- Problems 1 and 2 are worth 10 points each (instead of the usual 5).

- Recall: if $X \sim Gamma(a, b)$, then $\frac{1}{X} \sim InvGamma(a, b)$. So if you want to draw random samples from the Inverse Gamma, you could simply take draws from the corresponding Gamma and then take the reciprocal.

- Please use the associated `.Rmd` or `.qmd` document as there is some code in there that will be helpful to you! Note that some code chunks I've already designed as `echo = F` to hide in the knitted/rendered version.

1. **Hierarchical Normal model, common variance**. Recall the hierarchical model we derived in class:

$$Y_{ij}|\theta_j, \sigma^2 \stackrel{\text{iid}}{\sim} N(\theta_j, \sigma^2), \qquad i = 1, \ldots, n_j, j = 1 \ldots, J$$
$$\theta_j|\theta, \tau^2 \stackrel{\text{iid}}{\sim} N(\theta, \tau^2), \qquad j = 1, \ldots, J$$
$$\theta \sim N(\mu_0, \sigma_0^2)$$
$$1/\tau^2 \sim \text{Gamma}(a, b)$$
$$1/\sigma^2 \sim \text{Gamma}(c, d)$$

For this problem, we will be analyzing tennis serve speed data, provided in the `tennis_serves.Rda` file. These data contain serve speeds from 6 different male tennis players. The Y_{ij} represents player the i-th observation of player j's serve speed (miles per hour). As described by the model above, we will assume each tennis player has their own mean serve speed θ_j, but we are assuming there may be some dependence/shared information in the mean speeds across the J players.

(a) Specifying priors: For this model, we need to specify the following values for the prior distributions: $\mu_0, \sigma_0^2, a, b, c, d$. Specify values of your own choosing. Feel free to talk to your friends, your family, consult the internet, etc. Think about what each of these prior parameters represents in order to set some sensible values.

(b) Gibbs sampler: Approximate the joint posterior $p(\theta_1, \ldots, \theta_J, \mu, \tau^2, \sigma^2|\mathbf{y}_1, \ldots, \mathbf{y}_J)$ using a Gibbs sampler. In particular, using your choice of priors from (a), run your sampler with a burn-in of 5000 iterations, and obtain/retain samples for another 5000 iterations. **You should be able to use the code from class!**

(c) Assessing convergence: We should perform some MCMC diagnostics for all the parameters to justify that the overall MCMC has converged. Display traceplots for each parameter, and assess convergence. If it hasn't converged, go back to the previous part and run the chain for longer! *At this point, we should also do some PPCs, but that's not the focus of this week's material...*

(d) Posterior summaries: Once you've determined that your chain converged, obtain posterior means and 95% credible intervals for each θ_j, as well as for θ and the standard deviations σ and τ. Present these summary quantities in a single table (feel free to use ChatGPT to help make the table). Who has the slowest average serve speed? The fastest?

(e) Shrinkage: We will demonstrate what is called the *shrinkage property* of the shared prior for the θ_j. Using the provided starter code, create a plot that shows the observed sample mean serve speed for each player compared to the approximated posterior mean. Discuss what you see, and if/why it does or does not make sense.

(f) Sources of variability: recall that our models induces two sources of variability among the observed Y_{ij}'s: σ^2 and τ^2. To compare these sources of variation, one can obtain posterior samples of $R = \frac{\tau^2}{\tau^2+\sigma^2}$ using the posterior samples of σ^2 and τ^2. Plot a histogram of the approximate posterior distribution of R, and interpret in context what the plot tells us about the sources of variation in the data, according to our model.

(g) Extending the model: The magnitude of within-player variability could be attributed to two other reasons. First, we assumed that all players have the same variation in service speed. Second, if you look at the data, you'll notice that each player has a mix of fast and slow serves; this is probably due to the fact that players get two opportunities to serve. As a result, the first serve tends to be fast, and if they mess up, the player will typically take a slower, more conservative serve. Suppose we have data on if the serve was a first or second serve. Can you write a model that allows for player-specific variability, as well as different models for first and second serves for the players?

2. **Hierarchical Normal model, known but different sampling variance.** The North American Breeding Bird Survey (BBS) is a yearly survey to monitor bird populations. Regression models were used to estimate the change in population size for many species of birds between 1966 to 1999. Negative trend estimates correspond to estimated declines in population abundances, and positive trend estimates correspond to increased population abundance over time. For 28 particular grassland species of birds, for bird i we have the corresponding trend estimate $\hat{\beta}_i$ and its corresponding standard error σ_i. We assume the following "data" model (data is in quotes because we don't actually observe the $\hat{\beta}_i$ as they themselves are estimates):

$$\hat{\beta}_i|\beta_i, \sigma_i^2 \overset{\text{indep}}{\sim} N(\beta_i, \sigma_i^2) \qquad i = 1, \ldots, 28$$

where the σ_i^2 are assumed known from the standard errors. *Be careful: σ_i are standard errors, but the Normal is parameterized by variance σ_i^2.*

The data are stored in the file `bbs.Rda`. Take a look at the data before continuing!

