Special Topics: Bayesian Approaches to Statistics and Modeling

Mark Kurzeja

Overview

- In this course, we have primarily used maximum likelihood to fit models ("pick the 'best' values to fit the data")
- Probability, in a frequentist setting, is a long-run frequency
- Bayesian methods think about problems in a fundamentally different way

Frequentist Inference → Bayesian Beliefs → Bayesian Modeling

Beliefs and Bayesians

Average IQ

- Imagine we are trying to determine average IQ of students at the University of Michigan
- Someone asks me, "What do you think the average IQ score could be of University of Michigan Students?"

Beliefs and Bayesians

Average IQ

<u>Belief</u>
Michigan IQ scores
$\sim Norm(100, 10)$

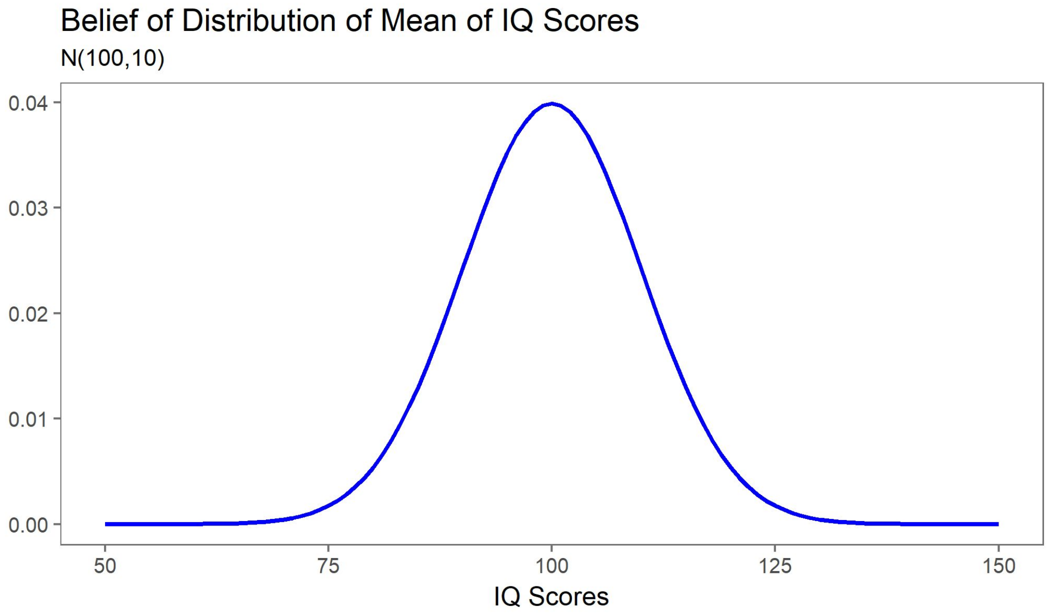
- Imagine we are trying to determine average IQ of students at the University of Michigan
- Someone asks me, "Hey, what do you think the average IQ could be?"
- I look online, and see that the IQ test scores, for the U.S. population, are, by design, normally distributed with $\mu = 100, \sigma = 10$. I'll start out with that as an "educated guess", knowing that it may not be the best belief for the mean but will be a good place to start

Beliefs and Bayesians

Average IQ

<u>Belief</u>
Michigan IQ scores
$\sim Norm(100, 10)$

- What does this starting belief look like?



