Bayesian Inference

Ryan Henning
Frank Burkholder

galvanize

1. Frequentists vs. Bayesian
2. Bayes' Rule
3. Prior, likelihood, posterior distributions

What is the probability that it rained in my city last night?

(No info is given about which city I'm currently in.)

$$P(\text{rain}) = 0.1$$

What is the probability that it rained in my city last night given that I'm in Seattle?

$$P(\text{rain}|\text{Seattle}) = 0.65$$

What is the probability that it rained in my city last night?

(No info is given about which city I'm currently in.)

$$P(\text{rain}) = 0.1$$

What is the probability that it rained in my city last night <u>given</u> that I live in Seattle <u>and</u> I see that the road is wet?

$$P(\text{rain}|\text{Seattle, wet roads}) = 0.97$$

Defining probability

<u>Frequentist Probability</u>

“Long Run” frequency of an outcome

<u>Subjective Probability</u>

A measure of degree of belief

Bayesians consider both types

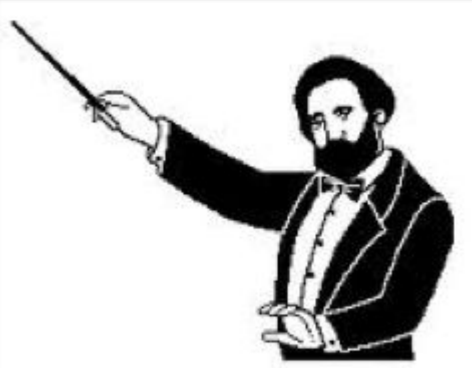
Experiment 1:

A fine classical musician says he’s able to distinguish Haydn from Mozart.
Small excerpts are selected at random and played for the musician.
Musician makes 10 correct guesses in exactly 10 trials.

Experiment 2:

Drunken man says he is psychic and can correctly guess what face of the coin will fall down, mid-air. Coins are tossed and the drunken man shouts out guesses while the coins are mid-air. Drunken man correctly guesses the outcomes of the 10 throws.

Adapted example from Jim Berger’s book, <u>The Likelihood Principle</u>. Also adapted from Tammy Lee’s slides.



Frequentist: “They’re both so skilled! I have **as much confidence** in musician’s ability to distinguish Haydn and Mozart as I do the drunk’s psychic ability to predict coin tosses”

Bayesian: “I’m not convinced by the drunken man…”

The Bayesian approach is to incorporate prior knowledge into the experimental results.

Bayes' Rule

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

Bayes' Rule

posterior probability

likelihood

prior probability

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

marginal likelihood

Bayes' Rule: Example

$$P(\text{psychic}|\text{correct}) = \frac{P(\text{correct}|\text{psychic})P(\text{psychic})}{P(\text{correct})}$$

