All about the Gibbs sampler! I apologize that the set-up of these problems are long, but I hope you find them educational!

1. In class, we demonstrated Gibbs sampling for a two-parameter Normal model, where both μ and σ^2 are unknown. In fact, the Gibbs sampling algorithm works for any two-parameter model, or multi-parameter model when the number of parameters is more than one. We will see another one here.

   Recall the Gamma-Poisson conjugate model, where the sampling model is $Y_1, \cdots, Y_n | \theta \overset{i.i.d.}{\sim}$ Poisson(θ), and the conjugate prior for θ is $\theta \sim \text{Gamma}(a, b)$. In this set-up, when we treat a and b as known, we have the posterior exactly.

   Now, suppose that a is known but b is unknown. In this case, we will treat b as an unknown parameter. So it needs a prior. Let's use the following prior: $b \sim \text{Gamma}(1, 1)$.

   We come to the following statistical model:

$$\begin{aligned} Y_1, \cdots, Y_n \mid \theta, b &\overset{\text{iid}}{\sim} \text{Poisson}(\theta) && (1) \\ \theta \mid b &\sim \text{Gamma}(a, b) && (2) \\ b &\sim \text{Gamma}(1, 1) && (3) \end{aligned}$$

   Notice in (1) and (2) that we now condition on b to demonstrate that it's an unknown parameter!

   How will we sample the joint posterior distribution of (θ, b)? Gibbs sampler!

   (a) Write out the joint posterior distribution $f(\theta, b \mid \mathbf{y})$ up to proportionality using the $\propto$ symbol, and specifying above the symbol which parameters you are being proportional to.

   (b) Obtain the full conditional distribution for each unknown parameter. In this problem, you should be able to spot kernels. If you do, then you know exactly what the distribution should be! *If you don't spot a kernel, then we actually cannot drop things up to proportionality; we would have to carry all the terms with us to ensure that our distribution is a proper PDF that integrates to 1.*

2. (R) **Dependent gamma priors**: A cancer laboratory is estimating the rate of tumorigenesis in two strains of mice, A and B. They have tumor count data for 10 mice in strain A and 13 mice in strain B. Type A mice have been well studied, and information from other laboratories suggests that type A mice have tumor counts that are approximately Poisson-distributed with a mean of 12. Tumor count rates for type B mice are unknown, but type B mice are related to type A mice. The observed tumor counts for the two populations are:

$$\mathbf{y}_A = (12, 9, 12, 14, 13, 13, 15, 8, 15, 6) \qquad \mathbf{y}_B = (11, 11, 10, 9, 9, 8, 7, 10, 6, 8, 8, 9, 7)$$

We will model $Y_{A,i}|\theta_A \overset{\text{iid}}{\sim} \text{Poisson}(\theta_A)$ independent of $Y_{B,i}|\theta_B \overset{\text{iid}}{\sim} \text{Poisson}(\theta_B)$.

While we view the mice counts as independent (given the rates), we will now formulate a prior to reflect our belief that tumor rates may be related across the two groups. Define $\theta_A = \theta$ and $\theta_B = \theta \times \gamma$, for unknown θ and γ. Our priors for these parameters will be $\theta \sim \text{Gamma}(a_\theta, b_\theta)$ independent of $\gamma \sim \text{Gamma}(a_\gamma, b_\gamma)$.

(a) Set $(a_\theta, b_\theta) = (120, 10)$ and $(a_\gamma, b_\gamma) = (10, 10)$. Obtain a Monte Carlo sample of size $S = 1000$ from the prior distribution of θ_A and θ_B (i.e. obtains samples from the prior), and plot the sample as a scatterplot. Describe the prior dependence between θ_A and θ_B.

(b) Find the full conditionals $p(\theta|\gamma, \mathbf{y}_A, \mathbf{y}_B)$ and $p(\gamma|\theta, \mathbf{y}_A, \mathbf{y}_B)$. Think about the posterior means of these full conditional distributions and make sure they make sense.

(c) Run a Gibbs sampler by iteratively simulating from the full conditional distributions of θ and γ at least 5000 times. Be sure to set a seed first! After you've run the Gibbs sampler, throw away half of your observations as burn-in, then provide traceplots of your parameters. If you feel like your chain has converged, continue to the next step. If not, re-run your sampler for more iterations.

(d) Make a scatterplot illustrating the posterior joint dependence between θ_A and θ_B, and obtain an MCMC estimate of the posterior mean of $\theta_B - \theta_A$, as well as a 95% posterior credible interval for $\theta_B - \theta_A$. Based on these posterior summaries, does it appear that the two mice strains are similar? Why or why not?

