

Applied Bayesian Modeling for Natural Resource Management

FW 891

Christopher Cahill

28 August 2023



Quantitative Fisheries Center
MICHIGAN STATE UNIVERSITY

Outline: Part I

- Introductions
- Syllabus and recommended background
- Course organization
- Some Bayesian examples from my own work
- Philosophy and course themes
- Software and website
- Expectations and disclaimers
- Break

Introduction to FW 891

- Welcome
- Housekeeping
 - Washrooms, Zoom, UPLA, office hours, etc.
- Round-table introductions
 - Background, area of study, why are you here, ice breaker

Syllabus

- Everything through Github
 - <https://github.com/QFCatMSU/FW-Bayes>
- Syllabus, links to readings, etc.
- All presentations and code
- Office hours by appointment
 - cahill11@msu.edu

Course organization

- Monday lectures where we generally cover theory/concepts
- Wednesday labs where we work on implementation and in-class exercises
- 4 homework assignments
- 1 class project
 - Preliminary draft due 22 Sept
 - Final presentation/write-up 4 and 6 Dec

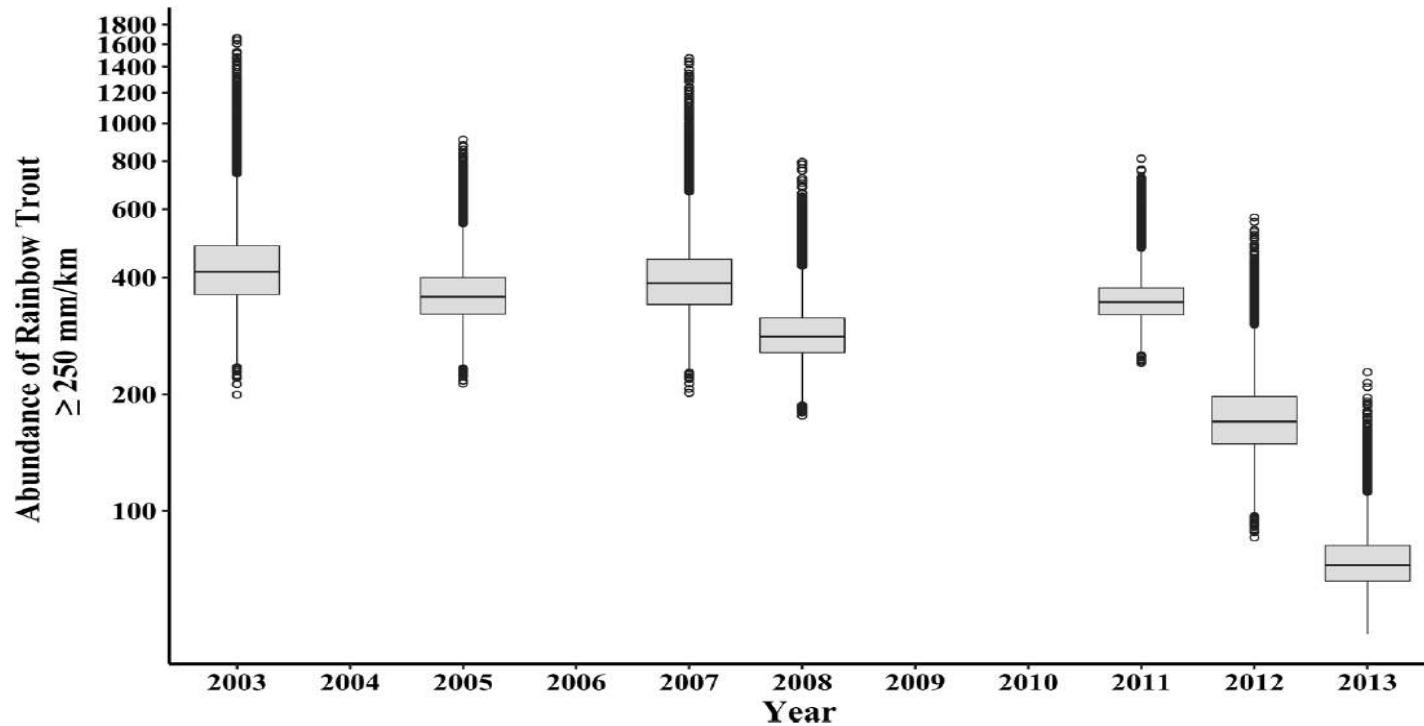
Recommended background

- Previous experience with statistics
- Previous experience programming
 - for-loops, functions, data visualization
- You can get by without this background, but you should expect to put in more time

My background



Bayesian examples from my work



- Decline was *not* statistically significant (frequentist)
- 88% chance of population decline (Bayesian)

Bayesian examples from my work

50% drop in rainbow trout puts Bow River's 'world-renowned' sport fishery at risk



