# Introduction to Bayesian Statistical Inference: Comparison and Assumptions

Henry Overos

University of Maryland, College Park

November 8, 2019

# Goals for Today's Workshop

- Increased understanding of the differences between Frequentist and Bayesian approaches to Inference and Modeling

# Goals for Today's Workshop

- Increased understanding of the differences between Frequentist and Bayesian approaches to Inference and Modeling
- Highlight the possible utility of Bayesian approaches in political science research

# Goals for Today's Workshop

- Increased understanding of the differences between Frequentist and Bayesian approaches to Inference and Modeling
- Highlight the possible utility of Bayesian approaches in political science research
- Provide resources for further study or work on this topic

# Goals for Today's Workshop

- Increased understanding of the differences between Frequentist and Bayesian approaches to Inference and Modeling
- Highlight the possible utility of Bayesian approaches in political science research
- Provide resources for further study or work on this topic

What do you think of when you hear
about Bayesian Statistics?

# Common Answers

- Terms
  - prior

# Common Answers

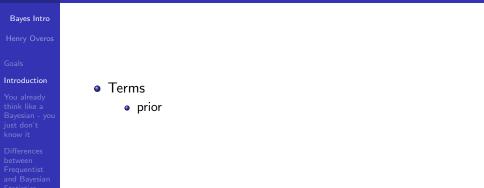Bayes Intro

Henry Overos

Goals

Introduction

You already
think like a
Bayesian - you
just don't
know it

Differences
between
Frequentist
and Bayesian
Statistics

The Simple
Mechanics of
Bayesian
Estimation

Questions?

- Terms
    - prior
    - posterior

# Common Answers

- Terms
    - prior
    - posterior
    - subjective

# Common Answers

- Terms
  - prior
  - posterior
  - subjective
- Controversy?
  - "I'm a closeted Bayesian"

# Common Answers

- Terms
  - prior
  - posterior
  - subjective
- Controversy?
  - "I'm a closeted Bayesian"
  - "It is too hard to explain, especially to Reviewer 2"

# Example Taking the Bus to Campus

- You read that the bus comes to your stop at 9 am

# Example Taking the Bus to Campus

- You read that the bus comes to your stop at 9 am
- Day 1 - the bus shows up at 9:10

# Example Taking the Bus to Campus

Bayes Intro

Henry Overos

Goals

Introduction

You already
think like a
Bayesian - you
just don't
know it

Differences
between
Frequentist
and Bayesian
Statistics

The Simple
Mechanics of
Bayesian
Estimation

Questions?

- You read that the bus comes to your stop at 9 am
- Day 1 - the bus shows up at 9:10
- Day 2 - the bus shows up at 9:08

- You read that the bus comes to your stop at 9 am
- Day 1 - the bus shows up at 9:10
- Day 2 - the bus shows up at 9:08
- Day 3 - the bus shows up at 9:16

# Example Taking the Bus to Campus

- You read that the bus comes to your stop at 9 am
- Day 1 - the bus shows up at 9:10
- Day 2 - the bus shows up at 9:08
- Day 3 - the bus shows up at 9:16
- What do you start doing?

- Your **prior** expectation before day 1 is to expect the bus to arrive at exactly 9:00 am

# Bus Example Cont'd

- Your **prior** expectation before day 1 is to expect the bus to arrive at exactly 9:00 am

- This prior is largely uninformed - you have no data to back up the validity of the claim

# Bus Example Cont'd

- Your **prior** expectation before day 1 is to expect the bus to arrive at exactly 9:00 am
- This prior is largely uninformed - you have no data to back up the validity of the claim
- As the days progress, your expectation becomes more informed by data and you update your expected bus arrival, creating a new understanding of the probability distribution of when the bus is **most likely** to arrive

# Bayes' Theorem

What you did with the bus, put more formally

$$P(B_j|A) = \frac{P(B_j)P(A|B_j)}{\sum_{i=1}^{k} P(B_i)P(A|B_i)}$$

# Bayes' Theorem

What you did with the bus, put more formally

$$P(B_j|A) = \frac{P(B_j)P(A|B_j)}{\sum_{i=1}^{k} P(B_i)P(A|B_i)}$$

- Named after the Rev. Thomas Bayes (1700-1761) who used conditional probability to estimate unknown parameters
- Also Pierre-Simon Laplace independently formulated a similar relationship between prior and conditional probabilities around the same time

# Bayes' Theorem

What you did with the bus, put more formally

$$P(B_j|A) = \frac{P(B_j)P(A|B_j)}{\sum_{i=1}^{k} P(B_i)P(A|B_i)}$$

- Named after the Rev. Thomas Bayes (1700-1761) who used conditional probability to estimate unknown parameters
- Also Pierre-Simon Laplace independently formulated a similar relationship between prior and conditional probabilities around the same time

The probability that a theory or hypothesis is true **if** some event has occurred

How likely is it that someone intentionally flies a plane into a skyscraper in New York?