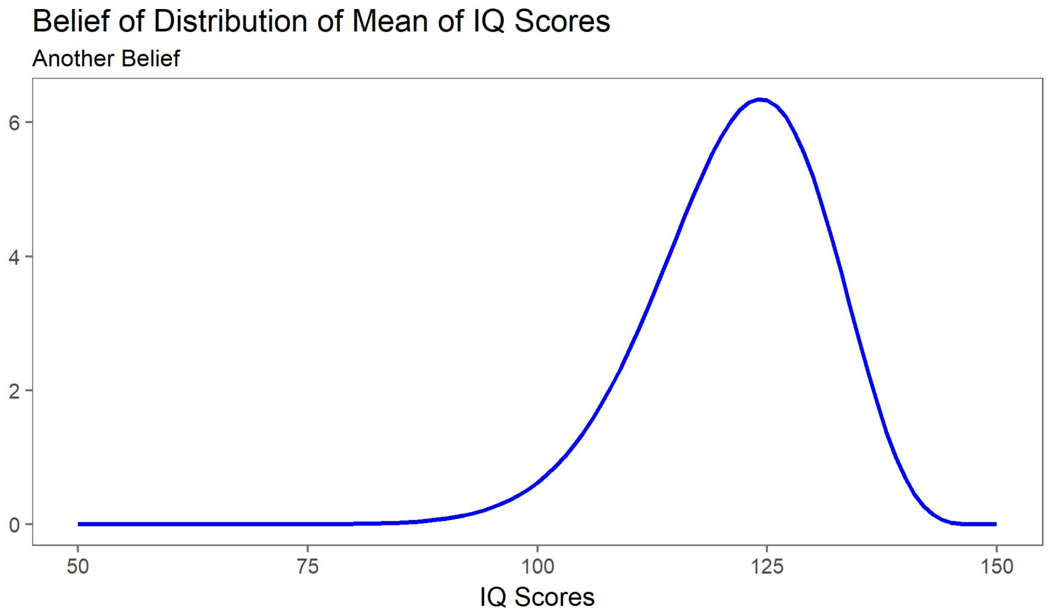
Beliefs and Bayesians

Average IQ

Belief

Michigan IQ scores $\sim Norm(100, 10)$

- Note: This is just a belief – it can be anything – skewed, multimodal, …

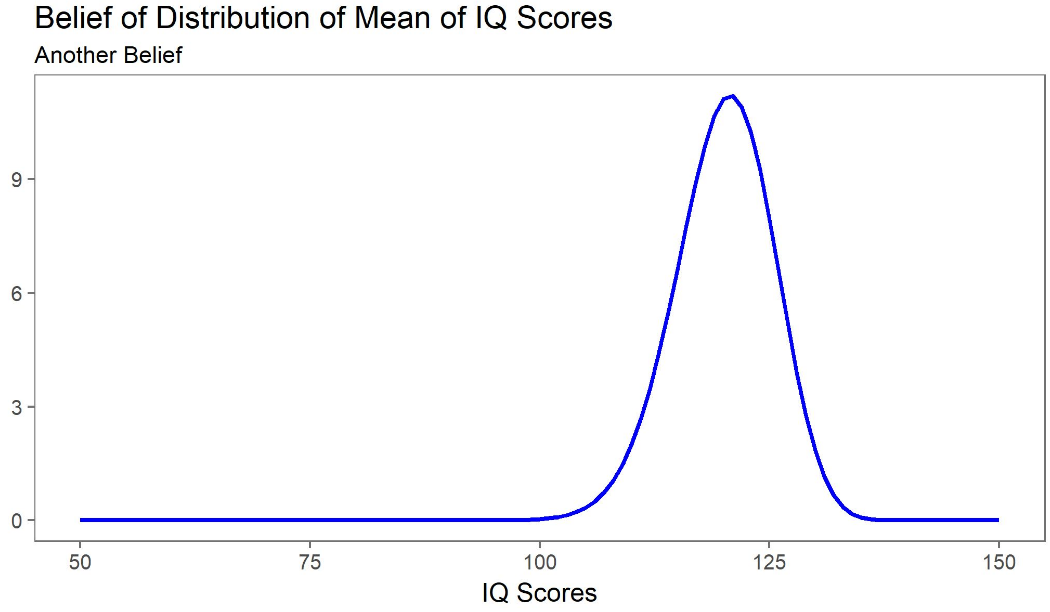


Beliefs and Bayesians

Average IQ

Belief
Michigan IQ scores $\sim Norm(100, 10)$

- Note: This is just a belief – it can be anything – skewed, multimodal, or have a smaller variance



