STAT 352: Data Modeling and Inference

Homework 11. Due Friday April 15th. I encourage you to type all of your solutions, though this is not necessary. However, you must scan (or photograph) any handwritten portions and upload the files to Canvas. For questions that require `R` code, you must turn in your `R` code on Canvas. Your code must be in a .Rmd file that must compile.

Q1 (Compound Bayes Models Theory): *Preamble:* This question follows the same structure as Q1 from Homework 10, however in each problem some aspect of the sampling distribution of the data conditioned on the parameter or the marginal distribution of the parameter is changed. The basic setup of each problem will be data $y_1, \ldots, y_n$ that are conditionally independent when conditioning on a parameter θ. There will be some assumed prior distribution for the parameter θ. In the first three problems, θ is a single parameter and the marginal distribution for θ has been changed from that in Homework 10. This change can be accounted for by expanding the number of parameters in the problem and introducing a parameter ϕ. In these three problems, the data is independent of ϕ when conditioning on θ and both θ and ϕ are conditionally conjugate. In the fourth there are two parameters and in the fifth there are three parameters. For these latter two problems, the parameter expansion modifies the sampling distribution of the data. This is achieved by introducing a z_i for each y_i and giving z_i an appropriate distribution. Once again, the distributions are chose to be conditionally conjugate.
A warning: The paragraphs setting up the questions are kind of long, but I hope your tasks are relatively easy. I could have just given you the math to do, but I think it is a better (more instructive) set of problems if the reasoning behind each problem was explained. Your task for the problem will be set off in a separate paragraph started with **Task:**. For each problem, you are given two tasks. The first is to explain something we should be able to conclude easily based on the work from Homework 10. The second is to do a new calculation involving the added parameters.

A note on Conditionally Conjugate Sampling/Prior systems: This problem involves conditionally conjugate systems after the parameter expansions are introduced. In the first three problems (where we introduce ϕ to change the prior for θ), this means that the distribution for $\theta|\phi, y_1, \ldots, y_n$ is in the same family as $\theta|\phi$ and that the distribution for $\phi|\theta$ is in the same family as ϕ. In the second two problems (where we introduce z_i to changes the sampling distribution for $y_i|\theta$), it means that $\theta|z_1, \ldots, z_n, y_1, \ldots, y_n$ is in the same family as θ and that $z_i|y_i, \theta$ is in the same family as z_i. This will be very useful in Question 2 when we do sampling for these five problems.

Your Tasks:

a) In this problem we will assume that $y_1, \ldots, y_n|\theta \overset{iid}{\sim} \text{Bernoulli}(\theta)$ and that θ has prior (marginal) distribution that is a mixture of two Beta distributions. For convenience, we will assume that θ is a 50/50 mixture of a Beta(α_0, β_0) and a Beta(β_0, α_0) and we are going to assume that $\alpha_0 < \beta_0$. The prior density θ is given by

$$f(\theta) = \frac{1}{2}\frac{\Gamma(\alpha_0 + \beta_0)}{\Gamma(\alpha_0)\Gamma(\beta_0)} \times \left[\theta^{\alpha_0 - 1}(1-\theta)^{\beta_0 - 1} + \theta^{\beta_0 - 1}(1-\theta)^{\alpha_0 - 1}\right].$$

The idea behind this prior is that the first part of it will concentrate most of its mass on values less than one half while the second part if it will concentrate most of its mass on values greater than one half. The sampling mass function of the data is given by

$$f(y_1, \dots, y_n|\theta) = \prod_{i=1}^{n} f(y_i|\theta) = \prod_{i=1}^{n} \theta^{y_i}(1-\theta)^{1-y_i} = \theta^{\sum y_i}(1-\theta)^{\sum(1-y_i)} = \theta^{s_n}(1-\theta)^{f_n}$$

where $s_n = \sum y_i$ is the number of successes in the Bernoulli trials and $f_n = n - s_n$ is the number of failures. The posterior density of θ (the conditional distribution $\theta|y_1, \dots, y_n$) is given (up to proportion) by

$$f(\theta|y_1, \dots, y_n) \propto \theta^{s_n+\alpha_0-1}(1-\theta)^{f_n+\beta_0-1} + \theta^{s_n+\beta_0-1}(1-\theta)^{f_n+\alpha_0-1}$$