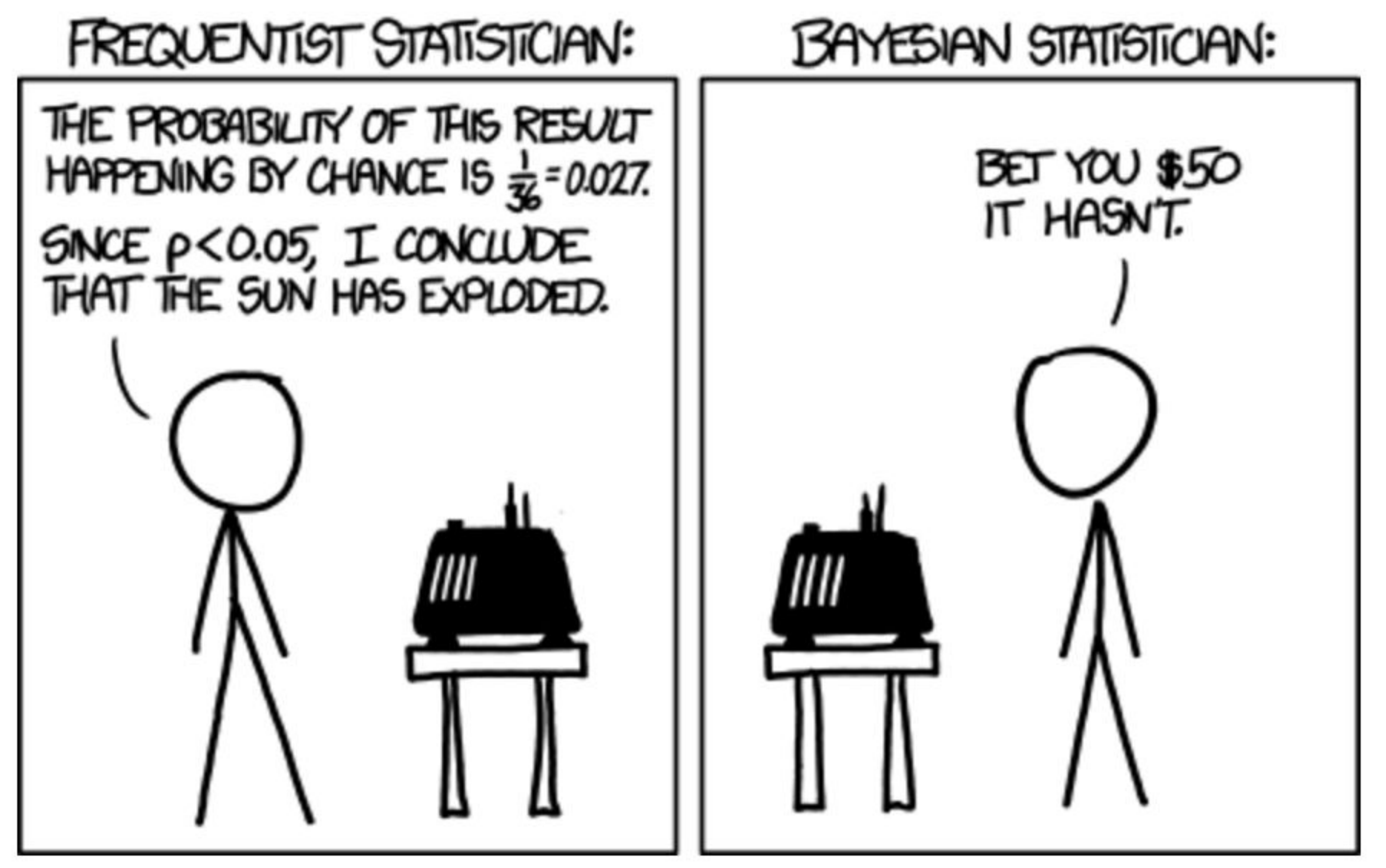
$$= \frac{1.0 * 0.0001}{0.5^{10}}$$

Very subjective!

$$= 10.2\%$$

xkcd: Frequentists vs. Bayesians (#1132)

xkcd: Frequentists vs. Bayesians (#1132)



Is this coin fair?
(i.e. is the probability of Heads = 0.5?)

Bayesian Updates

A: the distribution associated with the probability of flipping heads (success)

B: the results of our flips (number of heads, number of trials)

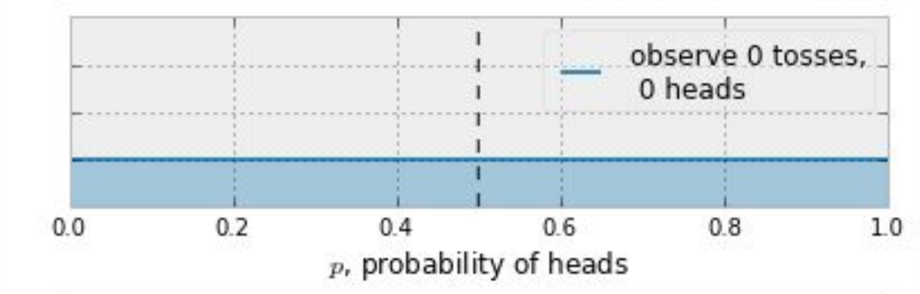
likelihood (the data gathered, #H #T)

