YData: An Introduction to Data Science

Lecture 37: Decisions

Jessi Cisewski-Kehe
Statistics & Data Science, Yale University
Spring 2020

Credit: data8.org

Announcements

Decisions

Decisions Under Uncertainty

Interpretation by Physicians of Clinical Laboratory Results (1978)

"We asked 20 house officers, 20 fourth-year medical students and 20 attending physicians, selected in 67 consecutive hallway encounters at four Harvard Medical School teaching hospitals, the following question:

"If a test to detect a disease whose prevalence is 1/1000 has a false positive rate of 5%, what is the chance that a person found to have a positive result actually has the disease, assuming that you know nothing about the person's symptoms or signs?"

Interpretation by Physicians of Clinical Laboratory Results (1978)

"Eleven of 60 participants, or 18%, gave the correct answer. These participants included four of 20 fourth-year students, three of 20 residents in internal medicine and four of 20 attending physicians. The most common answer, given by 27, was that [the chance that a person found to have a positive result actually has the disease] was 95%."

Conditional Probability

Round One

- Scenario:
  - Class consists of second years (60%) and third years (40%)
  - 50% of the second years have declared their major
  - 80% of the third years have declared their major
  - I pick one student at random.
- Which is more likely: Second year or third year?
  - Second year, because they are 60% of the class

Round Two

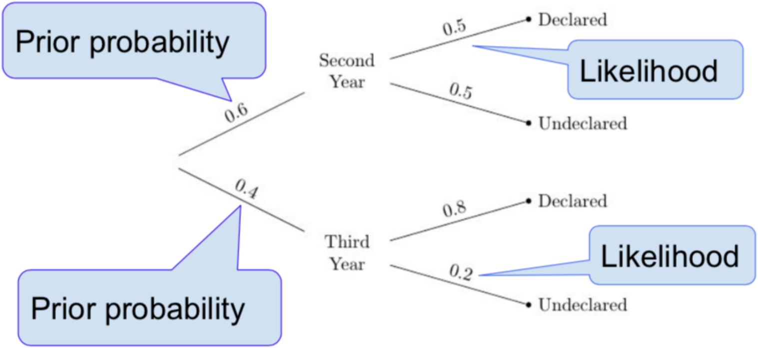
- Slightly different scenario:
    - Class consists of second years (60%) and third years (40%)
    - 50% of the second years have declared their major
    - 80% of the third years have declared their major
    - I pick one student at random...
      That student has declared a major!

- Second Year or Third Year?

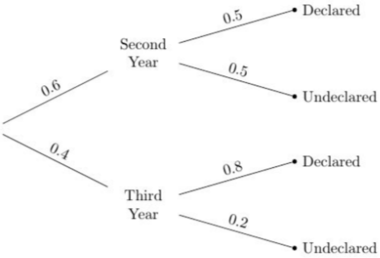
(DEMO)

Bayes' Rule

Diagram and Terminology



Bayes' Rule



Posterior probability:

P(Third Year | Declared)

$$= \frac{0.4 \times 0.8}{(0.6 \times 0.5) + (0.4 \times 0.8)}$$

$$= 0.5161\ldots$$

Pick a student at random.

(DEMO)

Purpose of Bayes' Rule

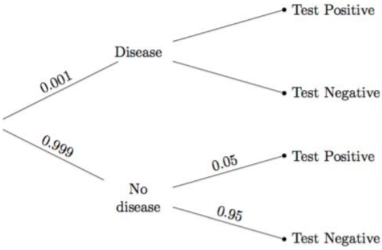
- Update your prediction based on new information
- In a multi-stage experiment, find the chance of an event at an earlier stage, given the result of a later stage

Decisions Under Uncertainty

Interpretation by Physicians of Clinical Laboratory Results (1978)

"We asked 20 house officers, 20 fourth-year medical students and 20 attending physicians, selected in 67 consecutive hallway encounters at four Harvard Medical School teaching hospitals, the following question:
"If a test to detect a disease whose prevalence is 1/1000 has a false positive rate of 5%, what is the chance that a person found to have a positive result actually has the disease, assuming that you know nothing about the person's symptoms or signs?"

Example: Doctors & Clinical Tests

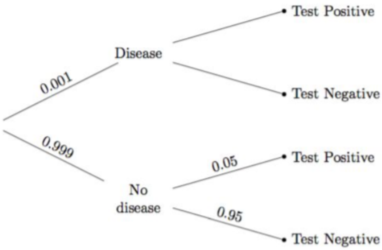


Problem did not give the *true positive* rate.

That's the chance the test says "positive" if the person has the disease.

It was assumed to be 100%.

Data and Calculation



$$P(\text{Disease given Test} +)$$

$$= \frac{0.001 * 1}{(0.001 * 1) + (0.999 * 0.05)}$$

$$= 0.0196270...$$

(DEMO)

Subjective Probabilities

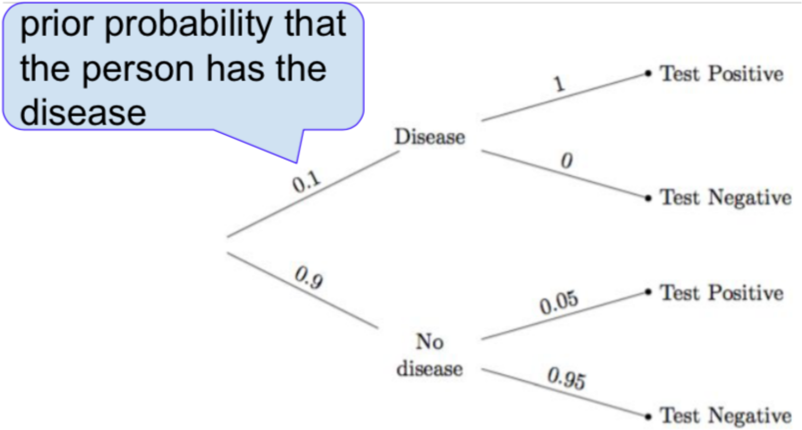
A probability of an outcome is...

- The frequency with which it will occur in repeated trials, or
- The subjective degree of belief that it will (or has) occurred

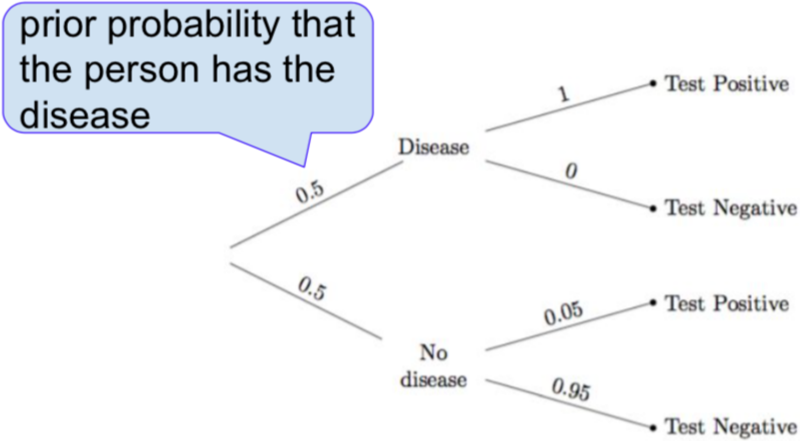
Why use subjective priors?

- In order to quantify a belief that is relevant to a decision
- When the subject of your prediction was not selected randomly from the population

A Subjective Opinion



A Different Subjective Opinion



(DEMO)