STA623 - Bayesian Data Analysis - Assignment 1

28 October - 1 November 2024

Marc Henrion

Assigment

Please email your typed or scanned solutions before 23:59 on Monday 2 December 2024 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
Hlungumazi Ngwira	MSC-BIO-STAT-03-22	2412
Abdul Hamza	MSC-BIO-STAT-14-23	2304
Francisco Kawonga	MSC-BIO-STAT-15-23	824
Blessings Chirambo	MSC-BIO-STAT-18-23	1092
Funny Osward	MSC-BIO-STAT-22-23	1296
Christopher Phiri	MSC-BIO-STAT-23-23	1025
Brian Mtofu	MSC-BIO-STAT-24-23	1344
Tereza Mwanavava	MSC-BIO-STAT-J-01-24	1408
Weldon Chihana	MSC-BIO-STAT-J-03-24	1050
Wongani Luhanga	MSC-BIO-STAT-J-04-24	2321
Joseph Kenneth	MSC-BIO-STAT-J-05-24	1792
Harry Milal	MSC-BIO-STAT-J-08-24	1206
Eneles Mponda	MSC-BIO-STAT-J-10-24	1736
Harriet Mchira	MSC-BIO-STAT-J-11-24	1791
Germue Gbawoquiya	MSC-BIO-STAT-J-17-24	2616
Marion Maganga	MSC-MAT-03-23	2460

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 $y_1, \dots, y_n$ 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. [20 marks]
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 2 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 lambda value used for your dataset and the mean of the `y` vector in the `dat` list object. [5 marks]

```r
set.seed(0000) # REPLACE 0000 with your individual seed value!
# Solutions using the seed value 123 will not be accepted.
```

```r
n<-rpois(n=1,lambda=25)
lambda<-rgamma(n=1,10,0.5)

dat<-list(
  N=n,
  y=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 $\text{Exp}(\lambda)$ sampling model to the data `dat`. You can choose your own values `a,b` for the prior. [20 marks]

4. 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? [20 marks]

5. 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 $\lambda|y_1, \dots, y_n$. [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]