Updating Undergraduate Courses to Include Bayesian Inference

Matthew Brems
General Assembly
matthew.w.brems@gmail.com

Timothy Book
General Assembly
timothykbook@gmail.com

About Matt

Global Lead Data Science Instructor, General Assembly

Managing Partner, ROC AUC, LLC

Business Consultants Chair, Statistics Without Borders

- Previously:
    - Data Science Fellow, Øptimus Consulting
    - Enterprise Analytics Intern, Smucker's
    - M.S., Statistics, The Ohio State University
    - B.A., Math & Economics, Franklin College

- Recommended Reads:
    - Data-Driven Thinking: "Factfulness"
    - Data Visualization: "Storytelling with Data"
    - Data Science: "An Introduction to Statistical Learning with Applications in R"

About Tim

BS in Math, Statistics from Penn State
MS in Statistics from Ohio State

Currently

- Lead Data Science Instructor at General Assembly. We help aspiring data scientists change their careers in an immersive 12-week program.

Previously

- Data Scientist for Summit Consulting
- Statistics Lecturer for Ohio State

Recommended Reads

- “Automate the Boring Stuff with Python” by Al Sweigert
- “Advanced R” by Hadley Wickham
- “The Way of Kings” by Brandon Sanderson

This Afternoon’s Agenda

1. What is Bayesian statistics?
2. Why teach this to undergraduates?
3. Freshmen/Sophomores: The Introductory Curriculum
4. Juniors/Seniors: The Job-Ready Curriculum
    - Plus examples!

What is Bayesian Statistics?

Simply put, **Bayesian statistics is the branch of statistics that deals with incorporating prior beliefs in your methodology**. These prior beliefs may or may not be completely subjective.

Contrast this with the classical **frequentist techniques, which use only the data collected**. Oftentimes, frequentist methods rely on strong parametric assumptions, usually normality.

Why allow subjectivity into our scientific pursuit?

There are multiple good reasons for this:

1. We can conduct statistical analyses with small sample sizes.

Why allow subjectivity into our scientific pursuit?

I conduct a psychological study on randomly sampled students at our university.

- In order to generate a frequentist confidence interval for the mean, what should our sample size be?

- In order to interpret the slope of a frequentist simple linear regression model relating X and Y, what should our sample size be?

Why allow subjectivity into our scientific pursuit?

I conduct a psychological study on randomly sampled students at our university.

- In order to generate a frequentist confidence interval for the mean, what should our sample size be?
- In order to interpret the slope of a frequentist simple linear regression model relating X and Y, what should our sample size be?
- In Bayesian inference, we can generate and interpret these with a sample size of 1!

Why allow subjectivity into our scientific pursuit?

There are multiple good reasons for this:

1. We can conduct statistical analyses with small sample sizes.

2. By having certain beliefs in what our results will be before looking at the data, we get to make stronger probability statements and interpretations.

Why allow subjectivity into our scientific pursuit?

In frequentist analysis, we apply no prior knowledge and let the data speak only for itself. Every time we want to fit a regression coefficient or estimate the mean of some distribution, every value from negative infinity to positive infinity is equally likely.

- **Our data needs to exclude these values that you and I could reasonably exclude without relying on our data!**

In Bayesian inference, we apply our prior knowledge. This enables our data to explore the range of values we've deemed reasonable/likely!

- In many cases, this leads to more precise inferences (i.e. smaller confidence intervals).

Why allow subjectivity into our scientific pursuit?

There are multiple good reasons for this:

1. We can conduct statistical analyses with small sample sizes.

2. By having certain beliefs in what our results will be before looking at the data, we get to make stronger probability statements and interpretations.

3. For the third advantage, we need to rethink what probability means!

What is probability, anyway?

According to Merriam-Webster:

> ***(1)*** *the ratio of the number of outcomes in an exhaustive set of equally likely outcomes that produce a given event to the total number of possible outcomes*
> ***(2)*** *the chance that a given event will occur*

The people at Merriam-Webster are ***frequentists!***

Bayesian Probability

According to Wikipedia,

> *[Bayesian probability] is interpreted as reasonable expectation representing a state of knowledge or as quantification of personal belief.*

What that means is, **Bayesian probability is interpreted as degrees of belief.**

Why does this matter?

Example: Will it rain tomorrow?

