Bayesian Inference

Darren Reger Lecture for Galvanize DSI

Frequentist v. Bayesian

Adapted example from Jim Berger's book, the Likelihood Principle

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 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.

Frequentist v. Bayesian

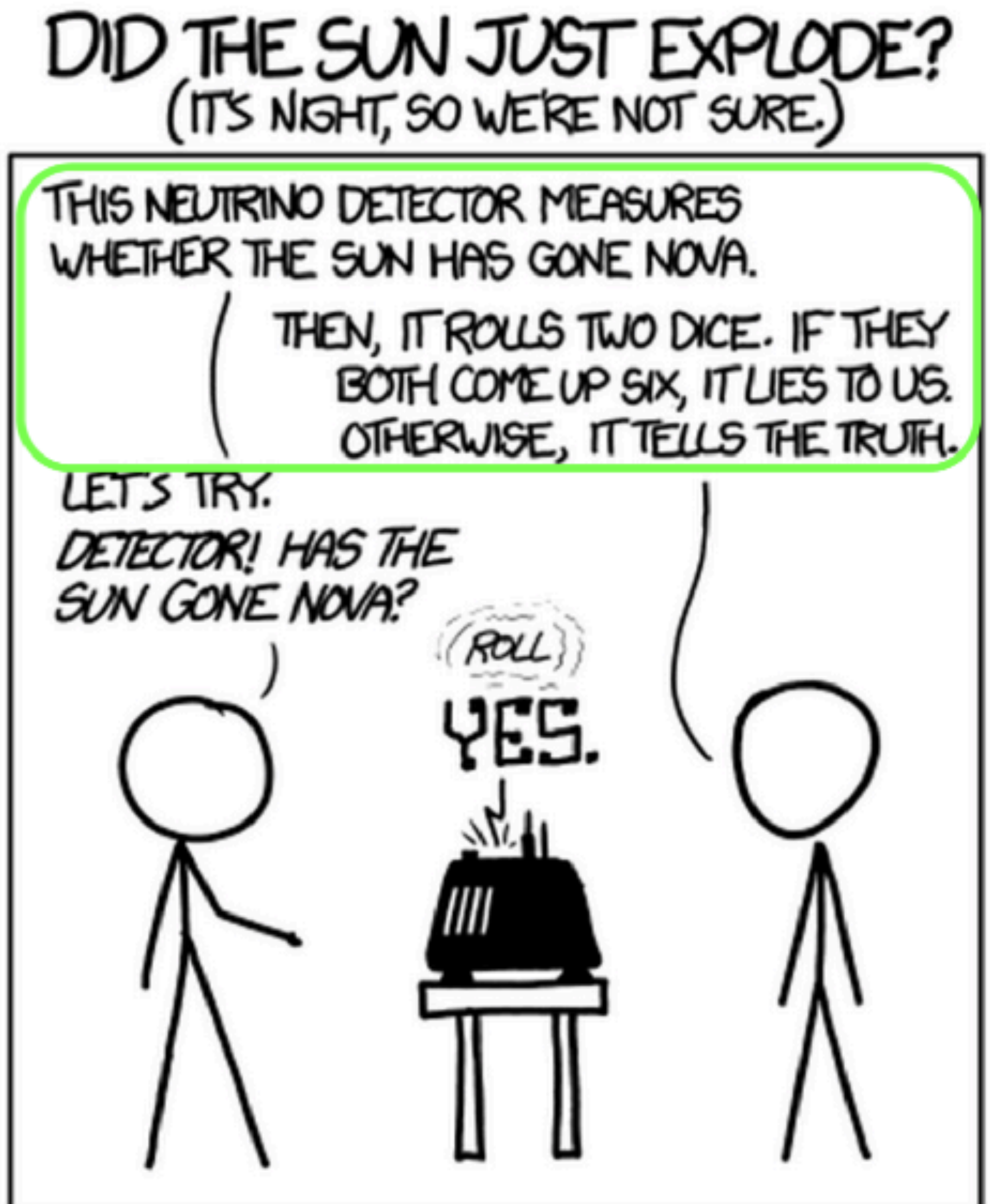
Frequentist: "They're both so skilled! I have **<u>as much confidence</u>** in musician's ability to distinguish Haydn and Mozart as I do the drunk's to predict coin tosses"

Bayesian: "I don't know man…"

- A Bayesian would incorporate some prior confidence about the musician's ability and the drunk's.

