

1 Class 1

FIX: Next time ...

- Draw out dots on $y = \alpha + \beta(x)$ versus with error (off line)
- Point out today or tomorrow that memorizing $\hat{y} = \alpha + \beta(x)$ and $y_i \sim \text{normal}(\hat{y}, \sigma^2)$ will make it easier to learn out distributions that focusing on $y_i = \alpha + \beta(x) + \epsilon$ and $\epsilon \sim \text{normal}(0, \sigma^2)$

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 Round-robin of quick intros

Who is in the class...

1.3 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

1.4 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.

1.5 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.

1.6 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

1.6.1 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.

1.7 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).

1.8 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!

1.9 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.

1.10 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 ...

1.11 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

(May want to have start of class 2 (MCMC) prepped!)

Stuff to have prepped for this class Part 1 (**finished at 4:15pm first time, but depends on questions!**)

- TWO or three articles to show ... (Flynn & Wolkovich, OSPREE bb ms and Morales-Castilla)
- Code that goes with some of them
- <https://www.countbayesie.com/blog/2015/2/18/hans-solo-and-bayesian-priors>
- The html from <https://github.com/lizzieinvancouver/bayesianflowsexample> for prior predictive checks (but I could find/make better example)
- priorpost code (misspecify model)

Stuff to have prepped for this class Part 2 (MCMC)

- The html from <https://github.com/lizzieinvancouver/bayesianflowsexample>
- <https://chi-feng.github.io/mcmc-demo/app.html?algorithm=RandomWalkMH&target=banana>

Stuff to have prepped for this class HOMEWORK review at the end!

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-THREE examples of this in papers (post code for the test data later).

- Flynn & Wolkovich: I stopped using 2-way interactions
- Ettinger et al. 2020: No interactions given uneven data and following the workflow we realized this was not good, main effects average over interactions
- Morales-Castilla et al. 2024: We ask different questions, no need for big models, but still writing important papers (could mention decsens)

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

Note to LIZZIE: Think about order below

Two take-homes I want them to get ... This is repeated later in this document ...

1. It's not that hard to come up with potentially informative priors that you should feel good about. You have them most often, you just need to get better at thinking about them.
2. In many models, data can overwhelm your priors.

Types of priors (informative, non-informative, weakly informative)... in 2024 I wrote them out and crossed off non-informative, I taught weakly as 3X and draw out on board what happens if prior constrains posterior and then you go up 3X. Be sure to WRITE out the math of the prior here (what I mean by 3X) and DRAW it.

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).

Could give my example of coming up with priors for synchrony (days per decade known of 3-5 days per decade, also give extremes we know are too big – 100 days per year or decade is clearly NOT happening, for example – and that y axis limits me some here). Should also draw and WRITE out example here. Could even open class with this as a STORY?

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.

Two take-homes: Make sure you say this! From above...

1. It's not that hard to come up with potentially informative priors that you should feel good about. You have them most often, you just need to get better at thinking about them.
2. In many models, data can overwhelm your priors.

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

2.5 OR: Start on MCMC...

Moved my notes from here to next class

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

Stuff to have prepped for this class (MCMC)

- HOMEWORK!
- Gelman blog posts on interactions ...
- <https://chi-feng.github.io/mcmc-demo/app.html?algorithm=RandomWalkMH&target=banana>
- `nep_needed.png`
- `diagnostiscsday.R` and make sure it's running well

Review the homework!

Have the class count off by 4-6 or such to end up in groups of 3-4 and then compare their homework

During this I should get a tally on which datasets people picked.

In small groups cover:

- Decide which parameters fit best or worst and discuss WHY?
- What common issues did they have in coding?

As a whole group, cover the homework to discuss ...

- What was hard?
- What parameters fit better or worse?
- Get to what they learned about interactions ... (16X) – review what you need to calculate power:
 - Sample size
 - error (σ_y)
 - effect size ('sledgehammer' experiments are used for a reason)
- I will post my version of the homework ...

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

Announce to class: You do not need to understand MCMC (or HMC) super well to do Bayesian, but you need to understand the diagnostics and having some sense of the algorithm can help with that (but don't panic if you don't get the algorithm, focus on the diagnostics and keep learning).

To get a posterior in Bayesian, we don't compute it, we instead (generally) get samples from it using an algorithm.