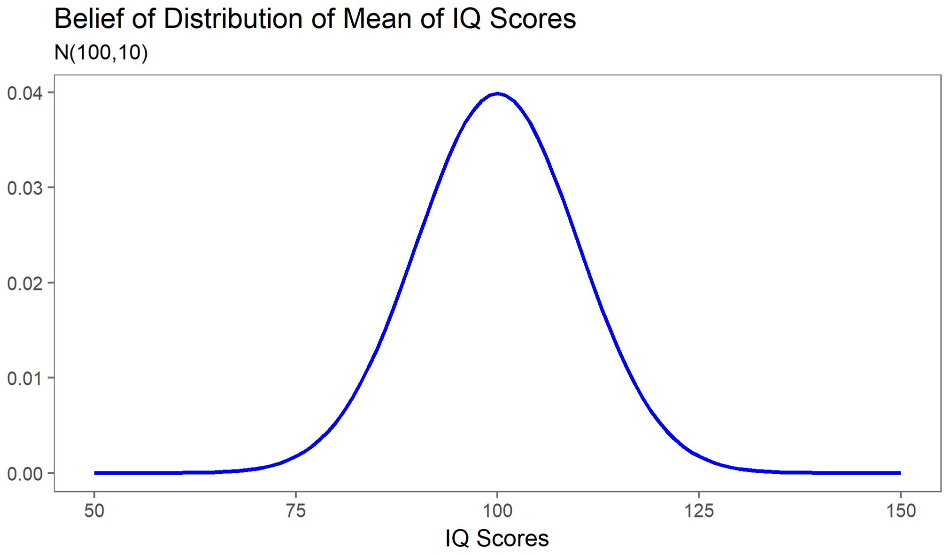
Beliefs and Bayesians

Average IQ

Belief
Michigan IQ scores $\sim Norm(100, 10)$

Observations
125

- Now, I go and test someone's IQ on campus. They have an IQ of 125. How does this change my belief?



- Does my belief shift left, right, or stay the same?