How likely is it that someone intentionally flies a plane into a skyscraper in New York?
Before the first plane hit the tower - our prediction that terrorists would fly a plane into a skyscraper was small, say 1 in 20,000 or 0.005 percent. It's possible but not something we would ever really consider.
This is our **prior**

How likely is it that someone intentionally flies a plane into a skyscraper in New York?

Before the first plane hit the tower - our prediction that terrorists would fly a plane into a skyscraper was small, say 1 in 20,000 or 0.005 percent. It's possible but not something we would ever really consider.

This is our **prior**

It is also the probability of a terrorist attack given the frequency of occurrences in the observed sample data (20,000 days).

# Example Cont'd

Next, we have to examine the probability of *event, conditioned on the hypothesis being true*. That is, what is the probability that a plane hit a skyscraper, **if** there was a terrorist attack? After the first plane hits, this number is 100 percent. If there are terrorists attacking New York, they're flying the plane into the skyscraper.

We then examine the probability of a plane hitting the tower **if terrorists are not attacking (i.e. accidentally**. Empirically, we can assume the number is 0.008 this because only two planes had hit skyscrapers in New York prior (1945 and 1946 but these were accidents).

# Example Cont'd

We now have everything we need to predict the likelihood that
there is a terror attack occurring when the plane hits:

$$P(\text{atk}) = 0.00005$$

Bayes Intro

Henry Overos

Goals

Introduction

You already
think like a
Bayesian - you
just don't
know it

Differences
between
Frequentist
and Bayesian
Statistics

The Simple
Mechanics of
Bayesian
Estimation

Questions?

# Example Cont'd

$$P(\text{atk}) = 0.00005$$

$$P(crash|atk) = 1$$

$$P(\text{atk}) = 0.00005$$

$$P(crash|atk) = 1$$

$$P(crash|\neg\text{atk}) = 0.00008$$

# Example Cont'd

$$P(\text{atk}) = 0.00005$$

$$P(crash|atk) = 1$$

$$P(crash|\neg\text{atk}) = 0.00008$$

$$P(\text{atk}|\text{Crash})$$

# Example Cont'd

$$P(\text{atk}) = 0.00005$$

$$P(crash|atk) = 1$$

$$P(crash|\neg\text{atk}) = 0.00008$$

$$P(\text{atk}|\text{Crash})$$

$$= \frac{P(\text{atk})P(\text{Crash}|\text{atk})}{P(\text{atk})P(\text{Crash}|\text{atk}) + P(\neg\text{atk})(\text{Crash}|\neg \text{atk})}$$

# Example Cont'd

Bayes Intro

Henry Overos

Goals

Introduction

You already
think like a
Bayesian - you
just don't
know it

Differences
between
Frequentist
and Bayesian
Statistics

The Simple
Mechanics of
Bayesian
Estimation

Questions?

$$P(\text{atk}) = 0.00005$$

$$P(crash|atk) = 1$$

$$P(crash|\neg\text{atk}) = 0.00008$$

$$P(\text{atk}|\text{Crash})$$

$$= \frac{P(\text{atk})P(\text{Crash}|\text{atk})}{P(\text{atk})P(\text{Crash}|\text{atk}) + P(\neg\text{atk})(\text{Crash}|\neg\ \text{atk})}$$

$$= \frac{0.00005 * 1}{0.00005 * 1 + 0.00008 * (1 - 0.00005)}$$

# Example Cont'd

$$P(\text{atk}) = 0.00005$$

$$P(crash|atk) = 1$$

$$P(crash|\neg\text{atk}) = 0.00008$$

$$P(\text{atk}|\text{Crash})$$

$$= \frac{P(\text{atk})P(\text{Crash}|\text{atk})}{P(\text{atk})P(\text{Crash}|\text{atk}) + P(\neg\text{atk})(\text{Crash}|\neg \text{atk})}$$

$$= \frac{0.00005 * 1}{0.00005 * 1 + 0.00008 * (1 - 0.00005)}$$

$$= \frac{0.00005}{.00005 + 0.0000796} \approx 0.385$$

Update our prior:

$$P(\text{atk}) = 0.385$$

# Second Plane?

Update our prior:

$$P(\text{atk}) = 0.385$$

$$P(crash|atk) = 1$$

Update our prior:

$$P(\text{atk}) = 0.385$$

$$P(crash|atk) = 1$$

$$P(crash|\neg\text{atk}) = 0.00008$$

## Second Plane?

Update our prior:

$$P(\text{atk}) = 0.385$$

$$P(crash|atk) = 1$$

$$P(crash|\neg\text{atk}) = 0.00008$$

$$P(\text{atk}|\text{Crash})$$

## Second Plane?

Update our prior:

$$P(\text{atk}) = 0.385$$

$$P(crash|atk) = 1$$

$$P(crash|\neg\text{atk}) = 0.00008$$

$$P(\text{atk}|\text{Crash})$$

$$= \frac{0.385 * 1}{0.385 * 1 + 0.00008 * (1 - 0.385)}$$

# Second Plane?

Update our prior:

$$P(\text{atk}) = 0.385$$

$$P(crash|atk) = 1$$

$$P(crash|\neg\text{atk}) = 0.00008$$

$$P(\text{atk}|\text{Crash})$$

$$= \frac{0.385 * 1}{0.385 * 1 + 0.00008 * (1 - 0.385)}$$

$$= \frac{0.385}{.385 + 0.3850492} \approx 0.999$$

# Why do we use Statistics?

Political Scientists make inferences about the
world using a comparatively small sample of data

Political Scientists make inferences about the
world using a comparatively small sample of data

- Statistics is used to make inferences about
  **parameters**, which tend to describe the
  data-generating process
- We **have and observe data** but we **model
  the data-generating process**

# Comparing Assumptions of Frequentist and Bayesian Statistical Inference

## Frequentist

- Probability: the relative frequency of an event given a large number of trials

# Comparing Assumptions of Frequentist and Bayesian Statistical Inference

## Frequentist

- Probability: the relative frequency of an event given a large number of trials
- Unknown parameters have a fixed value, we just don't know it

# Comparing Assumptions of Frequentist and Bayesian Statistical Inference

## Frequentist

- Probability: the relative frequency of an event given a large number of trials

- Unknown parameters have a fixed value, we just don't know it

- Correct model output provides a level of confidence that, if we were to repeat the experiment or run the model more times, the **true** value of a parameter falls within a certain interval

# Comparing Assumptions of Frequentist and Bayesian Statistical Inference

## Frequentist

- Probability: the relative frequency of an event given a large number of trials

- Unknown parameters have a fixed value, we just don't know it

- Correct model output provides a level of confidence that, if we were to repeat the experiment or run the model more times, the **true** value of a parameter falls within a certain interval

# Comparing Assumptions of Frequentist and Bayesian Statistical Inference

## Bayesian

- Probability: Given the evidence and prior knowledge on an event, what degree of certainty do we have that an event will occur?

# Comparing Assumptions of Frequentist and Bayesian Statistical Inference

## Bayesian

- Probability: Given the evidence and prior knowledge on an event, what degree of certainty do we have that an event will occur?

- Unknown parameters are not fixed and their "true" value is uncertain

# Comparing Assumptions of Frequentist and Bayesian Statistical Inference

## Bayesian

- Probability: Given the evidence and prior knowledge on an event, what degree of certainty do we have that an event will occur?

- Unknown parameters are not fixed and their "true" value is uncertain

- Model provides our level of certainty that the parameter value is inside an interval given the data we have now (Credible Interval)

# Comparing Assumptions of Frequentist and Bayesian Statistical Inference

## Bayesian

- Probability: Given the evidence and prior knowledge on an event, what degree of certainty do we have that an event will occur?

- Unknown parameters are not fixed and their "true" value is uncertain

- Model provides our level of certainty that the parameter value is inside an interval given the data we have now (Credible Interval)

# Estimating Parameters: Comparing Frequentist and Baysian Approaches

Given a parameter $\beta$, that describes the data-generating
process, how do we make inferences?

## Frequentist

- The Key: $P(\text{Data}|H_0)$

Given a parameter $\beta$, that describes the data-generating process, how do we make inferences?

## Frequentist

- The Key: $P(\text{Data}|H_0)$
- Reject $H_0$ if $P(\text{Data}|H_0) < 0.05$

# Estimating Parameters: Comparing Frequentist and Baysian Approaches

Given a parameter $\beta$, that describes the data-generating process, how do we make inferences?

## Frequentist

- The Key: $P(\text{Data}|H_0)$
- Reject $H_0$ if $P(\text{Data}|H_0) < 0.05$
- Fail to Reject $H_0$ if $P(\text{Data}|H_0) \geq 0.05$

# Estimating Parameters: Comparing Frequentist and Baysian Approaches

Given a parameter $\beta$, that describes the data-generating process, how do we make inferences?

