Statistical Inference, pt. I

ECON 3640–001

Marcio Santetti
Spring 2022

Motivation

Housekeeping

Notes based on `Johnson et al. (2022)`:

- Chapters 1 and 2
- Available `here`

The road so far

After studying *descriptive statistics* and *probability theory*, it is time **combine** these two areas and produce **informed conclusions** about a problem of interest.

This is the process of **statistical inference**.

Over the next few weeks, we will approach this topic from **both** *Bayesian* and *classical* (*frequentist*) perspectives.

We start with the *Bayesian*.

The Bayesian process

The Bayesian process

Regardless of the approach one adopts for inference, the main interest is to **learn from data**.

This process allows one to make *predictions*, evaluate *hypotheses*, fit statistical *models*, and so on.

Both approaches will also draw conclusions based on **sample data**.

The key distinction lies in the **logic** and **interpretations** derived from each point of view.

The Bayesian process

While for a frequentist a probability means a *long-run relative frequency* of a repeatable event, a Bayesian defines it as a *measure of relative plausibility* of an event.

Moreover, a frequentist assumes that the data **alone** drives any further information.

A Bayesian, on the other hand, uses data **along** with incoming (*prior*) information.

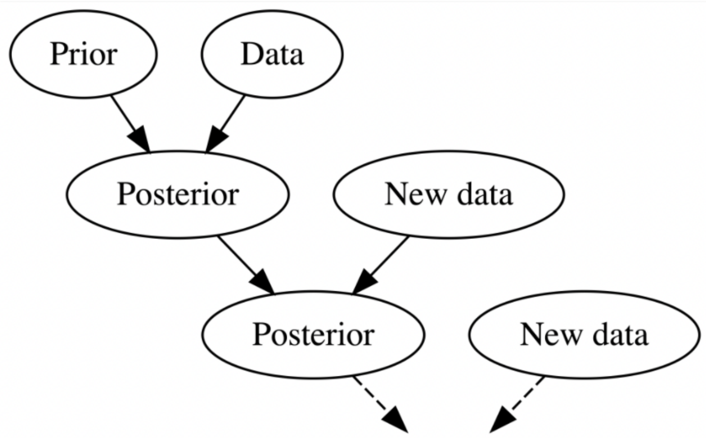
This way, while a Bayesian asks: "*in light of these data, what is the plausibility of my hypothesis being correct?*"; a frequentist asks: "*in case my hypothesis is not correct, what are the odds of having observed such data?*"

The Bayesian process

As we accumulate life experiences (i.e., incorporate new information), we **update** our knowledge about our interests and surroundings.

Thus, it is almost impossible not to use some **previous background** (prior information) when trying to answer a question.

Whether one's previous knowledge will be *overwhelmed* by reality or not is at the crux of the Bayesian inferential process.



The Bayesian process

Before we move on, let us recall some **key** probability concepts that we will use in a moment.

- **Marginal probability**: an *unconditional* probability, describing the behavior of a single variable.
  - $P(A)$
- **Joint probability**: a statement about the *simultaneous* behavior of two (or more) variables.
  - $P(A \text{ and } B)$
- **Conditional probability**: describes how one event is dependent upon another event.
  - $P(A \mid B)$

The Bayesian process

Now, recall the **mathematical definition** of a conditional probability:

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

And since $P(A \cap B) = P(B \cap A)$, we can rewrite the above as

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

... which is **Bayes' theorem**.

The Bayesian process

Now, we can replace events *A* and *B* with other elements. The usual notation is as follows:

$$P(\theta|y) = \frac{P(\theta)\ P(y|\theta)}{P(y)}$$

where θ may be a population *parameter*, a specific *hypothesis*, or whatever the researcher's *interest* may be.

And y is the available *data/information* they have.

$P(\theta \mid y)$ is known as a **posterior distribution**.

The Bayesian process

Each element in Bayes' theorem has a special *name* and *function*.

- $P(\theta)$ is the **prior** belief/information we incorporate into our analysis.
  - Notice that it is a *marginal probability*, thus reflecting what we know about the subject **before seeing any data**.
- $P(y|\theta)$ is the **likelihood** function.
  - It describes the ways in which we can observe the data (y), ***given that our hypothesis*** (θ) ***is true***. It quantifies the extent to which the evidence supports our proposition.
- $P(y)$ is the **marginal likelihood** (*aka denominator*, *normalizing constant*, *probability of the data*).
  - It can be interpreted as the probability of observing the data **under *all* possible scenarios**.