Frequentist v. Bayesian

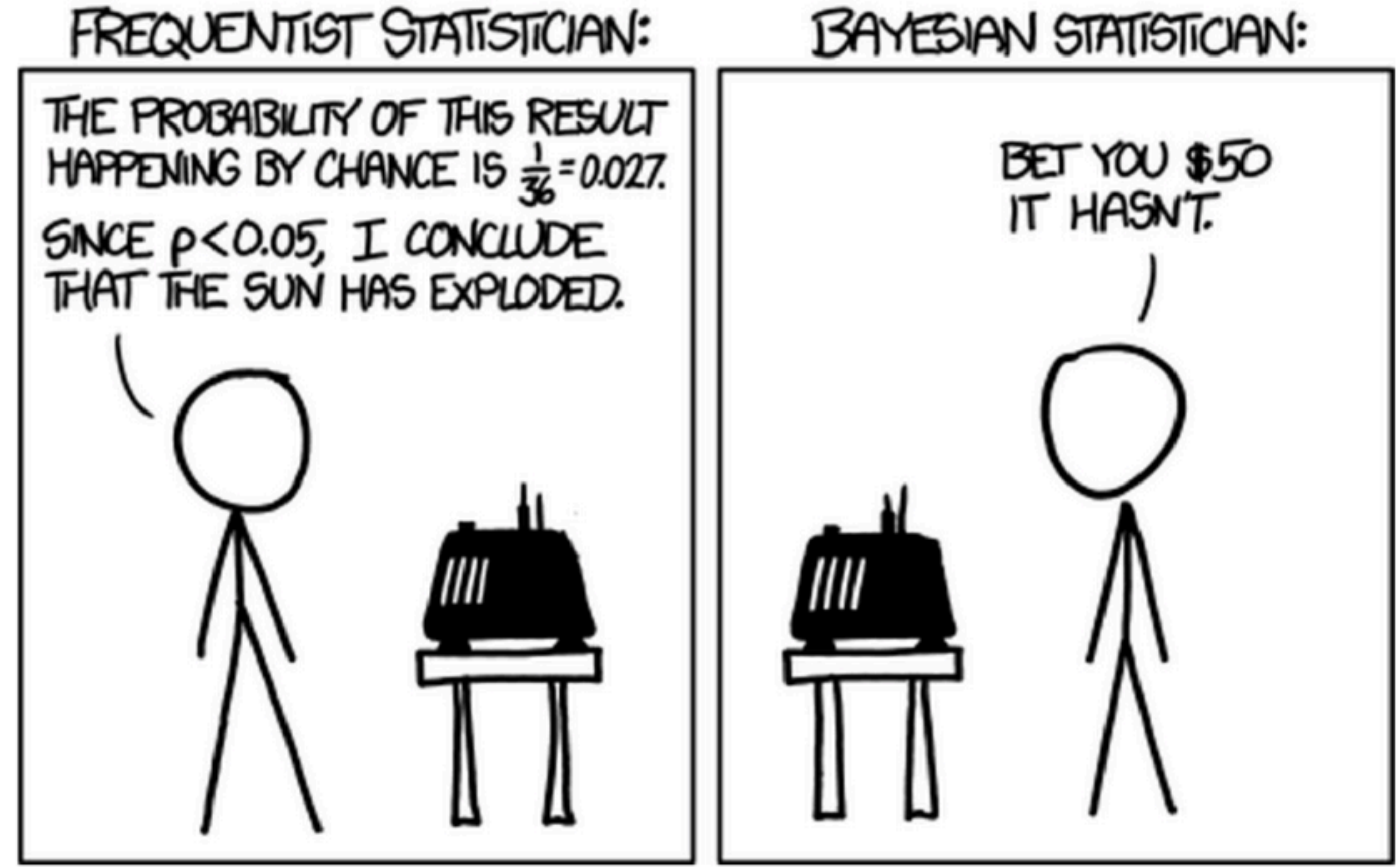
H_0: Sun has not exploded...



Evidence collecting process

Evidence!