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

prior probability (our belief about the probability of heads, initially, but after collecting data it's the old posterior)

posterior probability (the probability of heads given the flips)

the marginal probability of the flips (ensures that total probability of posterior sums to 1)



No data

Uniform prior -> Uniform posterior

A: the distribution associated with the probability of flipping heads (successes)

B: the results of our flips (number of heads, number of trials

likelihood (the data gathered, #H #T)

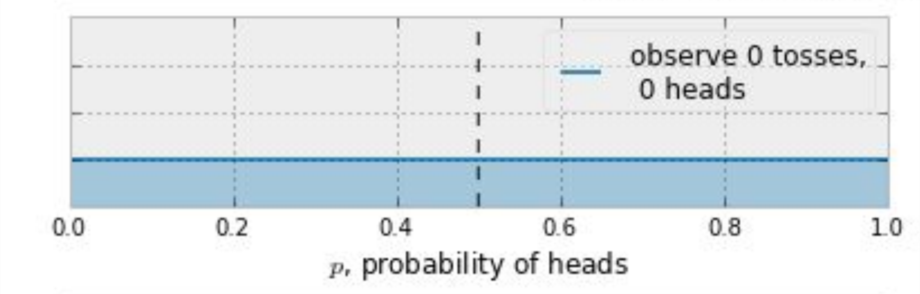
$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

prior probability (our belief about the probability of heads, initially, but after collecting data it's the old posterior)

posterior probability (the probability of heads given the flips)

the marginal probability of the flips (ensures that total probability of posterior sums to 1)

Bayesian Updates



No data

Uniform prior -> Uniform posterior
Note that they are both continuous.

A: the distribution associated with the probability of flipping heads (successes)

B: the results of our flips (number of heads, number of trials

likelihood (the data gathered, #H #T)

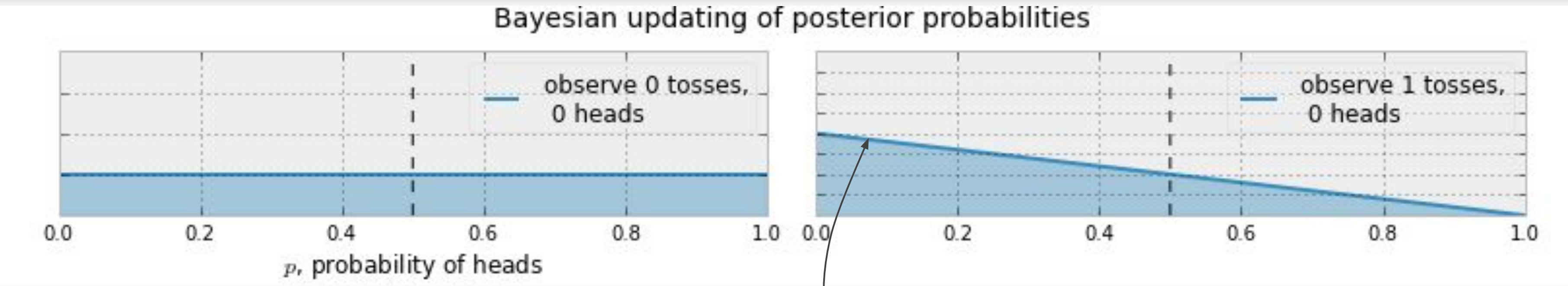
$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

prior probability (our belief about the probability of heads, initially, but after collecting data it's the old posterior)

posterior probability (the probability of heads given the flips)

the marginal probability of the flips (ensures that total probability of posterior sums to 1)

Bayesian Updates



Likelihood: 0 successes, 1 trial (Binomial)
Posterior = Likelihood * Prior / Marginal