The Bayesian process

Since the denominator in Bayes' theorem is simply a *normalizing constant*, we can **rewrite** it as

$$P(\theta|y) \propto P(\theta)\ P(y|\theta)$$

where the $\propto$ symbol means "proportional to."

In other words, the normalizing constant's function is to *ensure* that the posterior is a *proper probability distribution*.

- i.e., sums/integrates to 1.

This is especially *convenient* since, for complex situations, properly calculating the denominator may be *computationally expensive*.

Bayes' theorem in practice

Bayes' theorem in practice

Suppose we are interested in the underlying *proportion* of Bayesians *vs.* frequentists in the Social Science fields.

We then gather some data from surveys, asking people whether they would define themselves as *frequentists* or *Bayesians*.

In terms of Bayes' theorem, we have

$$P(\theta|y) = \frac{P(\theta)\ P(y|\theta)}{P(y)}$$

where θ is the entire population proportion of Bayesians (B) to frequentists (F), and y is the data from the survey.

Bayes' theorem in practice

Let us understand this setting **logical** and **graphically**.

Assume we interview **4** people, and you do not know how many are B and how many are F.

What are the **possible outcomes** from this experiment?

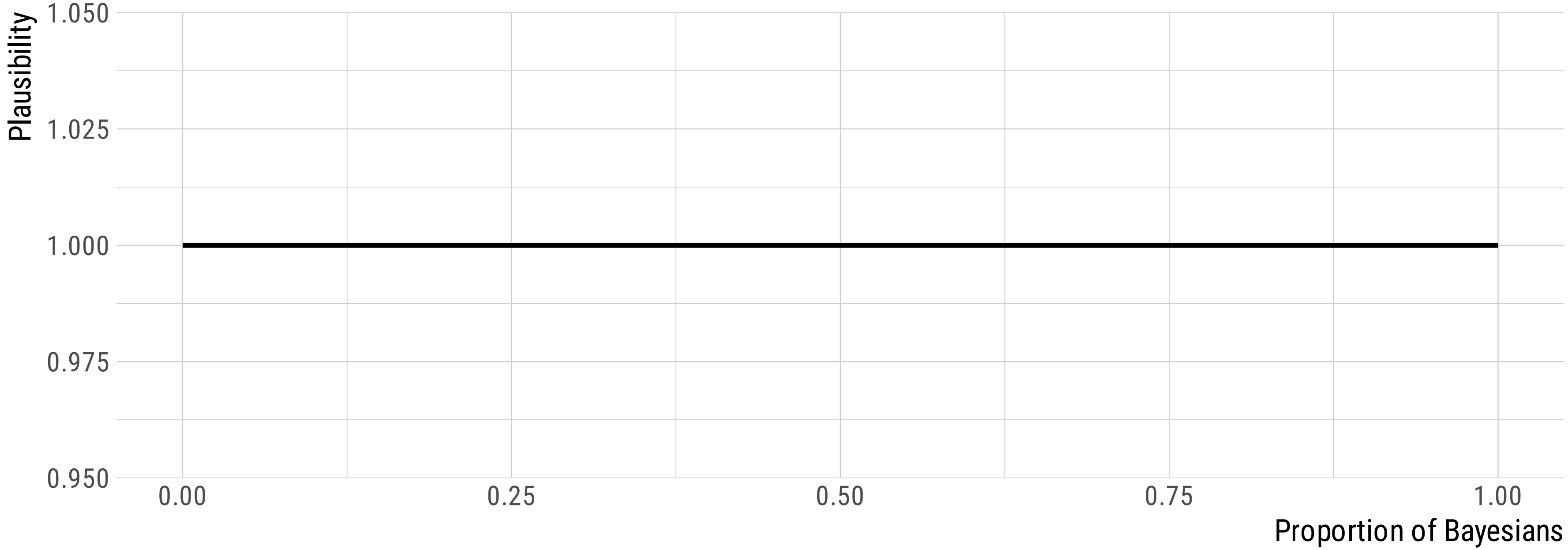
After we know the ways in which we can observe the outcomes, how do we use this information for **inference** about the proportion of interest?

