Bayesian Statistics for Freshmen and Sophomores

The Setting

College students are expected to take a **general statistics**, **methods** or **data analysis** course while in undergrad.

This course is typically **algebra-based**, often concludes with inferential procedures, and may be geared toward a particular subject or field.

We imagine Bayesian statistics into **two to three weeks** of a typical fifteen-week semester course.

Sample Syllabus of Bayesian Statistics

Week One

- Bayes' Theorem
- Jumping from probabilities to distributions
- Defining prior, likelihood, and posterior

Sample Syllabus of Bayesian Statistics

Week Two

- Recap of parameters and inference
- Interpreting posterior distributions

Prerequisites (Week Zero?)

Depending on the setup of the course, probability rules (including Bayes' Theorem!) may have been covered earlier in the course or in a previous course.

For the purposes of this setup, we'll assume that students are already comfortable with certain probability rules, including:

- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $P(A|B) = P(A \cap B)/P(B)$
- $P(A \cap B) = P(A|B) \times P(B)$
- $P(A) = \sum_i P(A \cap B_i)$

This content *may* need to be reviewed prior to diving into Bayes' Theorem.

Programming Note

Keeping in mind the diverse set of students who often take this class, programming can be either included or excluded.

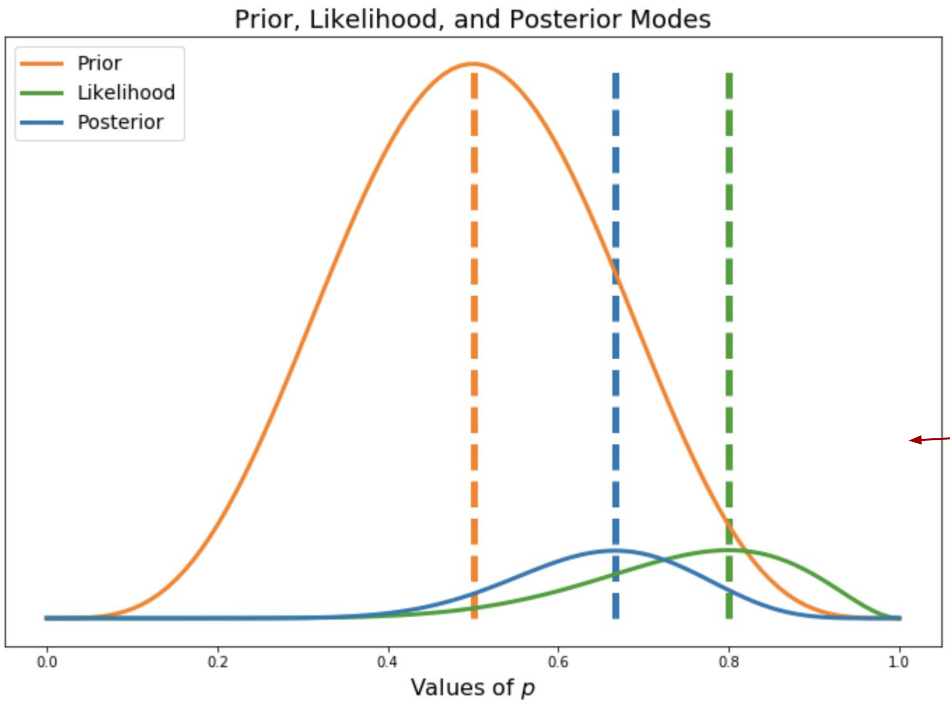
- **If programming is not desired**: You can generate the outputs for students to interpret and there are a wealth of applets available for students to use that require no coding background.

- **If programming is desired**:
    - Provide the students with code (scripts, functions) and allow them to run the code themselves.
    - Expect students to learn coding with syntax. *(More on this later.)*

Programming Note

```
plot_beta_binomial(a = 5,            # Hyperparameter alpha for Beta(alpha, beta)
                   b = 5,            # Hyperparameter beta for Beta(alpha, beta)
                   n_trials = 10,    # Hyperparameter n for Binomial(n, k)
                   n_successes = 8) # Hyperparameter k for Binomial(n, k)
```

students fill in function arguments



students immediately see the output

Week 1

Introducing Bayes' Theorem

When covering Bayes' Theorem, it can help to derive the theorem from other probability rules:

$$P(A \cap B) = P(B \cap A)$$

$$\Rightarrow P(A|B)P(B) = P(B|A)P(A)$$

$$\Rightarrow P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Introducing Bayes’ Theorem

When introducing Bayes’ Theorem, we find it helpful to start with a few examples that scaffold in a particular order:

1. A basic Bayes’ Theorem problem.

2. A Bayes’ Theorem problem with a surprising result. (We find that Monty Hall works well here.)

3. A Bayes’ Theorem problem that allows for a subjective assignment to the prior probability. (i.e. sampling objects from jars.)

Introducing Bayes' Theorem: Problem 1

Suppose that someone goes to the doctor for a check-up. A test indicates that the person does have high blood pressure. 10% of the population has high blood pressure. Among those who have high blood pressure, there is a positive test 70% of the time. Among those who do not have high blood pressure, there is a positive test 20% of the time. Given the positive test, what is the posterior probability that the person has high blood pressure?