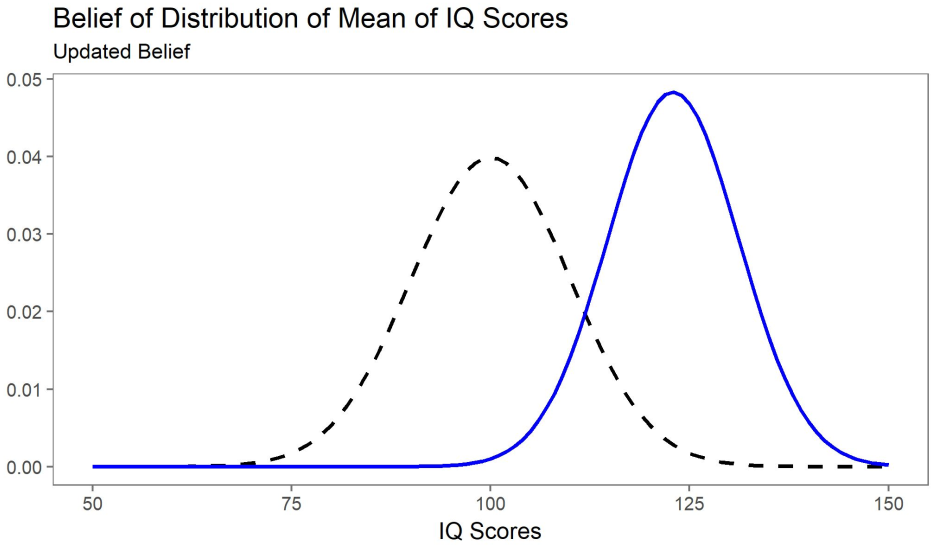
Beliefs and Bayesians

Average IQ

Belief
Michigan IQ scores $\sim Norm(100, 10)$

Observations
125

- It should shift right. We had a belief that the mean was 100…



- Seeing a value of 125 indicates the mean might be higher

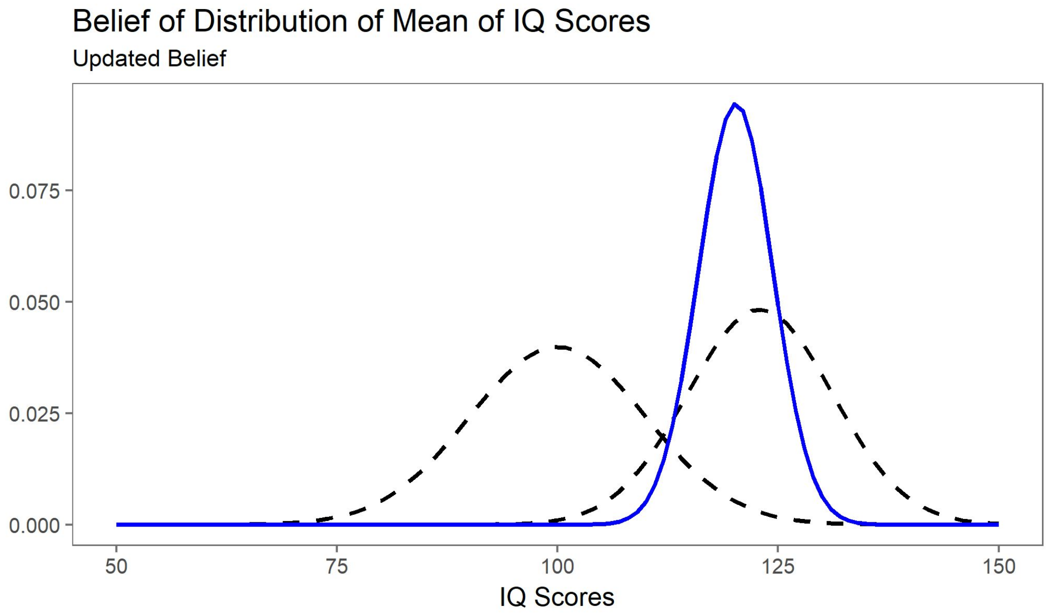
Beliefs and Bayesians

Average IQ

<u>Belief</u>
Michigan IQ scores $\sim Norm(100, 10)$

<u>Observations</u>
125, 115

- Now we observe a student with a 115 IQ. How does this change our belief?