Bayes' theorem in practice

For simplicity, let us assume that, before seeing any data, any proportion is **equally likely**.

What does this imply?

$$\theta \sim \mathrm{Unif}(0, 1)$$

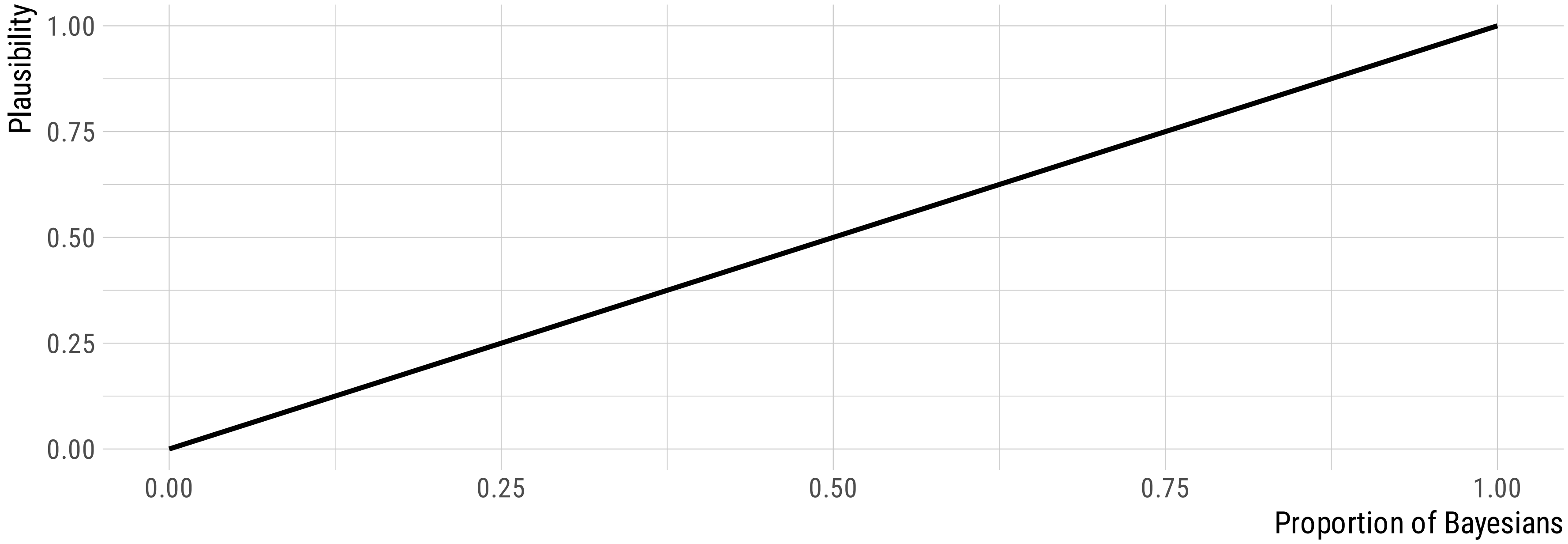


Bayes' theorem in practice

How do we define $P(y|\theta)$?