- Check for understanding of defining various probabilities.
- General practice.
- Introduce terminology "prior" and "posterior."
- https://seeing-theory.brown.edu/bayesian-inference/index.html

Introducing Bayes’ Theorem: Problem 2

You’re on a game show called “Let’s Make A Deal.” There are three doors (A, B, or C), with a car behind one and a goat behind the other two. The host asks you to pick a door. Among the doors you do not choose, the host opens one to reveal a goat. The host then asks if you want to stick with your original door or switch to the other unopened door. Find the posterior probability of winning the car if you switch doors.

- Check for understanding of defining various probabilities.
- Reinforce terminology “prior” and “posterior.”
- Introduce the counterintuitive nature of probability and the need for caution.

Introducing Bayes’ Theorem: Problem 3

In your office, there are two jars of candies. One jar (jar A) contains 40 candies, of which 10 contain caramel. The other jar (jar B) contains 40 candies, 20 of which contain caramel. You select candy from one of the jars and find that it contains caramel. Given the evidence, what is the probability that you pulled the caramel from jar A?

- Check for understanding of defining various probabilities.
- Reinforce terminology “prior” and “posterior.”
- Introduce the counterintuitive nature of probability and the need for caution.
- This third problem allows us to dig deeper!

Introducing Bayes’ Theorem: Problem 3

This third problem involving jars allows us to start getting into the subjective nature of Bayesian inference without explicitly calling it subjective. Instead, a discussion can prompt this!

- “We assumed that each jar was equally likely. What would happen if jar A was located at the front of the office, where more people walked by it?”

- You can provide numbers, but by encouraging a discussion where students identify reasonable values for the priors without an explicit number can help them grow comfortable with the subjective nature of Bayesian probabilities.

From Probabilities to Distributions

- Suppose I want to better understand **the average height of college students**.

- If I'm using Bayesian inference, I want to understand the posterior probability of the average height, but we run into an issue.

- Earlier, we could simply list out our options (the car was behind one of three doors, the candy came from one of two jars) and calculate the posterior probability of each option.

From Probabilities to Distributions

$$P(\mu = 0|\text{data}) = \frac{P(\text{data}|\mu = 0)P(\mu = 0)}{P(\text{data})}$$

$$P(\mu = 0.1|\text{data}) = \frac{P(\text{data}|\mu = 0.1)P(\mu = 0.1)}{P(\text{data})}$$

$$P(\mu = 0.2|\text{data}) = \frac{P(\text{data}|\mu = 0.2)P(\mu = 0.2)}{P(\text{data})}$$

$$\vdots$$

From Probabilities to Distributions

- Obviously, there are limitations to this:
  - It gets computationally intensive to calculate these probabilities directly.
  - We can't enumerate all possible posterior probabilities!

- Recall that a distribution is **the set of all values of a variable and how frequently we observe each value**.
  - $P(\mu = 0), P(\mu = 0.01), P(\mu = 0.02), \ldots$ is *approximately* the distribution of μ
  - $P(\mu = 0|data), P(\mu = 0.01|data), P(\mu = 0.02|data), \ldots$ is *approximately* the distribution of $\mu|\text{data}$

From Probabilities to Distributions

$$P(\mu = 0|\text{data}) = \frac{P(\text{data}|\mu = 0)P(\mu = 0)}{P(\text{data})}$$

$$\Downarrow$$

$$f(\mu|\text{data}) = \frac{f(\text{data}|\mu)f(\mu)}{f(\text{data})}$$

From Probabilities to Distributions

$$P(\mu = 0|\text{data}) = \frac{P(\text{data}|\mu = 0)P(\mu = 0)}{P(\text{data})}$$

$$\Downarrow$$

$$f(\mu|\text{data}) = \frac{f(\text{data}|\mu)f(\mu)}{f(\text{data})} \propto f(\text{data}|\mu)f(\mu)$$

From Probabilities to Distributions

$$P(\mu = 0|\text{data}) = \frac{P(\text{data}|\mu = 0)P(\mu = 0)}{P(\text{data})}$$

$$\Downarrow$$

posterior

likelihood

prior we chose

$$f(\mu|\text{data}) = \frac{f(\text{data}|\mu)f(\mu)}{f(\text{data})} \propto f(\text{data}|\mu)f(\mu)$$

Terminology

Prior Distribution: the set of all values of the parameter and how frequently we believe the parameter takes on each value, **before seeing the data.**

Terminology

Prior Distribution: the set of all values of the parameter and how frequently we believe the parameter takes on each value, **before seeing the data.**

- Based on your own knowledge, what do you think the true value of μ might look like?

Terminology

Prior Distribution: the set of all values of the parameter and how frequently we believe the parameter takes on each value, **before seeing the data.**

- Based on your own knowledge, what do you think the true value of μ might look like?

Likelihood: the function that measures **how likely we are to observe our data for all values of the parameter**.

Terminology

Prior Distribution: the set of all values of the parameter and how frequently we believe the parameter takes on each value, **before seeing the data.**

- Based on your own knowledge, what do you think the true value of μ might look like?

