Skimpy Introduction to Bayesian Analysis

Thomas Bayes

- 1702–1761

Frequentist (Traditional) Statistics

- Most popular approach in the behavioral sciences
  - Probability is defined in terms of the long run frequency of a fixed event
    - E.g., we want to know if a coin is biased or not
      - We flip it 1000 times and record how many times we obtain a head and how many times we obtain a tail
      - If the probability of observing a head/tail is near .5 then we infer that the model is correct (coin is not biased)

Frequentist (Traditional) Statistics

- Let's say we want to know if males and females differ on depression
  - We assume there is a fixed difference (parameter) that we are searching for (i.e., $\mu_M - \mu_F$), and each time we make an observation (i.e., sample males and females and record depression scores, $M_M - M_F$) we can make a statement regarding probability
    - i.e., compute the *p*-value from the sampling distribution of the statistic

Frequentist (Traditional) Statistics

- An important thing to keep in mind with the frequentist approach is that the probability we are calculating relates to the probability *of our data* given some fixed (but unknown) parameter in the world
  - E.g., the probability of observing a difference in depression scores of 5 points, if the difference in the population is actually 0

Bayesian Statistics

- Subjectivist View
  - Combines subjective probabilities regarding an event (priors) with observations regarding the event (data)

  - Probability of an Event =

    Prior Probability + Observed Probability

Bayesian Statistics

- For example, let's say we want to know if males and females differ on depression
  - We start with any knowledge we have regarding the difference (prior), and combine that with observations regarding the difference (data)
  - We weight our final "answer" regarding the difference on the relative precision of our prior information and our data
    - How "informative" or "precise" is the prior?
    - Is the data based on 5 observations or 5000 observations?

Components of Bayesian Statistics

- Prior Information
  - Background knowledge on the parameters of the model being tested
    - Expressed as a distribution
  - E.g., let's say we are looking at the difference in depression scores between males and females, and we know from past studies that females tend to score about 6 points higher than males, with a sd of about 2
    - Our prior distribution for the difference could be a normal distribution with a mean of 6 and sd of 2
      - If we had no information we might use a flat (e.g., uniform) distribution (noninformative prior)

Components of Bayesian Statistics

- Data/New Information
  - What have we learned from our observations?
    - E.g., mean and variance in a one sample problem
    - Expressed as a likelihood
      - What the most likely values are for the unknown parameters, given the data

- Posterior
  - Prior knowledge is updated by the current data to yield updated knowledge in the form of the posterior distribution
  - Posterior = prior + data

Bayes Theorem

This is the prior: i.e. what you believed before you saw the evidence.