Multiple 'smoking gun' stressors behind the drastic population decline

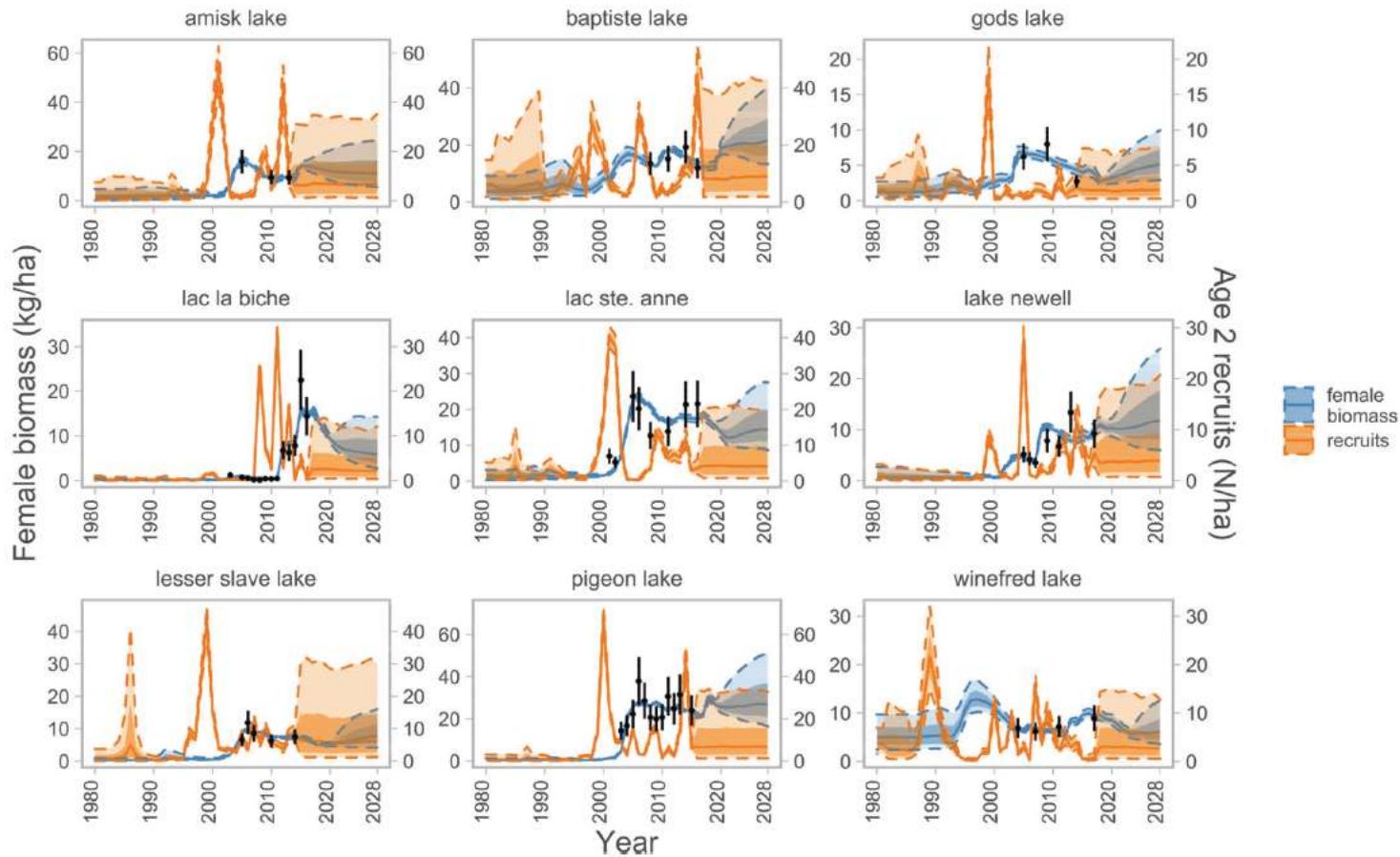
Sarah Rieger - CBC News - Posted: Nov 26, 2018 7:18 PM EST | Last Updated: November 26, 2018



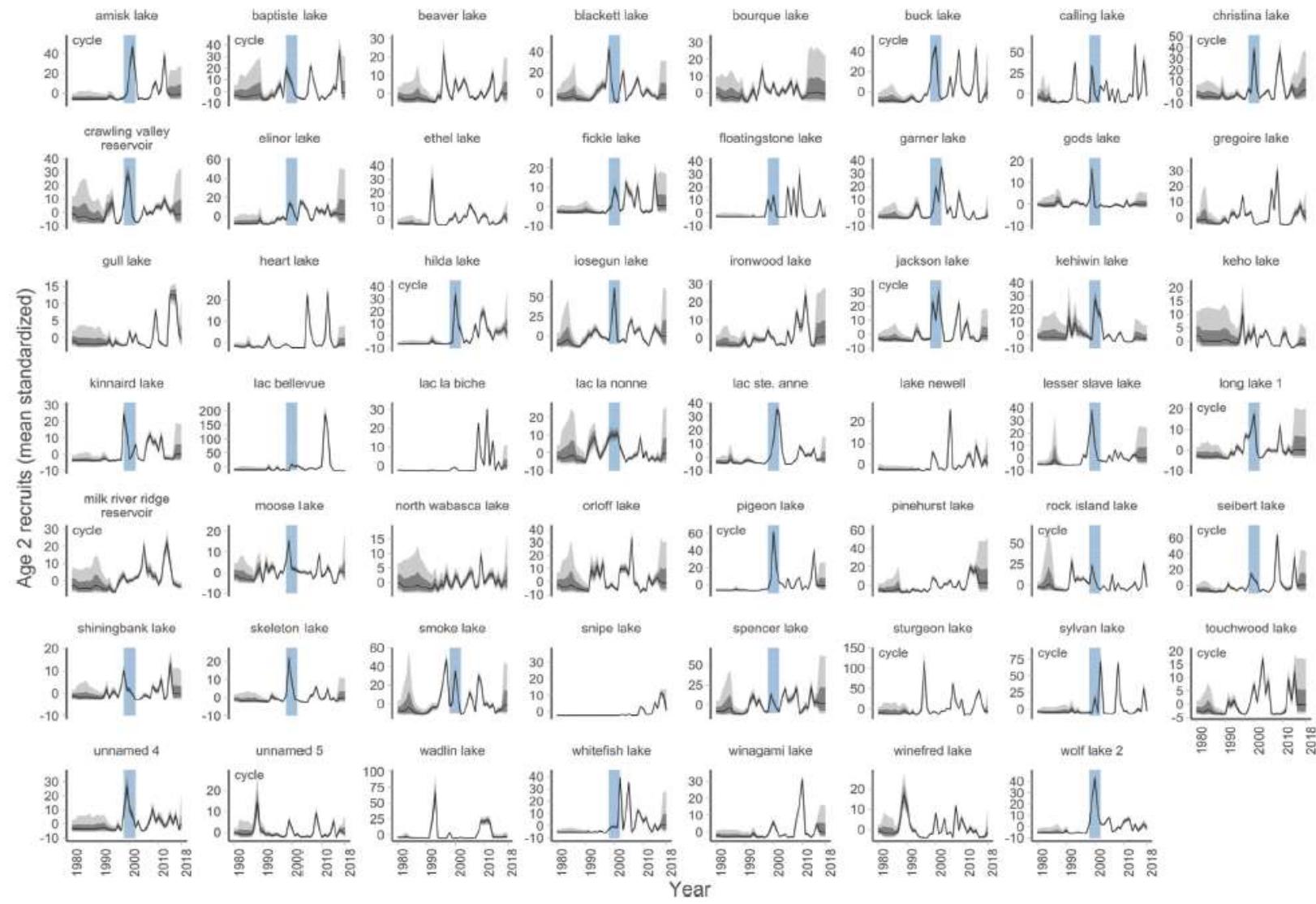
Provincial fisheries data from the lower Bow River shows the population of rainbow trout dropped between 43 and 50 per cent over a 10-year period. (Associated Press)

