QUEEN MARY, UNIVERSITY OF LONDON

MTH6102: Bayesian Statistical Methods

Exercise sheet 6

2023-2024

This assignment counts for 4% of the module total. The deadline for submission is **Monday the 13th November at 11am.**

Submit the R code used as an R script file (with extension .R). But you need to write the answers in a separate file. This can be a Word document, pdf or a clearly legible image of hand-written work. So you need to submit two files.

The first two questions use a dataset on QMPlus.

1. **(25 marks)**. For input data use the first column in the dataset, labelled `x`. Suppose that the observed data $x_1, \ldots, x_n$ follow a normal distribution $N(\mu, \sigma^2)$, where σ is assumed to be known and equal to 2.

   (a) Let the last two digits of your ID number be BC.
   As a prior distribution for μ, take a normal distribution $N(\mu_0, \sigma_0^2)$. We want the prior mean to be $B + 15$ and the prior probability $P(\mu \leq C)$ to be 0.1. Find the prior parameters μ_0 and σ_0^2 that satisfy this.

   (b) Using this prior distribution, find the posterior distribution for μ. (You may use the formulas)

   (c) Calculate the posterior probability $P(\mu \leq 2)$.

   (d) Calculate a 95% equal-tail credible interval for μ.

   (e) Calculate a 95% HDP credible interval for μ

2. **(30 marks)**. Suppose we observe iid data $y_1, \ldots, y_n$ from Poisson distribution with parameter λ. Let λ have the Gamma(α, β) distribution, the conjugate prior distribution for the Poisson likelihood, where α and β are known prior parameters.

   (a) Find the posterior distribution for λ.
   Now, an ecologist counts the numbers of centipedes in each of $n = 20$ twenty one-metre-square quadrats. The numbers $y_1, \ldots, y_{20}$ are in the second column labelled as y in the dataset.

   (b) Let the last three digits of your ID number be ABC. Suppose we want the prior mean for λ to be $5 + A$ and the prior standard deviation to be $5 + B$. Find the prior distribution parameters that satisfy this.

   (c) Using the prior distribution from (b), find the posterior distribution for λ.

   (d) Calculate the posterior median and a 95% credible interval for λ.

(e) Calculate the posterior median and a 95% credible interval for θ, where

$$\theta = 1 - \exp(-\lambda).$$

3. **(45 marks).** This question continues exercise sheet 5, question 4. Now there are two machines, each with a different precision of measurement, τ_1 and τ_2. They each take a number of measurements with the same known mean $\mu = 1000$. The measurements are $x = (x_1, \ldots, x_m)$ on the first machine can be modelled as a random sample from a normal distribution with known mean $\mu = 1000$ and precision τ_1 and $y = (y_1, \ldots, y_n)$ on the second machine can be modelled as a random sample from a normal distribution with known mean $\mu = 1000$ and precision τ_2.

   The observed data are $m = 10$, $\sum_{i=1}^{m}(x_i - 1000)^2 = 0.12$, $n = 8$ and $\sum_{i=1}^{n}(y_i - 1000)^2 = 0.09$. Use independent gamma prior distributions for τ_1 and τ_2 with same parameters α and β as for τ in exercise sheet 5, that is $\alpha = 5$ and $\beta = 0.05$.

   (a) Find the joint posterior density of τ_1 and τ_2.

   (b) What are the marginal posterior distributions for τ_1 and τ_2?

   (c) Using R, generate a sample of size $10,000$ from the joint posterior density of τ_1 and τ_2.

   We are in fact interested in the standard deviations $\sigma_1 = \dfrac{1}{\sqrt{\tau_1}}$ and $\sigma_2 = \dfrac{1}{\sqrt{\tau_2}}$.

   (d) Use R to denerate samples from the posterior distributions of σ_1 and σ_2, by transforming the τ_1, τ_2 samples that you generated in (c). (There is no need to work out on paper the posterior distributions of σ_1, σ_2.)

   (e) Use R to find the posterior median and a 95% credible interval for each of σ_1 and σ_2.

   (f) Using the samples of σ_1 and σ_2 that you generate in (d) estimate the difference in $d = \sigma_1 - \sigma_2$ and the posterior probability that $d < 0$.