Bayesian Statistics
Exercises 5.

1. Let $x_1, \ldots, x_n$ be independent and identically distributed $N(\mu, 1)$. Assume a prior for μ that is $N(0, 3^2)$. Explain what the following R code does - and, particularly, explain i) the vectorization being performed in the function `loglik`, and ii) the problems caused when computing the weights `w1`, and how the program solves this problem for computing `w2`. Compare the numerical results with analytical calculations for the same data x.

```
set.seed(111)
m=10000
n=1000
x = rnorm(n,0,1)

musamp = rnorm(m,0,3)

loglik = function(x,musamp){
  n=length(x)
  m = length(musamp)
  xx = rep(x,rep(m,n))
  ll.matrix=matrix(dnorm(xx,musamp,1,log=TRUE),nrow=m)
  rowSums(ll.matrix)
}

logl=loglik(x,musamp)
normalize=function(x){x/sum(x)}
w1 = normalize(exp(logl))
w2 = normalize(exp(logl-max(logl)))

sum(w1[musamp<0])
sum(w2[musamp<0])
sum(w2*musamp)
sum(w2*musamp^2) - sum(w2*musamp)^2
```

2. Let $x_1, \ldots, x_n$ be independent and identically distributed observations from the mixture density

$$f_X(x; \pi_0, \mu_0, \mu_1, \tau_0, \tau_1) = \pi_0 N(x; \mu_0, 1/\tau_0) + (1 - \pi_0) N(x; \mu_1, 1/\tau_1) \quad (1)$$

where $N(\cdot; \mu, 1/\tau)$ denotes the density of a normal distribution with mean μ and variance $1/\tau$.

Assume that the true values of the parameters are $\pi_0 = 0.2, \mu_0 = 0, \mu_1 = 4, \tau_0 = \tau_1 = 1$.

(a) The mixture model (1) can also be written using the following "latent variable representation": $z_i \sim \text{Bernoulli}(1 - \pi_0)$; $x_i | z_i = k \sim N(\mu_k, 1/\tau_k)$. This is called the latent variable representation because z_i is unobserved (latent) so the probability density of the x_i requires you to marginalize $p(x_i, z_i)$ over z_i. Show how marginalizing the joint distribution over z_i yields the density (1) for x_i.

(b) Simulate 1000 observations from this mixture distribution. (Make sure to set a seed so that your results are reproducible.) Plot a histogram of the results. [Hint: simulating from a mixture can be done by exploiting the latent variable representation.]

(c) Write a function to compute the log likelihood for a vector x at any given $\pi_0, \mu_0, \mu_1, \tau_0, \tau_1$. (You might also like to vectorize this function to allow the log-likelihood to be efficiently computed for any vectors of these parameters, analogous to the code in Q1)

(d) Assume now that you know the true values of the means and precisions $(\mu_0, \mu_1, \tau_0, \tau_1)$ and wish to estimate π_0. Assume that your prior is

$$\pi_0 \sim \text{Beta}(0.5, 0.5) \quad (2)$$

i) Write a function that uses important sampling, with the prior distribution as your importance sampling distribution, to obtain an approximation to the posterior distribution of π_0. Apply this to your simulated data set (created in part b) to obtain an approximate posterior mean and an approximate 90% posterior CI for π_0. (Unless you are unlucky the posterior CI should cover the true value of π_0! If not, try again with a new seed, to check whether there might be a problem with your implementation.)

ii) Use your log-likelihood function to compute the log-likelihood on a grid of values of π_0 from 0 to 1. Use this grid-based approach to obtain a discrete approximation to the posterior distribution for π_0. Again, obtain a posterior mean for π_0 and a 90% posterior CI.

iii) Use your answer to the previous part to develop an improved importance sampling function. (Note that, to satisfy the requirements that the support of q contains the support of p, the importance sampling function must be continuous, so you have to convert your discrete approximation to a continuous approximation.) Use this improved IS function to obtain another estimate for the posterior mean and an approximate 90% posterior CI for π_0. Comment on how the variance of the importance weights differs between this IS function and using the prior as an IS function.

3. Repeat part d) of question 1, considering the precisions τ_0, τ_1 also to be unknown with independent priors

$$\tau_k \sim \Gamma(\text{shape}{=}0.1, \text{rate}{=}0.1). \tag{3}$$

[Hint 1: the likelihood will now of course be $p(x_1, \ldots, x_n|\pi_0, \tau_0, \tau_1)$. Hint 2: when considering a grid of values of τ you probably want to work with an equal-spaced grid on $\log(\tau)$ in some finite range.]

Do you feel that any of the three methods you try are able to get reliable estimates of the joint posterior distribution in this case? Give reasons for your answer.

4. Repeat part d) of question 1, considering the precisions τ_0, τ_1 unknown with prior as above, and means μ_0, μ_1 to be unknown, with independent priors

$$\mu_k \sim N(0, 3^2). \tag{4}$$

(NOTE: remember that in R the normal functions are parameterized in terms of the standard deviation – here 3 – not the variance.)

Do you feel that any of the three methods you try are able to get reliable estimates of the joint posterior distribution in this case? Give reasons for your answer. Comment on differences or similarities between the posterior distributions of μ_1 and μ_2.