Posterior becomes the *new prior* for the next posterior calculation

A: the distribution associated with the probability of flipping heads (successes)

B: the results of our flips (number of heads, number of trials)

likelihood (the data gathered, #H #T)

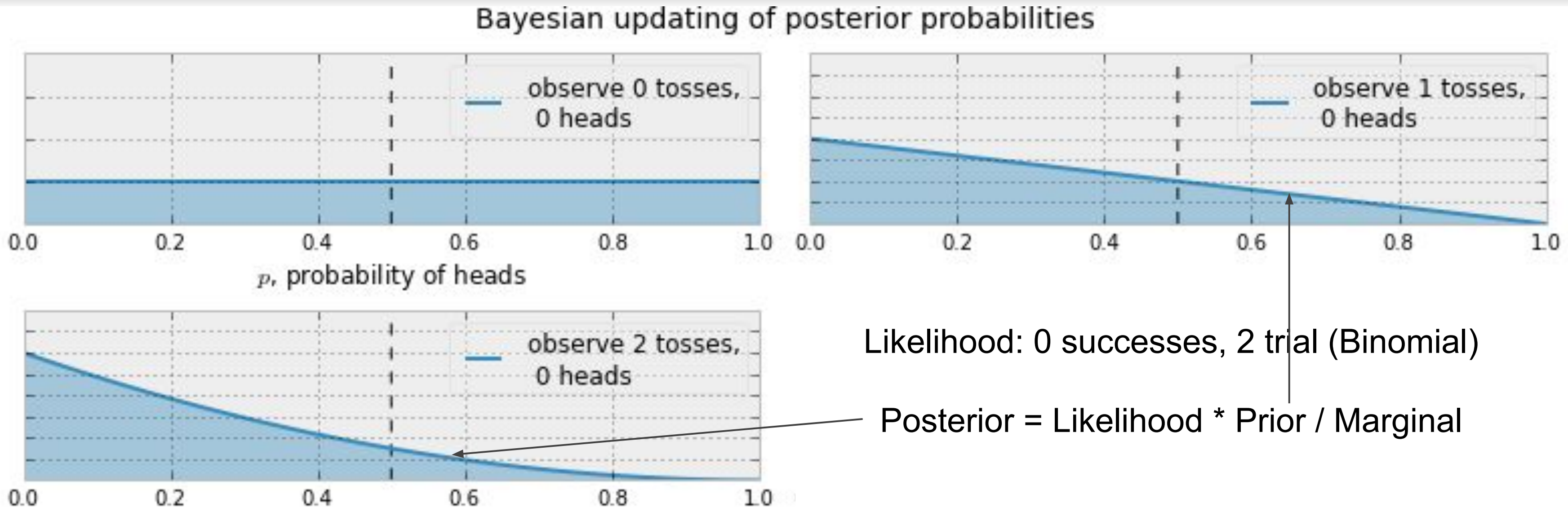
prior probability (our belief about the probability of heads, initially, but after collecting data it's the old posterior)

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

posterior probability (the probability of heads given the flips)

the marginal probability of the flips (ensures that total probability of posterior sums to 1)

Bayesian Updates



Likelihood: 0 successes, 2 trial (Binomial)

Posterior = Likelihood * Prior / Marginal

A: the distribution associated with the probability of flipping heads (successes)

B: the results of our flips (number of heads, number of trials)

likelihood (the data gathered, #H #T)