Frequentist v. Bayesian



Frequentist v. Bayesian

Frequentist Probability
Long-run frequency of an outcome

+

Subjective Probability
A measure of degree of belief

=

Bayesian Probability
Consider both subjective belief and long-run frequency

Bayesian Inference

- So we have our **beliefs**
- And we have our **evidence**
  - A Frequentist considers the **evidence**
  - A Bayesian considers **both**
    - A Bayesian isn't forced to keep beliefs the same forever; she can <u>**update**</u> **her** <u>**beliefs**</u> **based on** <u>**evidence**</u>

Bayesian vs. Frequentist

Boils Down to What is Fixed?

- Frequentist:
  - Data are a repeatable random sample $\longrightarrow$ there is a **frequency**
  - Underlying parameters remain constant during this repeatable process
  - $\longrightarrow$ **Parameters are fixed**
- Bayesian:
  - Data are observed from a realized sample
  - Parameters are unknown and described probabilistically
  - $\longrightarrow$ **Data are fixed**

Bayesian vs. Frequentist

General Inference

- Frequentist:
  - Point estimates and standard errors
  - Deduction from $P(data|H_0)$, by setting α in advance
  - P-value determines acceptance of H_0 or H_1
- Bayesian:
  - Start with a prior $\pi(\theta)$ and calculate the posterior $\pi(\theta|data)$
  - Broad descriptions of the posterior distribution such as means and quantiles

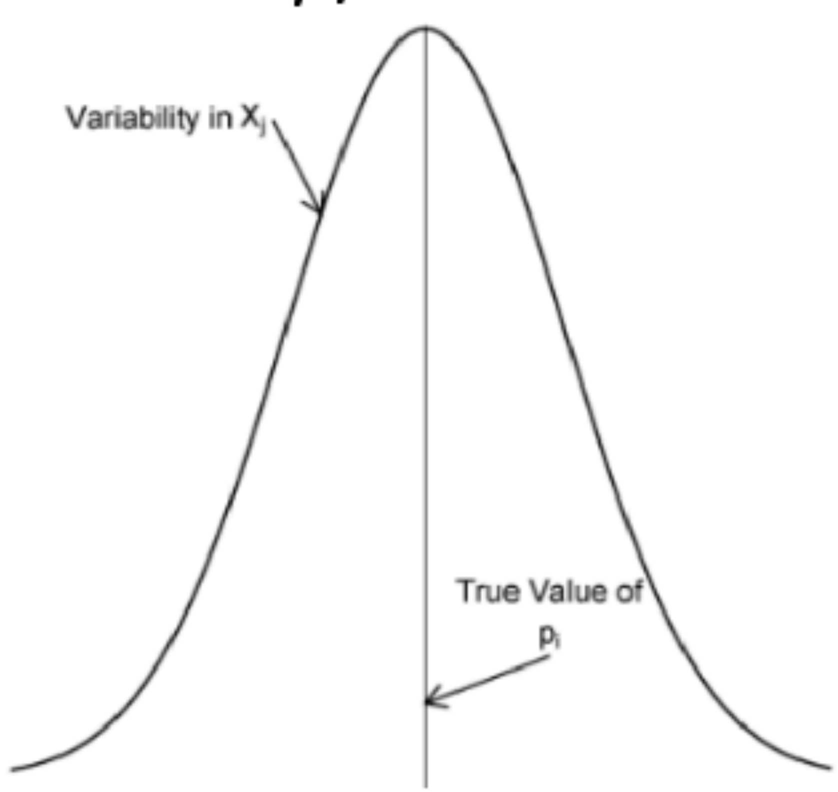
Frequentist: $P(data|H_0)$ is the sampling distribution given fixed parameter

Bayesian: $P(\theta)$ is the prior distribution of the parameter (before the data are seen), $P(\theta|data)$ is the posterior distribution of the parameter after seeing the data.

