Bayesian Statistics – Lecture 2

Thomas J. Faulkenberry, Ph.D.

May 28, 2024

Goals for today

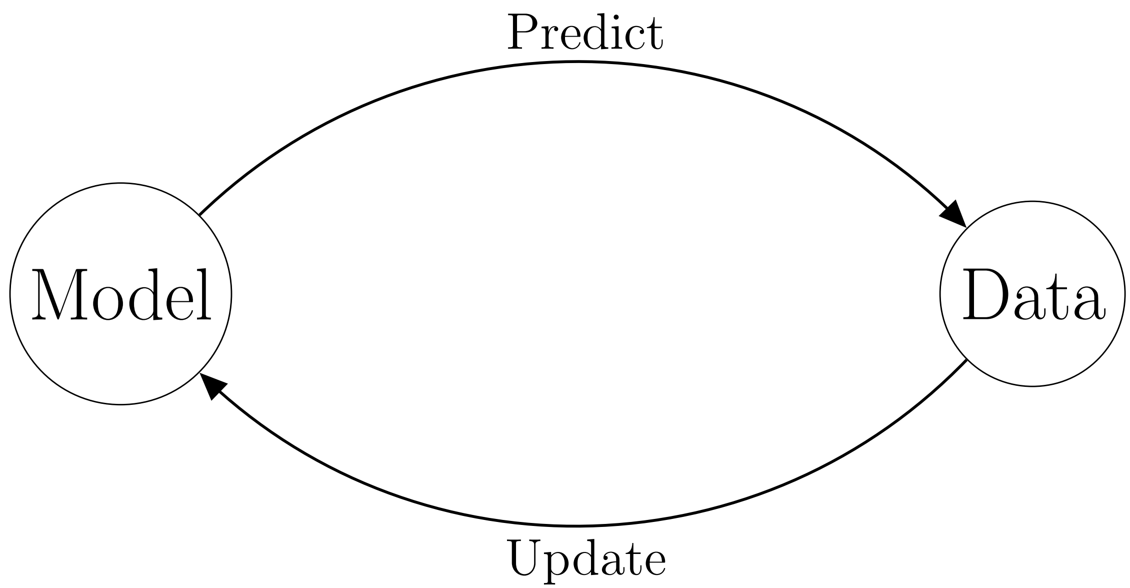
1. Introduce and understand Bayes' Theorem

2. Learn some new vocabulary

   - *Prior*
   - *Posterior*
   - *Predictive*
   - *Marginal*

3. Introduce the *Bayes factor* as a tool for model comparison

What is "Bayesian" statistics?



This is really all there is to it, but it is hard to see at first. The more you do it, the better you'll get at "seeing through" the complicated bits. . .

It all comes from Bayes' Theorem

You may have seen this somewhere before:

$$P(A \mid B) = P(A) \times \frac{P(B \mid A)}{P(B)}$$

It all comes from Bayes' Theorem

In the context of statistics, you'll more often see it like this:

$$\pi(\theta \mid \text{data}) = \pi(\theta) \times \frac{p(\text{data} \mid \theta)}{p(\text{data})}$$

It all comes from Bayes' Theorem

In the context of statistics, you'll more often see it like this:

$$\underbrace{\pi(\theta \mid \text{data})}_{\substack{\text{Posterior}\\ \text{distribution}\\ \text{for } \theta}} = \underbrace{\pi(\theta)}_{\substack{\text{Prior}\\ \text{distribution}\\ \text{for } \theta}} \times \underbrace{\frac{p(\text{data} \mid \theta)}{p(\text{data})}}_{\substack{\text{predictive updating}\\ \text{factor}}}$$

It all comes from Bayes' Theorem

In the context of statistics, you'll more often see it like this:

$$\underbrace{\pi(\theta \mid \text{data})}_{\substack{\text{Posterior}\\ \text{distribution}\\ \text{for } \theta}} = \underbrace{\pi(\theta)}_{\substack{\text{Prior}\\ \text{distribution}\\ \text{for } \theta}} \times \underbrace{\frac{p(\text{data} \mid \theta)}{p(\text{data})}}_{\substack{\text{predictive updating}\\ \text{factor}}}$$

Informally – to get the **posterior** distribution for your model parameter θ:

1. start with a **prior** distribution for θ

2. change/update the prior distribution at each value of θ by an **updating factor** that depends on the fit between θ and your observed data

An example

Let's make these concepts a bit more concrete with an example.