Examples from my work cont'd

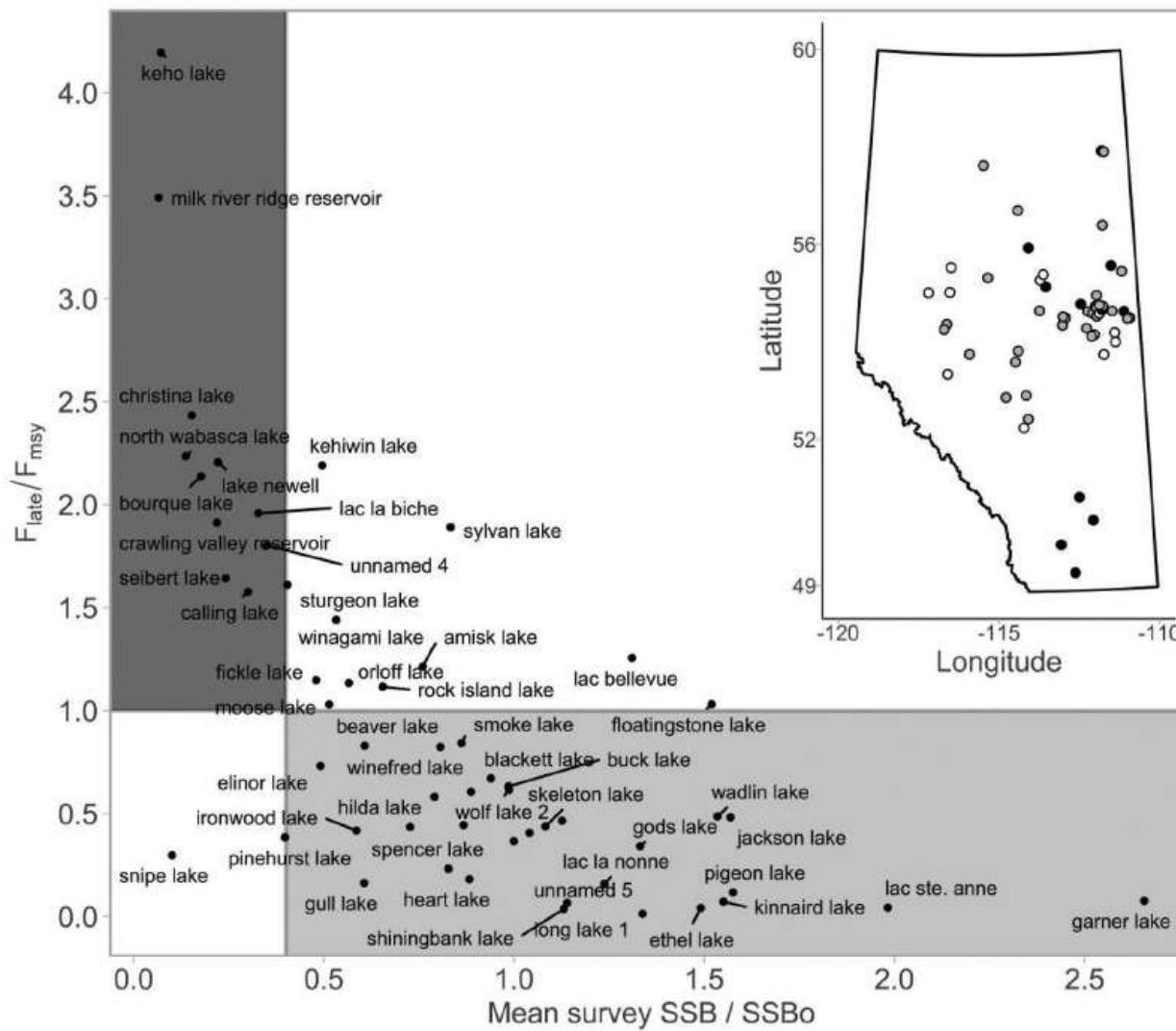
- Reconstructing population dynamics across landscapes



Examples from my work cont'd



Examples from my work cont'd



My philosophy on statistical modeling

- The reliability of inferences matters when *management of public-trust natural resources* might be affected by your analyses
 - We will spend *a lot* of time evaluating models
- Intuition is critical in applied ecological analyses
 - We will be trying to improve your intuition with Bayesian modeling
- There are rarely cookbook solutions to hard problems
 - All about improving your statistical toolbox

Course themes

- Development of fairly advanced implementation skills using R and Stan
 - Why?
- There are easier ways to do some of the analyses we will learn
- By learning to do easy things the hard way, the hard things become easy (Royle and Dorazio 2008)
- Open source and reproducible science

Software, implementation, website

- We are primarily going to use R and Stan
- Recommend Rstudio
- Lectures and code will be available through GitHub
 - You do not need to know how to use GitHub, but that is where you can find code and presentations



mc-stan.org



Stan is not



What is Stan, why use it?