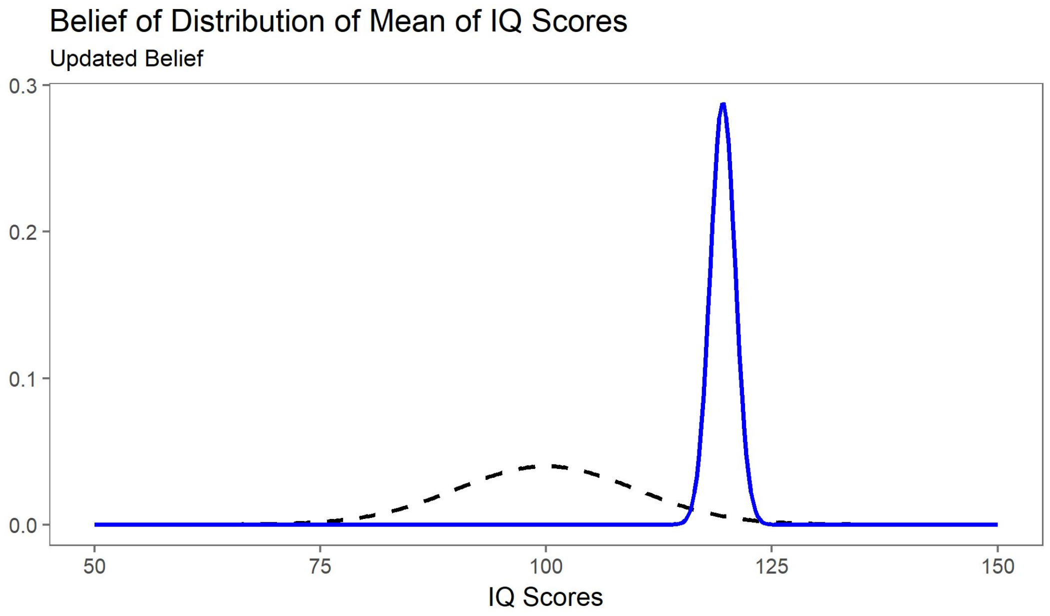
Beliefs and Bayesians

Average IQ

<u>Belief</u>
Michigan IQ scores $\sim Norm(100, 10)$

<u>Observations</u>
125, 115, 115, 120, 125, 117

- What happens if we observe more and more data? I observe IQ scores of 115, 120, 125, and 117



Beliefs and Bayesians

Average IQ

Belief
Michigan IQ scores
$\sim Norm(100, 10)$

Observations
125, 115, 115, 120, 125, 117

Updating
Posterior Distribution

- What happens if we observe more and more data?
- We can see that more and more data allows us to better focus in our belief about the mean of IQ scores at U of M
- The process that we went through is called *Bayesian Updating.* This provides a distribution on the quantity of interest, called the **posterior**
- The posterior allows us to update our beliefs and answer questions about the quantity of interest