In 1996, the New Mexico State Legislature passed a House Joint Memorial declaring "Red or Green?" as the official state question. This refers to the question always asked whether one prefers red or green chile when ordering New Mexican cuisine.

An example

For our example, we are interested in the percentage of “green chile people” in my department.

If we assume that each of my $N = 10$ colleagues' choice is independent of the others, then we can use a **binomial** model, which gives the probability of observing a certain number of "successes" out of a set of N trials. It treats each choice as a "coin flip" where the probability of "success" (i.e., choosing green chile) on each trial is θ

Binomial model

Some notation:

- Data to observe: $x =$ number of people who prefer green chiles
- Parameter of interest: $\theta =$ (population) proportion of people who prefer green chile
- Probability distribution:

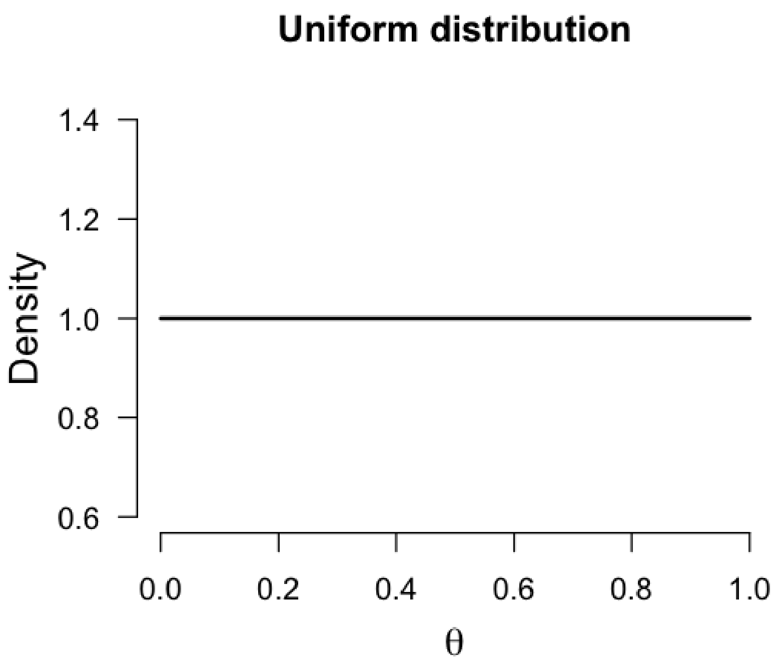
$$p(x \mid \theta) = \frac{N!}{x!(N-x)!} \cdot \theta^x \cdot (1-\theta)^{N-x}$$

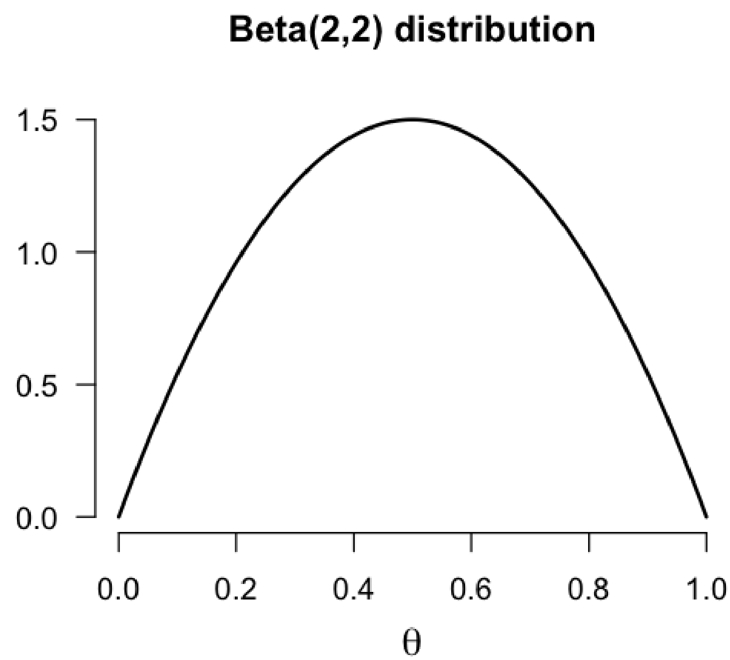
Prior – the key Bayesian ingredient

The start of any Bayesian analysis is to choose a **prior distribution** for your model parameter(s). This is a probability distribution that *mathematically encodes* your prior belief about θ *before observing any data*.