- Stan is a state-of-the-art platform for statistical modeling and high-performance statistical computation
- Named after [Stanislav Ulam](#)
- Unique class of algorithms (HMC, NUTS), scales well
- Cannot do everything (discrete unknowns)



Expectations

- Be kind
- Make an honest effort to learn this stuff
- Share your code
 - I share my code with all of you, so please do the same
 - Sharing code is critical for reproducible science
- This course lies at the intersection of mathematics, statistics, ecology, and numerical computing
 - Failure is okay and to be expected

Some disclaimers

1. Bayesian statistics is a vast and difficult topic
2. This is not a mathematical statistics course
3. I will do my best to track down answers
4. I don't know everything
5. Please ask questions
6. When in doubt, see disclaimers 3-5

Break

- Return in 10 minutes

Outline: Part II

- Introduction to statistical modeling
- Random variables in words and math
- The Bernoulli and other common distributions
- Conducting a statistical inference
- Bayes' rule
- Frequentist and Bayesian sneak turtles 
- Pros and Cons of Bayesian inference
- Summary and outlook

A brief introduction to statistical modeling

- This class is about model-based inference
 - Focus on the development of arbitrarily abstract statistical models
- These models always contain:
 - Deterministic (i.e., systematic) components
 - Random (i.e., stochastic) components
- Typically use a model to estimate parameters of interest, conduct an inference, or predict something
- This section draws heavily from Royle and Dorazio (2008), particularly chapter 2

Random variables: words

- Any observation can be viewed as the realization (or outcome) of a stochastic process
 - Phrased differently: chance plays a part in what we observe
- Simplest example: binary outcome
 - Alive or dead, mature or not mature, detected or sneaky (not-detected)
- Element of chance in all of outcomes and we need a way to deal with that

Random variables: maths

Consider a model of the binary random variable Y :

$$\text{suppose } p = \Pr(Y = 1)$$

$$\Pr(Y = 1) + \Pr(Y = 0) = 1$$

$$\Pr(Y = 0) = 1 - p$$

- \Pr = probability
- Can express a statistical model of the observable outcomes

Bernoulli probability mass function (pmf)

$$\Pr(Y = y) = p^y(1 - p)^{1-y}$$

- This thing is called a probability mass function (pmf), and this equation represents pmf of a Bernoulli distributed random variable
- y is observed datum
 - pmf = discrete outcomes
- Note: if this was a continuously distributed random variable we would use probability *density* function (pdf)

Bernoulli probability mass function (pmf)

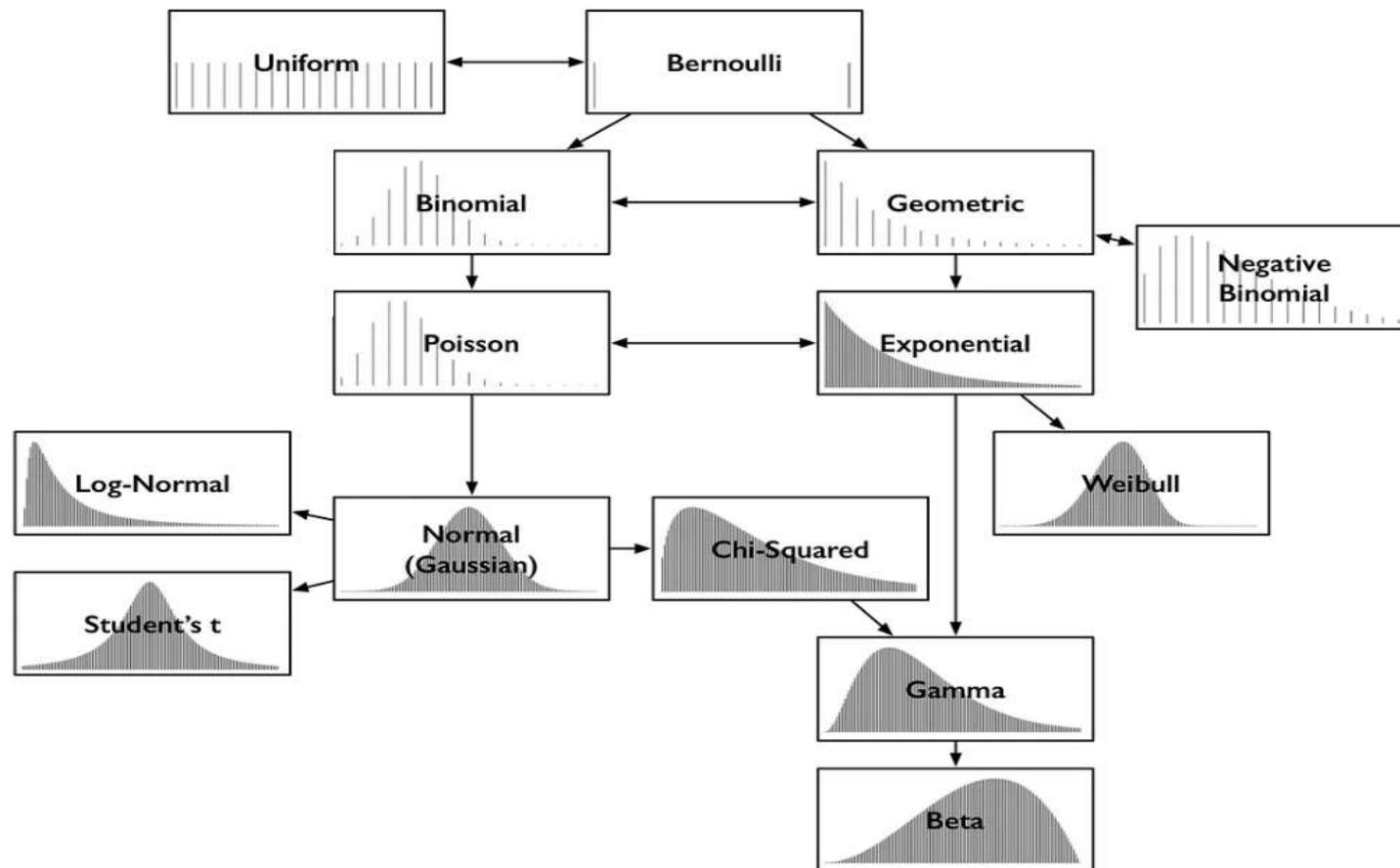
- A conventional notation for this stuff is

$$f(y \mid \theta)$$

- Intended to indicate that the probability of an observed value y depends on parameter(s) θ used to specify the distribution of the random variable Y :

$$f(y \mid \theta) = p^y(1 - p)^{1-y}$$

Some common statistical distributions



Common probability distributions and some key relationships

Figure Reference

More notation notes for everyone's sanity

- Will try provide its pmf (or pdf) expressed as a function of y and the parameters of the distribution.
- For example, will use $y \sim N(\mu, \sigma)$ to indicate a random variable Y is normally distributed with mean μ and standard deviation σ
- Regular font for scalars and a bold font for **vectors** or **matrices**

More notation notes for everyone's sanity

- May write it this way, too:

$$\text{Normal}(y \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2} \left(\frac{y - \mu}{\sigma}\right)^2\right).$$

Conducting a statistical inference

- Given some data, a mathematical model, and the desire to infer something what do we do?
- **Frequentist methods:**
 - Uncertainty is expressed in terms of the variability of hypothetical replicate data sets
 - What is the meaning of a 95% *confidence* interval?
- **Bayesian methods:**
 - Uncertainty expressed through the lens of probability
 - What is the meaning of a 95% *credible* interval?
 - Does this through the use of Bayes' rule

What is Bayes' rule?