which is a mixture of two Beta distributions (after we divide by the normalizing constant to get area 1 underneath the curve). We could figure out the relative areas under these two Beta distributions and try to sample from the posterior directly, but it is easier to introduce another parameter $\phi \in \{0, 1\}$ which identifies the part of the mixture that θ came from in the prior. The conditional prior for θ given ϕ is

$$\theta|\phi \sim \begin{cases} \text{Beta}(\alpha_0, \beta_0) & \text{if } \phi = 0 \\ \text{Beta}(\beta_0, \alpha_0) & \text{if } \phi = 1 \end{cases}$$

and the marginal prior for ϕ is $\phi \sim \text{Bernoulli}(0.5)$. There is an implied assumption there that the data and ϕ are conditionally independent when conditioning on θ.

Task: Using the result for the conjugate system from Homework 10 Q1a, explain why

$$\theta|\phi, y_1, \dots, y_n \sim \begin{cases} \text{Beta}(s_n + \alpha_0, f_n + \beta_0) & \text{if } \phi = 0 \\ \text{Beta}(s_n + \beta_0, f_n + \alpha_0) & \text{if } \phi = 1 \end{cases}.$$

Additionally, use Bayes' Rule to show that

$$\phi|\theta \sim \text{Bernoulli}\left(\frac{\theta^{\beta_0}(1-\theta)^{\alpha_0}}{\theta^{\beta_0}(1-\theta)^{\alpha_0} + \theta^{\alpha_0}(1-\theta)^{\beta_0}}\right).$$

b) In this problem we will assume that $y_1, \dots, y_n|\theta \overset{iid}{\sim} \text{Poisson}(\theta)$ and that θ has prior (marginal) distribution that is Beta-Prime$(\alpha_0, \alpha_1, \beta_0)$. The prior density for θ is given by

$$f(\theta) = \frac{\Gamma(\alpha_0 + \alpha_1)}{\Gamma(\alpha_0)\Gamma(\alpha_1)} \frac{\beta_0^{\alpha_0}\theta^{\alpha_0-1}}{(1+\beta_0\theta)^{\alpha_0+\alpha_1}}.$$

The idea behind this prior is that it has a polynomial rate at both 0 and ∞ which are controlled by α_0 and α_1, respectively. The sampling mass function of the data is given by

$$f(y_1, \dots, y_n|\theta) = \prod_{i=1}^{n} f(y_i|\theta) = \prod_{i=1}^{n} \frac{\theta^{y_i}}{y_i!}\exp(-\theta) = \frac{\theta^{\sum y_i}}{\prod y_i!}\exp(-n\theta) \propto \theta^{s_n}\exp(-n\theta)$$

where $s_n = \sum y_i$ is the total count from the Poisson random variables. The posterior density of θ (the conditional density for $\theta|y_1, \dots, y_n$) is given (up to proportion) by

$$f(\theta|y_1, \dots, y_n) \propto \frac{\theta^{s_n+\alpha_0-1}}{(1+\beta_0\theta)^{\alpha_0+\alpha_1}}\exp(-n\theta).$$

We could try to sample from this distribution directly, it is not clear how to do so. We can introduce and additional parameter ϕ into the prior distribution for θ that separates the control at the origin from the control at infinity. The conditional prior for θ given ϕ is

$$\theta|\phi \sim \text{Gamma}(\alpha_0, \phi\beta_0)$$

and the marginal prior for ϕ is $\phi \sim \text{Gamma}(\alpha_1, 1)$. There is an implied assumption that the data and ϕ are conditionally independent when conditioning on θ.

Task: Using the result for the conjugate system from Homework 10 Q1b, explain why

$$\theta|\phi, y_1, \ldots, y_n \sim \text{Gamma}(s_n + \alpha_0, n + \phi\beta_0).$$

Additionally, use kernel matching to show that

$$\phi|\theta \sim \text{Gamma}\left(\alpha_0 + \alpha_1, \theta\beta_0 + 1\right).$$

c) In this problem we will assume that $y_1, \ldots, y_n|\theta \overset{iid}{\sim} \text{Gamma}(\nu_0, \theta)$ where the shape ν_0 is fixed and that θ has prior (marginal) distribution that is Weibull$(0.5, \beta_0)$. The prior density for θ is given by

$$f(\theta) = 0.5\frac{\sqrt{\beta_0}}{\sqrt{\theta}} \exp\left(-\sqrt{\beta_0\theta}\right).$$

