YData: An Introduction to Data Science

Lecture 37: Decisions

Elena Khusainova & John Lafferty
Statistics & Data Science, Yale University
Spring 2021

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%."

Decisions Under Uncertainty

https://opinionator.blogs.nytimes.com/2010/04/25/chances-are/

"In one study, Gigerenzer and his colleagues asked doctors in Germany and the United States to estimate the probability that a woman with a positive mammogram actually has breast cancer, [...] the doctors were told to assume the following statistics — couched in terms of percentages and probabilities — about the prevalence of breast cancer among women in this cohort, and also about the mammogram's sensitivity and rate of false positives:

The probability that one of these women has breast cancer is 0.8 percent. If a woman has breast cancer, the probability is 90 percent that she will have a positive mammogram. If a woman does not have breast cancer, the probability is 7 percent that she will still have a positive mammogram. Imagine a woman who has a positive mammogram. What is the probability that she actually has breast cancer? "

Decisions Under Uncertainty

"When Gigerenzer asked 24 other German doctors the same question, their estimates whipsawed from 1 percent to 90 percent. Eight of them thought the chances were 10 percent or less, 8 more said 90 percent, and the remaining 8 guessed somewhere between 50 and 80 percent. Imagine how upsetting it would be as a patient to hear such divergent opinions.

As for the American doctors, 95 out of 100 estimated the woman's probability of having breast cancer to be somewhere around 75 percent."

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

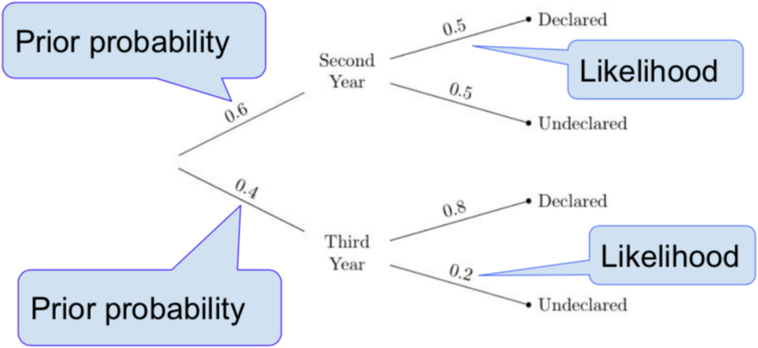
Round Two

- Slighty 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?

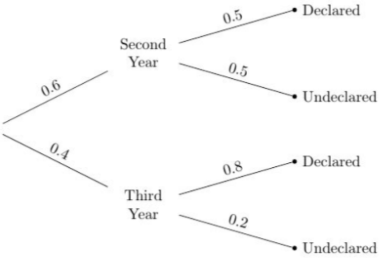
(DEMO)

Bayes' Rule

Diagram and Terminology



Bayes' Rule



Posterior probability:

$$P(\text{Third Year} \mid \text{Declared})$$

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

$$= 0.5161\ldots$$

Pick a student at random.

(DEMO)

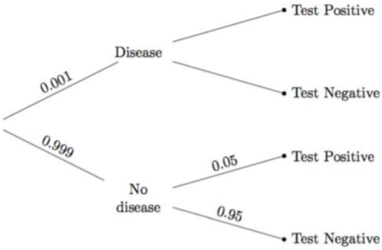
Purpose of Bayes' Rule

- 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

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

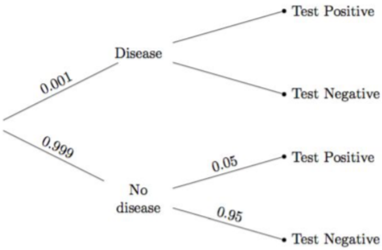


Problem did not give the *true positive* rate.

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

It was assumed to be 100%.

Data and Calculation



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

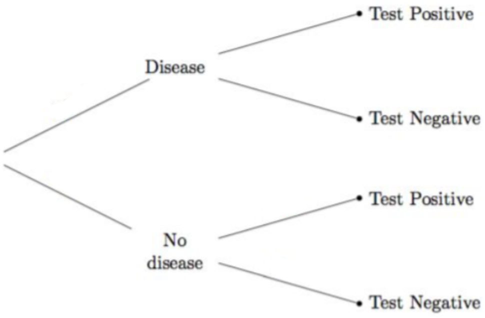
$$= \frac{0.001*1}{(0.001*1)+(0.999*0.05)}$$

$$= 0.0196270...$$

(DEMO)

Discussion

The probability that one of these women has breast cancer is 0.8 percent. If a woman has breast cancer, the probability is 90 percent that she will have a positive mammogram. If a woman does not have breast cancer, the probability is 7 percent that she will still have a positive mammogram. Imagine a woman who has a positive mammogram. What is the probability that she actually has breast cancer?



Natural Frequencies

Eight out of every 1,000 women have breast cancer. Of these 8 women with breast cancer, 7 will have a positive mammogram. Of the remaining 992 women who don't have breast cancer, some 70 will still have a positive mammogram. Imagine a sample of women who have positive mammograms in screening. How many of these women actually have breast cancer?

When Gigerenzer tested another set of 24 doctors, this time using natural frequencies, nearly all of them got the correct answer, or close to it.

(DEMO)

Subjective Probabilities

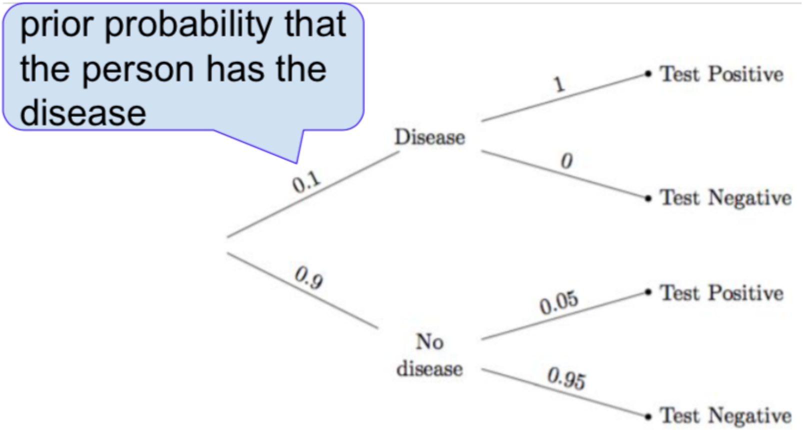
A probability of an outcome is...

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

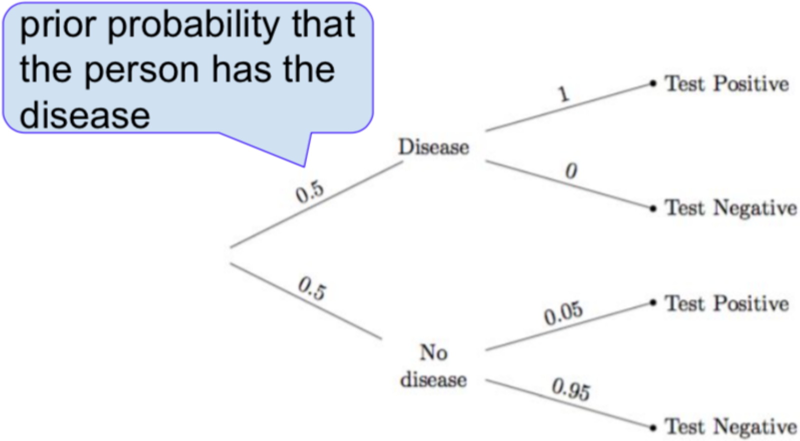
Why use subjective priors?

- 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)