- Thomas Bayes (1702 - 1761) was a mathematician and Presbyterian minister in England
 - Developed a simple algebraic rule that has wide-ranging ramifications for statistical inference
 - Pierre-Simon Laplace independently rediscovered and extensively developed the methods

$$p(A \mid B) = \frac{p(B \mid A)p(A)}{p(B)}$$

Bayes' rule

$$p(A \mid B) = \frac{p(B \mid A)p(A)}{p(B)}$$

- The conditional probability of observing A given that B happened equals the conditional probability of observing B given A, times the marginal probability of A divided by the marginal probability of B
- A and B are *observable* events

Statistical inference using Bayes' rule

$$p(\theta | D) = \frac{p(D | \theta)p(\theta)}{p(D)}$$

- θ is parameter of interest
- D is data
- $p(D | \theta)$ is the likelihood
- $p(\theta)$ is the prior
- $p(\theta | D)$ is the posterior
- $p(D)$ is the marginal probability of the data
 - normalization constant

Bayes' rule can be paraphrased as

$$\text{posterior} \propto \text{likelihood} \cdot \text{prior}$$

- Posterior distribution is *always* a result of both the information contained in the data *and* our prior knowledge (assumptions) about model unknowns
- Must formally express our *a priori* knowledge in the form of a prior probability distribution for each parameter
- Bayes' rule is a mathematical formalization of how we update beliefs when new information becomes available

An example to start us off

- Determine the detection probability of invasive Alabama sneak turtles by releasing some into a small pond and counting them on a later occasion
- We release $n = 30$ and later count $y = 2$ sneakers
 - successes vs. trials
- What's probability of detecting a turtle, i.e., θ ?

$$p(y \mid \theta) = \frac{n!}{y!(n-y)!} \theta^y (1-\theta)^{n-y}$$

- There are both frequentist and Bayesian ways to solve this binomial detection probability problem

Analytical maximum likelihood estimate

$$L(\theta \mid y) \propto \theta^y (1 - \theta)^{n-y}$$

$$L(\theta \mid y) \partial \theta = \theta^y (1 - \theta)^{n-y} \left(\frac{y}{\theta} - \frac{n-y}{1-\theta} \right)$$

$$\theta^y (1 - \theta)^{n-y} \left(\frac{y}{\theta} - \frac{n-y}{1-\theta} \right) = 0$$

$$\hat{\theta} = \frac{y}{n}$$

- This is frequentist
- Note the “flip”

Frequentist Alabama sneak turtles



- Now let's do it numerically in R to learn:

$$L(y \mid \theta) \propto \theta^y (1 - \theta)^{n-y}$$

```
1 library(tidyverse) # for plotting
2 # install.packages("devtools")
3 # devtools::install_github("QFCatMSU/gg-qfc")
4 library(ggqfc) # formatting, run previous 2 lines for first gg-qfc install
5
6 y <- 2 # successes
7 n <- 30 # trials
8 eps <- 1e-6 # buffer because theta must lie within [0,1]
9 theta <- seq(eps, 1 - eps, length.out = 1e4) # sequence to thetas to try
10 logLike <- y * log(theta) + (n - y) * log(1 - theta) # plug thetas into pmf
11 MLE <- theta[which.max(logLike)] # maximum likelihood estimate
12 my_data <- data.frame(logLike, theta) # for plotting
```

Analytical vs. numerical MLEs (frequentist)

```
1 y / n # analytical theta
```

```
[1] 0.06666667
```

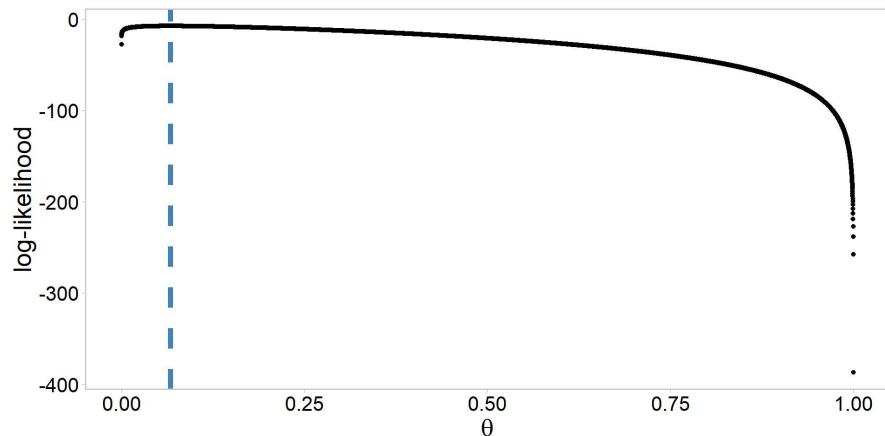
```
1 MLE # numerical approximation theta
```

```
[1] 0.06670754
```

