STA623 - Bayesian Data Analysis - Assignment 1

22 - 26 September 2025

Marc Henrion

Assigment

Please email your typed or scanned solutions before 23:59 on Monday 24 November 2025 to BOTH `mhenrion@mlw.mw` and `biostat-unima@unima.ac.mw`.

Please include `STA623 - Assignment 1` in the subject line. 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
- $X \sim F$ - X distributed according to distribution function F
- $E[X]$, $E[Y]$, $E[Z]$, $E[T]$ - the expectation of X, Y, Z, T respectively

Table 1: Please use the random seed associated with your name / ID. Solutions using other data than those generated using your seed will not be accepted.

Student	ID	Seed
Eric Mangani	MSC/BIO/STAT/08/23	1899
Satiel Ngwira	MSC/BIO/STAT/17/23	1845
Ausbin Kutumani	MSC/BIO/STAT/J/01/25	1846
Chikondi Moyo	MSC/BIO/STAT/J/03/25	1608
Kenneth Kachiphaphi	MSC/BIO/STAT/J/04/25	1316
Steven Kaunda	MSC/BIO/STAT/J/06/25	1408
Felix Msamira	MSC/BIO/STAT/J/07/24	1005
Eliams Moyo	MSC/BIO/STAT/S/02/24	2616
Loveness Soko	MSC/BIO/STAT/S/04/24	2587
Filudi Nakutuwa	MSC/BIO/STAT/S/07/24	2472
Ephat Chitsulo	MSC/BIO/STAT/S/08/24	2100
Alex Kachitsa	MSC/BIO/STAT/S/09/24	1970
Steven Chiyembe	MSC/BIO/STAT/S/10/2024	2387
Charity Hamuza	MSC/BIO/STAT/S/12/24	2268
Cassim Nanyumba	MSC/BIO/STAT/S/13/24	1935
Hastings Malunga	MSC/BIO/STAT/S/14/24	1296
Osward Kaposa	MSC/BIO/STAT/S/15/24	1472
Seti Evance	MSC/BIO/STAT/S/16/24	1344
Edward Kamphongwe	MSC/BIO/STAT/S/17/24	2184
Chikondi Banda	MSC/BIO/STAT/S/19/24	2688
Steven Nanga	MSC/BIO/STAT/S/23/24	1920
Chikumbutso Banda	NA	1560

Exercise

For the exercise below, you will need to specify a seed value. You will be given individual seed numbers according to the table on the previous page. **You have to use your own individual seed value** – your data (and hence your results) will be unique to you and different from those of your colleagues.

Assume you observe some data $y_1, \ldots, y_n$ for the waiting times (in hours) from arrival to be seen by a doctor at a large hospital's A&E department.

1. Why could an Exponential(λ) sampling model be a reasonable assumption? [5 marks]

For the rest of this exercise, assume that the data are exponentially distributed:

$$Y_1, \ldots, Y_n \sim \text{Exp}(\lambda)$$

2. Run the code below to generate the `dat` data frame. In the first line, you have to specify a random seed. You are each given a different seed value (meaning no two of you have the same dataset). **Be sure to change the first line to include your individual seed value!** Print out the number of data observations in your dataset, the average lambda value used for your dataset, the averae waiting time $\bar{y}$ (as per the `wait` column in the `dat` data frame) and the number of male patients (as per variable `sex`). [5 marks]

```
set.seed(0000) # REPLACE 0000 with your individual seed value!
# Solutions using the seed value 0000 will not be accepted.

n<-rpois(n=1,lambda=100)
es<-rnorm(1,mean=1.2,sd=0.075)

sex<-sample(x=c("Male","Female"),size=n,prob=c(0.5,0.5),replace=TRUE)
lambda<-rgamma(n=n,shape=10,rate=ifelse(sex=="Male",1.5,1.5*es))

dat<-data.frame(
  sex=sex,
  wait=rexp(n=n,rate=lambda)
)
```

3. Write computer code (and submit a print-out of this code with your assignment) that fits the model resulting from a $\Gamma(a, b)$ prior and an Exp(λ) sampling model to the data `dat`. You can choose your own values `a,b` for the prior. Make sure the model estimates WAIC while sampling. [20 marks]

4. Do some diagnostic checks on the results: show the trace plot for $\lambda|y_1, \ldots, y_n$ 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? [20 marks]

5. Interpret your results:

- What is the posterior mean of $\lambda|y_1, \ldots, y_n$? [5 marks]
- What is the posterior median of $\lambda|y_1, \ldots, y_n$? [5 marks]
- Compute a 95% Bayesian confidence interval for your posterior estimate of $\lambda|y_1, \ldots, y_n$. [15 marks]
- How does your prior compare to your posterior? [5 marks]
- Do your computational results agree with the theoretical posterior distribution from question 1 above? [5 marks]

6. Now assume that the rate parameter λ depends on the patient's sex:

$$\log(\lambda) = \beta_0 + \beta_1 \cdot x_{female}$$

where $x_{female} = 1$ if the patient is female, 0 otherwise. Assume a $\mathcal{N}(0.5, 5)$ prior for β_0 and a $\mathcal{N}(0, 5)$ prior for β_1. Write code that fits the resulting model using MCMC and compare the posterior distributions of $\lambda_{male} = \exp(\beta_0)$ and $\lambda_{female} = \exp(\beta_0 + \beta_1)$. Compare the WAIC from this model to the WAIC from the model you fitted in question 3. Which model do you recommend? [15 marks]