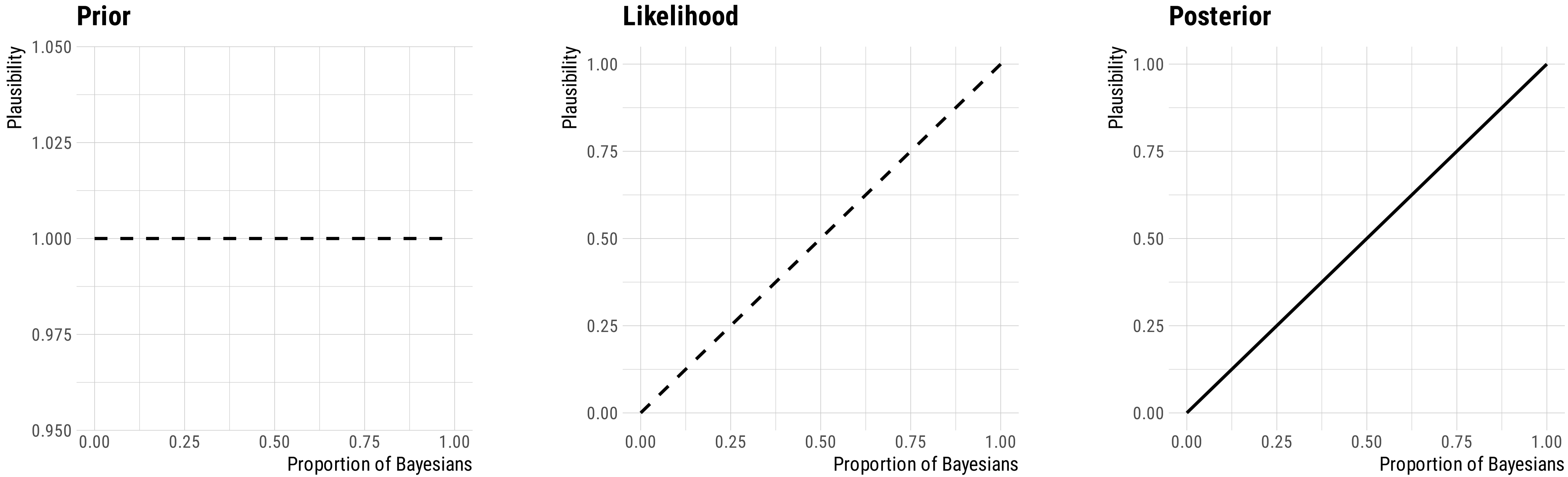
The survey follows a **binomial experiment**.

Now, assume we ask the first person, and she answers "Bayesian."



