Bayes' Theorem

Gregory M. Shinault

Goal for this Lecture

Learn about the famous rule of Thomas Bayes. We will see

1. the statement and proof of the Bayes formula,
2. learn how to use the Bayes formula, and
3. see how significant proper use of the Bayes formula is in decision making.

The material of this lecture roughly corresponds to Section 2.2 of the textbook.

Importance of Bayes' Formula

Bayes' Theorem gives us a way to update probabilistic beliefs in the face of new information.

Statement

Let A be an event, and $B_1, \ldots, B_n$ a partition of Ω. Then

$$\mathbb{P}(B_k|A) = \frac{\mathbb{P}(A|B_k)\mathbb{P}(B_k)}{\sum_{\ell=1}^{n} \mathbb{P}(A|B_\ell)\mathbb{P}(B_\ell)}.$$

Comment: This can be memorized, but understanding the derivation is more important.

Example

We have three numbered urns.
Urn 1 has 2 red marbles and 3 blue marbles.
Urn 2 has 2 red marbles and 1 yellow marble.
Urn 3 has 3 red marbles and 2 blue marbles.

Behind a curtain I conduct an experiment.
I roll a die.
If the die is 1, 2, 3 then I draw a marble from Urn 1.
If the die is 4, 5 then I draw a marble from Urn 2.
If the die is 6 then I draw a marble from Urn 3.

Now I reveal to you that I drew a red marble. What is the probability that I drew from Urn 2?

Example

Suppose we randomly select a person from the population and apply a diagnostic test for leukemia. Leukemia affects roughly 1/10000 people. This diagnostic test is very accurate:

$$\mathbb{P}(+|L) = 0.995, \quad \mathbb{P}(-|L^c) = 0.999.$$

If the test comes back positive, what is the probability the person tested actually has leukemia?

Example

Let's repeat the previous diagnostic test with a variation. This time we will perform the test on someone with a symptom that is indicative of leukemia in 5% of known cases. This diagnostic test still has accuracy given by:

$$\mathbb{P}(+|L) = 0.995, \quad \mathbb{P}(-|L^c) = 0.999.$$

If the test comes back positive, what is the probability the person tested actually has leukemia?

Visualization of Diagnostic Testing (Sample Size 6000)

Parameters

Let's consider the diagnostic problem with

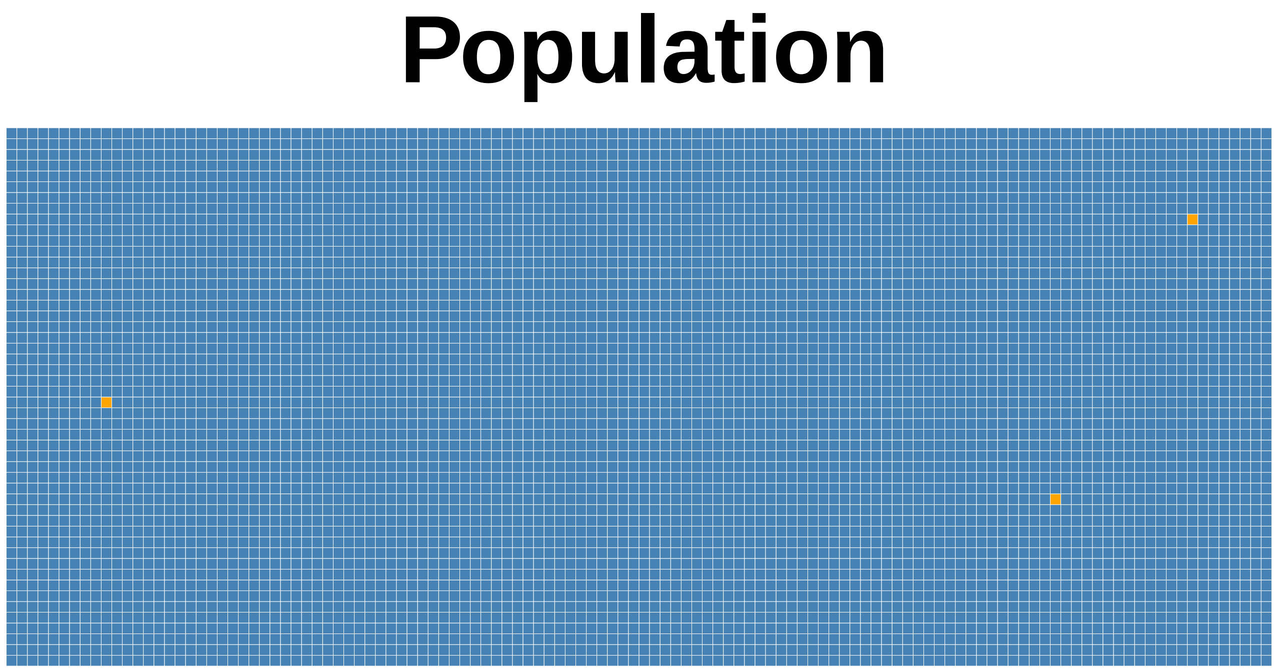
$$\begin{aligned} \mathbb{P}(+|\text{Sick}) &= 0.9995 \\ \mathbb{P}(-|\text{Healthy}) &= 0.999 \\ \mathbb{P}(\text{Sick}) &= 0.001 \end{aligned}$$

