean estimates of X, Y, Z
- $\hat{T}$, $\hat{t}$ - given a statistic T, estimator and estimate of T
- $P(A)$ - probability of an event A occurring
- $f_X(.)$, $f_Y(.), f_Z(.)$ - probability mass / density functions of X, Y, Z

- p(.) - used as a shorthand notation for pmfs / pdfs if the use of this is unambiguous (i.e. it is clear which is the random variable)
- $X \sim F$ - X distributed according to distribution function F
- $E[X]$, $E[Y]$, $E[Z]$, $E[T]$ - the expectation of X, Y, Z, T respectively

Exercise

Assume you observe some data $y_1, \dots, y_n$ for exponentially distributed random variables $Y_1, \dots, Y_n \sim \text{Exp}(\lambda)$.

1. Derive the conjugate prior distribution for this sampling model and state what the resulting posterior distribution is. [16 marks]
2. Derive the Jeffreys prior for this sampling model and state whether this is a proper prior distribution or not. [16 marks]
3. Now assume a $\Gamma(a, b)$ prior distribution. Derive the posterior predictive distribution for new data $\tilde{Y}$. [16 marks]
4. Write computer code (and submit a print-out of this code with your assignment), assuming you had a data object `dat` loaded in memory that would allow you to fit the model resulting from a $\Gamma(a, b)$ prior and an $\text{Exp}(\lambda)$ sampling model. Fit your model to the data $y_i, i = 1, \dots, 20$ given below: [16 marks]

$$0.3245, 0.1982, 0.2751, 0.0278, 0.3608, 0.3463, 0.1237, 0.0807, 0.0291, 0.0780$$

$$1.6237, 2.0366, 0.5749, 1.6216, 1.0214, 0.3318, 0.5985, 0.2696, 0.3420, 0.6647$$

5. Do some diagnostic checks on the results: show the trace plot for $\lambda | y_1, \dots, y_n$ and plot an estimate of the posterior based on the MCMC results. Compute the Gelman-Rubin potential scale reduction factor. Do you see evidence for non-convergence? [10 marks]
6. Interpret your results:

- What is the posterior mean of $\lambda | y_1, \dots, y_n$? [5 marks]
- What is the posterior median of $\lambda | y_1, \dots, y_n$? [5 marks]
- Compute a 95% Bayesian confidence interval for your posterior estimate of $\lambda | y_1, \dots, y_n$. [8 marks]
- How does your prior compare to your posterior? [4 marks]
- Do your computational results agree with the theoretical posterior distribution from question 1 above? [4 marks]