STAT 207, Spring 2023
Homework # 2
Released April 18. Due May 4 on Canvas

Instructions: You may discuss the homework problems in small groups, but you must write up the final solutions and codes yourself. You can either type your work in LaTeX or write down on papers and scan. Please make sure that all handwriting are visible and please combine all pages into a single PDF file (in the correct order).

For any problems that involve coding, you must provide written answers and also include your codes. You can either include your codes at the end of the homework and label which questions they correspond to, or include as part of the answer (e.g., in the R Markdown style). You will receive no credit if you submit only codes or only written answers.

1. BDA3 Problem 5.3. Instead of printing the table, you can plot the pair-wise probabilities of one being better than the other as an 8×8 matrix.

2. BDA3 Problem 5.8

3. BDA3 Problem 5.13

4. An ecologist wants to estimate how many deer there are in a valley. She catches $n_1 = 20$ deer, marks them, and sends them back. A few days later, she returns and catches $n_2 = 20$ deer again. She finds that in the second sample, 13 deer are new and 7 deer are marked in the first sample. Thus in total, $r = 33$ unique deer are observed. Suppose the total deer population N does not change during the period. Also assume the probability for a deer being captured in a sample is the same in both samples and for all deer, and denote this probability as π. Consider the following priors for the unknown parameters:

$$N \sim Poi(\lambda)$$

$$\pi \sim Beta(\alpha, \beta)$$

with some fixed λ, α, β.

   (a) Write down the posterior distribution $p(N, \pi | n_1, n_2, r)$ up to a constant.

   (b) Write down the steps to sample from the posterior conditional distributions. Please specify the name of the distribution whenever you can.

   (c) Suppose the ecologist wants to sample another deer, denote π_0 as the probability that this deer is not seen in either of the two samples. Write down the posterior predictive distribution of π_0. Then describe how you can compute the expectation of π_0 using the MCMC developed in part (b).

5. In this question, we will analyze data on 10 power plant pumps using a Poisson gamma model. The number of failures Y_i is assumed to follow a Poisson distribution

$$Y_i|\theta_i \sim_{ind} Poisson(\theta_i t_i), \quad i = 1, ..., 10$$

where θ_i is the failure rate for pump i and t_i is the length of operation time of the pump (in 1000s of hours). The data is as follows:

pump	1	2	3	4	5	6	7	8	9	10
t_i	94.3	15.7	62.9	126	5.24	31.4	1.05	1.05	2.1	10.5
y_i	5	1	5	14	3	19	1	1	4	22

We consider the conjugate gamma prior distribution for the failure rate

$$\theta_i|\alpha, \beta \sim_{ind} Gamma(\alpha, \beta), \quad i = 1, ..., 10$$

with hyperpriors $\alpha \sim Exp(1)$ and $\beta \sim Gamma(0.1, 1.0)$. All Gamma distribution uses the shape and rate parameters.

(a) Write out the steps of a Metropolis-Hastings within Gibbs sampling algorithm to analyze these data.

(b) Apply the algorithm to the data and show histograms of the posterior marginal distributions for α and β, and a scatter plot of the bivariate posterior distribution.

(c) Analytically integrate θ_i from the posterior and derive (up to proportionality) the posterior $p(\alpha, \beta|y)$.

(d) Construct a Metropolis-Hastings algorithm to sample from the posterior $p(\alpha, \beta|y)$ without sampling θ_i.

(e) Repeat part (b) using this reduced sampler.

(f) Describe how you can draw samples from $p(\theta_i|y)$ from the reduced sampler