The Posterior, Front and Center

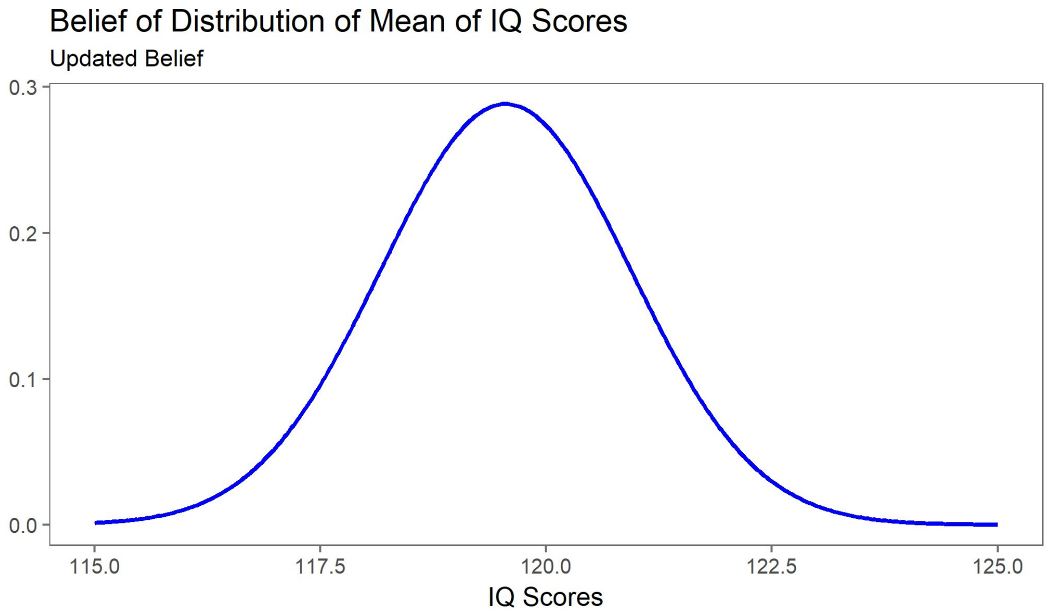
Average IQ

<u>Belief</u>
Michigan IQ scores
$\sim Norm(100, 10)$

<u>Observations</u>
125, 115, 115, 120, 125, 117

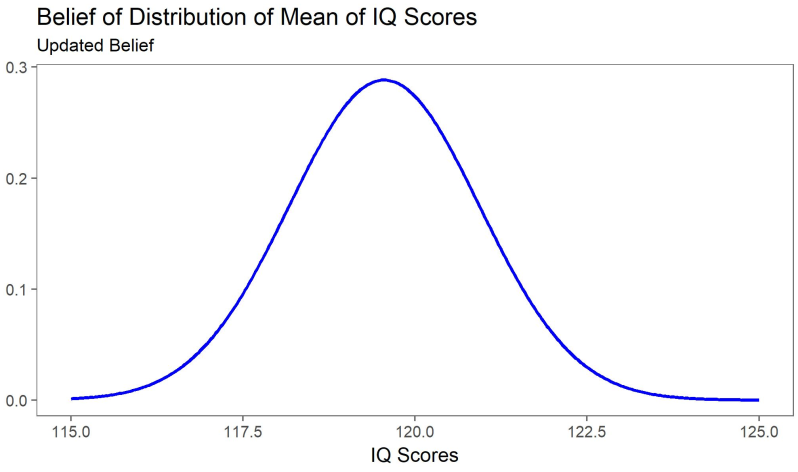
<u>Updating</u>
Posterior Distribution

- This provides a distribution on the quantity of interest, called the **posterior**