Likelihood: the function that measures **how likely we are to observe our data for all values of the parameter**.

- It connects our data with our beliefs. “For all possible values of mu, how likely is it that I observed this data?”
- Similar to a *p*-value!

Terminology

Posterior Distribution: the set of all values of the parameter and how frequently we believe the parameter takes on each value, **after taking the data into account**.

Terminology

Posterior Distribution: the set of all values of the parameter and how frequently we believe the parameter takes on each value, **after taking the data into account**.

- The posterior distribution is our prior distribution, but updated with our data.
- The posterior is a compromise between our prior beliefs and our observed data.
- https://rpsychologist.com/d3/bayes/

Week 2

Recap of Parameters and Inference

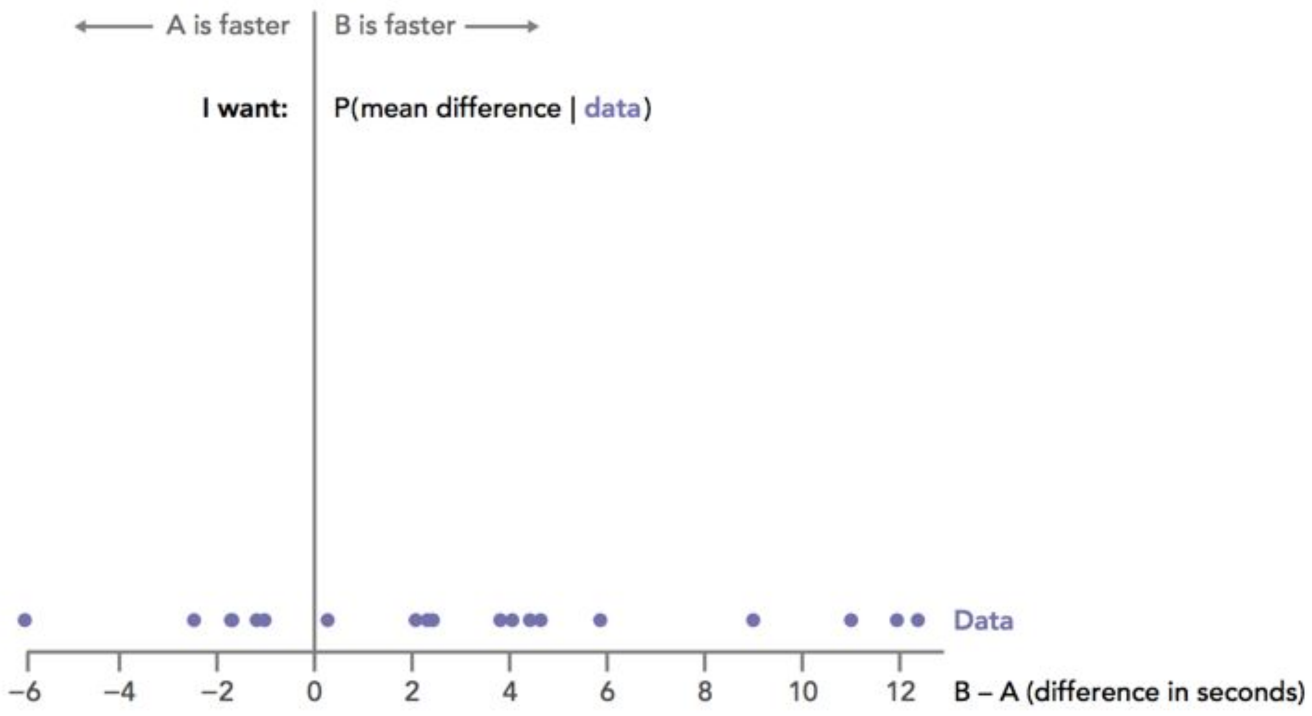
Spending part of a lesson on **defining parameters** (as measurements of the population) and **identifying parameters** is helpful when breaking into Bayesian inference.

Recap of Parameters and Inference

Spending part of a lesson on **defining parameters** (as measurements of the population) and **identifying parameters** is helpful when breaking into Bayesian inference.