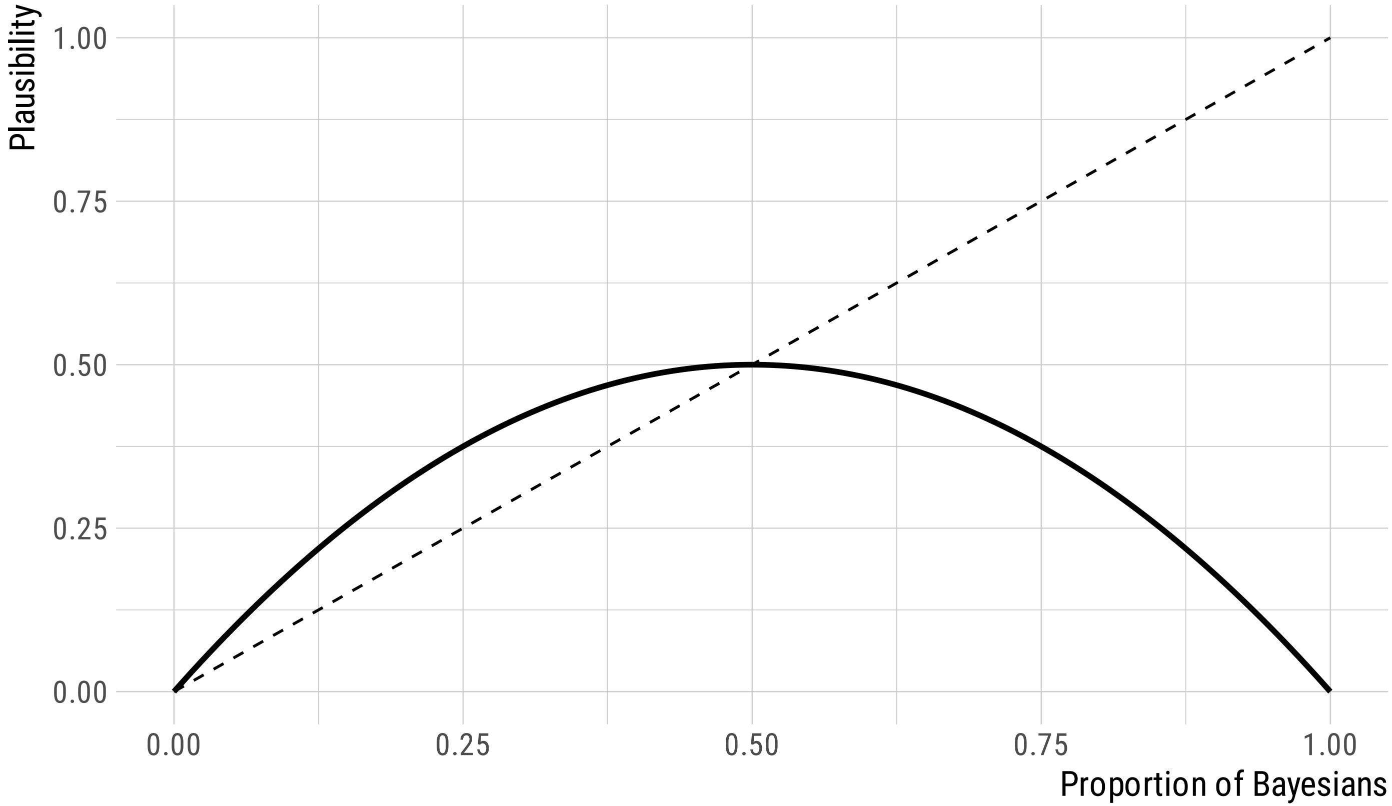
Bayes' theorem in practice

Combining prior and likelihood...



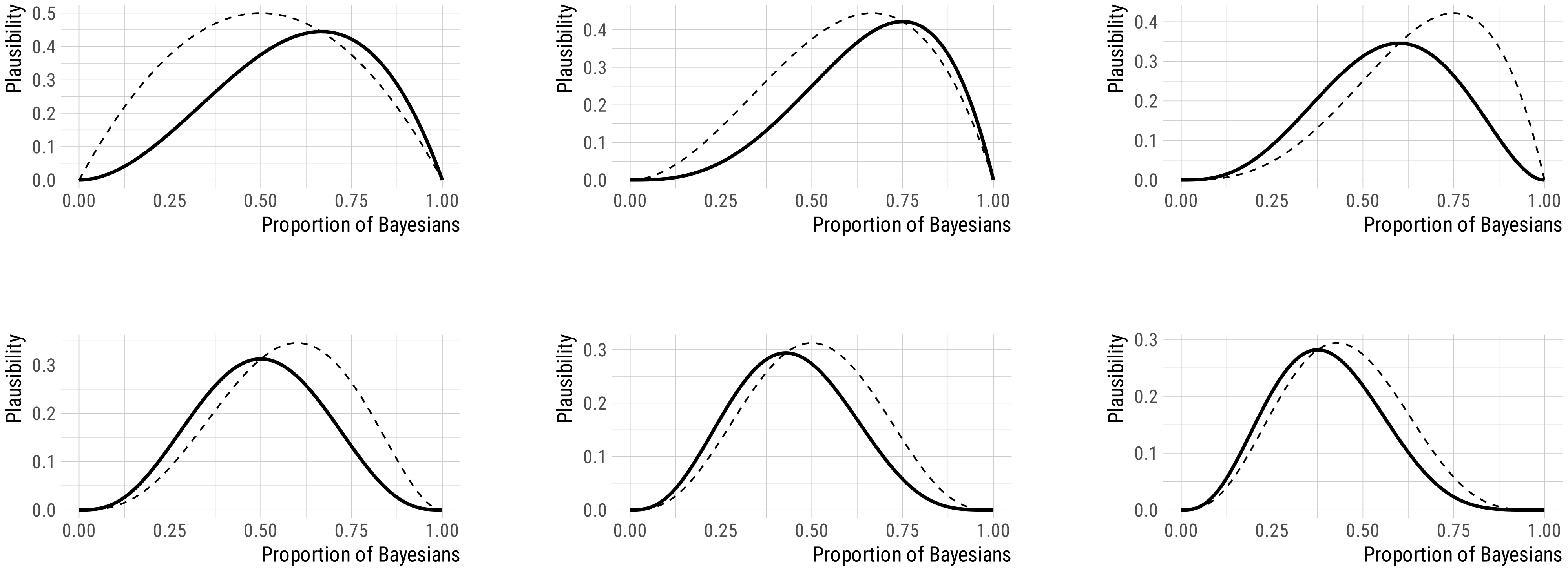
Bayes' theorem in practice

Now, we ask a second person. She answers "Frequentist."



Bayes' theorem in practice

And off we go...



Bayes' theorem in practice

What Bayes's theorem (and Bayesian inference) does is **updating** previous plausibilities in light of new data, producing a *new set* of plausibilities (the posterior distribution).

Next time: More examples!