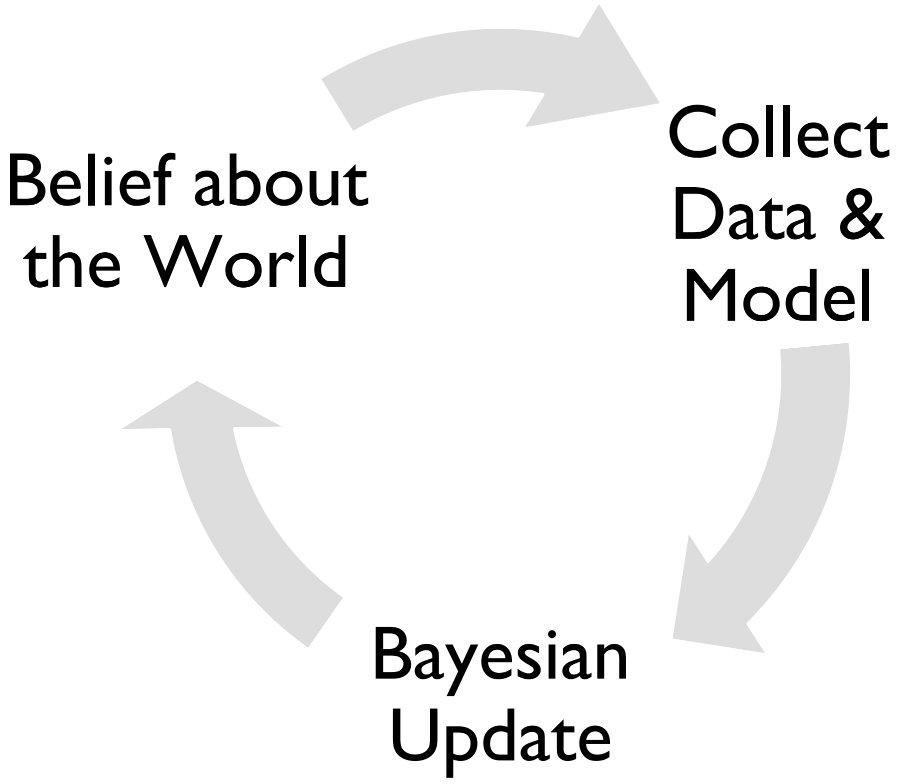


The Posterior, Front and Center

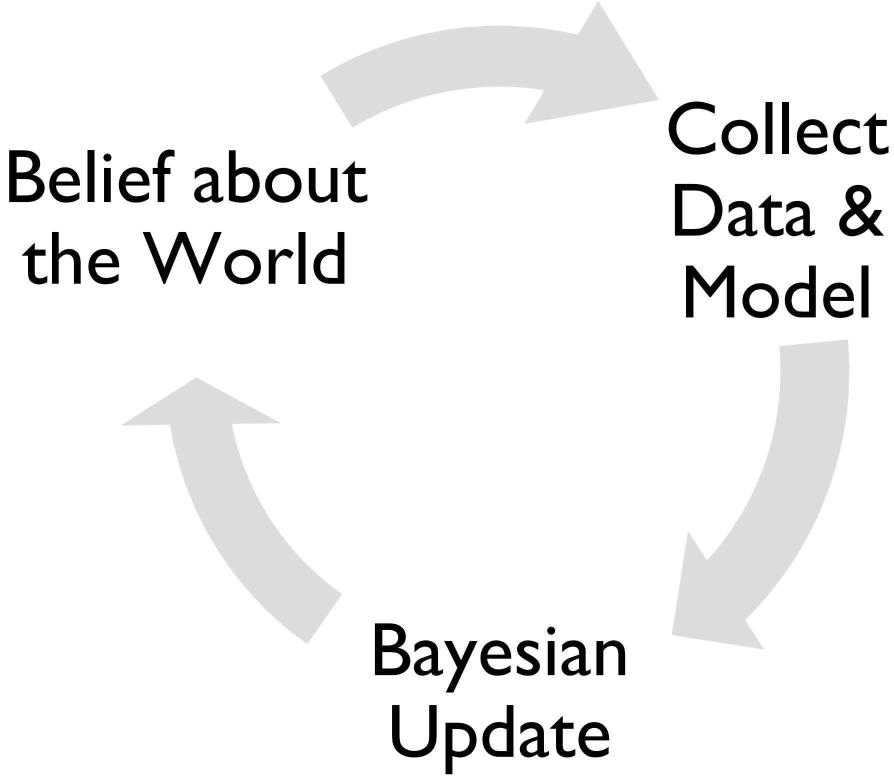
- This provides a distribution on the quantity of interest, called the **posterior**
- What is the mean of our belief?: 119.55
- What is the "most likely" value?: 119.55
- What is the median (50th percentile of the distribution)?: 119.55
- with 95% probability
- What is the 95% credible interval for the mean IQ score?: (116.84, 122.26)
- **All from the posterior distribution!**



