1 Class 1

1.1 Structure of today

- Few minutes for urgent course content questions ...
- Review what is Bayesian (briefly) and the workflow I will teach for it
- Focus on simulating data from a linear regression
- END: course content, grading etc..

Okay! You all should have received my email about the course, which means
you're here because ...

- Excited to learn Bayesian inference and modeling!
- Excited to work together in pairs or teams during class (even if you're auditing)
- Know enough R to code actively in class
- Have a laptop and note-taking devices

If you're **not sure** about any of these, stay here and come talk with me after.

1.2 Who I am and course aims ...

Who am I? Quick review of how I learned Bayesian and how it is all I basically used now. Like many died in the wool Bayesians I believe this makes me:

- Happier, free from p-values
- Think harder about the science
- Have a WAY better sense of how the models I use work and how well they work on data similar to my own.
- More suspicious of a lot of stats

1.3 Before I dive in ... reminder: don't panic

No one gets everything in a stats class the first time, but you need to keep listening and not zone out.

1.4 What I want you to get out of this class

- The basics of what Bayesian is and how to implement
- The importance of a workflow and understand one I use and recommend
- Give some example of what they'll be able to do at the end (will vary by student)
- Get some of your burning questions answered ... **Feel free to ask questions/interrupt!**

- You probably will not come out of this class ready to analyze cut-point ordinal models for your community ecology data, but you'll have the workflow skills to start to think about to approach such a problem.

1.5 What is Bayesian? Pros and cons

It's a way of getting estimates from a model based on the likelihood from data and your prior beliefs.
It's a way of fitting and inferring from models that is extremely flexible and relies on prior knowledge. (That's basically all you need to know for today.)

Ask the students to list out pros and cons. Make sure they hit the below.

Pros

- Very flexible!
- Optimally handles uncertainty
- Intuitive
- No assumptions! No iid, nothing to memorize!
- Stop worrying about what your p-value is or contorting yourself to accurately define a CI
- Get mechanistic insights!
- Have a better sense of your parameter estimates

Cons

- No assumptions, you must check your own model and know what you're doing ...
- Computationally heavy

1.5.1 Types of Bayesians

There are **many** types of Bayesians:

- Andy Royle Bayesians with specific beliefs about how you fit mark-recapture models
- People obsessed with DAGs
- Facultative Bayesians
- Andrew Gelman Bayesians (BDA)

I will teach you my style of Bayesian ... which is pretty close to a Gelman Bayesian with other ideas (Betancourt etc.) thrown in.

This does not matter! Except when you go out into the world and meet the other Bayesians.

1.6 What is Bayesian? A workflow

1. Come up with your model
2. Simulate data from your model to check it
3. Prior predictive checks
4. Run your model on empirical data
5. Retrodictive checks (aka PPCs)

This class will focus on most of this workflow!
Except step 1 and we won't dwell on step 2 (prior checks).

1.7 Simulate from a linear model: Part 1

We're going to use something that works with linear regression for our model, so continuous x and continuous y

Get class to come up with an example and DRAW it out on a graph

Options: Plant growth in response to soil nutrient concentration, biometric scaling etc.

Ask students equation for a line.

Write out various notations and differentiate **parameters** from **data** (ideally, skip the error here)

Okay, I want to simulate data from this equation, what do I do?

In this section be sure to ...

- Slope versus intercept
- Come up with parameter numbers to write on board
- Mention `rnorm`
- Get the ERROR onto the equation if you have not already
- mention n
- What is an effect size?

Students should pair up and work on doing this with the following rules ...

- You must BOTH end up with the code you come up with.
- Simulate, plot and then try to figure out a way to tell if you have done it right...
- You alert me when you're done, stuck or have a question ... [If they are done, they should check their work using `lm`, then try to simulate a LOGISTIC regression.]

Note to self: Give the class a 10 minute break by 3pm!

1.8 Simulate from a linear model: Part 2

Come together and review how they did. Live code with them the course example using `lm` and `stan_glm` to check work.

Discuss: How might this be valuable?

And be sure to discuss why this is critical in Bayesian approaches ...
NO assumptions; you must CHECK and UNDERSTAND your model.

1.9 Simulate from a linear model: Add interactions

Review this if time allows ....

- Intercept only model
- Adding an interaction to a model ...