The idea behind this prior is that it has an asymptote at 0 and a heavier tail at ∞ than any Gamma prior. The sampling density function of the data is given by

$$\begin{aligned} f(y_1, \ldots, y_n|\theta) &= \prod_{i=1}^n f(y_i|\theta) = \prod_{i=1}^n \tfrac{\theta^{\nu_0}}{\Gamma(\nu_0)} y_i^{\nu_0 - 1} \exp(-\theta y_i) = \tfrac{\theta^{n\nu_0}}{\Gamma(\nu_0)^n} (\prod y_i)^{\nu_0 - 1} \exp(-\theta \sum y_i) \\ &\propto \theta^{n\nu_0} \exp(-\theta s_n) \end{aligned}$$

where $s_n = \sum y_i$ is the sum from the Gamma random variables. The posterior density of θ (the conditional density for $\theta|y_1, \ldots, y_n$) is given (up to proportion) by

$$f(\theta|y_1, \ldots, y_n) \propto \theta^{n\nu_0 + \frac{1}{2} - 1} \exp\left(-\theta s_n - \sqrt{\beta_0\theta}\right).$$

We could try to sample from this distribution directly, it is not clear how to do so. We can introduce and additional parameter ϕ into the prior distribution for θ the origin and tail behavior of the Weibull. The conditional prior for θ given ϕ is

$$\theta|\phi \sim \text{Gamma}\left(\frac{1}{2}, \frac{\phi\beta_0}{2}\right)$$

and the marginal prior for ϕ is $\phi \sim$ Inverse-Exponential$(1/2)$. The marginal prior density for ϕ is given by

$$f(\phi) = \frac{1}{2\phi^2} \exp\left(-\frac{1}{2\phi}\right)$$

There is an implied assumption there that the data and ϕ are conditionally independent when conditioning on θ.

Task: Using the result for the conjugate system from Homework 10 Q1c, explain why

$$\theta|\phi, y_1, \ldots, y_n \sim \text{Gamma}\left(n\nu_0 + \frac{1}{2}, s_n + \frac{\phi\beta_0}{2}\right).$$

Additionally, use kernel matching to show that

$$f(\phi|\theta) \propto \frac{1}{\phi^{\frac{3}{2}}} \exp\left(-\left(\frac{\phi\beta_0\theta}{2} + \frac{1}{2\phi}\right)\right).$$

Remark: Though this might not seem very useful now, this distribution dates back to the turn of the twentieth century. It is an Inverse-Gaussian distribution and is related to Brownian motion (surely a teacher at some point made you put pollen in a glass of water and watch it move around, that is Brownian motion in three dimensions). The particular decomposition we are using to get the Weibull with power one half dates back to a paper by Andrews and Mallows in the Journal of the Royal Statistical Society Series B in 1974.

d) In this problem we will assume that $y_1, \ldots, y_n|\mu, \nu \overset{iid}{\sim} \text{Laplace}(\mu, \sqrt{\nu})$ where the mean μ and the scaling $\sqrt{\nu}$ are unknown (I am using the Greek letter ν, spelled "nu" and pronounced like "new," for the variance of a related Gaussian that we will get in this problem). We assume that (μ, ν) has prior (marginal) distribution that is Normal$(a_0, b_0) \times$ InverseGamma(δ_0, ζ_0). That is we assume that $\mu \perp\!\!\!\perp \nu$ in the prior and $\mu \sim$ Normal(a_0, b_0) and $\nu \sim$ Inverse-Gamma(δ_0, ζ_0). The sampling density function of the data is given by

$$\begin{aligned} f(y_1, \ldots, y_n|\mu, \nu) &= \prod_{i=1}^n f(y_i|\mu, \nu) = \prod_{i=1}^n \tfrac{1}{2\sqrt{\nu}} \exp\left(-\tfrac{|y_i-\mu|}{\sqrt{\nu}}\right) \\ &= \tfrac{1}{2^n \nu^{\frac{n}{2}}} \exp\left(-\sum \tfrac{|y_i-\mu|}{\sqrt{\nu}}\right) \end{aligned}$$

where we can see that there is no clean simplification because of the absolute value in the exponential. In particular, there are statistics to pull out that would allow us to simplify this further. The Normal(a_0, b_0) prior distribution has density

$$f(\mu) = \frac{1}{\sqrt{2\pi b_0}} \exp\left(-\frac{(\mu - a_0)^2}{2b_0}\right) \propto \exp\left(-\frac{\mu^2 - 2\mu a_0}{2b_0}\right)$$

and the Inverse-Gamma(δ_0, ζ_0) prior distribution has density

$$f(\nu) = \frac{\zeta_0^{\delta_0}}{\Gamma(\delta_0)} \nu^{-\delta_0 - 1} \exp\left(-\frac{\zeta_0}{\nu}\right) \propto \nu^{-\delta_0-1} \exp\left(-\frac{\zeta_0}{\nu}\right).$$

