QUEEN MARY, UNIVERSITY OF LONDON

MTH6102: Bayesian Statistical Methods

Solutions of 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

   Solution:

   (a) The prior mean is $\mu_0 = B + 15$. Also

$$
\begin{aligned}
P(\mu \leq C) &= 0.1 \\
P\left(\frac{\mu - \mu_0}{\sigma_0} \leq \frac{C - \mu_0}{\sigma_0}\right) &= 0.1 \\
\Phi\left(\frac{C - \mu_0}{\sigma_0}\right) &= 0.1 \\
\frac{C - \mu_0}{\sigma_0} &= \Phi^{-1}(0.1) \\
\sigma_0 = \frac{C - \mu_0}{\Phi^{-1}(0.1)}
\end{aligned}
$$

   $\Phi^{-1}(0.1)$ (inverse function of standard normal cdf) could be found using `qnorm (0.1)`, to give $\Phi^{-1}(0.1) = -1.282$.

(b) The posterior distribution is $N(\mu_1, \sigma_1^2)$, with formulae for the parameters as in the lectures. Suppose the data is in the vector `x`, and `sigma`, `mu0` and `sigma0` have been given the correct values.

```
n = length(x)
xbar = mean(x)
mu1 = (mu0/sigma0^2 + n*xbar/sigma^2)/(1/sigma0^2 + n/sigma^2)
sigma1 = sqrt(1/(1/sigma0^2 + n/sigma^2))
```

(c) This is given by `pnorm(2, mean=mu1, sd=sigma1)`.

(d) This is given by `c(qnorm(0.025,mean=mu1,sd=sigma1),qnorm(0.975,mean=mu1,sd=sigma1))`.

(e) The Gaussian posterior distribution $N(\mu_1, \sigma_1^2)$ is symmetric and unimodal. Hence, the 95% equal-tail credible interval derived in 1d) is, in fact, an HPD interval. Indeed, we want to find endpoints θ_L and θ_U such that

$$P(\theta_L < \mu < \theta_U) = 0.95$$
$$f_{N(\mu_1,\sigma_1^2)}(\theta_L) = f_{N(\mu_1,\sigma_1^2)}(\theta_U)$$

By standardising, the above happens by taking

$$\frac{\theta_L - \mu_1}{\sigma_1} = -1.96$$
$$\frac{\theta_U - \mu_1}{\sigma_1} = 1.96$$

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

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

Solution:

(a) The posterior distribution is Gamma(α_1, β_1) with $\alpha_1 = \alpha + S$ and $\beta_1 = n + \beta$, where n is the sample size and S is the sum of the data.

(b) The Gamma(α, β) distribution has mean $\dfrac{\alpha}{\beta}$ and variance $\dfrac{\alpha}{\beta^2}$.

We want to match these to $5 + A = E(\lambda)$ and $(5 + B)^2 = Var(\lambda)$. So we have to find α and β such that

$$5 + A = \frac{\alpha}{\beta}, \; (5 + B)^2 = \frac{\alpha}{\beta^2} = \frac{5 + A}{\beta}$$

Hence rearranging gives

$$\beta = \frac{5 + A}{(5 + B)^2}, \; \alpha = \frac{(5 + A)^2}{(5 + B)^2}.$$

(c) The posterior distribution is Gamma(α_1, β_1) with $\alpha_1 = n + \alpha$ and $\beta_1 = S + \beta$, where n is the sample size, S is the sum of the data, $\alpha = \frac{(5+A)^2}{(5+B)^2}$ and $\beta = \frac{5+A}{(5+B)^2}$.

(d) If the posterior parameters are `alpha1` and `beta1`, then the median and 95% credible interval are given by

```
qgamma(c(0.5, 0.025, 0.975), shape=alpha1, rate=beta1)
```

(e) The transformation $1 - \exp(-\lambda)$ is a one-to-one increasing function, so we can just transform each of the results from part (c). E.g. for the median:

```
med = qgamma(0.5, shape=alpha1, rate=beta1)
1-exp(-med)
```

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 $d < 0$.

Solution:

(a) The margina prior distributions are

$$\tau_1, \tau_2 \sim \text{ Gamma}(\alpha, \beta)$$

independently, with $\alpha = 5$, $\beta = 0.05$. The joint posterior density of τ_1 and τ_2, $p(\tau_1, \tau_2 \mid x, y)$, is, by independence,

$$\begin{aligned}
p(\tau_1, \tau_2 \mid x, y) &\propto p(\tau_1, \tau_2) p(x, y \mid \tau_1, \tau_2) \\
&= p(\tau_1) p(\tau_2) p(x \mid \tau_1) p(y \mid \tau_2) \\
&\propto \tau_1^{\alpha-1} \exp(-\beta\tau_1) \tau_2^{\alpha-1} \exp(-\beta\tau_2) \tau_1^{m/2} \exp\big(- \tau_1 \frac{1}{2} \sum_{i=1}^{m} (x_i - 1000)^2 \big) \\
&\quad \times \tau_1^{n/2} \exp\big(- \tau_2 \frac{1}{2} \sum_{i=1}^{n} (y_i - 1000)^2 \big) \\
&= \tau_1^{\alpha+m/2-1} \exp\big(- \tau_1 (\beta + \frac{1}{2} \sum_{i=1}^{m} (x_i - 1000)^2) \big) \\
&\quad \times \tau_2^{\alpha+n/2-1} \exp\big(- \tau_2 (\beta + \frac{1}{2} \sum_{i=1}^{n} (y_i - 1000)^2) \big)
\end{aligned}$$

Hence, $p(\tau_1, \tau_2 \mid x, y) = p(\tau_1 \mid x) p(\tau_2 \mid y)$, where $p(\tau_1 \mid x) \sim \text{Gamma}(\alpha_1, \beta_1)$ and $p(\tau_2 \mid y) \sim \text{Gamma}(\alpha_2, \beta_2)$
with

$$\alpha_1 = \alpha + m/2 = 10, \ \beta_1 = \beta + \frac{1}{2} \sum_{i=1}^{m} (x_i - 1000)^2 = 0.11,$$

$$\alpha_2 = \alpha + n/2 = 9, \ \beta_2 = \beta + \frac{1}{2} \sum_{i=1}^{n} (y_i - 1000)^2 = 0.095.$$

(b) From (a), the posterior distributions are independent, and

$$\tau_1 \sim \text{ Gamma}(\alpha_1, \beta_1)$$

$$\tau_2 \sim \text{ Gamma}(\alpha_2, \beta_2)$$

with

$$\alpha_1 = \alpha + m/2 = 10, \ \beta_1 = \beta + \frac{1}{2} \sum_{i=1}^{m} (x_i - 1000)^2 = 0.11,$$

$$\alpha_2 = \alpha + n/2 = 9, \ \beta_2 = \beta + \frac{1}{2} \sum_{i=1}^{n} (y_i - 1000)^2 = 0.095.$$

There is R code for the rest of the questions (c), (d), (e), (f) on QMPlus.