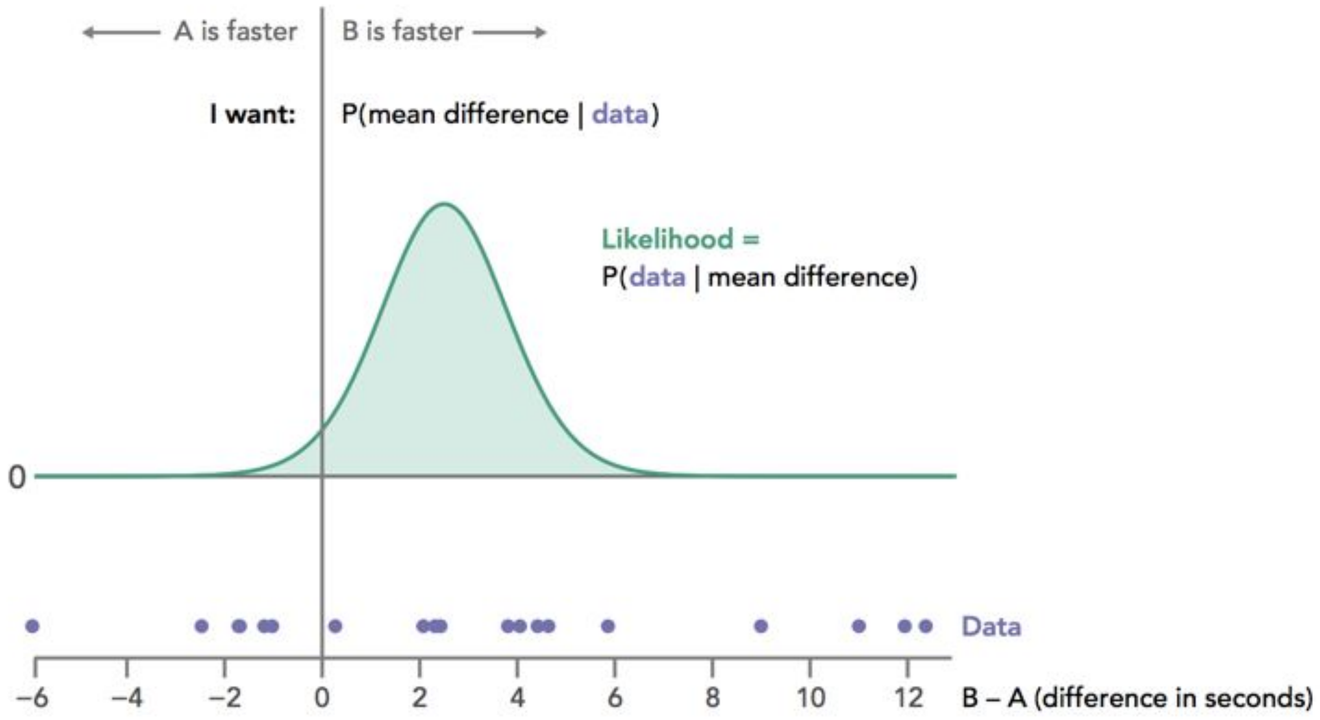
- "Suppose I develop a medicine reducing blood pressure on average."
- "Suppose I develop a medicine (Drug A) reducing blood pressure. I want to study whether or not Drug A reduces blood pressure more, on average, than Drug B."
- "Suppose I develop a medicine (Drug A) reducing blood pressure. I want to study whether or not Drug A reduces blood pressure in more people than Drug B."

Connecting it Visually



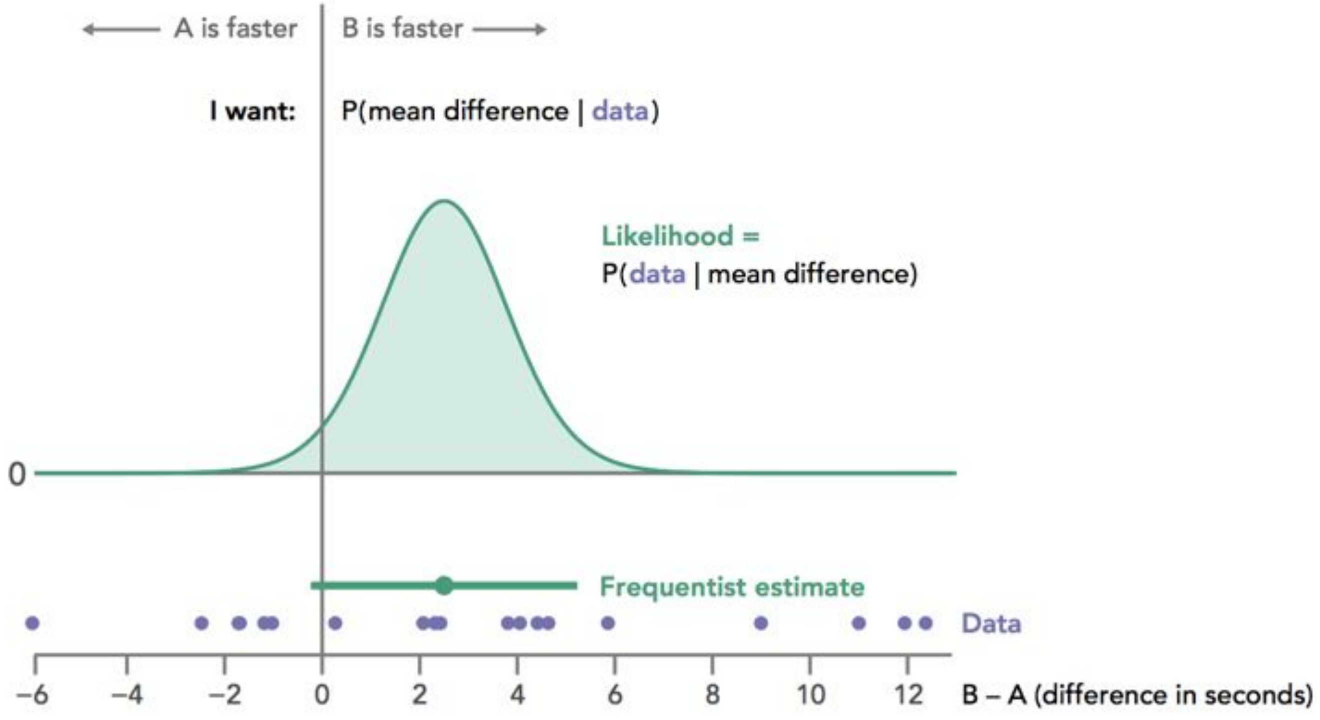
Source: http://www.mjskay.com/presentations/openvisconf2018-bayes-uncertainty-2.pdf