The joint posterior density of (μ, ν) is given up to proportion by

$$f(\mu, \nu|y_1, \ldots, y_n) \propto \nu^{-\frac{n}{2}-\delta_0-1} \exp\left(-\left(\sum \frac{|y_i - \mu|}{\sqrt{\nu}} + \frac{\mu^2 - 2\mu a_0}{2b_0} + \frac{\zeta_0}{\nu}\right)\right).$$

We could try to sample directly from this distribution, but it is not clear how to do so. We can resolve both issues (the absolute value and the $1/\sqrt{\nu}$ in the exponential) by introducing an extra variable for each y_i in a fashion similar to the use of ϕ in Q1c. In particular, we will introduce a z_i for each y_i and assume that the conditional sampling distribution of $y_i|z_i, \mu, \nu$ is normal with mean μ and variance ν/z_i. The z_i are assumed to be marginally independent of (μ, ν) and themselves independent and identically distributed following an Inverse-Exponential$(1/2)$ distribution. Formally, these assumptions are given by

$$\begin{aligned} f(y_1, \ldots, y_n, z_1, \ldots, z_n|\mu, \nu) &= \prod f(y_i, z_i|\mu, \nu) \\ f(y_i, z_i|\mu, \nu) &= f(y_i|z_i, \mu, \nu) f(z_i) \\ f(y_i|z_i, \mu, \nu) &= \sqrt{\tfrac{z_i}{2\pi\nu}} \exp\left(-\tfrac{z_i(y_i-\mu)^2}{2\nu}\right) \\ &= \sqrt{\tfrac{z_i}{2\pi\nu}} \exp\left(-\tfrac{z_i y_i^2 - 2z_i y_i \mu + z_i \mu^2}{2\nu}\right) \\ f(z_i) &= \tfrac{1}{2z_i^2} \exp\left(-\tfrac{1}{2z_i}\right) \end{aligned}.$$

Task: Using the result for the conjugate system from Homework 10 Q1d, explain why

$$\begin{aligned} \mu|\nu, y_1, \ldots, y_n, z_1, \ldots, z_n &\sim \text{Normal}\left(\frac{\sum z_i y_i/\nu + a_0/b_0}{\sum z_i/\nu + 1/b_0}, \frac{1}{\sum z_i/\nu + 1/b_0}\right) \\ \nu|\mu, y_1, \ldots, y_n, z_1, \ldots, z_n &\sim \text{Inverse-Gamma}\left(\frac{n}{2} + \delta_0, \frac{\sum z_i(y_i-\mu)^2}{2} + \zeta_0\right) \end{aligned}.$$

Additionally, use kernel matching to show that

$$f(z_i|y_i, \mu, \nu) \propto \frac{1}{z_i^{\frac{3}{2}}} \exp\left(-\left(\frac{z_i(y_i-\mu)^2}{2\nu} + \frac{1}{2z_i}\right)\right).$$

Remark: Just like in Q1c where we used this hierarchy for ϕ to change $\sqrt{\beta_0\theta}$ to $\phi\beta_0\theta$, we are using it here for z_i to change $|y_i - \mu|/\sqrt{\nu}$ to $z_i(y_i-\mu)^2/\nu$. We are changing the absolute value problem into a weighted squares problem where the weights are the z_i and the z_i are drawn from a specific distribution. We will do this again in Q1e, but with a different distribution for the z_i in order to do regression using a t distribution as the sampling distribution.
Hint: For $\mu|\nu, y_1, \ldots, y_n, z_1, \ldots, z_n$, think about writing the full conditional density as proportional to $\exp\left(-\frac{\mu^2 - 2\mu a_n}{2b_n}\right)$ for some a_n and b_n. Then you would know that the full conditional distribution for μ is normal with mean a_n and variance b_n.

