Applications: Bayes Rule

CSE 312 Spring 21
Lecture 9

Announcements

First "real-world mini-project" comes out tonight.

Due in 2 weeks.

Goal is to give you a chance to think about why what you're learning matters in the real world.

Video available to help Naive Bayes prog. question.

Today

Bayes’ Rule in the real world!

pollev.com/cse312

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}$$

Application 1: Medical Tests

Helping Doctors and Patients Make Sense of Health Statistics

A researcher posed the following scenario to a group of 160 doctors:

Assume you conduct a disease screening using a standard test in a certain region. You know the following information about the people in this region:

The probability that a person has the disease is 1% (prevalence) ("prior")

If a person has the disease, the probability that she tests positive is 90% (sensitivity)

If a person does not have the disease, the probability that she nevertheless tests positive is 9% (false-positive rate)

A person tests positive. She wants to know from you whether that means that she has the disease for sure, or what the chances are. What is the best answer?

A. The probability that she has the disease is about 81%.

B. Out of 10 people with a positive test, about 9 have the disease.

C. Out of 10 people with a positive test, about 1 have the disease.

D. The probability that she has the disease is about 1%

Let's do the calculation!

Let D be "the patient has the disease", T be "the test was positive."

LTP

$\mathbb{P}(T) = \mathbb{P}(T|D)\mathbb{P}(D) + \mathbb{P}(T|\bar{D})\mathbb{P}(\bar{D})$

$\rightarrow .9 \cdot .01 + .09 \cdot .99$

$$\mathbb{P}(D|T) = \mathbb{P}(T|D) \cdot \mathbb{P}(D)/\mathbb{P}(T)$$

$$= \frac{\boxed{.9 \cdot .01}}{.99 \cdot .09 + .01 \cdot .9} \approx 0.092$$

Calculation tip: for Bayes' Rule, you should see one of the terms on the bottom exactly match your numerator (if you're using the LTP to calculate the probability on the bottom)