We can make predictions, but the accuracy of our predictions can never be known since **tomorrow only occurs once!** We can't simply run "tomorrow" over and over again to find how often it rains.

The only tools we have at our disposal are historical weather data and subjectivity.

Example: Will it rain tomorrow?

The results of Bayesian analysis will yield us a probability. This probability won't represent the actual probability of rain, however. It will represent our *degree of belief*.

Historical data X *"I think it will rain tomorrow"* = *"There is a 60% chance of rain tomorrow"*

Can we teach this to undergraduates?

We believe that this interpretation of probability is actually *more* intuitive than the classical one.

The goal of this workshop is to correct for two fundamental falsehoods inherent in the way statistics is taught right now:

Fundamental Falsehood #1

Bayesian statistics is “too complicated” for students until they get to graduate school.

Fundamental Falsehood #2

Teaching undergraduate students about confidence intervals and *p*-values is a good idea.

(Let's focus more on this one.)

The Problem with *p*-values

Recently, *p*-values have been under attack, and for good reason. They're difficult to understand, interpret, and use. Consequently, other disciplines have taken to abusing them. Whether or not this is on purpose, **this is bad science!**

The Problem with *p*-values

$$
\begin{aligned}
p\text{-value} &= P(Z > |z| \,\big|\, H_0 \text{ true}) \\
&= P(\text{data is more extreme} \,\big|\, H_0 \text{ true}) \\
&\approx P(\text{data} \,\big|\, H_0 \text{ true})
\end{aligned}
$$

The Problem with p-values

$$
\begin{aligned}
p\text{-value} &= P\big(Z > |z| \big| H_0 \text{ true}\big) \\
&= P\big(\text{data is more extreme} \big| H_0 \text{ true}\big) \\
&\approx P\big(\text{data} \big| H_0 \text{ true}\big)
\end{aligned}
$$

Often, we interpret p-values as $P\big(H_0 \text{ true} \big| \text{data!}\big)$

The Problem with *p*-values

In 2016, the ASA put out a statement detailing the many misuses *p*-values suffer in modern day statistics and how to solve this problem.

“The p-value was never intended to be a substitute for scientific reasoning,” said Ron Wasserstein, the ASA’s executive director. “Well-reasoned statistical arguments contain much more than the value of a single number and whether that number exceeds an arbitrary threshold. The ASA statement is intended to steer research into a ‘post $p<0.05$ *era.’”*

How did we get here?

We believe the cause of many of these woes is the difficult interpretations of some key classical statistical concepts.

The Bayesian equivalents are direct and easy to understand and interpret.

Confidence Intervals

$$95\% \text{ Interval for } \mu\text{: } (80, 100)$$

Correct Interpretation	Common Mistake	Bayesian Credible Interval
We are 95% confident that the true population mean is between 80 and 100.	*There is a 95% probability that the true population mean is between 80 and 100.*	*There is a 95% probability that the true population mean is between 80 and 100.*

p-values

$$p < \alpha = 0.05$$

Correct Interpretation	**Common Mistake**	**Bayesian Approach**
Assuming the null hypothesis is true, the probability of seeing our data, or something more extreme than our data, is below 5%. Therefore, we reject the null hypothesis.	*The probability that the null hypothesis is true is below 5%.*	*The probability that the null hypothesis is true is below 5%.*

So, it's magic, huh?

Bayesian methods eschew the need for *p*-values by instead thinking in terms of pure probability. Once these methods are carried out, computing these probabilities is much easier than computing a *p*-value.

It's not all roses, though. There are some hurdles we need to pass in order to get up and running. Specifically, these are the biggest two:

Hurdle #1: Computers

Almost all modern Bayesian methods are carried out via computer. We believe in letting Juniors and Seniors program results themselves. We will probably feed Freshmen and Sophomore in intro classes pre-computed results.

This might feel wrong at first - but don't we already do this? Students are rarely computing two-sample proportion tests by hand. Almost always, they're using built-in calculator functions to do this for them. (As would be the case in a professional setting!)

Hurdle #2: The "Yuck" Factor

Science is objective! Why are we allowing subjectivity into our analyses?

We will take great care to assuage these fears. In most cases, Bayesian methods will require weaker (and fewer) assumptions than their frequentist counterparts.

Also... what part of statistics *isn't* subjective?