e) In this problem we will assume that $y_1, \ldots, y_n|\alpha, \beta, \nu, x_1, \ldots, x_n \overset{ind}{\sim} \mathrm{T}_k(\alpha + \beta x_i, \nu)$ where this is a shifted and scaled student T distribution with k degrees of freedom. We will assume that k is fixed and greater than 2 in this problem, though typically one would try to estimate k. The conditional mean is given by $\mathrm{E}[y_i|x_i, \alpha, \beta, \nu, k] = \alpha + \beta x_i$ and the conditional variance is given by $\mathrm{Var}[y_i|x_i, \alpha, \beta, \nu, k] = \nu\frac{k}{k-2}$. The intercept α, the slope β, and the variance multiplier ν are all unknown. We assume that (α, β, ν) has prior (marginal) distribution that is $\text{Normal}(a_0, b_0) \times \text{Normal}(c_0, d_0) \times \text{InverseGamma}(\delta_0\zeta_0)$. That is we assume that $\alpha \perp\!\!\!\perp \beta \perp\!\!\!\perp \nu$ in the prior and $\alpha \sim \text{Normal}(a_0, b_0)$, $\beta \sim \text{Normal}(c_0, d_0)$, and $\nu \sim \text{Inverse-Gamma}(\delta_0, \zeta_0)$. The sampling density function of the data is given by

$$\begin{aligned} f(y_1, \ldots, y_n|\alpha, \beta, \nu, k, x_1, \ldots, x_n) &= \prod_{i=1}^n f(y_i|\alpha, \beta, \nu, k, x_i) \\ &= \prod_{i=1}^n \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\sqrt{\pi\nu k}} \frac{1}{\left(1 + \frac{(y_i - \alpha - \beta x_i)^2}{\nu k}\right)^{\frac{k+1}{2}}} \\ &\propto \nu^{-\frac{n}{2}} \prod_{i=1}^n \frac{1}{\left(1 + \frac{(y_i - \alpha - \beta x_i)^2}{\nu k}\right)^{\frac{k+1}{2}}} \end{aligned}$$

where we can see that there is no clean simplification because of the product of terms in the exponential. In particular, there are statistics to pull out that would allow us to simplify this further. The $\text{Normal}(a_0, b_0)$ prior distribution has density

$$f(\alpha) = \frac{1}{\sqrt{2\pi b_0}} \exp\left(-\frac{(\alpha - a_0)^2}{2b_0}\right) \propto \exp\left(-\frac{\alpha^2 - 2\alpha a_0}{2b_0}\right)$$

the $\text{Normal}(c_0, d_0)$ prior distribution has density

$$f(\beta) = \frac{1}{\sqrt{2\pi d_0}} \exp\left(-\frac{(\beta - c_0)^2}{2d_0}\right) \propto \exp\left(-\frac{\beta^2 - 2\beta c_0}{2d_0}\right)$$

and the Inverse-Gamma(δ_0, ζ_0) prior distribution has density

$$f(\nu) = \frac{\zeta_0^{\delta_0}}{\Gamma(\delta_0)} \nu^{-\delta_0 - 1} \exp\left(-\frac{\zeta_0}{\nu}\right) \propto \nu^{-\delta_0-1} \exp\left(-\frac{\zeta_0}{\nu}\right)$$

The joint posterior density of (α, β, ν) is given up to proportion by

$$\begin{aligned} f(\alpha, \beta, \nu | k, y_1, \ldots, y_n, x_1, \ldots, x_n) \quad \propto \quad & \nu^{-\frac{n}{2}-\delta_0-1} \exp\left(-\tfrac{\alpha^2 - 2\alpha a_0}{2b_0}\right) \exp\left(-\tfrac{\beta^2 - 2\beta c_0}{2d_0}\right) \\ & \times \prod_{i=1}^n \frac{1}{\left(1+\frac{(y_i - \alpha - \beta x_i)^2}{\nu k}\right)^{\frac{k+1}{2}}} \end{aligned} .$$