Go through the math on the board, introduce dummy variables and then set them to try to simulate a model with an interaction and see if they can return the parameters.
Or, we get to this tomorrow more likely ...

1.10 Review of course (by 4:20pm)

- 3 weeks, 6 classes, we'll get to hierarchical modeling
- ... but I am not sure when! I reserve the right to move things around (small chance I will start hierarchical modeling next week).
- Grading is participation and homework
- There are TWO homework (end of each of the first 2 weeks). Please do them! Even if you are auditing.
- No project, you must use a provided dataset
- Course managed on GitHub; you can submit homework on GitHub or Canvas.
- GitHub has wiki with resources ... Review (if time allows)
- We will use rstanarm, which is a version of Stan – make sure you have it running before the next class.
- Remind me to give you a BREAK in the middle of class
- Questions?

2 Class 2

Stuff to have prepped for this class ....

- TWO articles to show ...
- `https://www.countbayesie.com/blog/2015/2/18/hans-solo-and-bayesian-priors`
- The html from `https://github.com/lizzieinvancouver/bayesianflowsexample`
- `https://chi-feng.github.io/mcmc-demo/app.html?algorithm=RandomWalkMH&target=banana`

Review the workflow! Write it up and point to where we are
...maybe tell them – yes! You will spend more of your life fitting models to not your empirical data if you properly use the Bayesian workflow.
Give TWO examples of this in papers (post code for the test data later).
Maybe touch on – I have ended up with simpler models.

Review equations from yesterday ... Be sure to encourage them to move towards the one without $normal(0, \sigma)$ use $normal(\mu, \sigma)$

2.1 What is Bayesian: Posterior

Go over it. Maybe on the chalkboard ...

Use the webpage eventually to come up with an example Give my Star Wars example.

Discuss in pairs: Another example (if time allows)

Other examples: Complete separation.... Dolph's Iraq war example (which is not great).

2.2 What is Bayesian: Prior

Types of priors (informative, non-informative, weakly informative)

Discuss in pairs: An example of a prior you would set on a parameter related to your system (or just a fun example). (Only time allows)

Round robin of how would you set a prior for your data ... (set model first).

2.3 How much do priors matter?

It depends ..... ask the class what they think it depends on.

... ask the class what they think it depends on.

EXAMPLE: Show code where likelihood overwhelms prior. Can go through Stan briefly if time allows and introduce MCMC.

EXAMPLE: Show prior predictive check from `https://github.com/lizzieinvancouver/bayesianflowsexample/blob/main/example.html`

Why we will not do prior checks ...
Because they are annoying in rstanarm, brms etc. They are easier in raw Stan code.

2.4 MAYBE: Simulate another example ... and fit it in rstanarm

2.5 What is MCMC? And why do we need it ...

To get a posterior in Bayesian, we generally get samples from it.
Sampling for Bayesian models almost always have 3 ingredients:

1. Monte Carlo – process to generate random draws (`rnorm`)
2. Markov chain – Monte carlo with correlated steps
3. Algorithm – e.g., Metropolis Hastings

ON THE BOAD: give an example for linear regression (alpha, beta, sigma) and walk through for Metropolis Hastings.
NEXT: Discuss proposal issue and mention GIBBS.
THEN: mention Stan and what it does conceptually.

Then spend a while looking at: `https://chi-feng.github.io/mcmc-demo/app.html?algorithm=RandomWalkMH&target=banana`

Maybe mention 'hill-climbing' – which is what this is in some ways.

How many dimensions is a posterior? As many dimensions as the number of parameters you have.
So it's a complex space to search!
Mention GIBBS and how it works (long runs, search for chain hangups) versus Stan (divergences).

2.6 Review the homework assignment!

- Go over the tasks (homework on board)
- Review the datasets briefly (mention hierarchical)
- How to submit
  - GitHub – do you all want write access?
  - Canvas (but I prefer GitHub)
- What to do if you get stuck ... ask classmates for help, use Piazza, move onto next step.
- A note on using ChatGPT

3 Class 3

Review the workflow! Maybe write it up fast or such and point to where we are ...maybe tell them – yes! You will spend more of your life fitting models to not your empirical data if you properly use the Bayesian workflow.

Review the homework!

- What was hard?
- What parameters fit better or worse?
- Get to what they learned about interactions ... (16X)