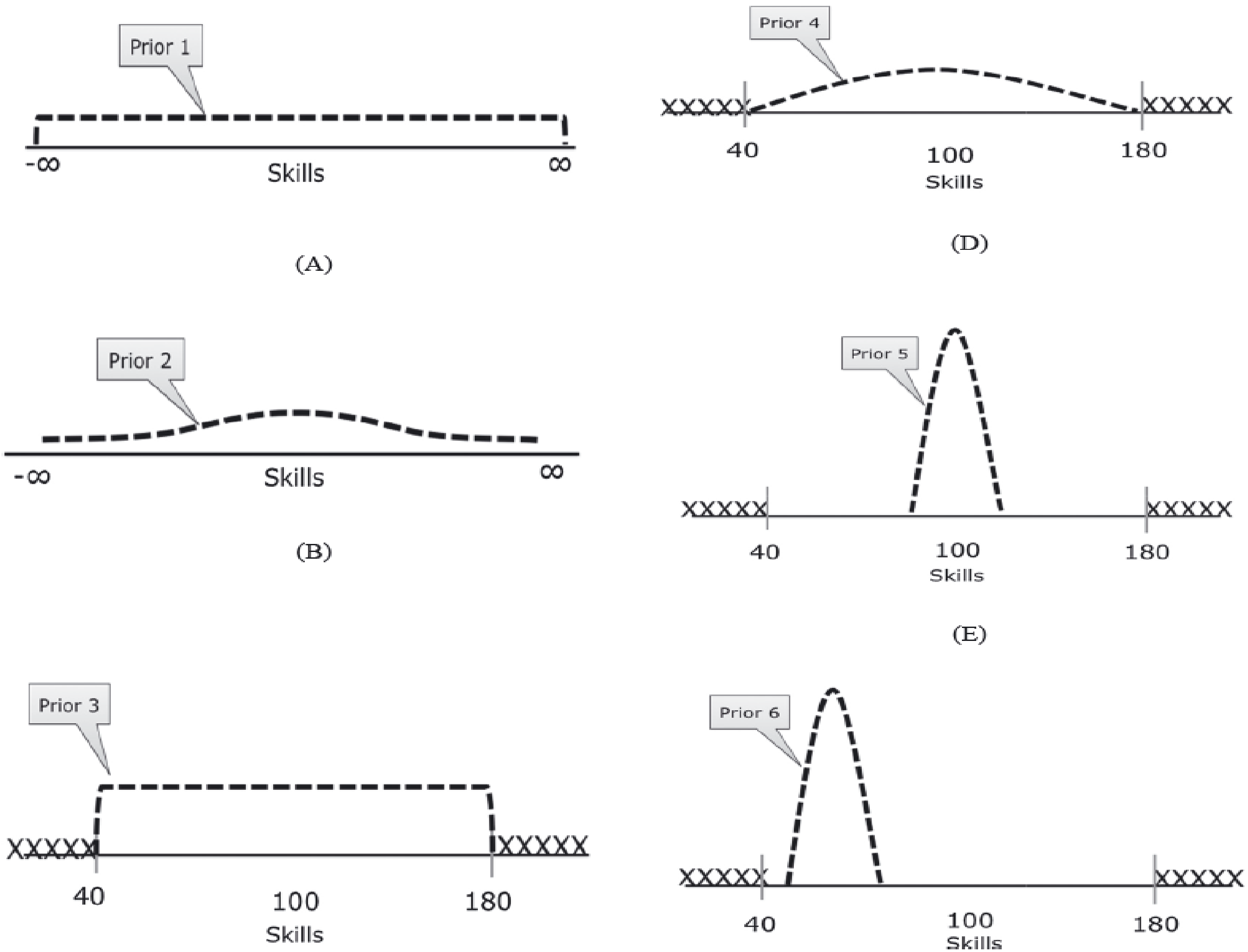
This is the likelihood of seeing that evidence if your hypothesis is correct.

This is the posterior

$$p(H \mid D) = \frac{p(H)\,p(D \mid H)}{p(D)}$$

This is the normalizing constant:
i.e. The likelihood of that evidence under any circumstances.