Prior – the key Bayesian ingredient

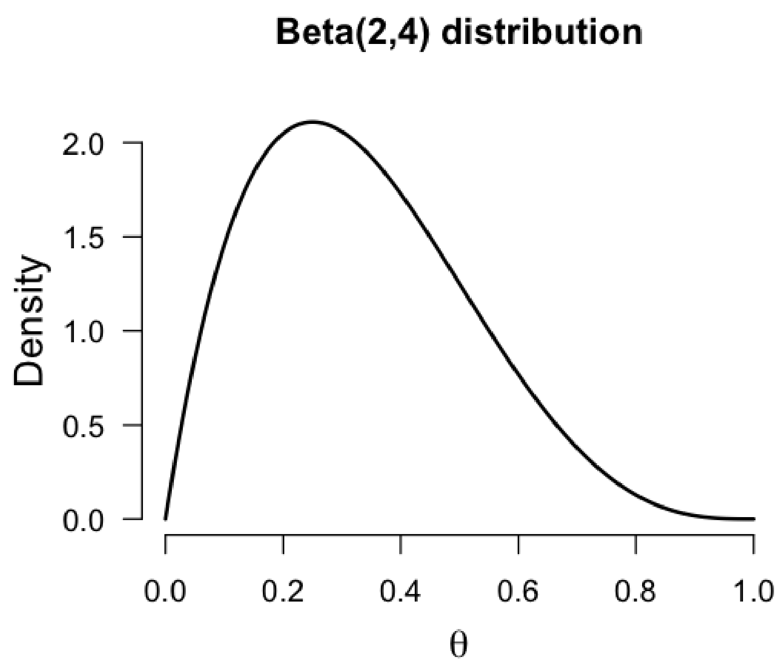
Here are two examples:

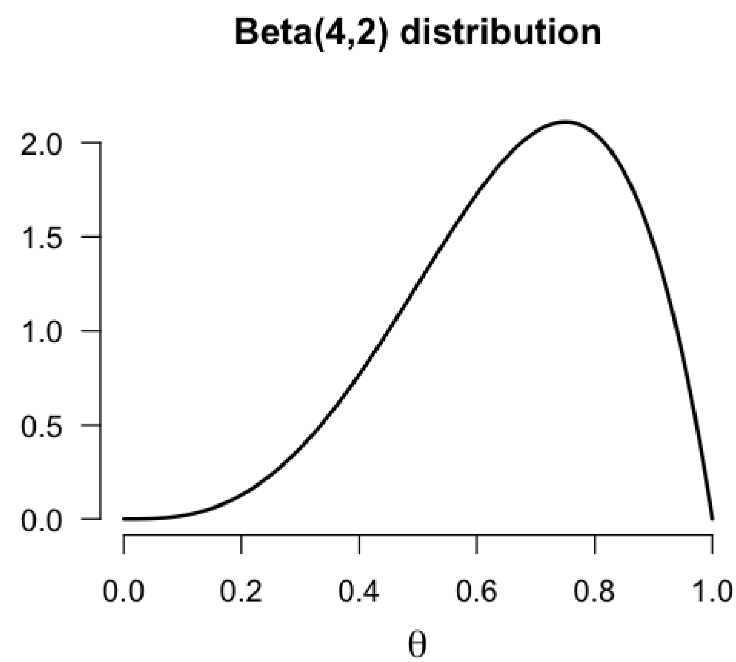




Prior – the key Bayesian ingredient

In fact, Beta distributions work really nicely with the binomial model. Here are two more examples:



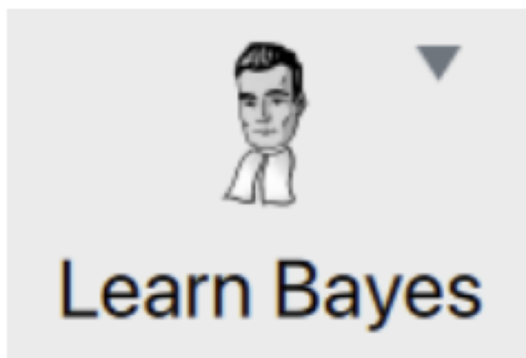


From prior to posterior

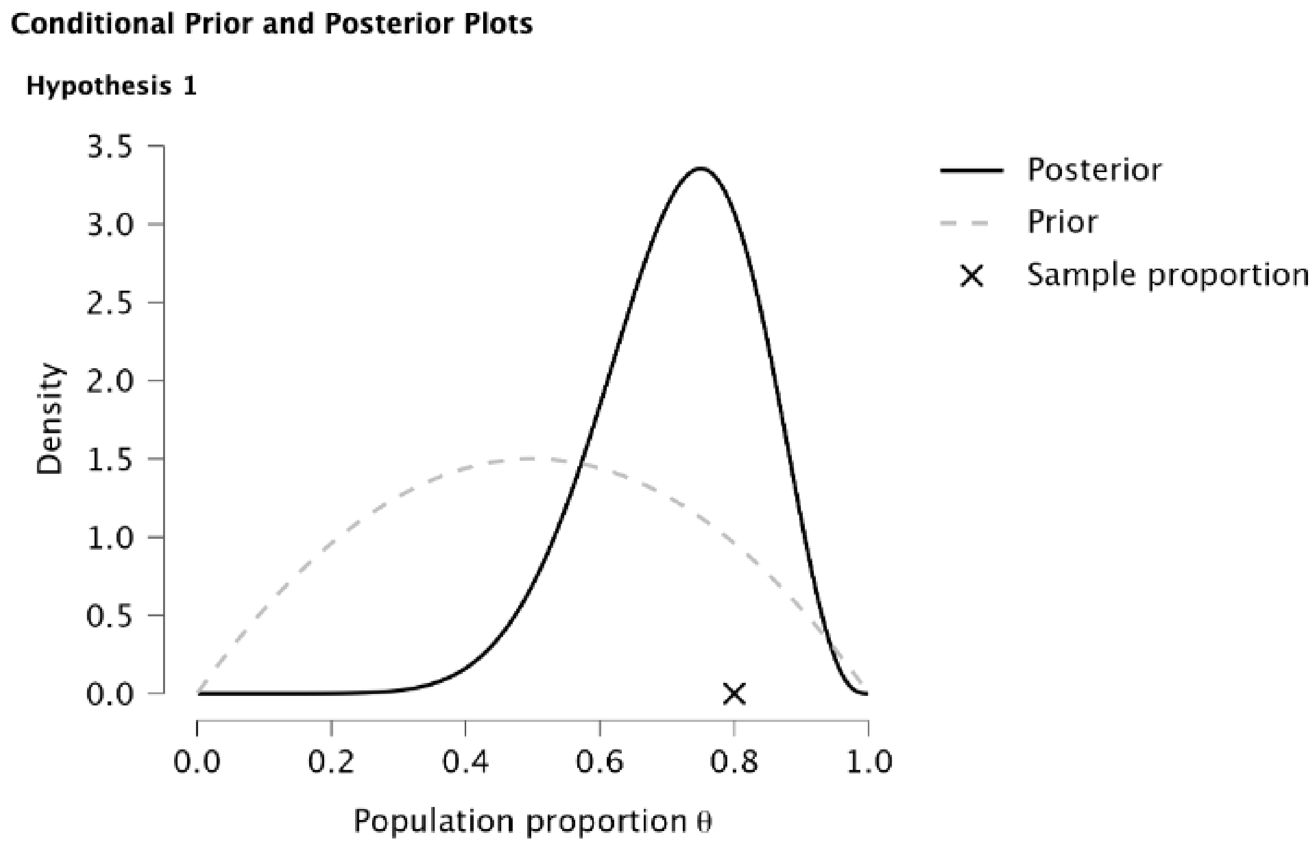
Observed data: 8 of my 10 colleagues preferred green chiles.

The "magic" of Bayes is that after you choose your prior, observing data $(x = 8)$ then *updates* the prior to a **posterior** distribution on θ.

You can work through the math of Bayes thereom, but first let's use JASP – specifically, the "Learn Bayes" add-on module



From prior to posterior



From prior to posterior

We'll now work through (some of) the math of Bayes' theorem, just to get some intuition about the updating process.