We could try to sample directly from this distribution, but it is not clear how to do so. We can resolve this problem by introducing an extra variable for each y_i in a fashion similar to the use of ϕ in Q1b. In particular, we will introduce a z_i for each y_i and assume that the conditional sampling distribution of $y_i | z_i, \alpha, \beta, \nu, k, x_i$ is normal with mean $\alpha + \beta x_i$ and variance ν / z_i. The z_i are assumed to be marginally independent of (α, β, ν) and themselves independent and identically distributed following a Gamma$(k/2, k/2)$ distribution. Formally, these assumptions are given by

$$\begin{aligned} f(y_1, \ldots, y_n, z_1, \ldots, z_n | \alpha, \beta, \nu, k, x_1, \ldots, x_n) &= \prod f(y_i, z_i | \alpha, \beta, \nu, k, x_i) \\ f(y_i, z_i | \alpha, \beta, \nu, k, x_i) &= f(y_i | z_i, \alpha, \beta, \nu, k, x_i) f(z_i | k) \\ f(y_i | z_i, \alpha, \beta, \nu, k, x_i) &= \sqrt{\tfrac{z_i}{2\pi\nu}} \exp\left(-\tfrac{z_i(y_i - \alpha - \beta x_i)^2}{2\nu}\right) \\ &= \sqrt{\tfrac{z_i}{2\pi\nu}} \exp\left(-\tfrac{z_i y_i^2 + \alpha^2 z_i + \beta^2 z_i x_i^2 - 2\alpha z_i y_i - 2\beta z_i y_i x_i + 2\alpha\beta z_i x_i}{2\nu}\right) \\ f(z_i | k) &= \tfrac{k^{\frac{k}{2}} z_i^{\frac{k}{2}-1}}{2^{\frac{k}{2}} \Gamma\left(\frac{k}{2}\right)} \exp\left(-\tfrac{z_i k}{2}\right) \end{aligned} .$$

Tasks: In order to save space, let $\vec{y} = (y_1, \ldots, y_n)$, $\vec{x} = (x_1, \ldots, x_n)$, and $\vec{z} = (z_1, \ldots, z_n)$. Using the result for the conjugate system from Homework 10 Q1e, explain why

$$\begin{aligned} \alpha | \beta, \nu, k, \vec{y}, \vec{x}, \vec{z} \quad &\sim \quad \text{Normal}\left(\tfrac{\sum z_i (y_i - \beta x_i)/\nu + a_0/b_0}{\sum z_i/\nu + 1/b_0}, \tfrac{1}{\sum z_i/\nu + 1/b_0}\right) \\ \beta | \alpha, \nu, k, \vec{y}, \vec{x}, \vec{z} \quad &\sim \quad \text{Normal}\left(\tfrac{\sum z_i x_i (y_i - \alpha)/\nu + c_0/d_0}{\sum z_i x_i^2/\nu + 1/d_0}, \tfrac{1}{\sum z_i x_i^2/\nu + 1/d_0}\right) \\ \nu | \alpha, \beta, k, \vec{y}, \vec{x}, \vec{z} \quad &\sim \quad \text{Inverse-Gamma}\left(\tfrac{n}{2} + \delta_0, \tfrac{\sum z_i (y_i - \alpha - \beta x_i)^2}{2} + \zeta_0\right) \end{aligned} .$$

Additionally, use kernel matching to show that

$$z_i | \alpha, \beta, \nu, k, y_i, x_i \sim Gamma\left(\frac{k+1}{2}, \frac{k}{2} + \frac{(y_i - \alpha - \beta x_i)^2}{2\nu}\right).$$

Remark: Just like in Q1d, we are once again trading the least squares problem for a weighted least squares problem. The difference is that the weights are being taken from a Gamma distribution, which produces the polynomial tail behavior (similar to how the Gamma distribution for ϕ in Q1b produced a polynomial tail for θ in that problem). Almost every robust regression problem is formulated as a weighted least squares problem where the weights are drawn from a particular distribution. The shape of the distribution of the weights dictates how much the loss cares about the contribution of small and large errors to the sum of squares. In fact, a lot of more general regression models (generalized linear models, which deal with conditional distributions for the data that might not even be supported on the whole real line) are able to be reformulated as weighted least squares problems. These are topics that would generally be covered in STAT 431 and STAT 432.

Hint: For $\alpha|\beta, \nu, k, y_1, \ldots, y_n, z_1, \ldots, z_n$, think about writing the full conditional density as proportional to $\exp\left(-\frac{\alpha^2 - 2\alpha a_n}{2b_n}\right)$ for some a_n and b_n. Then you would know that the full conditional distribution for α is normal with mean a_n and variance b_n.