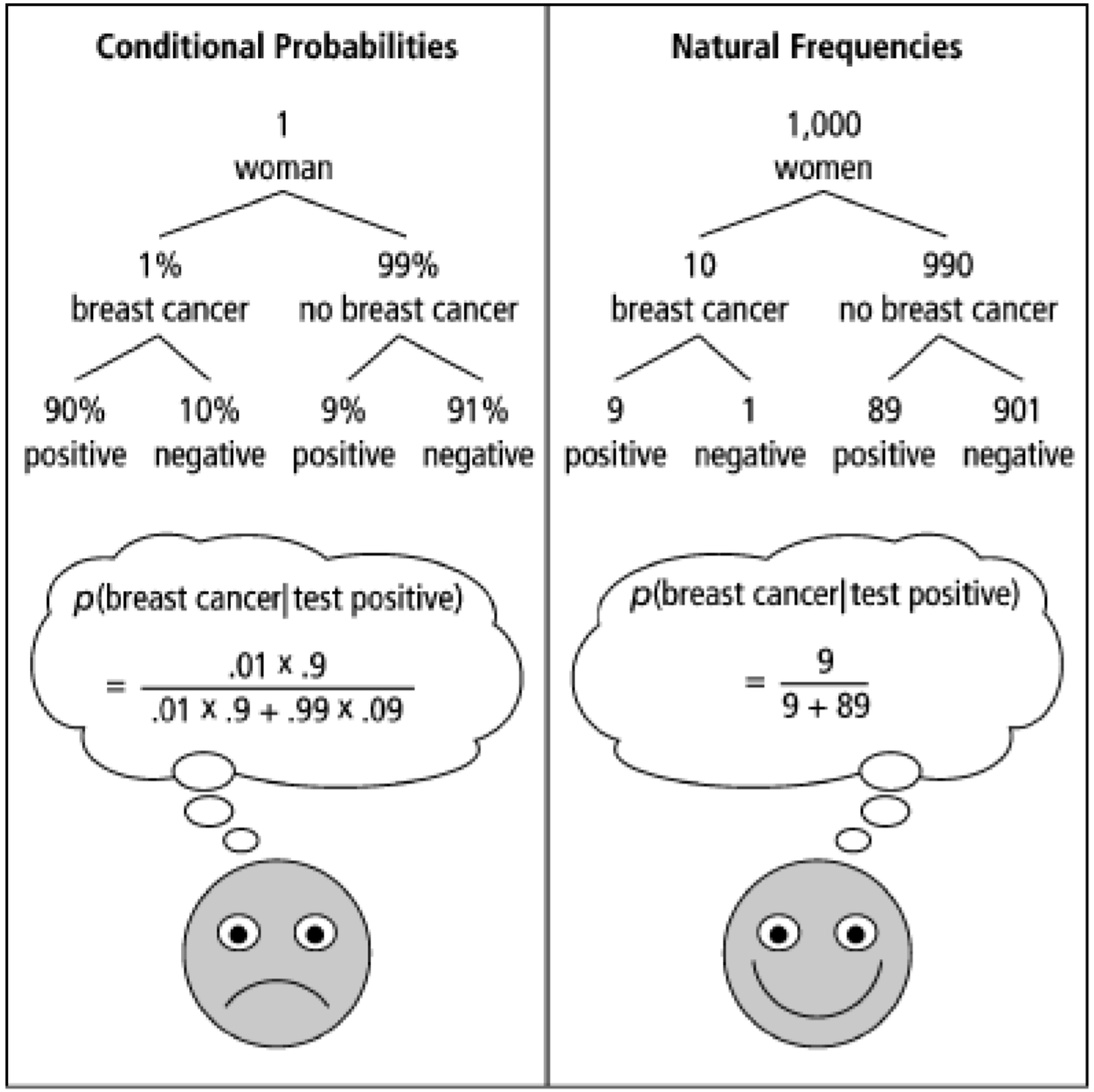
How did the doctors do

C (about 1 in 10) was the correct answer.

Of the doctors surveyed, less than ¼ got it right (so worse than random guessing – and worse than the class did on the polleverywhere last week).

After the researcher taught them his calculation trick, more than 80% got it right.

One Weird Trick!



Calculation Trick: imagine you have a large population (not one person) and ask how many there are of false/true positives/negatives.

What about the real world?

When you're older and have to do more routine medical tests, don't get concerned (yet) when they ask to run another test.*

It's usually fine.*

*This is not medical advice, Robbie is not a physician.

No More Medical Testing Examples

We’re living in a pandemic…

We’re not going to use COVID for any examples

(Robbie is too tired of seeing bad takes on twitter)

If you want to think about COVID and Bayes Rule, you’ll be able to on the real world assignment coming out today.

Bayes Factor

Bayes Factor

Another Intuition Trick: from 3Blue1Brown

When you test positive, you (**approximately**) multiply the prior by the "Bayes Factor" (aka likelihood ratio)

$$\frac{\text{sensitivity}}{\text{false positive rate}} = \frac{1 - FNR}{FPR}$$

Bayes Factor

Does it work?

Let's try it…

Find

$$\text{prior} \cdot \frac{\text{Sensitivity}}{FPR}$$

Wonka Bars

$\text{prior} \cdot \frac{\text{sensitivity}}{\text{FPR}}$

Willy Wonka has placed golden tickets on 0.1% of his Wonka Bars.

You want to get a golden ticket. You could buy a 1000-or-so of the bars until you find one, but that's expensive…you've got a better idea!

You have a test – a very precise scale you've bought.

- If the bar you weigh **does** have a golden ticket, the scale will alert you 99.9% of the time.
- If the bar you weigh does not have a golden ticket, the scale will (falsely) alert you only 1% of the time.

$\frac{1}{11}$

If you pick up a bar and it alerts, what is the probability you have a golden ticket?

$0.1\% \cdot \frac{99.9\%}{1\%} \approx 9.99\%$

9.09%

Wonka Bars

Bayes Factor

$\frac{99.9}{1}$

Prior: $.1\%$

Product: 9.99, so about 10%

About what Bayes Rule gets!