Examples of Priors



Posterior = Prior + Data

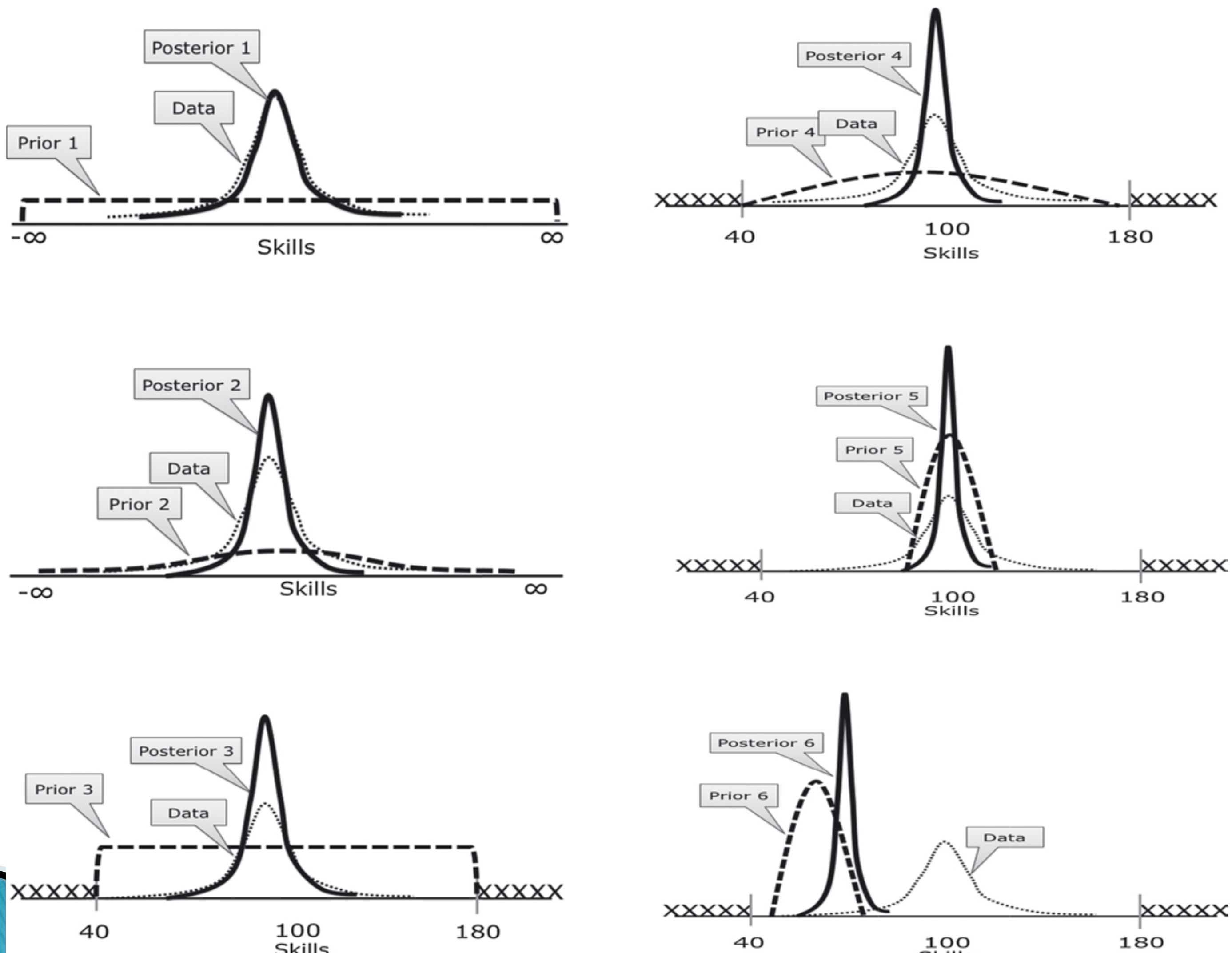


Table 1

Overview of the Similarities and Differences Between Frequentist and Bayesian Statistics

	Frequentist statistics	Bayesian statistics
Definition of the *p* value	The probability of observing the same or more extreme data assuming that the null hypothesis is true in the population	The probability of the (null) hypothesis
Large samples needed?	Usually, when normal theory-based methods are used	Not necessarily
Inclusion of prior knowledge possible?	No	Yes
Nature of the parameters in the model	Unknown but fixed	Unknown and therefore random
Population parameter	One true value	A distribution of values reflecting uncertainty
Uncertainty is defined by	The sampling distribution based on the idea of infinite repeated sampling	Probability distribution for the population parameter
Estimated intervals	Confidence interval: Over an infinity of samples taken from the population, 95% of these contain the true population value	Credibility interval: A 95% probability that the population value is within the limits of the interval