(a) Model 1 (complete pooling): let's assume that the 28 species have the same true trend, i.e. $\beta_1 = \beta_2 = \ldots = \beta_{28} = \beta$. We will use a Normal prior for β with a variance of 25, and a mean that corresponds to no change in abundance over time. Find the resulting posterior distribution of β. *This will look like a previous model, but you cannot copy-paste results!*

(b) Model 1 (cont.): Let's perform a PPC. Using your work from (a), obtain 20 posterior predictive datasets (PPDs). Create a plot that overlays the density

estimates of these 20 PPDs, along with the density estimate of the original data. Color the lines by PPD vs original data. (This is what I believe Prof. Christian Stratton taught you, but come ask me if you have questions about making these plots!) Based on what you see, do you believe this model is a good fit to the data? Why or why not?

(c) Model 2: Next we will assume that the population trend estimates are distinct (i.e. $\beta_i \neq \beta_j$ *a priori*). We will create a hierarchical model by assuming the β_i are themselves a random sample from a Normal distribution with unknown mean β and variance τ^2. Pick some priors for β and τ^2 that leverage semi-conjugacy (you do not need to pick out particular values of the hyperparameters a, b, etc. yet). Once you do, use statistical notation to write out the statistical model in three lines in the order of: sampling model, stage-1 prior, and stage-2 prior. Be as clear as possible (e.g. conditioning where necessary, writing iid/ind where appropriate).

(d) Model 2 (cont.): We will use Gibbs sampling to approximate the posterior. Obtain the full conditionals for all the unknown parameters.

(e) Model 2 (cont.): To the best of your abilities, pick some sensible but relatively weak priors for β and τ^2. Then implement your Gibbs sampler from the previous step. Be sure to check for convergence. **You don't need to show me trace-plots; please confirm for yourself!** Once you've converged, perform the same PPC as in part (b), this time using the hierarchical model. *How we obtain PPDs follows the same general procedure as before, even though we have a hierarchical model!* Do you believe this Model 2 is a better fit to the data than Model 1? Why or why not?

(f) Then, make a table that compares the observed $\hat{\beta}_i$ to the posterior mean of β_i for each species, in increasing order of $\hat{\beta}_i$. What do you notice about the posterior means for the species whose estimated trends estimates $\hat{\beta}_i$ are close to versus far from the sample mean of the trend estimates $\bar{\hat{\beta}}$? In particular, comment on how the posterior mean of β_i for the two species with the most negative $\hat{\beta}_i$ changed, and try to explain why they changed!

(g) Do birds seem to be on the decline? Do some inference on β and τ^2 (or some function of τ^2) to answer this question.

3. **Hierarchical Poisson-Gamma model**. Let's return to the Italy marriage rates! We will now fit a hierarchical model to these data as another way to allow for shared information across time-periods. We will use a Poisson sampling model (as you all did in your case study), but now $Y_{ij}|\theta_j \overset{\text{iid}}{\sim} \text{Poisson}(\theta_j)$ for $i = 1, \ldots, n_j$ and $j = 1, 2, 3$. Here, j represents the three time periods: before, during and after. This sampling model still allows the three time periods to have different rates.

(a) We believe that the θ_j should have the same prior, whose distributions is parameterized by an unknown parameter λ that will enable us to pool some information across the time periods. To leverage semi-conjugacy, we will use an Exponential distribution as the stage-1 prior for the θ_j. *Recall that the Exponential is a special*

case of the Gamma. To continue leveraging semi-conjugacy, we will use either an Inverse-Gamma(a, b) or a Gamma(a, b) stage-1 prior for λ. *The choice of Inverse Gamma vs Gamma will be determined based on how you parameterize your stage-1 prior.*

Using statistical notation, write out the statistical model in three lines in the order of: sampling model, stage-1 prior, and stage-2 prior. Be as clear as possible (e.g. conditioning where necessary, writing iid/ind where appropriate).

(b) Sketch the graphical representation of the hierarchical model you created in (a). Also in words, interpret what λ represents in your model.

(c) We can use Gibbs sampling to sample from the posterior of the θ_j and λ! Obtain the full conditionals.

(d) (Optional for extra credit) Given your interpretation of λ in (b), choose a and b that could reasonably reflect beliefs about λ. You might want to do Monte Carlo sampling in R to help you!

(e) (Optional for extra credit) Implement your Gibbs sampler in R for 10,000 iterations and then throw away the first half to burn-in. Examine the sample ACF of the parameters. How is the autocorrelation compared to your model from the case study?

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.