

Possible points: 100

1. **Problem 1 [50 pts]**

You are given the following dataset (Mendenhall et al. 1989):

$$x = c(21, 24, 25, 26, 28, 31, 33, 34, 35, 37, 43, 49, 51, 55, 25, 29, 43, 44, 46, 46, 51, 55, 56, 58)$$

$$y = c(rep(1, 14), rep(0, 10))$$

where y_i represents the response (0 or 1) for each of the 24 patients as a function of the number of days of radiotherapy (x_i) that patient received. (Note the above syntax means that the first 14 outcomes were ones and the last 10 were zeros.)

Consider a Bayesian logistic regression model:

$$\text{logit}(P(Y_i = 1)) = \alpha_0 + \alpha_1 X_i$$

with the following structure:

$$\alpha_0 \sim N(\delta_0, 2)$$

$$\alpha_1 \sim N(\delta_1, 2)$$

$$\delta_0 \sim N(0, 10)$$

$$\delta_1 \sim N(0, 10)$$

- Find the analytic expression for the full joint posterior density for the 4 parameters of the logistic regression model above.
- Find the mode of the above 4-dimensional density (you can use an optimization function, or write your own Newton-Raphson.)
- Find the estimated curvature at the above mode.
- Based on the mode and curvature, find the 4-dimensional normal approximation to the joint posterior.
- Find the 2-dimensional marginal posterior density of (α_0, α_1) using the Monte Carlo approximation
- Plot the bivariate contour plot of the 2-dimensional marginal posterior density of (α_0, α_1)
- Plot the 2 separate 1-dimensional marginal posteriors of α_0 and α_1
- Find the 1-dimensional normal approximations to the two marginals above, and plot them superimposed on the individual marginals.
- Do the individual 1-dimensional marginal modes for α_0 and α_1 differ from their 2-dimensional joint mode?
- Find the posterior mean and 95% HPD for α_1
- Using importance sampling, perform a sensitivity analysis of the posterior mean and 95% HPD for α_1 , under alternative priors:

$$\delta_0 \sim N(0, 100)$$

$$\delta_1 \sim N(0, 100)$$

2. Problem 2 [50 pts]

Recall the genetic linkage example: animals are distributed into 4 categories: $Y = (y_1, y_2, y_3, y_4)$ according to the genetic linkage model, where the probabilities of falling into the 4 categories are given as:

$$((2 + \theta)/4, (1 - \theta)/4, (1 - \theta)/4, \theta/4).$$

- (a) Use the importance sampling to obtain the posterior mean and posterior standard deviation for the data $Y = (125, 18, 20, 34)$. Use the matching Normal density as the importance function.
- (b) Compare your importance sampling estimates of the posterior mean and posterior standard deviation to those obtained by the Normal approximation.
- (c) Repeat the above two parts for the data $Y = (14, 0, 1, 5)$.
- (d) Compare the histograms of the importance weights for both data sets, and discuss the adequacy of the importance sampling estimate for each data set.
- (e) Repeat all of the above parts using Beta densities as the importance functions. (First, find the appropriate Beta densities - be creative, but mindful of what a good importance sampling density needs to have!)
- (f) Use importance sampling to obtain the normalized likelihood for the data $Y = (125, 18, 20, 34)$. Plot this curve and the true normalized likelihood in the same figure.
- (g) Repeat the previous step for the data $Y = (14, 0, 1, 5)$. Plot this curve and the true normalized likelihood in the same figure.