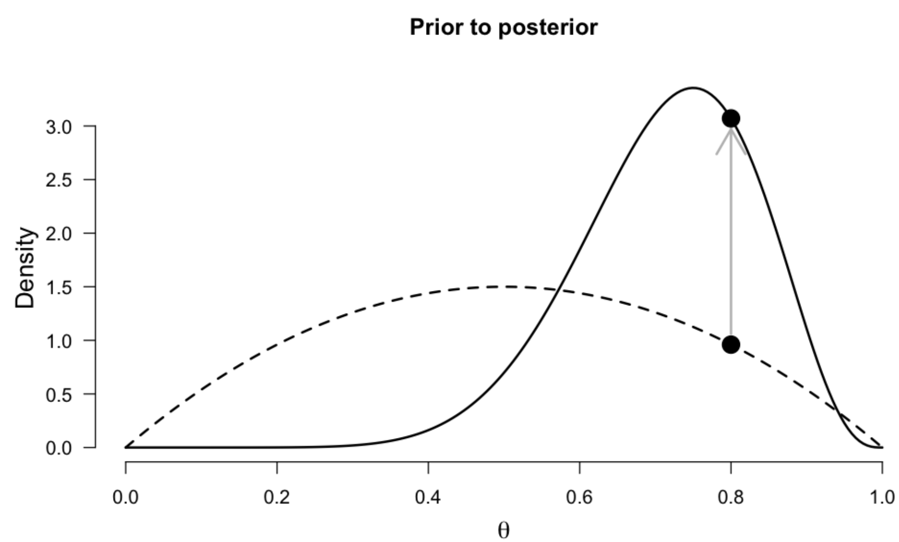
Recall:

$$\underbrace{\pi(\theta \mid \text{data})}_{\substack{\text{Posterior} \\ \text{distribution} \\ \text{for } \theta}} = \underbrace{\pi(\theta)}_{\substack{\text{Prior} \\ \text{distribution} \\ \text{for } \theta}} \times \underbrace{\frac{p(\text{data} \mid \theta)}{p(\text{data})}}_{\substack{\text{predictive updating} \\ \text{factor}}}$$

From prior to posterior

We view the posterior distribution of a *function* of θ. So, we need to track what happens to values of θ after the data ($x = 8$ successes) is observed.

For concreteness, let's track a specific value of θ, say $\theta = 0.8$, the observed proportion.



Calculating the Bayesian updating

Step 1: calculate $\pi(\theta)$

- Under a Beta(2,2) prior, the prior density of $\theta = 0.8$ is 0.96
- note: by hand, this calculation is difficult, but you can use R to calculate it quickly

Calculating the Bayesian updating

Step 2: calculate $p(\text{data} \mid \theta)$

- calculating this *likelihood* requires only the binomial model:

$$p(x = 8 \mid \theta = 0.8) = \frac{10!}{8! \cdot 2!} \cdot 0.8^8 \cdot 0.2^2 = 0.3020$$

Calculating the Bayesian updating

Step 3: calculate $p(\text{data})$

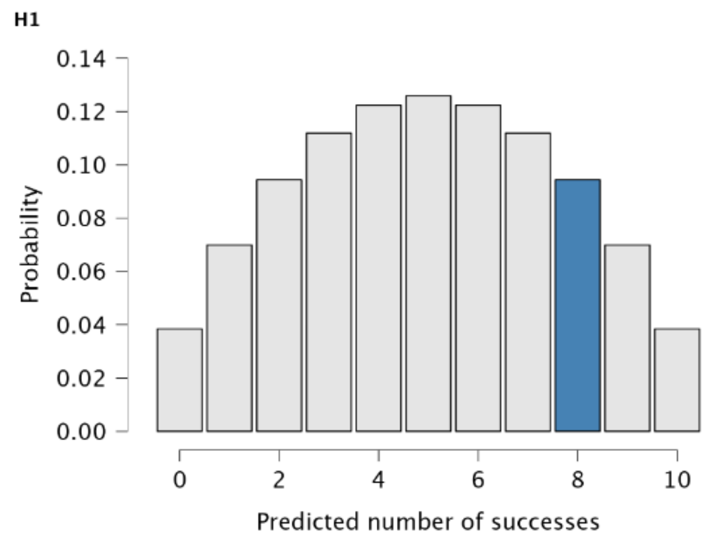
- this is called the *prior predictive* (or *marginal*) distribution

- just like the binomial distribution, it gives the probability of x successes, but instead of fixing a value of θ, it *averages over* all possible values of θ.