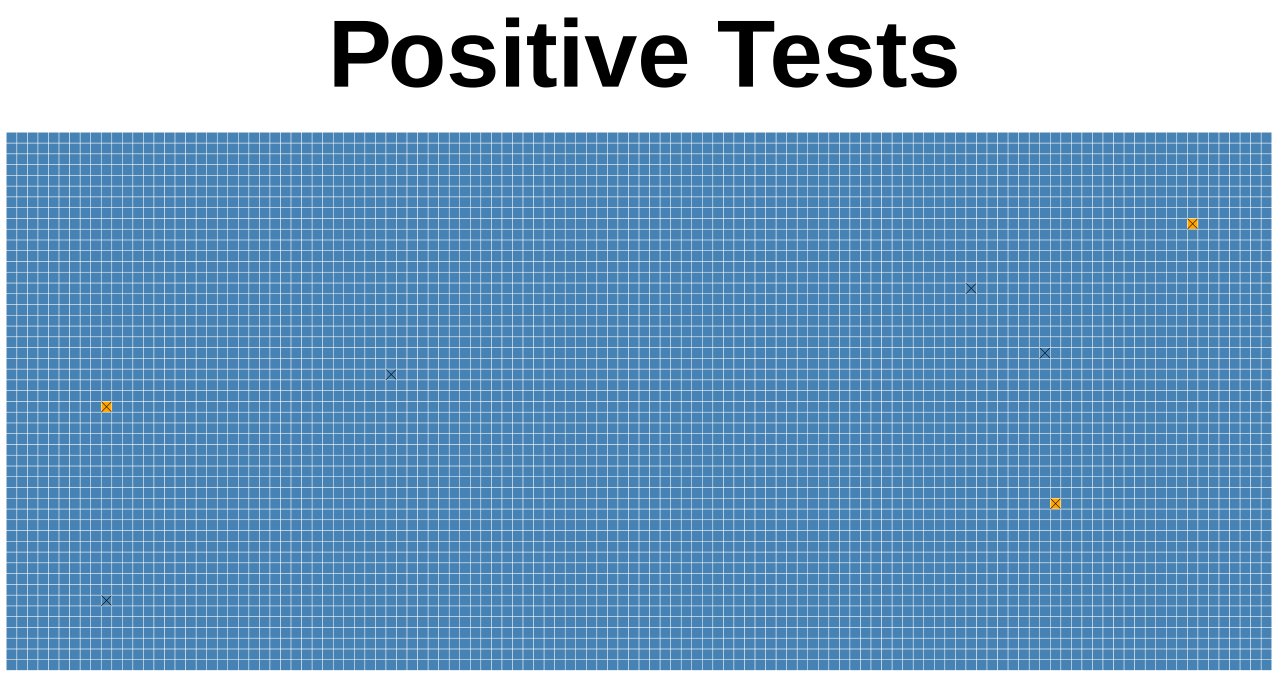
By Bayes'

$$\mathbb{P}(\text{Healthy}|+) = \frac{\mathbb{P}(+|\text{Healthy})\mathbb{P}(\text{Healthy})}{\mathbb{P}(+|\text{Healthy})\mathbb{P}(\text{Healthy}) + \mathbb{P}(+|\text{Sick})\mathbb{P}(\text{Sick})} = 0.4998749$$

Numerical Analysis of the Simulation

```
## [1] "Number of True Positives:  3"
```


Population


Positive Tests

```
## [1] "Number of False Negatives:  0"

## [1] "Number of False Positives:  4"

## [1] "Number of True Negatives:  5993"

## [1] "Empirical Probability of No Disease Given Positive Test:  0.571428571428571"

## [1] "Actual Probability of No Disease Given Positive Test:  0.499874906179635"
```

The Wrap Up

Summary

1. Bayes' formula is just a combination of the multiplication rule and the law of total probability.
2. Bayes' formula gives us a method to update probabilities for beliefs when faced with new evidence.

Next step

We have all the basic results for conditional probability now. Next we consider events for which conditioning tells us nothing.