Homework IV (2022)

1. Let $Y$ be a random variable from $N(0, r^{-1})$ and $r$ be a random variable from $Gamma\,(\nu/2, \nu/2)$, respectively. Note that the pdf of $Gamma\,(\nu/2, \nu/2)$ random variable is
$$p(r) = \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)} r^{\nu/2-1} e^{-\nu r/2}.$$
   (a) Show that the marginal distribution of $Y$ is Student's $t$ distribution with degree of freedom (df) $\nu$. Note that the pdf of distribution with df $\nu$ is
$$p(y) = \frac{\nu^{-1/2}\Gamma((\nu+1)/2)}{\sqrt{\pi}\Gamma(\nu/2)} \left(1 + \frac{y^2}{\nu}\right)^{-(\nu+1)/2}.$$
   (b) Present an algorithm to generate $t_\nu$ random variable with $\nu = 4$ and generate 10000 random numbers using above result. Then calculate mean of the 10000 random numbers and compare true mean 0.

   (c) Using a Metropolis-Hastings algorithm with the following candidate densities, $N(0, 1)$ and $t_2$, generate 10000 random numbers. Then compare two results (acceptance rates and mixing). Also, calculate the sample means using the random numbers in (b) and (c) and compare them.

2. Data file 'bikes.txt' has counts of the number of bicycles and other vehicles for 10 randomly selected residential streets with bike routes in Berkeley, CA.

   (a) Set up a hierarchical model for the observed number of bicycles on streets $j = 1, \ldots, 10$ that is binomial with unknown probability $\theta_j$ and sample size equal to the total number of vehicles (bicycles included) in that block. Assign a beta population distribution for the parameter $\theta_j$ ($\theta_j \sim^{iid} Beta(\alpha, \beta)$) and a non-informative hyperprior distribution ($p(\alpha, \beta) \propto (\alpha + \beta)^{-5/2}$). Write down the joint posterior distribution.

   (b) Compute the marginal posterior density of the hyperparameters and draw simulations from the joint posterior distribution of the parameters and hyperparameters.

   (c) Compare the posterior distributions of the parameters $\theta_j$ to the raw proportions in location $j$.

   (d) Give a 95% CI for the average underlying proportion of traffic that is bicycles.

   (e) A location on a new residential street with a bicycle route is sampled at random during which time 100 vehicles of all kinds go by. Give a 95% CI for the number of those vehicles that are bicycles.

3. Solve the following problems:

   (a) For a standard normal random variable $Z$, calculate $P(Z > 2.5)$ using Monte Carlo sums based on indicator function. How many simulated random variables are needed to obtain three digits of accuracy?

   (b) Using Monte Carlo sums verify that if $X \sim Gamma(1, 1)$, $P(X > 5.3) \approx .005$. Find the exact .995 cutoff to three digits of accuracy.

4. In this problem, we will analyze the Scottish lip cancer data available at icampus. $O_i$ and $E_i$ are the observed and expected number of cases of cancer in region $i = 1, \cdots, n = 56$. Assume that $O_i|\lambda_i \sim Poisson(E_i\lambda_i)$, independent over $i$ given $\lambda_1, \cdots, \lambda_n$. We consider two models for $\lambda_1, \cdots, \lambda_n$:
   1. $\lambda_i \sim Gamma(a, b)$, where $a \sim Gamma(0.1, 0.1)$ and $b \sim Gamma(0.1, 0.1)$.
   2. $\lambda_i = \exp(\gamma_i)$, where $\gamma_i \sim^{iid} N(\mu, \sigma^2)$, $\mu \sim N(0, 10^2)$, and $\sigma^2 \sim InvGamma(0.1, 0.1)$.
   Run MCMC algorithms for both models using WinBUGS (10000 iterations and 2000 burn-ins) and compare the posterior means and corresponding 95% confidence intervals for $\lambda$'s. Discuss whether inference on $\lambda$'s is affected by model choice.