Note that you must use “MockPractical.Rmd” for your solutions so you can produce a pdf of your solutions (and associated R output).

This test contains **four** questions. The maximum mark is 50.

Bayesian computation

1. Suppose that X follows a Cauchy distribution with density

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad -\infty < x < \infty.$$

We wish to estimate the probability $p = \mathbb{P}(X > 2)$. We can write this as the integral

$$p = \mathbb{P}(X > 2) = \int_{-\infty}^{\infty} I(x > 2) f(x) dx$$

where $I(x > 2)$ is the indicator function returning the value 1 if $x > 2$ and 0 otherwise.

1.1 A Monte Carlo estimator of p based on N independent realisations $X_1, \dots, X_N$ of X is

$$\hat{p}_N = \frac{1}{N} \sum_{i=1}^{N} I(X_i > 2).$$

Write an R function `mc` that takes a single argument `N` and returns a single realisation of $\hat{p}_N$. You may use the intrinsic function `rcauchy` for generating draws of X.

1.2 Run your `mc` function with $N = 1000$ and compare the resulting estimate to the true value of p. Hint: `pcauchy` might be helpful here.

1.3 Write a function `mcMany` with arguments `N` and `n` that returns a vector of length `n` whose elements are independent realisations of $\hat{p}_N$. Hint: use either a for loop or `replicate`.

1.4 Using your `mcMany` function and `n=1000`, estimate the mean and variance of $\hat{p}_N$ for $N = 500$. How do these compare (in terms of relative error) to the true mean and variance of $\hat{p}_N$ for this choice of N?

1.5 For $x > 2$ the behaviour of $f(x)$ is similar to

$$g(x) = \frac{2}{x^2}, \quad x > 2.$$

Hence, given independent draws $X_1, \dots, X_n$ from g, the **importance sampling** estimator of p is

$$\tilde{p}_N = \frac{1}{N} \sum_{i=1}^{N} \frac{X_i^2}{2\pi(1 + X_i^2)}.$$

(a) Briefly explain why the indicator function is not needed in the expression for $\tilde{p}_N$.

(b) The following R function takes a single argument `N` and returns a single realisation of $\tilde{p}_N$.

```
imp=function(N)
{
  x <- 2/runif(N)
  return((1/N)*sum(x^2/(2*pi*(1+x^2))))
}
```

Write a function `mcMany2` with arguments `N` and `n` that returns a vector of length `n` whose elements are independent realisations of $\tilde{p}_N$.

(c) Using your `mcMany2` function and `n=1000`, estimate the mean and variance of $\tilde{p}_N$ for $N = 500$. How do these compare to the mean and variance of $\hat{p}_N$?

2. A doctor is interested in the survival time of patients (in years) following a major surgery. It is proposed that the survival time, X, follows a Gamma distribution with shape parameter $\alpha = 2$ and an unknown rate parameter β. Hence, X has probability density function

$$f(x|\beta) = \beta^2 \exp\{-\beta x\}, \qquad x > 0.$$

Our prior distribution is $\beta \sim \text{Gamma}(a, b)$ with $a = b = 1$.

Suppose that the survival times of m patients are recorded precisely as $\boldsymbol{x}^o = (x_1^o, \ldots, x_m^o)'$. A further $n - m$ patients were only monitored for $s^* = 5$ years, i.e. their unknown survival times that exceeded $s^* = 5$ were recorded as s^*. Denote the censored observations by $\boldsymbol{x}^c = (x_1^c, \ldots, x_{n-m}^c)' = (s^*, \ldots, s^*)'$. Associated with each censored observation x_i^c is a latent variable z_i representing the true survival time before censoring. Denote the latent variables by $\boldsymbol{z} = (z_1, \ldots, z_{n-m})'$.

The joint posterior for β and $\boldsymbol{z}$ is then

$$\pi(\beta, \boldsymbol{z}|\boldsymbol{x}) \propto \beta^{a-1} \exp(-b\beta) \prod_{i=1}^{m} \beta^2 \exp(-\beta x_i^o) \prod_{j=1}^{n-m} f(x_j^c, z_j|\beta).$$

The term $f(x_j^c, z_j|\beta)$ is the probability density of obtaining z_j from a $\text{Gamma}(2, \beta)$ distribution *and* that it exceeds s^*; that is,

$$\beta^2 z_j \exp(-\beta z_j) I(z_j > s^*)$$

where the indicator $I(z_j > s^*)$ takes the value 1 if $z_j > s^*$ and 0 otherwise.

2.1 Using the *inverse sampling* method, write a function called `gammaT` that takes arguments `N`, `sstar` and `beta` and returns `N` draws from a left truncated (at `sstar`) $\text{Gamma}(2, \beta)$ distribution. Hint: the functions `pGamma(x,a,b)` and `qGamma(x,a,b)` evaluate the cdf and inverse cdf (resp.) of a Gamma random variable with shape and rate parameters given by `a` and `b`, at the value `x`.

2.2 Generate a (synthetic) data set by running the following R code:

```
set.seed(3421)
data <- rgamma(100,2,0.5)
#Truncate anything bigger than 5
data[data>5] <- 5
data <- sort(data) #x=(x^o,x^c)
```

How many precise observations are there?

