PROBLEMS:

1. **Improper hyperpriors in hierarchical models**: Problem 5.10 in BDA3. (Hints: First show that it suffices to prove that \( p(\tau|y) \) is integrable to establish the propriety of the full posterior distribution. Then examine \( p(\tau|y) \) for integrability near zero (for (a)) and infinity (for (b)). In each case find the limiting behavior of \( p(\tau|y) \), e.g., does it behave like a constant, \( \tau^{-1}, \tau^{-2} \) or something else. Remember \( \hat{\mu} \) and \( V_\mu \) both depend on \( \tau \.)

2. **Exponential/Gamma hierarchical model - theory**: The length of time that a light bulb lasts before burning out can be modeled as an exponential distribution, \( p(y|\theta) = e^{-\theta y} \), which has mean \( 1/\theta \) and variance \( 1/\theta^2 \). If we measure lifetimes in 100s of hours then we might imagine the mean is around 7.5 (750 hours). (Note: this is a common statistics example but hard to believe that light bulbs do fail according to this distribution. Why? There would be many early failures if true.) We get two light bulbs from each of 9 manufacturers. Let \( y_{i1} \) and \( y_{i2} \) represent the lifetimes of the bulbs from manufacturer \( i \) and let \( \theta_i \) represent the exponential parameter for manufacturer \( i \). If we assume the manufacturers are exchangeable, then we can model the 9 exponential lifetime parameters (the \( \theta_i \)'s) as samples from a population. A natural population distribution is the Gamma distribution with parameters \( (\alpha, \beta) \). To begin we take a flat hyperprior, \( p(\alpha, \beta) \propto 1 \).

(a) Given the assumptions above give the (unnormalized) joint posterior density \( p(\theta, \alpha, \beta|y) \).

(b) Derive the conditional posterior distribution of \( \theta \) given the hyperparameters, \( p(\theta|y, \alpha, \beta) \).

(c) Obtain the (unnormalized) marginal posterior distribution of the hyperparameters \( p(\alpha, \beta|y) \).

(d) Is \( p(\alpha, \beta) \propto 1 \) a suitable prior? (Consider the behavior of the marginal posterior distribution as \( \alpha, \beta \to \infty \) with \( \alpha/\beta \) constant.)

3. **Exponential/Gamma hierarchical model - analysis**: To obtain a suitable prior distribution I’ll repeat the argument that has been successful in other problems. The gamma prior distribution has limiting behavior of \( \phi \) with parameters \((\theta, \alpha, \beta) \). We take a prior that is flat on the mean and standard deviation (actually we use just the \( \beta^{-1/2} \) part of the standard deviation since the other term is just the mean), i.e., \( p(\alpha/\beta, \beta^{-1/2}) \propto 1 \). This prior distribution implies \( p(\alpha, \beta) \propto \beta^{-5/2} \). (Let \( \phi = \alpha/\beta \) and \( \gamma = \beta^{-1/2} \) apply the theory for transformation of variables.)

(a) For computing the marginal posterior distribution in this problem it is best to work on a transformation of the parameters, \( \phi_1 = \log(\alpha/\beta) \) and \( \phi_2 = \log(\beta) \). Show that on this scale our prior distribution is \( p(\phi_1, \phi_2) \propto e^{\phi_1 - 0.5\phi_2} \).

(b) Suppose that we observe data as follows: \( y_1 = (0.2, 2.1), \ y_2 = (3.0, 1.6), \ y_3 = (5.5, 26.8), \ y_4 = (2.3, 1.7), \ y_5 = (0.7, 6.5), \ y_6 = (6.2, 19.3), \ y_7 = (1.0, 0.7), \ y_8 = (4.0, 7.9), \ y_9 = (5.8, 8.0) \). Obtain 1000 simulations from the posterior distribution of \( \alpha, \beta, \theta \). Provide summaries of your simulations (i.e., the posterior distributions of \( \alpha, \beta, \) and the \( \theta \)'s. There are at least two ways to get the computation done. One way is to construct a grid approximation to \( p(\phi_1, \phi_2|y) \) as we have done elsewhere in the class. Don’t forget to transform \( \phi_1 \) and \( \phi_2 \) back to \( \alpha \) and \( \beta \). An alternative is to use Stan (through the Rstan interface).

(c) Manufacturer 3 (\( \bar{y} = 16.2 \)) appears to produce much better bulbs than manufacturer 7 (\( \bar{y} = 1.7 \)). But this is based on only two observations per manufacturer. Estimate the probability that the mean lifetime for bulbs from manufacturer 3 (\( 1/\theta_3 \)) is greater than the mean lifetime for bulbs from manufacturer 7 (\( 1/\theta_7 \)).
(d) Suppose we go back to the store and purchase a new bulb from the first manufacturer. Use your simulations to obtain a predictive distribution for the bulb’s lifetime.

(e) Suppose we choose a new manufacturer (not one of the nine for which we have data) and purchase a bulb from them. Use your simulations to obtain a predictive distribution for the bulb’s lifetime. How does this distribution compare to the distribution in the previous part. Explain.

4. **Normal approximation.** Here we consider the use of normal approximations for low-dimensional posterior distributions as an alternative to simulation. Consider a simple one-parameter logistic regression model as follows. Let $y_i$ be an indicator for the success on trial $i$ and let $x_i$ be a predictor for that trial. We model $y_i|\beta, x_i$ as a Bernoulli (0/1) trial with $\Pr(y_i = 1) = e^{\beta x_i}/(1 + e^{\beta x_i})$ and $\Pr(y_i = 0) = 1 - \Pr(y_i = 1)$. Such a model is only appropriate for situations when $x = 0$ is known to correspond to $\Pr(y_i = 1) = 0.5$. The prior distribution for $\beta$ is $p(\beta) \propto 1$. The observed data are 10 observations, $y = (1, 0, 0, 0, 1, 0, 1, 0, 0, 1)$ and $x = (1.1, -2.3, -1.0, -5.3, 2.5, -5.2, 2.6, -6.0, 2.6, 0.3)$.

(a) Create a grid approximation to the univariate posterior distribution of $\beta$ on the interval $[-1, 10]$. 

(b) Find the posterior mode and evaluate the second derivative of the log-posterior at the posterior mode. You can use any method you like to find the posterior mode (you can use glm code in R or any one-dimensional optimization strategy).

(c) Create a normal approximation to the posterior distribution of $\beta$ based on the information in (b). Graph the normal approximation over the same interval as you used for the grid. How well does the normal approximation work? Where does it perform worst?