Calculating the Bayesian updating

Step 3: calculate $p(\text{data})$

- for $x = 8$ successes, this marginal probability is 0.094



Calculating the Bayesian updating

Step 4: plug everything into Bayes' theorem:

$$
\begin{aligned}
\pi(\theta = 0.8 \mid \text{data}) &= \pi(\theta = 0.8) \cdot \frac{p(\text{data} \mid \theta = 0.8)}{p(\text{data})} \\
&= 0.96 \times \frac{0.3020}{0.094} \\
&= 0.96 \times 3.213 \\
&= 3.08
\end{aligned}
$$

What about model comparison?

Let's introduce a competitor model – one where there is absolutely no preference (i.e., $\theta = 0.5$). Now we have a hypothesis testing scenario, where we will compare:

- $\mathcal{H}_0 : \theta = 0.5$
- $\mathcal{H}_1 : \theta \sim \text{Beta}(2, 2)$

Note: compared to last time, where we wrote $\mathcal{H}_1 : \theta \neq 0.5$, we are now being much more explicit about what we mean by "$\theta \neq 0.5$". Here, we specify *how* θ *is distributed* under $\mathcal{H}_1$.

Let's compare predictive performance of the models

In JASP, we can define both models and look at their prior predictive distributions.

Let's compare predictive performance of the models

We can immediately see that our observed data ($x = 8$ successes) is much more likely under $\mathcal{H}_1$ than under $\mathcal{H}_0$.

How much more likely? Take the ratio:

$$\begin{aligned}\frac{p(\text{data} \mid \mathcal{H}_1)}{p(\text{data} \mid \mathcal{H}_0)} &= \frac{0.094}{0.044} \\ &= 2.14\end{aligned}$$

This ratio is called the **Bayes factor**.