Hint: For $\beta|\alpha, \nu, k, y_1, \ldots, y_n, z_1, \ldots, z_n$, think about writing the full conditional density as proportional to $\exp\left(-\frac{\beta^2 - 2\beta c_n}{2d_n}\right)$ for some c_n and d_n. Then you would know that the full conditional distribution for β is normal with mean c_n and variance d_n.

Question 2 (Posterior and Predictive Distribution Sampling): *Preamble:* We are going to think about the sampling for the Theory Question. The first three problems all involve posterior distributions for two parameters. The fourth problem involves $n+2$ parameters and the fifth problem involves $n+3$ parameters. So, for the first three we will have Gibbs Sampling with two stages and the second two will have Gibbs Sampling with three and four stages, respectively. We will compare the models from the Theory Question to those from Homework 10 using the predictive distribution. Skeleton code for all of the samplers is contained in the file `S352_HW11_Sp_2022_helper_code.R`. You will need to complete the skeleton code to sample from the predictive distributions from the Theory Question and modify the code as recommended to obtain draws from the predictive distributions for the models from Homework 10.

Your Tasks:

a) Load the `nodal` data using the code `library(boot); data(nodal); y = nodal$stage`. You will analyze the data as though the `y` values are conditionally iid following a Bernoulli(θ) distribution. We need to set α_0 and β_0, so let's take $\alpha_0 = 1.5$ and $\beta_0 = 8.5$. This makes the two components of the prior have means 0.15 and 0.85 and standard deviations that are about 0.11.

   - First, complete the Q2a skeleton code and obtain 10000 Gibbs samples from the posterior for (θ, ϕ) as well as 10000 predictions from the model from Q1a.
   - Second, modify this code to sample from the model from HW10 Q2a by setting $\alpha_0 = \beta_0 = 1$ and changing the ϕ sampling line so that every sampled ϕ is equal to 1. Use this modified code to obtain 10000 Gibbs samples from the posterior for θ as well as 10000 predictions from the model from HW10 Q2a.
   - Third, compare the predictive distributions from the two models you have just fit and taken predictions from. Do you think that the two models are making different predictions or essentially the same predictions.

b) Load the `fir` data using the code `library(boot); data(fir); y = fir$count`. You will analyze the data as though the `y` values are conditionally iid following a Poisson(θ) distribution. We need to set α_0, α_1, and β_0, so let's take $\alpha_0 = \alpha_1 = \beta_0 = 1$.

   - First, complete the Q2b skeleton code and obtain 10000 Gibbs samples from the posterior for (θ, ϕ) as well as 10000 predictions from the model from Q1b.
   - Second, modify this code to sample from the model from HW10 Q2b by changing the ϕ sampling line so that every sampled ϕ is equal to 1. Use this modified code to obtain

10000 Gibbs samples from the posterior for θ as well as 10000 predictions from the model from HW10 Q2b.

- Third, compare the predictive distributions from the two models you have just fit and taken predictions from. Do you think that the two models are making different predictions or essentially the same predictions.

c) Load the `poisons` data using the code `library(boot); data(poisons); y = poisons$time`. You will analyze the data as though the `y` values are conditionally iid following a Gamma(ν_0, θ) distribution where you will fix the shape a $\nu_0 = 4.31$ (this is the MLE for the shape parameter). We need to set the β_0, so let's take $\beta_0 = 1$.

- First, complete the Q2c skeleton code and obtain 10000 Gibbs samples from the posterior for (θ, ϕ) as well as 10000 predictions from the model from Q1c.
  Note: I have included loading the library `statmod` in the skeleton code. This library has a function, `rinvgauss`, for sampling from the Inverse Gaussian distribution. The full conditional for ϕ is Inverse Gaussian with mean $m = \frac{1}{\sqrt{\beta_0 \theta}}$ and shape $s = 1$. An Inverse Gaussian random variable with mean `m` and shape `s` can be sampled using the line `rinvgauss(1,mean=m,shape=s)`.
- Second, modify this code to sample from the model from HW10 Q2c by changing the ϕ sampling line so that every sampled ϕ is equal to 2 (this is to get rid of the division by 2 in the Gamma prior in the hierarchy). You also need to modify the θ sampling line so that the shape is $n\nu_0 + 1$ in order to match HW10 Q2c. Use this modified code to obtain 10000 samples from the posterior for θ as well as 10000 predictions from the model from HW10 Q2c.
- Third, compare the predictive distributions from the two models you have just fit and taken predictions from. Do you think that the two models are making different predictions or essentially the same predictions.