## Frequentist

- The Key: $P(\text{Data}|H_0)$
- Reject $H_0$ if $P(\text{Data}|H_0) < 0.05$
- Fail to Reject $H_0$ if $P(\text{Data}|H_0) \geq 0.05$

## Bayesian

- The Key: $P(\beta|\text{Data})$
- $P(\beta|\text{Data})$ is what we call the posterior distribution
- The result of a Bayesian model estimation is a probability distribution around the parameter $\beta$

# The Most Important Thing

## Deriving the Posterior Distribution of the Parameter

$$\pi(\theta|y) = \frac{p(\theta)L(\theta|y)}{\int_\theta p(\theta)L(\theta|y)}$$

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

# The Most Important Thing

## Deriving the Posterior Distribution of the Parameter

$$\pi(\theta|y) = \frac{p(\theta)L(\theta|y)}{\int_\theta p(\theta)L(\theta|y)}$$

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

## Symbol Key

- y: vector of the observed data/outcome variables

# The Most Important Thing

## Deriving the Posterior Distribution of the Parameter

$$\pi(\theta|y) = \frac{p(\theta)L(\theta|y)}{\int_\theta p(\theta)L(\theta|y)}$$

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

## Symbol Key

- y: vector of the observed data/outcome variables
- $\theta$: the parameter of the data y's distribution

# The Most Important Thing

## Deriving the Posterior Distribution of the Parameter

$$\pi(\theta|y) = \frac{p(\theta)L(\theta|y)}{\int_\theta p(\theta)L(\theta|y)}$$

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

## Symbol Key

- y: vector of the observed data/outcome variables
- $\theta$: the parameter of the data y's distribution
- $\pi$: posterior

# The Most Important Thing

Bayes Intro

Henry Overos

Goals

Introduction

You already
think like a
Bayesian - you
just don't
know it

Differences
between
Frequentist
and Bayesian
Statistics

The Simple
Mechanics of
Bayesian
Estimation

Questions?

## Deriving the Posterior Distribution of the Parameter

$$\pi(\theta|y) = \frac{p(\theta)L(\theta|y)}{\int_\theta p(\theta)L(\theta|y)}$$

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

## Symbol Key

- y: vector of the observed data/outcome variables
- $\theta$: the parameter of the data y's distribution
- $\pi$: posterior
- $\pi(\theta|y)$: The Posterior Distribution of $\theta$

# The Most Important Thing

## Deriving the Posterior Distribution of the Parameter

$$\pi(\theta|y) = \frac{p(\theta)L(\theta|y)}{\int_\theta p(\theta)L(\theta|y)}$$

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

## Symbol Key

- y: vector of the observed data/outcome variables
- $\theta$: the parameter of the data y's distribution
- $\pi$: posterior
- $\pi(\theta|y)$: The Posterior Distribution of $\theta$
- $p(\theta)$: The Prior Distribution of $\theta$

# The Most Important Thing

## Deriving the Posterior Distribution of the Parameter

$$\pi(\theta|y) = \frac{p(\theta)L(\theta|y)}{\int_\theta p(\theta)L(\theta|y)}$$

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

## Symbol Key

- y: vector of the observed data/outcome variables
- $\theta$: the parameter of the data y's distribution
- $\pi$: posterior
- $\pi(\theta|y)$: The Posterior Distribution of $\theta$
- $p(\theta)$: The Prior Distribution of $\theta$
- $L(\theta|y)$: The Likelihood function of the outcome y, given the parameter $\theta$

$$\pi(\theta|y) \propto p(\theta)L(\theta|y)$$

$$\pi(\theta|y) \propto p(\theta)L(\theta|y)$$

The posterior distribution of the parameter we are interested in is equivalent to the prior distribution of the parameter times the likelihood of getting that parameter given the data that we have

$$\pi(\theta|y) \propto p(\theta)L(\theta|y)$$

The posterior distribution of the parameter we
are interested in is equivalent to the prior
distribution of the parameter times the likelihood
of getting that parameter given the data that we
have

As we get more data, the importance of the prior
will diminish

# References

Bayes Intro

Henry Overos

Goals

Introduction

You already
think like a
Bayesian - you
just don't
know it

Differences
between
Frequentist
and Bayesian
Statistics

The Simple
Mechanics of
Bayesian
Estimation

Questions?

📄 Ryan Bakker & Johannes Karreth (2019)
Lecture Notes and Slides
*Introduction to Bayesian Modeling for the Social Sciences*, ICPSR
2019.

📄 Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki
Vehtari, & Donald B. Rubin (2013)
*Bayesian Data Analysis (Third Edition)*

📄 Nate Silver (2012)
*The Signal and the Noise: Why so many Predictions Fail - and Some
Don't*

# Thank You!