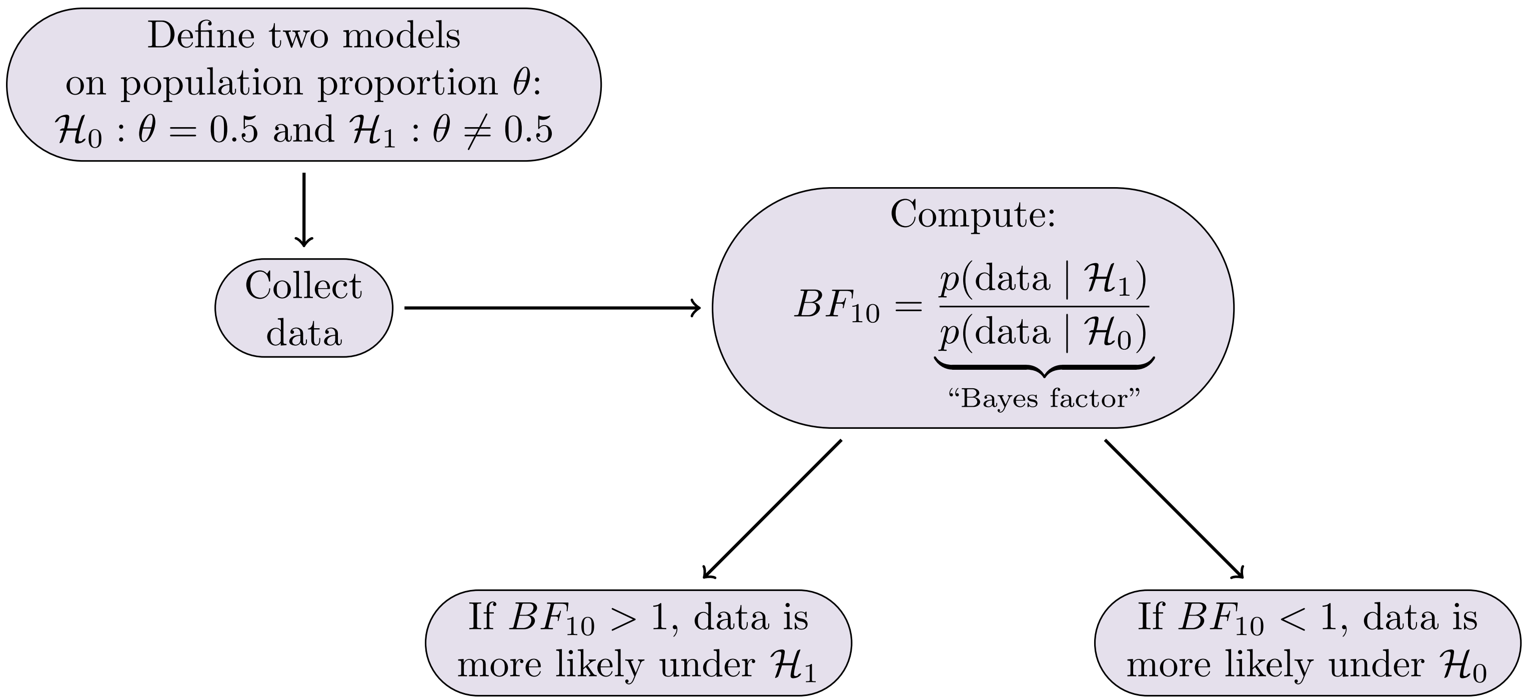
Bayes factor

The Bayes factor, denoted BF_{10}, measures the relative predictive performance of the two competing models $\mathcal{H}_1$ and $\mathcal{H}_0$:

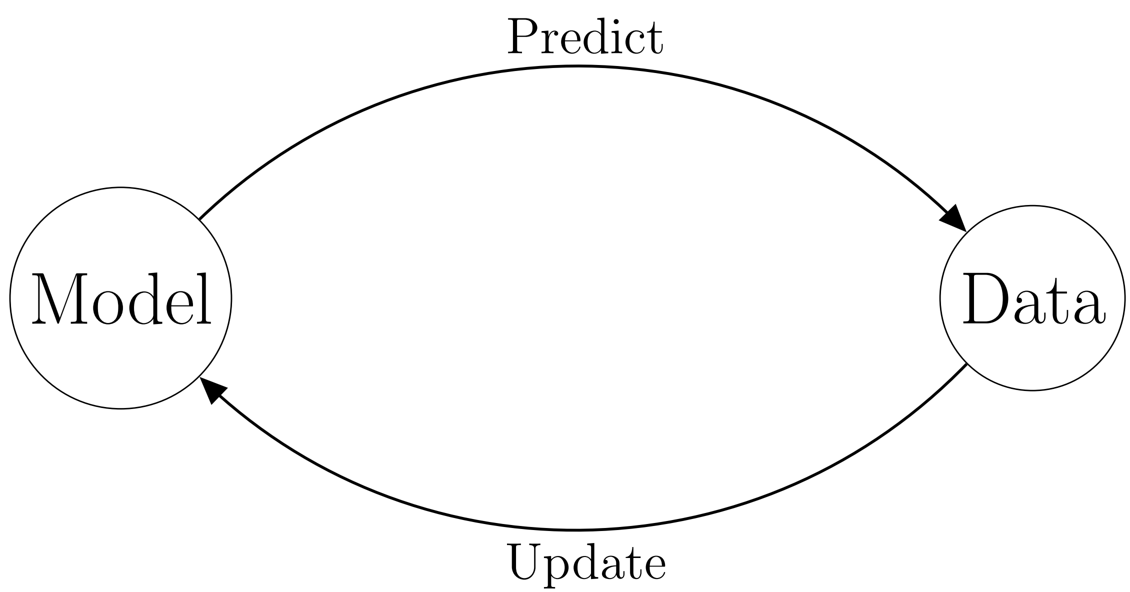
$$\mathsf{BF}_{10} = \frac{p(\text{data} \mid \mathcal{H}_1)}{p(\text{data} \mid \mathcal{H}_0)}$$

- if $\mathsf{BF}_{10} > 1$, then $\mathcal{H}_1$ is preferred.
- if $\mathsf{BF}_{10} < 1$, then $\mathcal{H}_0$ is preferred.

A Bayesian workflow



Epilogue – so what is "Bayesian" statistics?



That is really all there is to it!