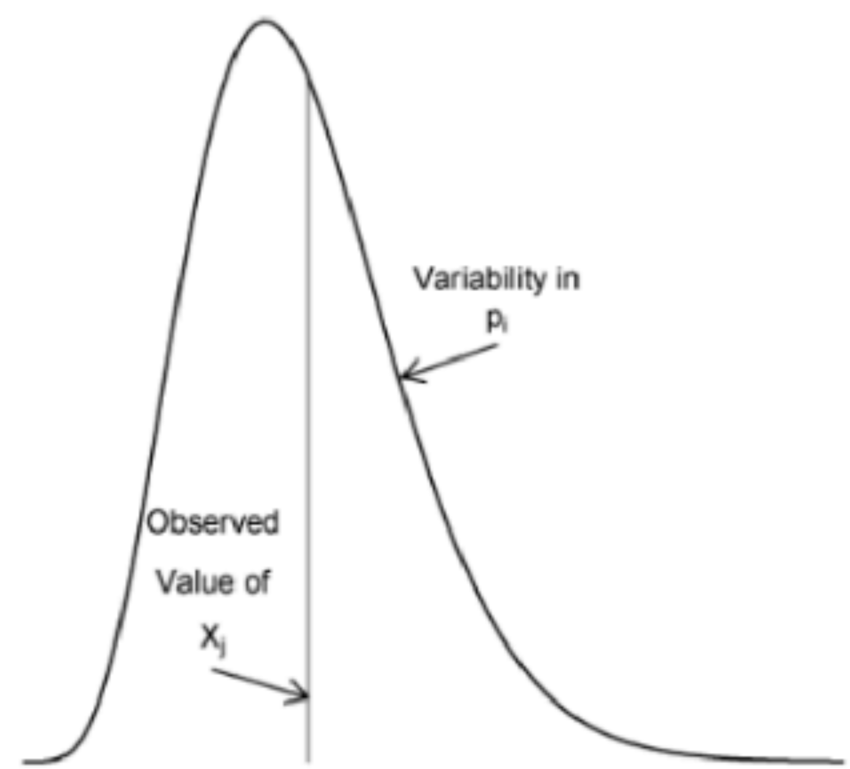
Bayesian vs. Frequentist

General Inference

Frequentist: Describe the variability in X_j for a fixed value of p_i

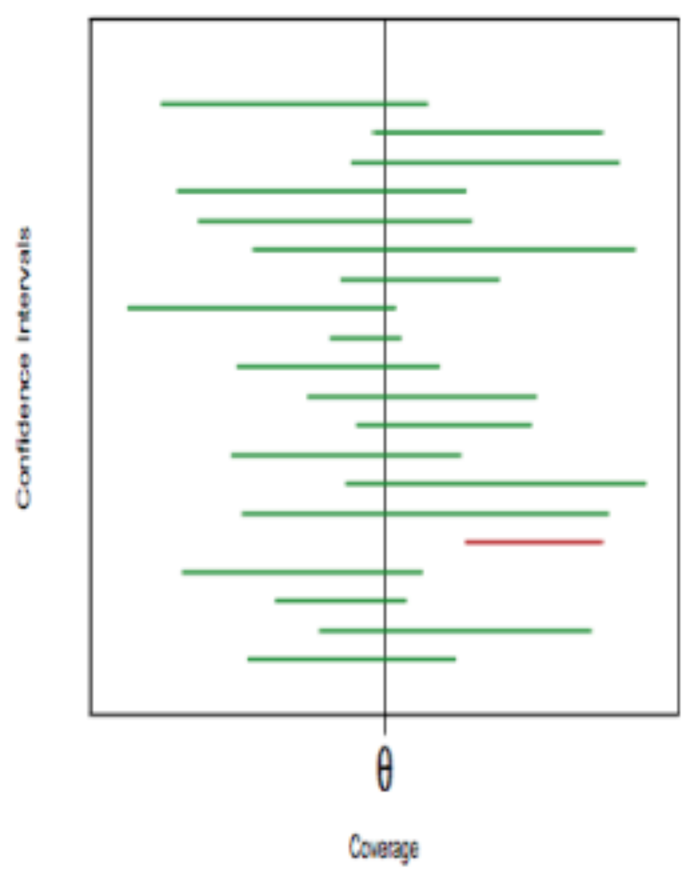


Bayesian: Describe the variability in p_i for fixed X_j

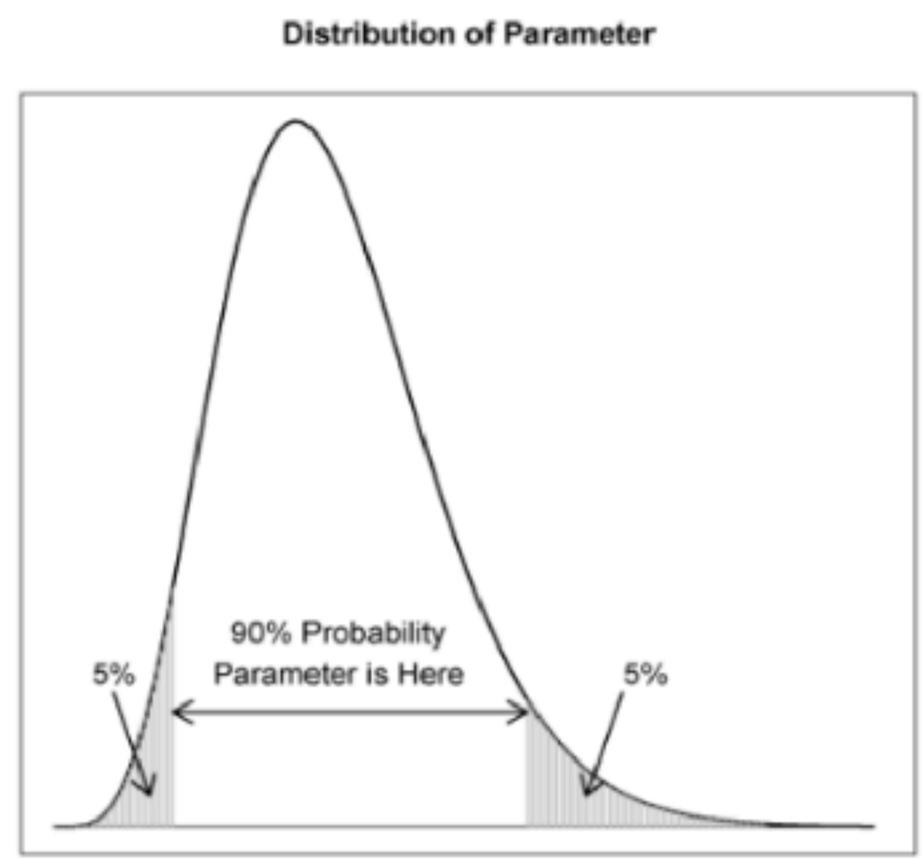


Intervals

Frequentist: A collection of intervals, 90% of them contain the true unknown parameter



Bayesian: An interval has a 90% chance of containing the true unknown parameter



Steps for Bayesian Inference

- Specify a probability model for unknown parameter values that includes some prior knowledge about the parameters if available
- Update knowledge about the unknown parameters by conditioning this probability model on observed data
- Evaluate the fit of the model to the data and the sensitivity of the conclusions to the assumptions

Bayesian Inference

Bayes Rule:

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

Posterior Distribution:

$$\pi(\theta|\boldsymbol{x}) = \frac{f(\boldsymbol{x}|\theta)\pi(\theta)}{\int_{\theta\in\Theta} f(\boldsymbol{x}|\theta)\pi(\theta)d\theta}$$

- prior $\pi(\theta)$ describes our current knowledge about θ
- likelihood $f(\boldsymbol{x}|\theta)$ is the distribution of the data for a given θ
- posterior $\pi(\theta|\boldsymbol{x})$ is our updated knowledge about θ after seeing the data

How do we get a prior?

- Previous studies, published research
- Researcher's intuition
- Expert opinion
- Can also use non-informative prior

$$P(A|B) = \frac{\overbrace{\boxed{P(A)}}^{\text{Weighted Average}} * \overbrace{\boxed{P(B|A)}}^{\text{Assumption}}}{\underbrace{\boxed{P(B)}}_{\text{Normalized}}}$$

To solve all of these problems we will follow these steps

1. Determine what we want the probability of, and what we are observing
2. Estimate initial probabilities for all of the possible answers
3. For each of the initial possible answers, assume it is true and calculate the probability of getting our observation with that possibility being true
4. Multiply the initial probabilities (Step 2) by the probabilities based on our observation (Step 3) for each of the initial possible answers
5. Normalize the results (divide the each probability by the sum of the total probabilities so that the new total probability is 1)
6. Repeat Steps 2-5 over and over for each new observation

Boardwork Examples:
1) Good barbers
2) Marbles in a bag interview question

Coin Flipping Priors