Ads and Disads of the Frequentist Approach

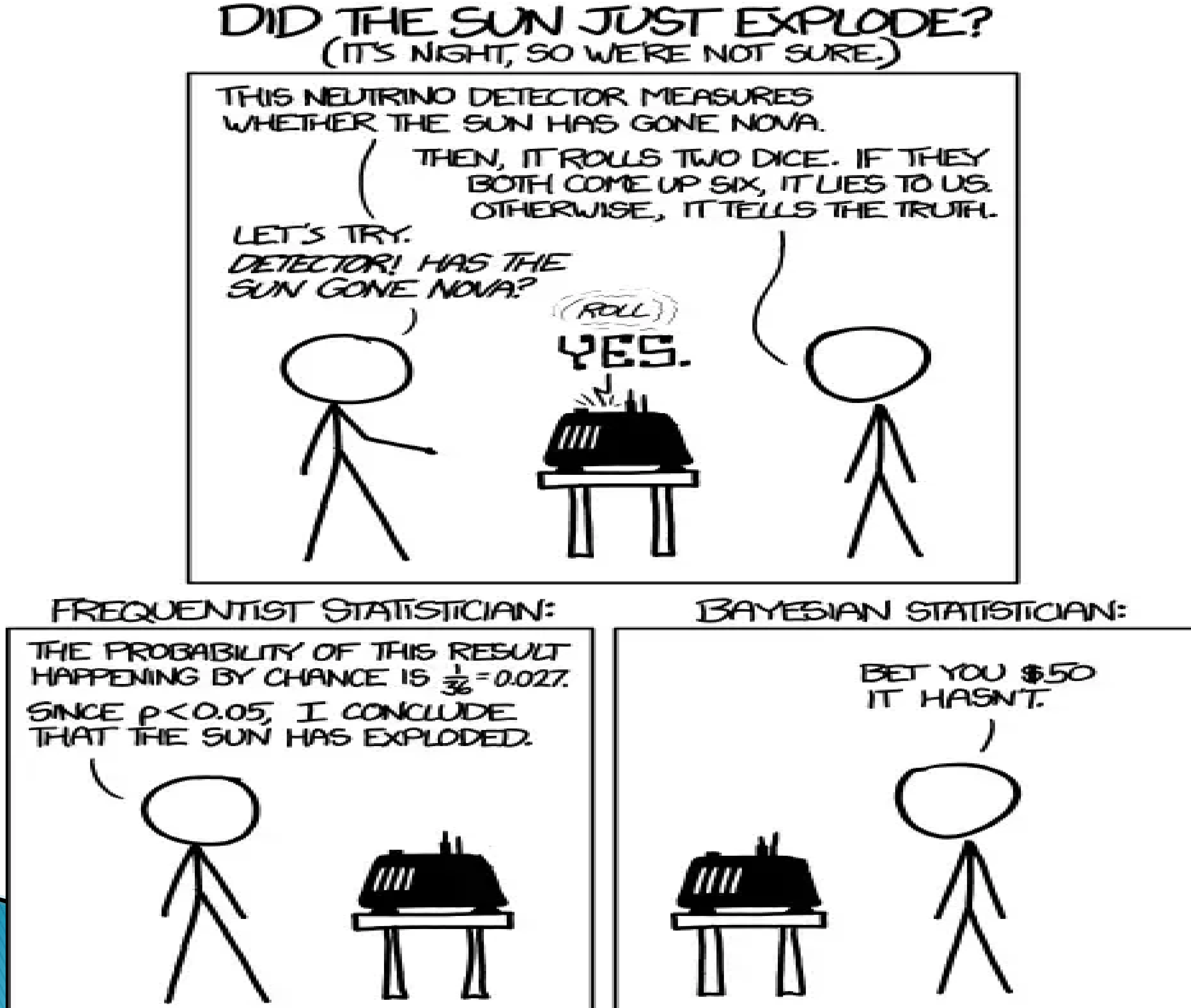
- Advantage of the Frequentist View
  - Objective and Unambiguous
    - Based on observed events and everyone looking at the data will come to the same conclusion
    - Good for checking models (e.g., CIs)
      - E.g., If we simulate data with given properties, will 95% of the CIs really include the population parameter?
- Disadvantage of the Frequentist View
  - Infinite sampling is not always realistic in the universe
    - E.g., What is the probability of there being life on Mars 100 billion years ago?
      - Does it make sense to make a statement like "if we studied life on Mars 100 billion years ago many times "?

Advantages of Bayesian Statistics

- Probabilities can be assigned to any event, not just events that are repeatable
- Prior information about a parameter can be combined with new information to provide "updated information"
- Inferences are conditional upon the data, not conditional upon the hypothesis
- Small sample inference proceeds in the same manner as if one had a large sample
- It provides interpretable answers, such as "the true parameter has a probability of 0.95 of falling in a 95% credible interval"
- It provides a convenient setting for more complex models
- Can be used to provide evidence for the null hypothesis

Disadvantage of the Bayesian Approach

- Subjectivity
  - E.g., Two individuals with different prior information can come to different conclusions regarding the probability of events


DID THE SUN JUST EXPLODE?
(IT'S NIGHT, SO WE'RE NOT SURE.)
THIS NEUTRINO DETECTOR MEASURES WHETHER THE SUN HAS GONE NOVA.
THEN, IT ROLLS TWO DICE. IF THEY BOTH COME UP SIX, IT LIES TO US. OTHERWISE, IT TELLS THE TRUTH.
LET'S TRY. DETECTOR! HAS THE SUN GONE NOVA?
ROLL
YES.
FREQUENTIST STATISTICIAN:
THE PROBABILITY OF THIS RESULT HAPPENING BY CHANCE IS 1/36 = 0.027. SINCE p < 0.05, I CONCLUDE THAT THE SUN HAS EXPLODED.
BAYESIAN STATISTICIAN:
BET YOU $50 IT HASN'T.