2.3 You are given the full conditional distributions

$$\beta|\boldsymbol{x}, \boldsymbol{z} \sim \text{Gamma}\left(a + 2n\,,\, b + \sum_{i=1}^{m} x_i^o + \sum_{j=1}^{n-m} z_j\right)$$

$$Z_j|\boldsymbol{x}, \beta, \boldsymbol{z}_{-j} \sim \text{Gamma}_{(z_j>s^*)}(2, \beta)$$

where $\text{Gamma}_{(z_j>s^*)}(2, \beta)$ denotes a gamma density truncated at the left at s^*.

Complete the following R function which takes as arguments the number of iterations `N`, a data set `x` and the left truncation limit `sstar`, and runs a Gibbs sampler to generate draws from the posterior $\pi(\beta, \boldsymbol{z}|\boldsymbol{x})$.

```
gibbs = function(N,x,sstar)
{
 a <- 1; b <- 1 #prior hyper-parameters
 n <- length(x)
 m <- #WRITE YOUR SOLUTION
 x0 <- x[1:m] #observed survival times
 betaVec <- rep(0,N) #store beta samples here
 zMat <- matrix(0,nrow=N,ncol=(n-m)) #sore z samples here
 #Initialise
 beta <- #WRITE YOUR SOLUTION
 z <- rep(sstar,n-m)
 betaVec[1] <- beta; zMat[1,] <- z
 for(i in 2:N)
 {
  #update beta
  beta <- #WRITE YOUR SOLUTION
  #update z
  z <- #WRITE YOUR SOLUTION
  #store
  betaVec[i] <- beta; zMat[i,] <- z
 }
 return(list(betaVec,zMat))
}
```

2.4 By running the sampler for 5,000 iterations, briefly investigate the mixing of the β chain and the $\boldsymbol{z}$ component chains. Remove any necessary burn-in, produce a kernel density estimate of the marginal β posterior and compare it to the ground truth value of β that generated the data.

Bayesian modelling

3. We have been given two competing models:

$$\begin{aligned} \mathcal{M}_1 : X|\mu,\phi &\sim \text{NegBin}\,(\mu,\phi)\,, \\ \mu &\sim \text{N}(5,4) \text{ truncated to have support } [0,\infty], \\ \phi &\sim \text{Gamma}(2,2); \\ \mathcal{M}_2 : X|\mu,\phi &\sim \text{NegBin}\,(\mu,\phi)\,, \\ \mu &\sim \text{N}(5,4) \text{ truncated to have support } [2,10], \\ \phi &\sim \text{Gamma}(2,2) \text{ truncated to have support } [10,\infty]. \end{aligned}$$

Note that, with this parameterisation of the negative binomial distribution, we have

$$\text{E}(X|\mu,\phi) = \mu \quad \text{and} \quad \text{Var}(X|\mu,\phi) = \mu + \mu^2/\phi.$$

We receive the following data:

$$\underline{x} = \{1,3,5,6,5,2,8,2,9,7,3,5,5,1,8,1,8,4,1,0\}\,.$$

3.1 Write some model code for Stan that covers each of the models (with two separate Stan files).

3.2 Write out R code to sample from the posterior for each model, and estimate a 95% credible interval for each of the model parameters.

3.3 Use the `loo` package to find pseudo-BMA weights for the models.

4. We have data about average student-to-teacher ratios for seven- and eight-year-olds for a number of towns, across different districts, for two different regions. If we let R_{ijk} be the average ratio for the ith town from the jth district for the kth region, we have the following model:

$$\begin{aligned}
R_{ijk}|\mu_{jk},\sigma_k^2 &\sim \text{log-N}\left(\mu_{jk},\sigma_k^2\right),\\
\sigma_k^2 &\sim \text{Exp}(\lambda),\\
\lambda &\sim \text{Gamma}(5,1),\\
\mu_{jk} &\sim \text{N}(\beta_k,1/\nu),\\
\beta_k &\sim \text{N}(\alpha,1/\tau),\\
\nu &\sim \text{Exp}(3),\\
\tau &\sim \text{Exp}(1)\\
\alpha &\sim ????.
\end{aligned}$$

where $i=1,\dots,n_{jk},\ j=1,\dots,n_k,\ k=1,2$, and the prior distribution for α has not yet been determined.

4.1 Using the bisection method for elicitation of expert knowledge, we elicit judgements of 2.3, 3 and 3.85 for the 25th, 50th and 75th percentiles for α respectively. Use https://jeremy-oakley.shinyapps.io/SHELF-single/ to determine a suitable prior distribution for α using the elicited judgements.

4.2 The collected data have been recorded in "Mock_practical.Rmd". How many distinct districts do we have data for?

4.3 Write model code for Stan that covers the given model and generates a sample from the predictive distribution for the ratio for an unsampled town in the second district in the first region.

4.4 Find the posterior means and standard deviations for all the model parameters.

4.5 Which town is best and which region is best overall in terms of having the lowest student-to-teacher ratio? Justify your answers.