- pretty darn close
- sneak turtles are devilish 

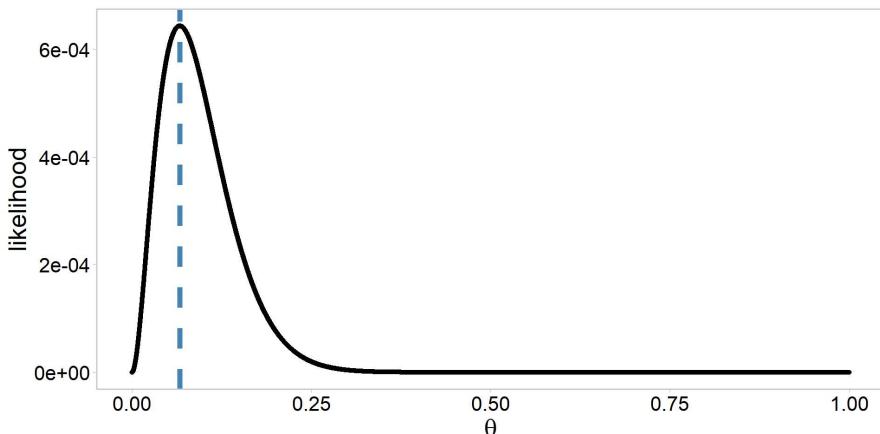
Alabama sneak turtles: log-likelihood

```
1 my_data %>%
2   ggplot(aes(x = theta, y = logLike)) +
3   geom_point(size = 1.4) +
4   ylab("log-likelihood") +
5   xlab(expression(theta)) +
6   geom_vline(
7     xintercept = MLE, linetype = 2, linewidth = 2,
8     color = "steelblue"
9   ) # add MLE
10  theme_qfc() +
11  theme(text = element_text(size = 20))
```



Moving from log-likelihood to likelihood

```
1 my_data %>%
2   ggplot(aes(x = theta, y = exp(logLike))) + # note exp()
3   geom_point(size = 1.4) +
4   ylab("likelihood") +
5   xlab(expression(theta)) +
6   geom_vline(
7     xintercept = MLE, linetype = 2, linewidth = 2,
8     color = "steelblue"
9   ) +
10  theme_qfc() +
11  theme(text = element_text(size = 20))
```



How might we do the sneak turtle analysis in a Bayesian framework?

- Remember:

$$\text{posterior} \propto \text{likelihood} \cdot \text{prior}$$

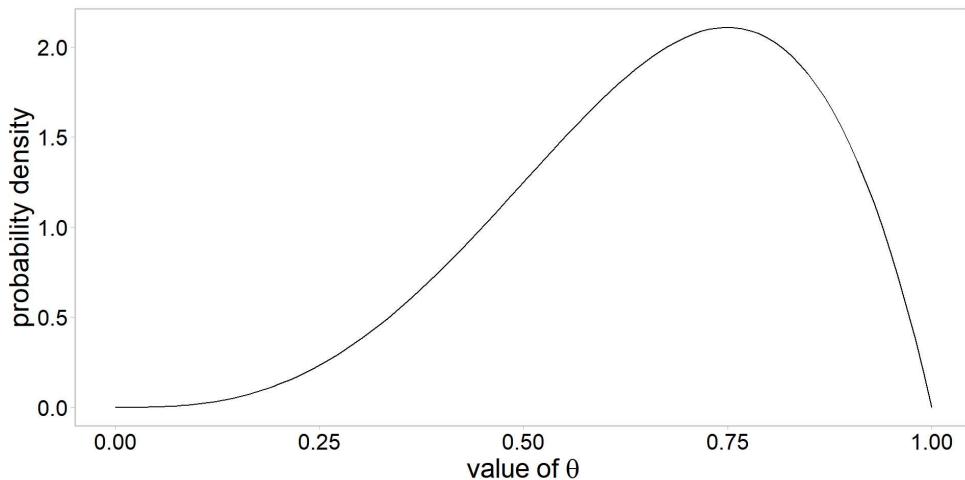
- What do we currently have, what do we know?
- Assume the following prior:

$$\theta \sim \text{beta}(\alpha, \beta)$$

- How to choose values for α and β ?

Visualizing a $\text{beta}(\alpha, \beta)$ prior

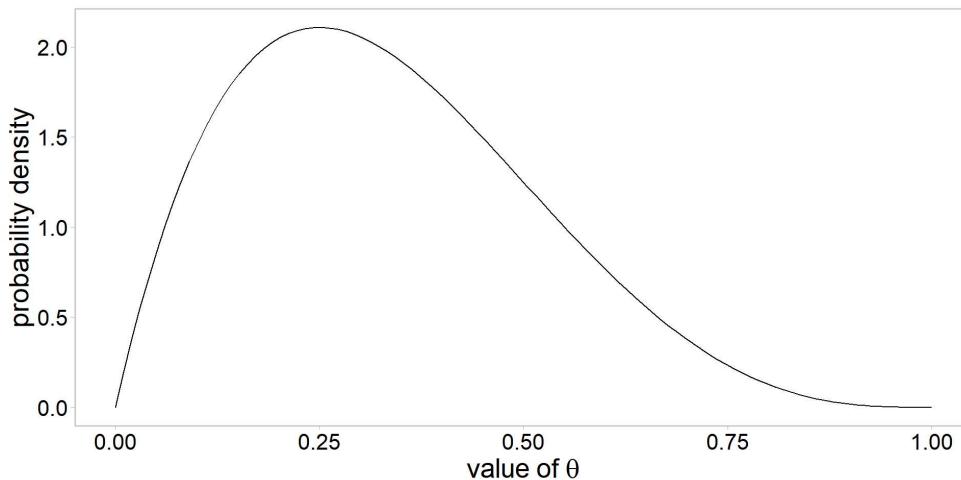
```
1 x <- seq(0, 1, len = 100) # create a sequence
2 db <- dbeta(x, shape1 = 4, shape2 = 2) # see ?dbeta
3 ggplot() +
4   geom_line(aes(x, db)) +
5   xlab(expression(value ~ of ~ theta)) +
6   ylab("probability density") +
7   theme_qfc() +
8   theme(text = element_text(size = 20))
```