Connecting it Visually



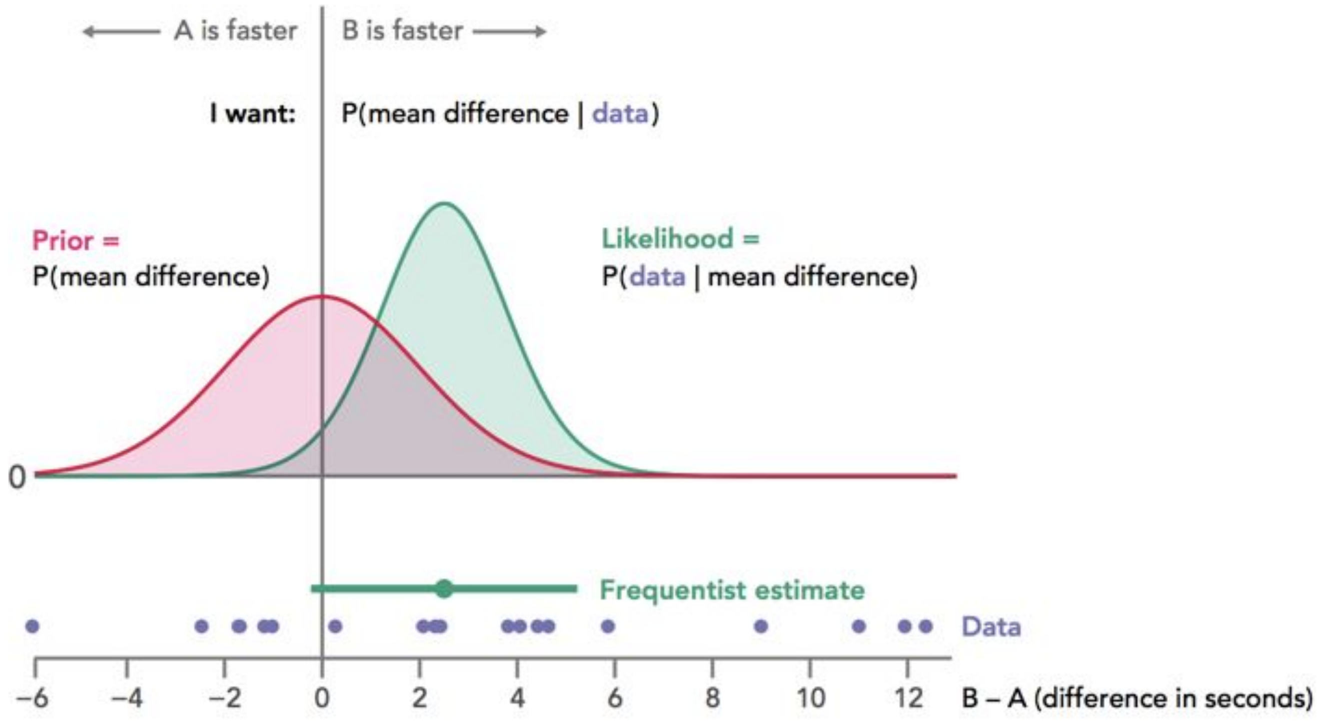
Source: http://www.mjskay.com/presentations/openvisconf2018-bayes-uncertainty-2.pdf