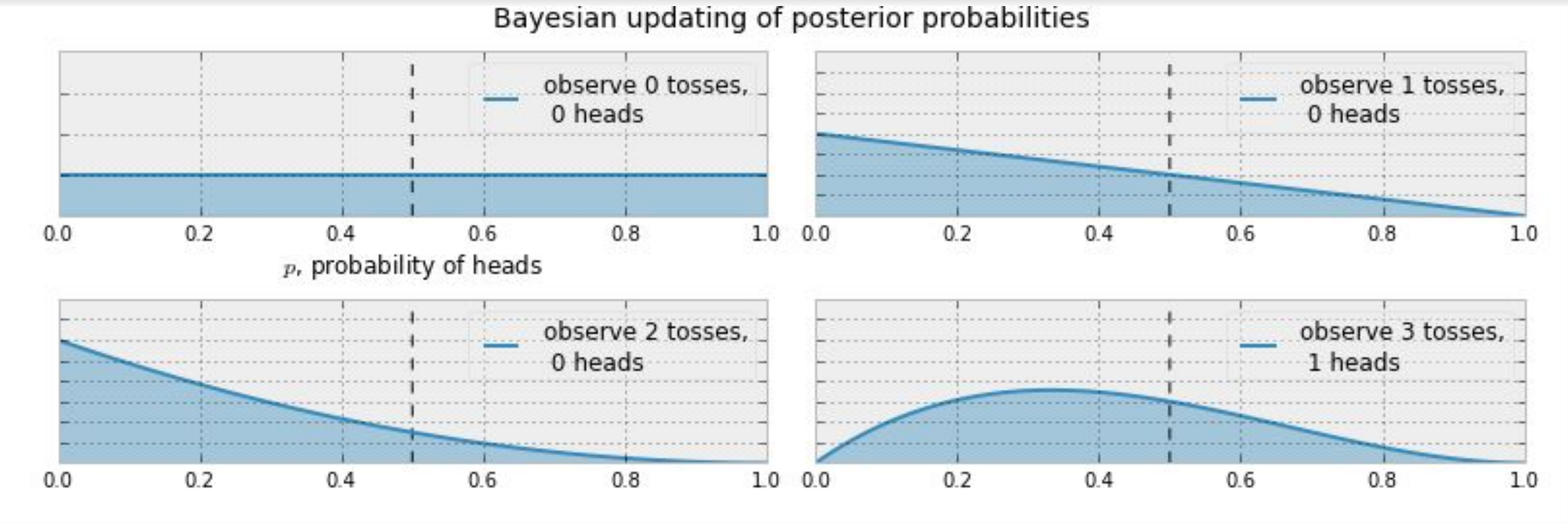
$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

prior probability (our belief about the probability of heads, initially, but after collecting data it's the old posterior)

posterior probability (the probability of heads given the flips)

the marginal probability of the flips (ensures that total probability of posterior sums to 1)

Bayesian Updates



A: the distribution associated with the probability of flipping heads (successes)

B: the results of our flips (number of heads, number of trials)

likelihood (the data gathered, #H #T)