Sampling for Bayesian models almost always have 3 ingredients:

1. Monte Carlo – process to generate random draws (**rnorm** or (θ, σ))
2. Markov chain – Monte Carlo with correlated steps (θ_{n-1}, σ)
3. Algorithm – e.g., Metropolis Hastings (FYI, you can use this for things other than getting samples from a posterior)

How many dimensions is a posterior? As many dimensions as the number of parameters you have.

ON THE BOARD: give an example for linear regression (alpha, beta, sigma) and walk through for Metropolis Hastings.

Steps are:

1. Start somewhere (initial conditions)
2. Use the Monte Carlo to get a random place to jump (θ_{new})
3. Now evaluate your posterior here relative to where you were
4. If your jump is better, accept and jump there! (θ_{new+1})
5. If your jump is no better, then you compare the ratio to a random draw from a uniform (0,1) distribution; if the ratio is higher than this value you jump anyway. (Acceptance probability.)
6. If the ratio is lower than this value you **don't jump** – your proposal is **rejected**; draw new random jump (step 2) and repeat

Ask the class: Why do you sometimes accept lower ratios?

To help your chain search fully ... Each random draw means you can go down the hill. Otherwise your chain gets into crevices and does not search fully, **SEARCHING fully the challenge**. Give the roomba vacuum example.

Problems: (1) The start of the chain is not good.

Problems: (2) The steps are correlated.

Solutions? One is multiple chains!

Others ...

- Drop the start of the chain
- Thin the samples
- Another is to run chains a LONG time (Gibbs)
- Get a better algorithm

NEXT: Discuss proposal issue and mention Gibbs (Gibbs uses conjugate priors to make more efficient proposals).

THEN: mention Stan and what it does conceptually. See page 247 of *Statistical Rethinking*: HMC runs a particle simulation so it can move quickly through flat posterior space and slower through complex space.

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.)

MENTION: Warmup and total iterations needed ...

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).

Divergent transitions: “A divergence arises when the simulated Hamiltonian trajectory departs from the true trajectory as measured by departure of the Hamiltonian value from its initial value. When this divergence is too high, the simulation has gone off the rails and cannot be trusted.” In other words, something is wrong in your model.

A few notes on divergences:

- You should not report estimates in a paper if you have divergent transitions! You must get rid of them ...
- Technically, any are bad and you should consider re-parameterizing your model to make your posterior a friendlier shape to search
- Some times smaller steps get rid of them (adapt delta command)
- Usually you can get rid of a small number (meh, under 30) – if you have hundreds for 4000 steps then something is DEEPLY wrong.

3.2 MCMC diagnostics

The three basic ones!

1. Rhat – compare variance with a chain to variance across chains
2. $n_{effective}$ – aim for 10% of your total iterations
3. msce – Markov Chain SE – I never use this as I have never seen it be uniquely helpful

And ... Divergences – pairs plots – LOOK up `npc_needed.png` And discuss.

Go through ShinyStan using `diagnostiscsday.R`

4 Class 4

Stuff to have prepped for this class

- Homework assignment! For the end
- OPSREE paper showing plot of random effects
- Code that simulates hierarchical data
- Plot showing partial pooling (shrinkage)

Review the workflow! Write it up fast with the help of the class and point to where we are Still simulating data, but a different type.

Ask the class, how do I know when I need a hierarchical model?

GOAL: Get to where I draw out a random intercepts model and a random slopes (only) model and a random slopes and intercepts model!

4.1 Math of a hierarchical model

Start with lmer syntax, then linear regression syntax, then move on ...

4.2 Pooling

Review: complete pooling, no pooling, partial pooling ...

Why do I compare these?

... what determines HOW MUCH pooling?

4.3 Let's simulate data from a hierarchical model!

Get the class to start discussing how to do this; then have them TRY it!

4.4 Fixed versus random effects, which is better?

Ask the class: Should you fit random effects when you really care or don't care about the effects?

4.5 Crossed versus nested

4.6 Some terminology

Hyperparameters – model parameters without group level subscripts (the μ and σ – overall averages)

4.7 Review the homework assignment!

- Datasets ...

5 Class 5

6 Class 6

Latent parameter examples?

Qualitative data (elongated fish example) – could briefly introduce MXTURE models here, as you could use the shapes as the way to possibly define different groups, where each group gets a different body-mass to length relationship.