Connecting it Visually



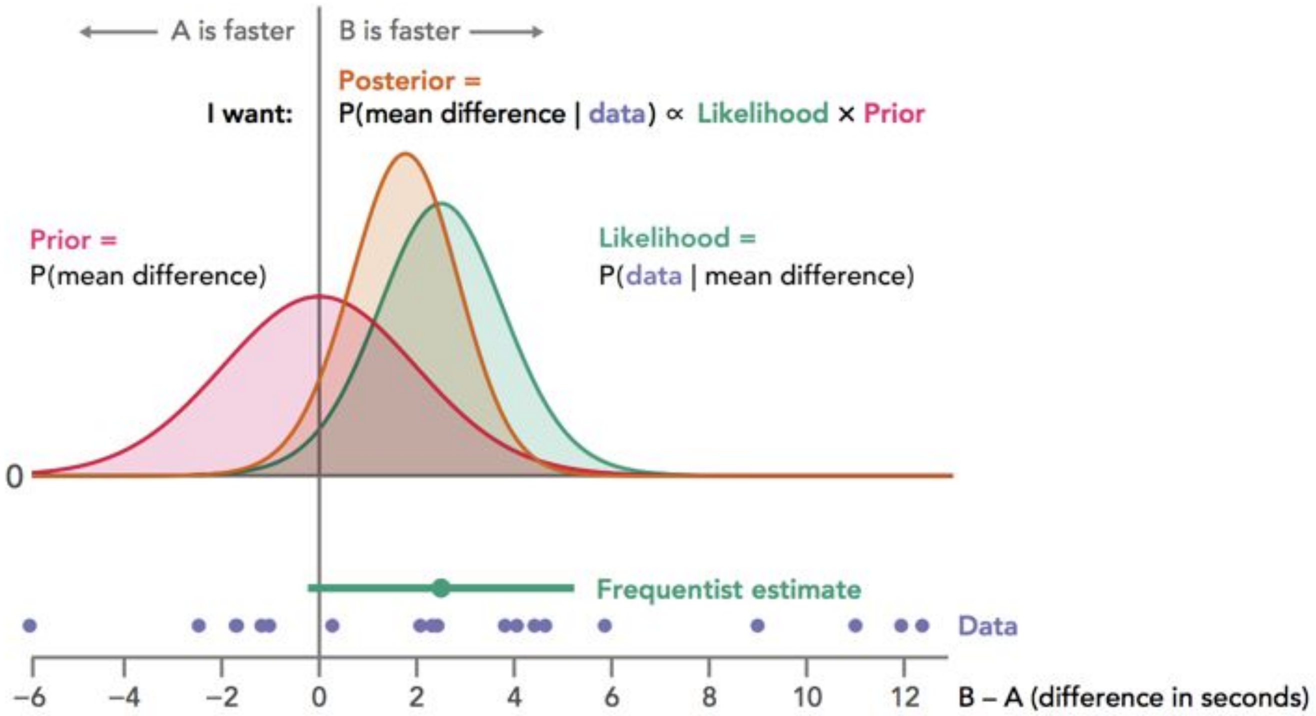
Source: http://www.mjskay.com/presentations/openvisconf2018-bayes-uncertainty-2.pdf

Connecting it Visually



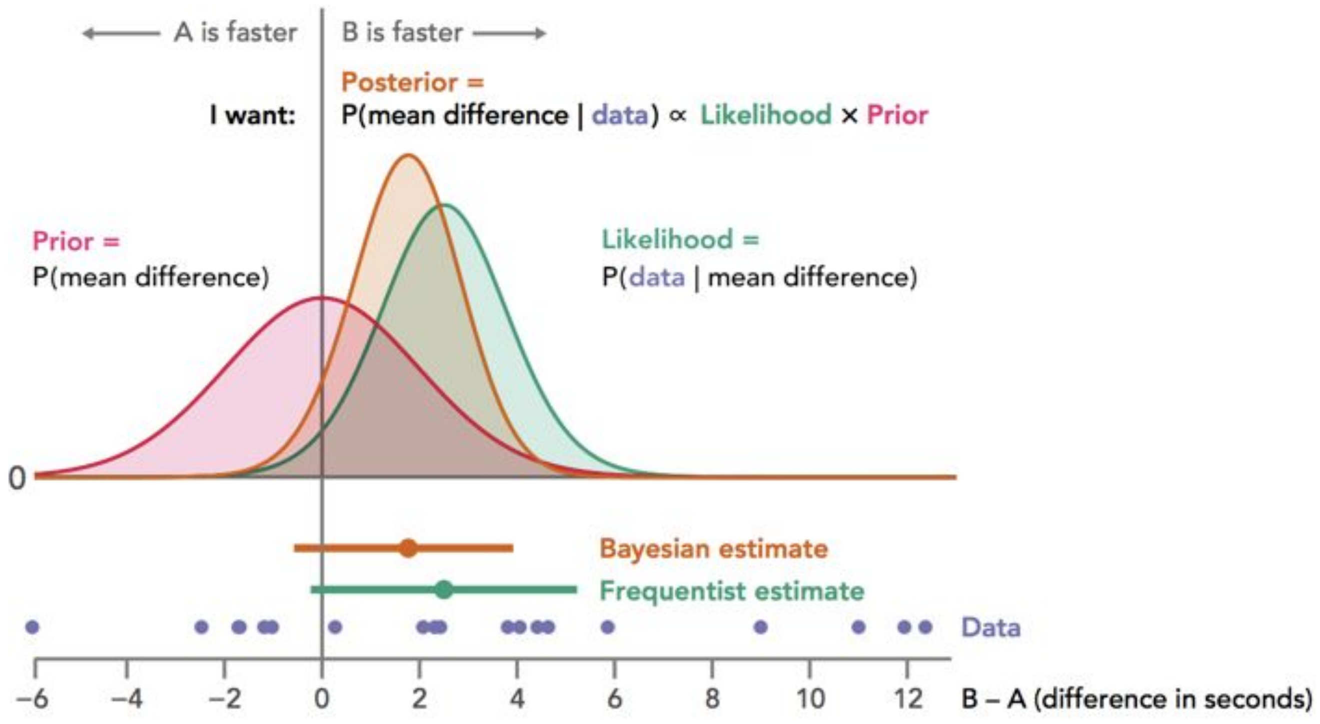
Source: http://www.mjskay.com/presentations/openvisconf2018-bayes-uncertainty-2.pdf