The Posterior, Front and Center



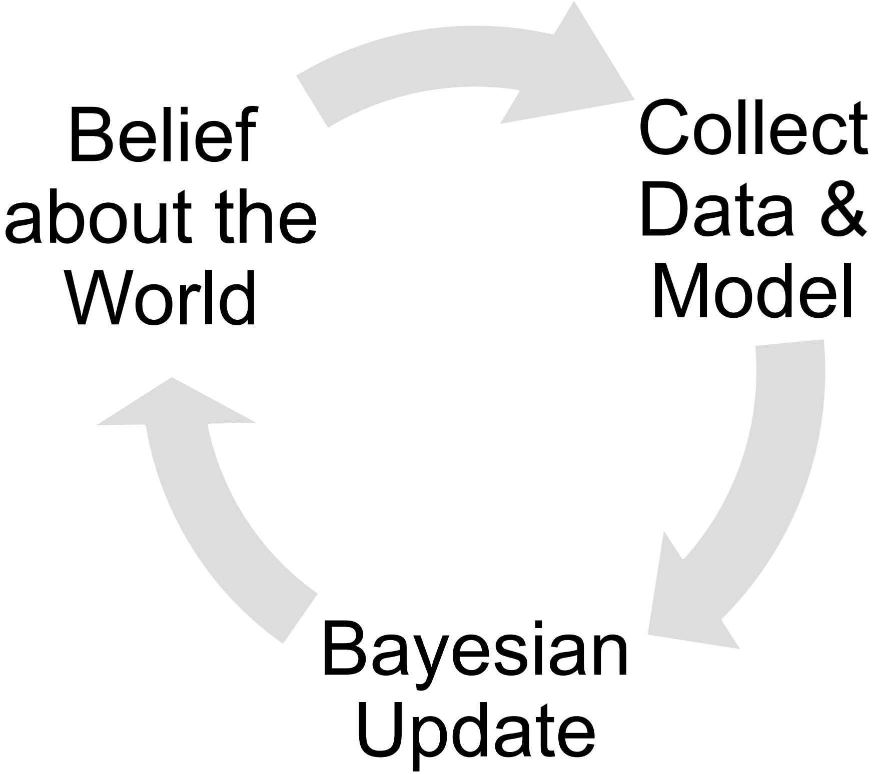
The Posterior, Front and Center



=

Bayesian Statistics

The Posterior, Front and Center



Steps to a Bayesian Update

1) Establish a belief – our first belief is a prior. We add data to update these beliefs
2) Collect Data & Model
3) Update our Beliefs using Data to get a posterior
4) Repeat steps 2 & 3 using the posterior from 3 as our new prior

The Posterior, Front and Center

Bayesian Methods & Posteriors

- All questions about our beliefs of our quantity of interest can be found via the posterior
- We can combine this with loss functions for optimal decision making under uncertainty given our model of the world
- This is a VERY powerful idea

- Like any good method though, it has downsides
- We have to rewrite our definition of probability to work with Bayesian methods
- Mathematically, the process can be difficult, intractable, or highly computationally expensive depending on the model

The Posterior, Front and Center

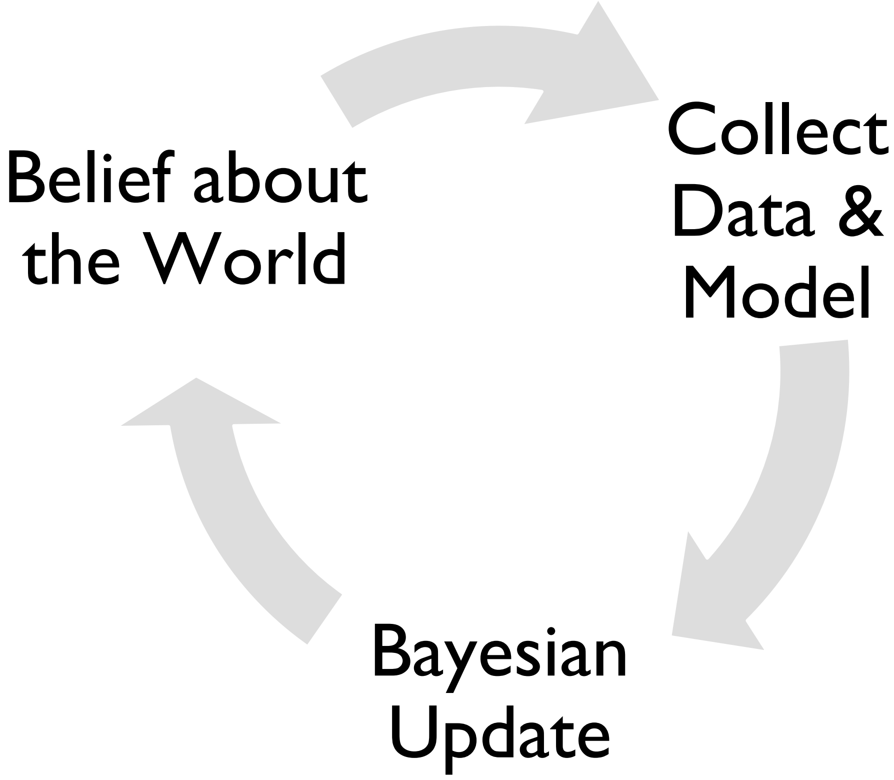
Flipping Probability it on its Head

- Frequentists view probability as a long-run limiting frequency; parameters are fixed unknowable constants and data is random
- Bayesians view probability as a "degree of belief", treat data as fixed, and posit parameters are random variables

Nothing is free

- Mathematically, unless you work with very specific families of beliefs (priors) from the beginning, the math can get difficult quickly
- Most models are so complex that we have no choice but to use sampling methods to estimate results

Where are we going?



Case Study

- To illustrate the flexibility and power of these methods, we will walk through a case study on Bayesian Regression
- Focus will be on the application, interpretation, and modeling and not on the theoretical side of Bayesian Computation