

Possible points: 100

1. Problem 1 [60 pts]

Use the same dataset as in the previous homework (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.) We have a Bayesian logistic regression model:

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

with the following structure:

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

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

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

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

- (a) Find the analytic expression for the 4-dimensional joint posterior density for the 4 parameters of the logistic regression model above (this was done in the previous homework, so just repeat this step).
- (b) Based on the joint posterior above, find the individual one-dimensional full-conditional densities for each of the 4 parameters.
- (c) Using the above full-conditional densities, implement the Gibbs sampler. Be specific about how to sample from each individual 1-dimensional full-conditional. Provide a detailed pseudo code for this implementation.
- (d) Run the Gibbs sampler for 10,000 iterations (please write the entire sampler yourself). Present the results, and diagnostics (you can use CODA or some other diagnostics library.)
- (e) Describe what you did to determine the burn-in period for the sampler above.
- (f) Show the ACF and describe what you did to choose the thinning parameter
- (g) Would you say your sampler converged after 10,000 iterations? If not, run the sampler until convergence seems to have been achieved.
- (h) Using the final thinned set of 4-dimensional samples, plot 4 separate histograms (estimated marginal densities), one for each of the 4 parameters.
- (i) Find the marginal posterior modes, means and 95% central credible intervals for each of the 4 parameters.
- (j) For α_0 and α_1 , comment on how similar the results above are to the results you got on the previous homework using the normal and importance sampling approximations

2. Problem 2 [40 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:

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

For data $Y = (125, 18, 20, 34)$, do the following:

- (a) Design the random-walk Metropolis-Hastings algorithm to obtain a Monte Carlo approximation to the posterior density of θ . Provide a detailed pseudo-code used for this sampler.
- (b) Implement the above sampler, using 2 different MH transition kernels: $N(0, 0.1)$ and $N(0, 0.01)$. (The numbers in parentheses are the mean and standard deviation.) Please write the entire sampler yourself.
- (c) Comment on the performance of each: compare trace plots and convergence diagnostics (You can use CODA or some other diagnostics library.). What are the acceptance rates like?
- (d) Suggest an alternative transition kernel that should perform better than any of the kernels above.
- (e) Using the kernel of your choice, run the MH algorithm for 10,000 iterations. Report on the results and diagnostics. Again, please write the sampler yourself, though you can use CODA or some other diagnostics library.
- (f) What was the acceptance rate?
- (g) Describe what you did to determine the burn-in period for the sampler above.
- (h) Show the ACF and describe what you did to choose the thinning parameter
- (i) Would you say your sampler converged after 10,000 iterations? If not, run the sampler until convergence seems to have been achieved.
- (j) Using the final thinned set of samples, plot the histogram of the estimated posterior density of θ .
- (k) Find the posterior mode, mean and 95% central credible interval for θ .