(e) Repeat (c)-(d) where now $(a_\gamma, b_\gamma) = (45, 45)$. Describe how changing the prior changed the results.

3. (R) *(Variation of Hoff 6.3).* ***Your first interesting model!***

A panel (i.e. longitudinal) study followed 25 married heterosexual couples over a period of five years. The study was interested in learning about if the probability of divorce was related to the age differential in the couple (husband's age - wife's age). The outcome for each couple i is binary, where $Y_i = 1$ if the couple did divorce by the end of the study, and $Y_i = 0$ if the couple did not divorce. Let x_i denote the age differential for the couple i. We will implement a **Probit Regression** model (which is simply a regression model for binary data, like logistic regression but we take a different distributional approach).

The Probit Regression model in Bayes world takes a *latent variable* approach, which is a really powerful tool in statistical modeling. A latent variable is a variable that is not actually observed, but if we learned its value, would facilitate sampling. You already saw your first latent variable approach in last week's homeowrk where you sampled from a mixture distribution simulating a coin flip as a "mixture component indicator". There is no true notion of a coin flip, but framing the problem in terms of a latent coin made your sampling really easy!

The Probit regression model is as follows:

$$Y_i|Z_i = \begin{cases} 1 & Z_i > 0 \\ 0 & Z_i \leq 0 \end{cases}$$
$$Z_i = \beta x_i + \epsilon_i$$

where β is an unknown coefficient that relates x_i to Y_i. The errors are $\epsilon_i \overset{\text{iid}}{\sim} N(0,1)$. Here, the Z_i are the latent variables ($i = 1, \ldots, n$) that turn the problem continuous, which is typically nice than discrete world. We never truly observe the Z_i, but maybe we can sample them along the way.

(a) Let's understand the model a bit better: using standard normal CDF notation Φ, what is $\Pr(Y_i = 1)$? Based on what you find, for fixed $x_i > 0$, are positive or negative values of β associated with increased probability of divorce?

(b) Assuming a prior $\beta \sim N(0, \sigma_\beta^2)$, find the full conditional distribution for β: $f(\beta|\mathbf{y}, \mathbf{z})$.

(c) Show that the full conditional of z_i, $f(z_i|\mathbf{y}, \mathbf{z}_{[-i]}, \beta)$, is a constrained Normal density (i.e. proportional a Normal density but constrained to lie in an interval instead of $\mathbb{R}$). Specifically, show that the full conditional is constrained to be either above 0 or below 0, depending on y_i. *The notation $\boldsymbol{z}_{[-i]}$ means all values of $\boldsymbol{z}$ except the i-th value.*

(d) Time to implement! Download the `divorce` data from the website. Letting $\sigma_\beta^2 = 16$, implement a Gibbs sampler that approximates the joint posterior of $(\beta, \mathbf{Z})$. *I will post code on the website about how to sample from the constrained Normal.* Don't store your $\mathbf{Z}$ values, as we're not actually interested in them (they are simply mechanisms for us). Run your code for 10000 iterations, throwing away the first 7500 as burnin. Provide a traceplot for β. Does it appear that we've converged?

(e) Now, let's actually answer the research question: how does age differential relate to divorce? Obtain a 95% posterior credible interval for β as well as the posterior probability that $\beta > 0$. What are your conclusions?

General rubric

Points	Criteria
5	The solution is correct *and* well-written. The author leaves no doubt as to why the solution is valid.
4.5	The solution is well-written, and is correct except for some minor arithmetic or calculation mistake.
4	The solution is technically correct, but author has omitted some key justification for why the solution is valid. Alternatively, the solution is well-written, but is missing a small, but essential component.
3	The solution is well-written, but either overlooks a significant component of the problem or makes a significant mistake. Alternatively, in a multi-part problem, a majority of the solutions are correct and well-written, but one part is missing or is significantly incorrect.
2	The solution is either correct but not adequately written, or it is adequately written but overlooks a significant component of the problem or makes a significant mistake.
1	The solution is rudimentary, but contains some relevant ideas. Alternatively, the solution briefly indicates the correct answer, but provides no further justification.
0	Either the solution is missing entirely, or the author makes no non-trivial progress toward a solution (i.e. just writes the statement of the problem and/or restates given information).
Notes:	For problems with multiple parts, the score represents a holistic review of the entire problem. Additionally, half-points may be used if the solution falls between two point values above.
Notes:	For problems with code, well-written means only having executed lines of code that are necessary to solving the problem (you're welcome to comment out code for yourself to keep), as well as presenting the solution for the reader to easily see.