- always a good idea to visualize

Visualizing a $\text{beta}(\alpha, \beta)$ prior

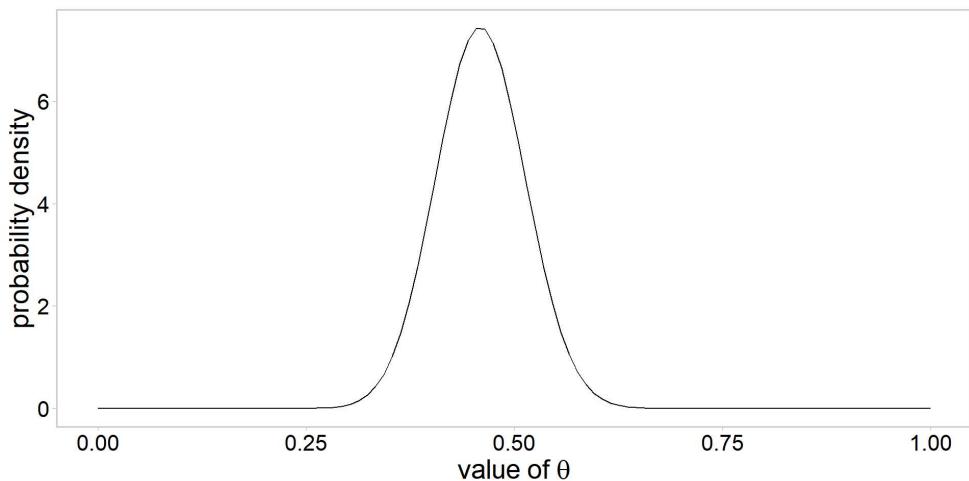
```
1 x <- seq(0, 1, len = 100) # create a sequence
2 db <- dbeta(x, 2, 4) # see ?dbeta
3 ggplot() +
4   geom_line(aes(x, db)) +
5   xlab(expression(value ~ of ~ theta)) +
6   ylab("probability density") +
7   theme_qfc() +
8   theme(text = element_text(size = 20))
```



- always a good idea to visualize

Visualizing a $\text{beta}(\alpha, \beta)$ prior

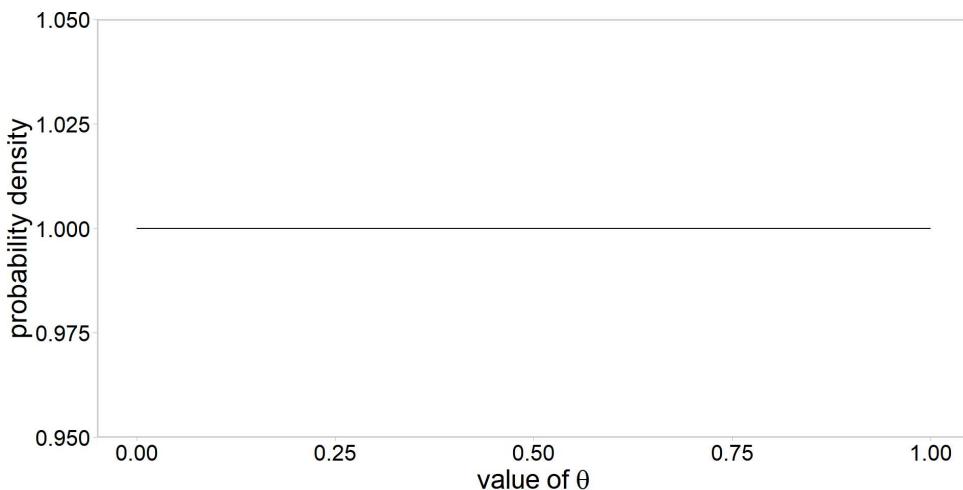
```
1 x <- seq(0, 1, len = 100) # create a sequence
2 db <- dbeta(x, 40, 47) # see ?dbeta
3 ggplot() +
4   geom_line(aes(x, db)) +
5   xlab(expression(value ~ of ~ theta)) +
6   ylab("probability density") +
7   theme_qfc() +
8   theme(text = element_text(size = 20))
```



- always a good idea to visualize

Visualizing a $\text{beta}(\alpha, \beta)$ prior

```
1 x <- seq(0, 1, len = 100) # create a sequence
2 db <- dbeta(x, 1, 1) # see ?dbeta
3 ggplot() +
4   geom_line(aes(x, db)) +
5   xlab(expression(value ~ of ~ theta)) +
6   ylab("probability density") +
7   theme_qfc() +
8   theme(text = element_text(size = 20))
```



- always a good idea to visualize

Do math to get an analytical posterior

the beta prior: $p(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$

the binomial likelihood: $p(y | \theta) \propto \theta^y(1-\theta)^{n-y}$

$$p(\theta | y) \propto p(y | \theta) \cdot p(\theta)$$

$$p(\theta | y) \propto \theta^y(1-\theta)^{n-y} \cdot \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$$p(\theta | y) \propto \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}$$

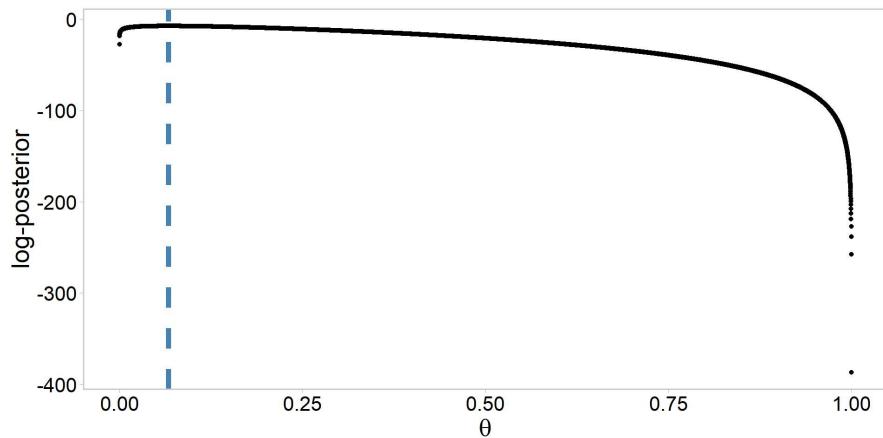
- Last line represents an analytical solution to the posterior

Let's do that in code

```
1 y <- 2 # successes
2 n <- 30 # trials
3 eps <- 1e-6 # buffer because theta must lie within [0,1]
4 a <- b <- 1 # alpha, beta parameters of the beta(shape1, shape2) prior
5 theta <- seq(eps, 1 - eps, length.out = 1e4) # sequence to thetas to try
6
7 # calculate log posterior:
8 logPost <- (y + a - 1) * log(theta) + (n - y + b - 1) * log(1 - theta)
9 MAP <- theta[which.max(logPost)] # maximum a posteriori estimate
10 my_data <- data.frame(logPost, theta) # for plotting
```