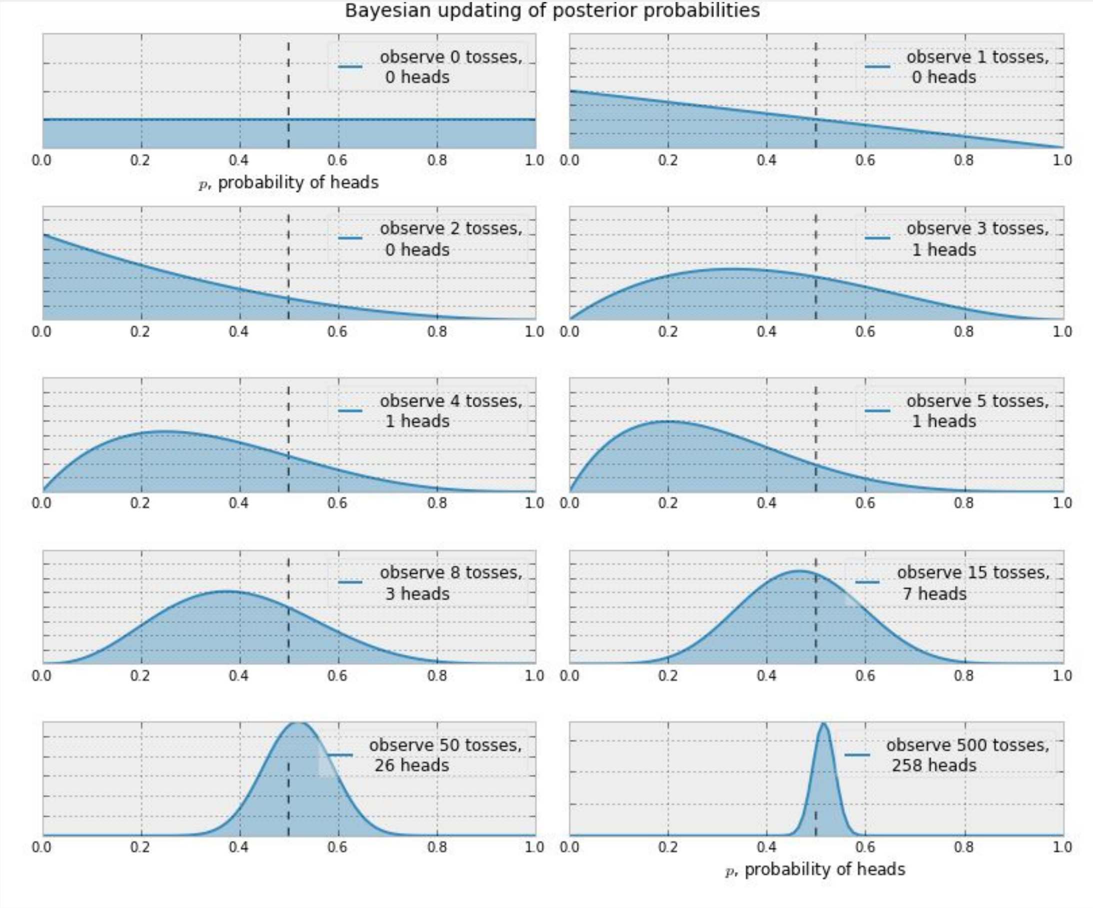
prior probability (our belief about the probability of heads, initially, but after collecting data it's the old posterior)

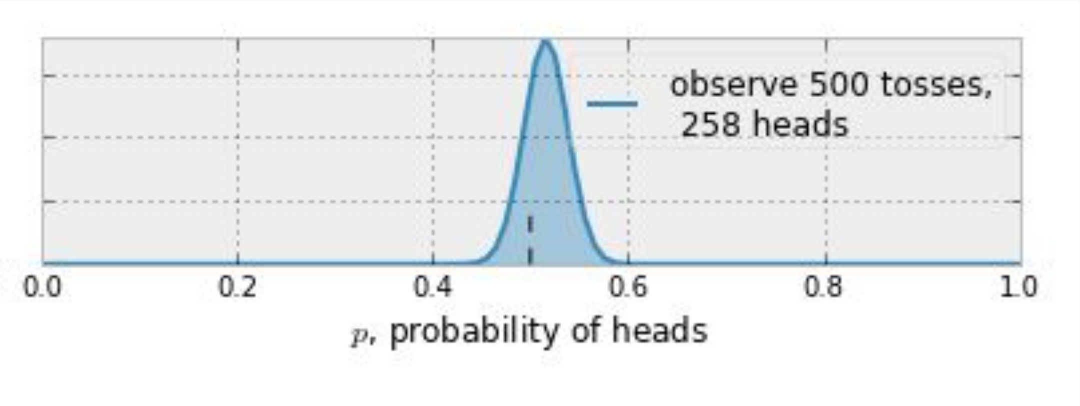
$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

posterior probability (the probability of heads given the flips)

the marginal probability of the flips (ensures that total probability of posterior sums to 1)

Bayesian Updates




observe 500 tosses,
258 heads
0.0
0.2
0.4
0.6
0.8
1.0
p, probability of heads

Bayesian Updating in Code

bayesian_updating_simple_example.ipynb