Introduction to Bayesian Inference

Felipe José Bravo Márquez

October 1, 2020

Some Critics to the Frequentist Approach

- The statistical methods that we have discussed so far are known as frequentist (or classical) methods.
- The frequentist approach requires that all probabilities be defined by connection to the frequencies of events in very large samples.
- This leads to frequentist uncertainty being premised on imaginary resampling of data.
- If we were to repeat the measurement many many times, we would end up collecting a list of values that will have some pattern to it.
- It means also that parameters and models cannot have probability distributions, only measurements can.
- The distribution of these measurements is called a sampling distribution.
- This resampling is never done, and in general it doesn't even make sense.

Bayesian Inference

There is another approach to inference called Bayesian inference [**?**], which is based on the following postulates:

- Probability describes **degree of belief**, not limiting frequency.
  - We can make probability statements about lots of things, not just data which are subject to random variation.
  - For example, I might say that ”the probability that Albert Einstein drank a cup of tea on August 1, 1948” is .35.
  - This does not refer to any limiting frequency.
  - It reflects my strength of belief that the proposition is true.
- We can make probability statements about parameters, even though they are fixed constants.
- We make inferences about a parameter θ by producing a probability distribution for θ. Inferences, such as point estimates and interval estimates, may then be extracted from this distribution.

Bayesian Inference

- In modest terms, Bayesian data analysis is no more than counting the numbers of ways the data could happen, according to our assumptions [**?**].
- In Bayesian analysis all alternative sequences of events that could have generated our data are evaluated.
- As we learn about what did happen, some of these alternative sequences are pruned.
- In the end, what remains is only what is logically consistent with our knowledge [**?**].
- Warning: understanding the essence of Bayesian inference can be hard.
- The following toy example tries to explain it in a gentle way.

Counting Possibilities

- Suppose there's a bag, and it contains four marbles.
- These marbles come in two colors: blue and white.
- We know there are four marbles in the bag, but we don't know how many are of each color.
- We do know that there are five possibilities:
  (1) [○○○○], (2) [●○○○], (3) [●●○○], (4) [●●●○], (5) [●●●●]
- These are the only possibilities consistent with what we know about the contents of the bag. Call these five possibilities the **conjectures**.
- Our goal is to figure out which of these conjectures is most **plausible**, given some evidence about the contents of the bag.
- We do have some evidence: A sequence of three marbles is pulled from the bag, one at a time, replacing the marble each time and shaking the bag, in that order. These before drawing another marble.
- The sequence that emerges is: ●○●, in that order. These are the data.

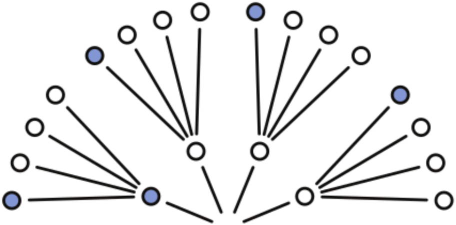
Counting Possibilities

- Now, let's see how to use the data to infer what's in the bag.
- Let's begin by considering just the single conjecture, [●○○○], that the bag contains one blue and three white marbles.
- On the first draw from the bag, one of four things could happen, corresponding to one of four marbles in the bag.

- Notice that even though the three white marbles look the same from a data perspective we just record the color of the marbles, after all they are really different events.
- This is important, because it means that there are three more ways to see ○ than to see ●.

Counting Possibilities

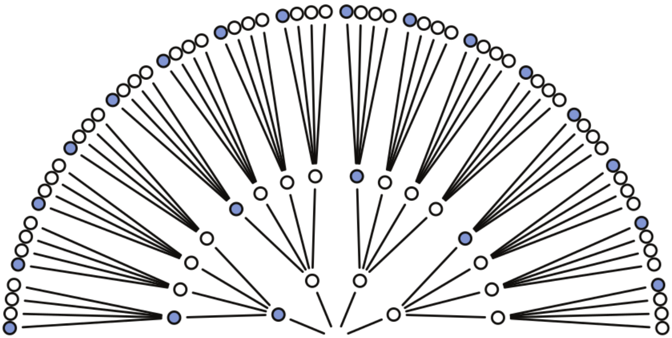
- Now consider the garden as we get another draw from the bag. It expands the garden out one layer:



- Now there are 16 possible paths through the garden, one for each pair of draws.
- On the second draw from the bag, each of the paths above again forks into four possible paths. Why?

Counting Possibilities

- Because we believe that our shaking of the bag gives each marble a fair chance at being drawn, regardless of which marble was drawn previously.
- The third layer is built in the same way, and the full garden is shown below:



- There are $4^3 = 64$ possible paths in total.

Counting Possibilities

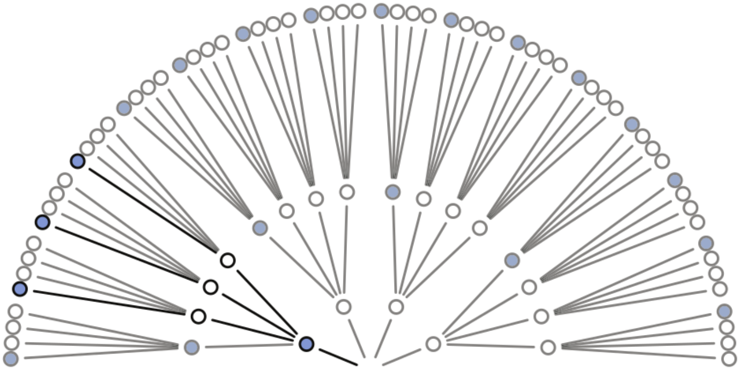
- As we consider each draw from the bag, some of these paths are logically eliminated.
- The first draw tuned out to be , recall, so the three white paths at the bottom are eliminated right away.
- If you imagine the real data tracing out a path, it must have passed through the one blue path near the origin.
- The second draw from the bag produces , so three of the paths forking out of the first blue marble remain.

Counting Possibilities

- As the data trace out a path, we know it must have passed through one of those three white paths (after the first blue path).
- But we don't know which one, because we recorded only the color of each marble.
- Finally, the third draw is ●.
- Each of the remaining three paths in the middle layer sustain one blue path, leaving a total of three ways for the sequence ●○● to appear, assuming the bag contains [●○○○].

Counting Possibilities

- The figure below shows the forking paths again, now with logically eliminated paths grayed out.



Counting Possibilities

- We can't be sure which of those three paths the actual data took.
- But as long as we're considering only the possibility that the bag contains one blue and three white marbles, we can be sure that the data took one of those three paths.
- Those are the only paths consistent with both our knowledge of the bag's contents (four marbles, white or blue) and the data ().

References I