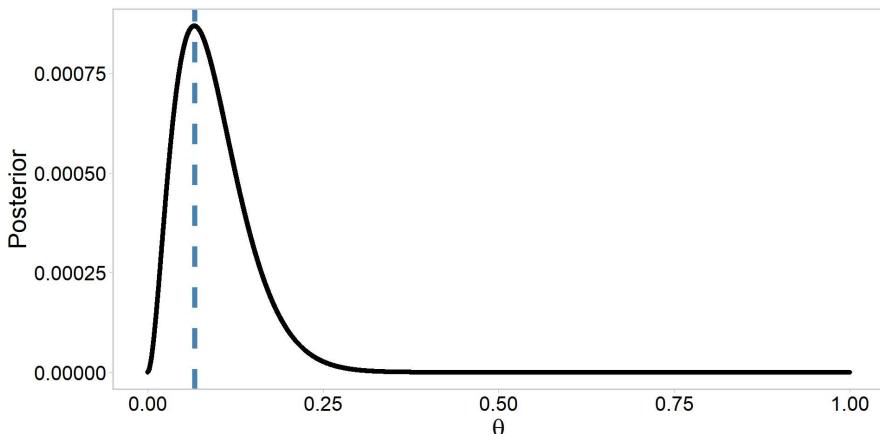
Plot the unstandardized log-posterior

```
1 my_data %>%
2   ggplot(aes(x = theta, y = logPost)) +
3   geom_point(size = 1.4) +
4   ylab("log-posterior") +
5   xlab(expression(theta)) +
6   geom_vline(
7     xintercept = MAP, linetype = 2, linewidth = 2,
8     color = "steelblue"
9   ) # add MAP estimate
10 theme_qfc() +
11 theme(text = element_text(size = 20))
```



Plot the standardized posterior

```
1 my_data %>%
2   ggplot(aes(x = theta, y = exp(logPost) / sum(exp(logPost)))) + # note exp
3   geom_point(size = 1.4) +
4   ylab("Posterior") +
5   xlab(expression(theta)) +
6   geom_vline(
7     xintercept = MAP, linetype = 2, linewidth = 2,
8     color = "steelblue"
9   ) + # add MAP estimate
10  theme_qfc() +
11  theme(text = element_text(size = 20))
```



Frequentist MLEs vs. Bayesian MAP

```
1 y / n # analytical theta (frequentist)
```

```
[1] 0.06666667
```

```
1 MLE # numerical approximation theta (frequentist)
```

```
[1] 0.06670754
```

```
1 MAP # Bayesian theta solution
```

```
[1] 0.06670754
```

- Why is this convergence among methods happening?
 - When might this not happen?
- While we get the same answer, the results have different interpretations
 - What do these findings mean for monitoring sneak turtles?

Alabama sneak turtles: a dead simple example chosen for a reason



- Allows us to easily look at how different priors for θ or amounts of data (n) influence the resulting Bayesian answer
 - Good exercise if you want to develop some intuition in this simple model
- Computing a posterior distribution analytically like we did here is often impossible
- Simulation techniques such as Markov Chain Monte Carlo (MCMC) can be used to draw samples from a posterior instead of solving the equations



Pros of Bayesian inference

- Pros
 - Probability statements are exact for any sample size
 - Way to formally incorporate prior information
 - Easy to make probability statements about things you care about
 - Sounds cool 

Cons of Bayesian inference

- Cons
 - Must specify a prior, and priors influence results
 - Easy to screw up (hence emphasis on simulation and model checks in this class)
 - Computing the posterior is difficult
 - In all but the simplest cases, requires the approximation of high dimensional integrals
 - Many algorithms are guaranteed to converge in the mathematical limit, but not in your graduate degree timeline (or even lifetime)

Summary and outlook

- A crash course in statistical inference
 - Frequentist and Bayesian paradigms
- Talked about some commonly used distributions
- Simple turtle example
 - This example had analytical solutions, but used it to better understand what is actually going on ‘under the hood’ aka in the math
- Most Bayesian analyses require the approximation of a posterior distribution via simulation techniques (i.e., algorithms like MCMC)

Before next class

- Make sure you have R, Rstudio, and Stan and cmdstanr installed and working
 - Rstudio and R:
 - <https://posit.co/download/rstudio-desktop/>
 - Stan and cmdstanr
 - <https://mc-stan.org/users/interfaces/rstan>
 - <https://mc-stan.org/cmdstanr/>

References

- Gelman et al. 2015. Stan: A probabilistic programming language for Bayesian inference and optimization
- Images on Stanislav Ulam slide courtesy of the [Atomic Heritage Foundation](#)
- Common probability distributions [figure](#)
- For a great refreshers of these and other introductory statistical inference topics, see:
 - Kery and Schaub. 2012. Bayesian Population Analysis using WinBUGS. Chapters 1-2. (Easier)
 - Royle and Dorazio 2008. Hierarchical modeling and inference in ecology. Chapters 1-2. (Advanced)

Exercises (optional)

1. Prior sensitivity test

Your supervisor is skeptical about these new-fangled Bayesian statistics, as they believe the prior is ridiculous and subjective. To convince them otherwise, conduct a prior sensitivity test for the Alabama sneak turtle example. Look at how varying the prior for θ shifts the resulting posterior distribution (and hence your inferences). What (if anything) can you do or say to relieve their concerns?

2. Is the coin biased?

You are people watching at a bar while reading about Bayesian statistics because you are a nerd, and you watch someone flip a coin with the following outcomes: TTHHHTHTHHHTHHH. This human then comes over to you and asks if you want to bet 100 dollars on the next coin flip being H. Can you use Bayes' theorem to determine the probability that they are tossing an unfair coin? If you conducted a similar analysis in a frequentist paradigm, what would you conclude?

3. Sneak turtles continued...

Suppose the example we used in class was criticized by your advisor and the resource management agency tasked with managing Alabama sneak turtles because you tried to count them on a cloudy day, and everyone thinks sneak turtles don't like cloudy days.

Thus, resource managers undertake another experiment analogous to yours, except this time they release $n = 100$ and later detect $y = 17$ turtles.

Using the posterior you calculated from the cloudy day as the prior for a new analysis, can you update your beliefs via Bayes' rule accordingly and develop a new posterior distribution for sneak turtle detection probability θ ?