d) Load the `acme` data using the code `library(boot); data(acme); y=acme$acme`. You will analyze the data as though the `y` values are conditionally iid following a Laplace($\mu, \sqrt{\nu}$) distribution (I am using the Greek letter ν, spelled "nu" and pronounced like "new," for a scaled version of the variance). We need to set a_0, b_0, δ_0, and ζ_0, so let's take $a_0 = 0$ and $b_0 = \delta_0 = \zeta_0 = 1$.

- First, complete the Q2d skeleton code and obtain 10000 Gibbs samples from the posterior for (μ, ν) as well as 10000 predictions from the model from Q1d.
  Note: I have included loading the library `statmod` in the skeleton code. This library has a function, `rinvgauss`, for sampling from the Inverse Gaussian distribution. The full conditional for z_i is Inverse Gaussian with mean $m_i = \frac{\sqrt{\nu}}{|y_i - \mu|}$ and shape $s_i = 1$. A vector of Inverse Gaussian random variables of length `n` with mean vector `m` and shape vector `s` can be sampled using the line `rinvgauss(n,mean=m,shape=s)`. The mean and shape inputs are recycled as necessary to get to length n. Also, notice that the prediction sample is achieved by sampling an Inverse Exponential z_{new} from its prior and then a Gaussian y_{new} instead of using the Laplace distribution directly. Both are valid methods for making a predictive draw. As in HW10, I have included a `rinvgamma` function for sampling from the Inverse Gamma distribution that is parameterized in terms of the shape δ and the scale ζ.

- Second, modify this code to sample from the model from HW10 Q2d by changing the z sampling line so that every sampled vector z is equal to a vectors of 1s of length n and $z_{new} = 1$. Use this modified code to obtain 10000 Gibbs samples from the posterior for (μ, ν) as well as 10000 predictions from the model from HW10 Q2d.
- Third, compare the predictive distributions from the two models you have just fit and taken predictions from. Do you think that the two models are making different predictions or essentially the same predictions.

e) Load the `penguins` data using the code `library(palmerpenguins); data("penguins"); y = penguins$flipper_length_mm; x = penguins$body_mass_g; ind = !(is.na(y) | is.na(x)); y = y[ind]; x = x[ind]`. You will analyze the data as though the `y` values are conditionally independent with $y_i|\alpha, \beta, \nu, k, x_i$ following a $\mathrm{T}_k(\alpha + \beta x_i, \nu)$ distribution (I am using the Greek letter ν, spelled "nu" and pronounced like "new," for a scaled version of the variance). In this problem, we will assume that k is fixed and equal to 7, although one would typically try to estimate k. We need to set a_0, b_0, c_0, d_0, δ_0, and ζ_0, so let's take $a_0 = c_0 = 0$, $b_0 = 10^6$, and $d_0 = \delta_0 = \zeta_0 = 1$.

- First, complete the Q2d skeleton code and obtain 10000 Gibbs samples from the posterior for (α, β, ν) as well as 10000 predictions for $x_{new} = 4200$ from the model from Q1d. Note that the prediction sample is achieved by sampling a Gamma z_{new} from its prior and then a Gaussian y_{new} conditioned on x_{new} instead of using the Shifted and Scaled T distribution directly. Both are valid methods for making a predictive draw. As in HW10, I have included a `rinvgamma` function for sampling from the Inverse Gamma distribution.
- Second, modify this code to sample from the model from HW10 Q2d by changing the z sampling line so that every sampled vector z is equal to a vectors of 1s of length n and $z_{new} = 1$. Use this modified code to obtain 10000 Gibbs samples from the posterior for (α, β, ν) as well as 10000 predictions for $x_{new} = 4200$ from the model from HW10 Q2d.
- Third, compare the predictive distributions for $x_{new} = 4200$ from the two models you have just fit and taken predictions from. Do you think that the two models are making different predictions or essentially the same predictions for $x_{new} = 4200$?
- Fourth, compare the estimated conditional means (`mean(alpha_draws)+mean(beta_draws)*x`) for each model. Do you think that the two models are providing different estimates of the conditional means or essentially the same estimates of the conditional means?