Application 1: Medical Tests

$\text{prior} \cdot \frac{\text{sensitivity}}{\text{FPR}}$

$1\% \cdot \frac{90\%}{9\%} \approx 10\%$

[Helping Doctors and Patients Make Sense of Health Statistics]

A researcher posed the following scenario to a group of 160 doctors:

Assume you conduct a disease screening using a standard test in a certain region. You know the following information about the people in this region:

The probability that a person has the disease is 1% (prevalence)

If a person has the disease, the probability that she tests positive is 90% (sensitivity)

If a person does not have the disease, the probability that she nevertheless tests positive is 9% (false-positive rate)

A person tests positive. She wants to know from you whether that means that she has the disease for sure, or what the chances are. What is the best answer?

A. The probability that she has the disease is about 81%.

B. Out of 10 people with a positive test, about 9 have the disease.

C. Out of 10 people with a positive test, about 1 have the disease

D. The probability that she has the disease is about 1%

Bayes Factor

What about with the doctors?

$$1\% \cdot \frac{90\%}{9\%} = 10\%$$

Again about right!

Caution

$$\frac{\text{prior} \cdot \text{sensitivity}}{\text{FPR}}$$

Multiplying by the Bayes Factor is an **approximation**

It gives you the exact numerator for Bayes, but the denominator is "the number of false positives if the prevalence (/prior) were 0"

When the prior is close to 0, this is a fine approximation!

But plug in a prior of 15% on the last slide, and we get 150% chance.

What about negative tests?

For negative tests, the Bayes Factor is

$$\frac{FNR}{\text{Sensitivity (false negative rate)}}$$

Careful Surveys

Application 2: An Imbalanced Survey

In 2014, a paper was published

"Do non-citizens vote in U.S. elections?"

This is a real paper (peer-reviewed). It claims that

1. In a survey, about 4% (of a few hundred) of non-U.S.-citizens surveyed said they voted in the 2008 federal election (which isn't allowed).

2. Those non-citizen voters voted heavily (estimate 80+%) for democrats.

3. "It is likely though by no means certain that John McCain would have won North Carolina were it not for the votes for Obama cast by non-citizens"

Application 2: What is this survey?

The “Cooperative Congressional Election Study” was run in 2008 and 2010.

It interviews about 20,000 people about how/whether they voted in federal elections.

Two strange observations:

1. The noncitizens are a very small portion of those surveyed. Feels a little strange.

2. Those people…maybe accidentally admitted to a crime?

Application 2: Another Red Flag

A response paper (by different authors)

[“The perils of cherry picking low frequency events in large sample surveys”]

Table 1
Response to citizenship question across two-waves of CCES panel.

Response in 2010	Response in 2012	Number of respondents	Percentage
Citizen	Citizen	18,737	99.25
Citizen	Non-Citizen	20	0.11
Non-Citizen	Citizen	36	0.19
Non-Citizen	Non-Citizen	85	0.45

An Explanation

Suppose 0.1% of people check the wrong check-box on any individual question (independently)

≈ 20 people

Suppose you really interviewed $20{,}000$ people, of whom 300 are really non-citizens (none of whom voted), and the rest are citizens, of whom 70% voted. What is the probability someone appears to have voted

$$\mathbb{P}(say\ V|say\ NC) = \frac{\mathbb{P}(say\ NC|say\ V)\cdot\mathbb{P}(say\ V)}{\mathbb{P}(say\ NC)} = \frac{.001\cdot .7}{.999\cdot\left(\frac{300}{20000}\right)+.001\cdot\left(\frac{19700}{20000}\right)} \approx 4.38\%$$

Conclusion

The authors of the original paper did know about response error…

…and they have an appendix that argues the population of "non-citizen" voters isn't distributed exactly like you'd expect.

But with it being such a small number of people, this isn't surprising.

And even they admit response bias played more of a role than they initially thought.

Though they still think they found some evidence of non-citizens voting (but not enough to flip North Carolina anymore).