STA623 - Bayesian Data analysis - Assignment 2

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 2` 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

Download the file `hospitalWaitTimes.csv` from the STA623 GitHub page.

This dataset contains information on waiting times from the A&E department in several hospitals.

The dataset contains the following columns:

- `pid` - this is just an anonymised patient identification number
- `sex` - this records the biological sex of each patient
- `triage` - records the category that the patients were triaged into by an admission nurse (`emergency`, `priority` or `queue`); the idea is that emergencies get seen without delay, priority cases get seen more quickly than normal cases and then the third category is for all other cases
- `hospital` - this records an identification code for the hospital where each patient was seen
- `wait` - this records the waiting time (in hours) that each patient had to wait before being seen by a A&E doctor

Use JAGS to fit the following model, choosing priors of your own choosing for each parameter, writing $Y_{i,j}$ for the waiting time variable for patient $i = 1, \dots, n$ seen in hospital $j = 1, \dots, k$:

$$Y_{i,j} = \beta_0 + \beta_1 \cdot male\_sex_i + \beta_2 \cdot triage\_emergency_i + \beta_3 \cdot triage\_priority + \mu_j + \epsilon_i$$

where $\mu_j \sim \mathcal{N}(0, \rho^2), j = 1, \dots, k$ and $\epsilon_i \sim \mathcal{N}(0, \sigma^2), i = 1, \dots, n$

1. Explain the choice of prior distributions for all model parameters $(\beta_0, \beta_1, \beta_2, \beta_3, \rho^2, \sigma^2)$. [15 marks]
2. Write JAGS model code to fit the model. [40 marks]
3. Fit the model, then show and summarise (as a point estimate + confidence interval) the posterior distributions for the various parameters. Explain your choice of Bayesian estimators you report. [15 marks]

4. Show trace plots and histograms for all model parameters and compute the effective sample size and Gelman-Rubin potential scale reduction factors. Discuss the results you are getting. [20 marks]

5. Discuss other model checks you could do. [10 marks]