Connecting it Visually



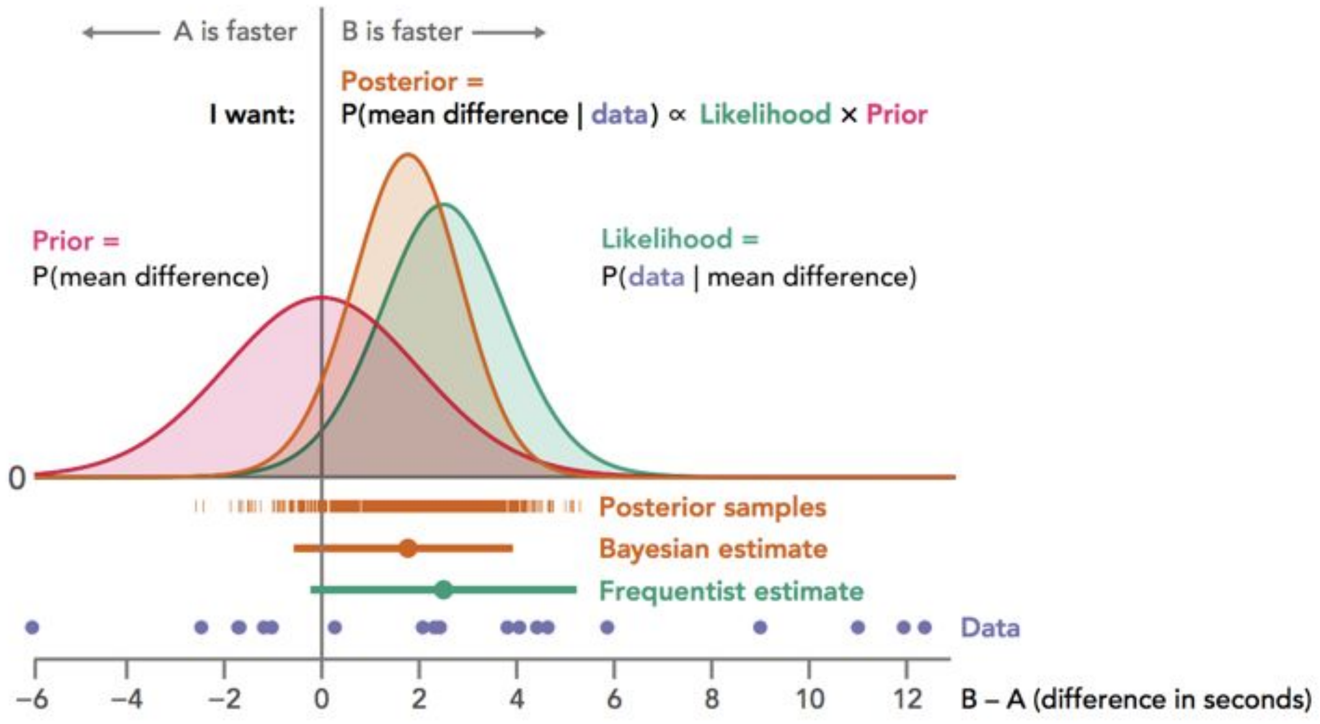
Source: http://www.mjskay.com/presentations/openvisconf2018-bayes-uncertainty-2.pdf

Connecting it Visually



Source: http://www.mjskay.com/presentations/openvisconf2018-bayes-uncertainty-2.pdf

Connecting it Visually



Source: http://www.mjskay.com/presentations/openvisconf2018-bayes-uncertainty-2.pdf

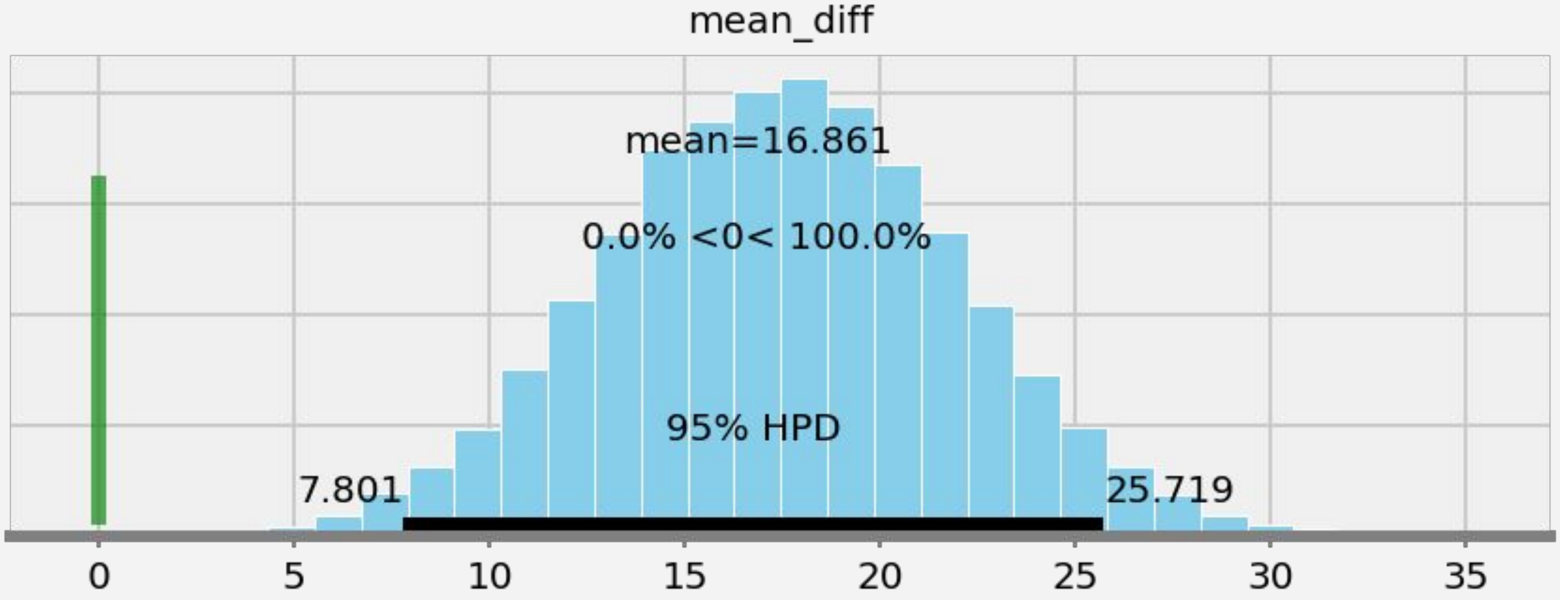
Interpreting the Posterior Distribution

The posterior distribution is a complete summary of the parameter of interest!

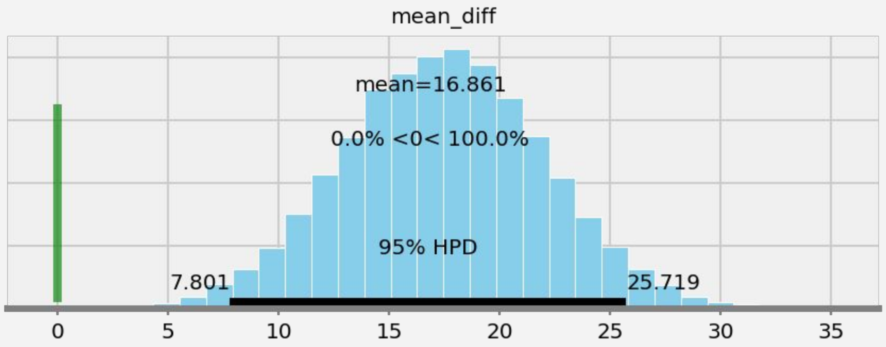
Example: Suppose we are comparing the effect of two drugs (A and B) on blood pressure. We specify a prior and a likelihood and get a posterior distribution of our parameter $\mu_A - \mu_B$.

Interpreting the Posterior Distribution

$$\mu_A - \mu_B$$



Interpreting the Posterior Distribution

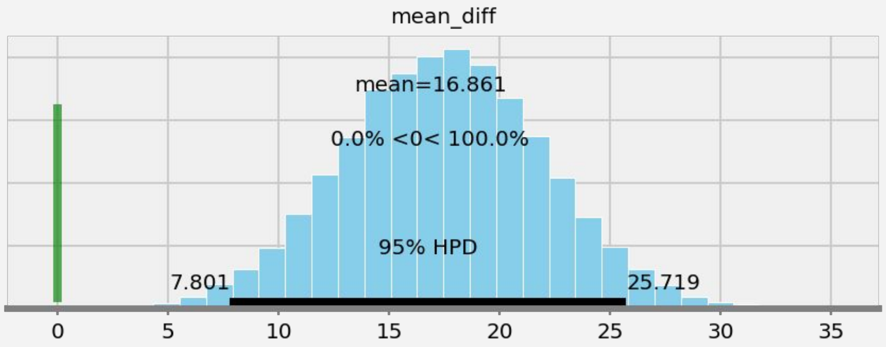


The extent to which you talk about the samples from the posterior distribution is up to you! *(More on this later.)*

If we want to answer questions, we can answer them directly:

- What is the probability that Drug A causes a higher drop in blood pressure than Drug B? (100%.)
- What is the probability that Drug A causes a drop at least 15 points higher than drug B? (Find the area/percentage of samples greater than or equal to 15.)
- What is the likeliest mean difference in blood pressure caused by drugs A and B? (Find the mode.)

Interpreting the Posterior Distribution



Special attention should be given to:

- credible intervals
- Bayesian estimation.

- Credible Interval: Find the 95% highest posterior density.
  - "There is a 95% chance that the true average difference in blood pressure caused by drug A and drug B is between 7.801 and 25.719." (Most common approach.)
  - Alternatively, you can grab the samples corresponding to the 2.5th and 97.5th percentiles.
- Bayesian Estimation: Instead of a p-value, quantify $P(H|\text{data})$.
  - "The probability that the effect of Drug A is less than Drug B is effectively 0%."

Sample Syllabus of Bayesian Statistics

Week One

- Bayes' Theorem
- Jumping from probabilities to distributions
- Defining prior, likelihood, and posterior

Week Two

- Recap of parameters and inference
- Interpreting posterior distributions