Exercise: Bayes' Theorem and Total Probability

Bookmark this page

Exercise: Bayes' Theorem and Total Probability

8 points possible (graded)

A new test is developed to determine whether a patient has an antibiotic-resistant bacterial infection. The test is correct 99% of the time: that is, if a patient has the infection, there is a 99% chance the test will return positive, and if a patient doesn't have the infection, there is a 99% chance the test will return negative. Suppose 0.1% of the general population has this infection. If a randomly selected person is administered this test and gets a positive result, what is the probability that she or he actually has the infection?

To solve this problem, let $\mathcal{T}$ be the event that the patient has a positive test result, and $\mathcal{S}$ be the event that the patient has the associated bacterial infection. Thus, what the problem is asking for is precisely the quantity $\mathbb{P}(\mathcal{S}|\mathcal{T})$. So somehow we have to figure out what this probability is.

What do we have access to? Well, from the problem statement, we are directly given the following quantities. What are they? **Please provide the exact answer for these three quantities.**

- $\mathbb{P}(\mathcal{T}|\mathcal{S}) =$ **Answer:** 0.99
- $\mathbb{P}(\mathcal{T}^c|\mathcal{S}^c) =$ **Answer:** 0.99

- $\mathbb{P}(\mathcal{S}) =$ **Answer:** 0.001

Let's see if Bayes' rule can help us:

$$\mathbb{P}(\mathcal{S}|\mathcal{T}) = \frac{\mathbb{P}(\mathcal{T}|\mathcal{S})\mathbb{P}(\mathcal{S})}{\mathbb{P}(\mathcal{T})}.$$

There's a slight problem. We don't know $\mathbb{P}(\mathcal{T})$, i.e., the probability that the patient has a positive test result. But let's break this into two cases. The patient either has the infection or not. So

$$\begin{aligned}\mathbb{P}(\mathcal{T}) &= \mathbb{P}(\textbf{patient has positive test result}) \\ &= \mathbb{P}(\textbf{patient has positive test result and patient has infection}) \\ &\quad + \mathbb{P}(\textbf{patient has positive test result and patient does not have infection}) \\ &= \mathbb{P}(\mathcal{T} \cap \mathcal{S}) + \mathbb{P}(\mathcal{T} \cap \mathcal{S}^c).\end{aligned}$$

- What result did we just use?

- ○ Bayes' Theorem
- ○ Product Rule
- ○ Law of Total Probability
- ○ Marginalization

We can go one step further:

$$\begin{aligned}\mathbb{P}(\mathcal{T}) &= \mathbb{P}(\mathcal{T} \cap \mathcal{S}) + \mathbb{P}(\mathcal{T} \cap \mathcal{S}^c) \\ &= \mathbb{P}(\mathcal{T}|\mathcal{S})\mathbb{P}(\mathcal{S}) + \mathbb{P}(\mathcal{T}|\mathcal{S}^c)\mathbb{P}(\mathcal{S}^c).\end{aligned}$$

- In the last equality, what result did we use?

- Bayes' Theorem
- Product Rule
- Law of Total Probability
- Marginalization

So what we have is:

$$\mathbb{P}(\mathcal{S}|\mathcal{T}) = \frac{\mathbb{P}(\mathcal{T}|\mathcal{S})\mathbb{P}(\mathcal{S})}{\mathbb{P}(\mathcal{T})} = \frac{\mathbb{P}(\mathcal{T}|\mathcal{S})\mathbb{P}(\mathcal{S})}{\mathbb{P}(\mathcal{T}|\mathcal{S})\mathbb{P}(\mathcal{S}) + \mathbb{P}(\mathcal{T}|\mathcal{S}^c)\mathbb{P}(\mathcal{S}^c)}.$$

There are just two probabilities that we haven't figured out. But we can from the information we are given! Determine what the following are. **Provide exact answers for these.**

- $\mathbb{P}(\mathcal{S}^c) =$ **Answer:** 0.999

- $\mathbb{P}(\mathcal{T}|\mathcal{S}^c) =$ **Answer:** 0.01

Now we can plug in all the probabilities into our very last big equation above!

- What is $\mathbb{P}(\mathcal{S}|\mathcal{T})$, the probability that given a randomly selected person has a positive test result, she or he has the infection? (Please be precise with at least 3 decimal places, unless of course the answer doesn't need that many decimal places. You could also put a fraction.)

  Answer: 0.09016393443

Does the probability seem low given how accurate the test is? Be sure to check out the solutions for a detailed explanation.

Solution:

Since the test is 99% accurate, this means that

$$\begin{aligned} \mathbb{P}(\mathcal{T}|\mathcal{S}) &= 0.99, \\ \mathbb{P}(\mathcal{T}^c|\mathcal{S}^c) &= 0.99. \end{aligned}$$

Since 0.1% of the general population has the infection, this means that

$$\mathbb{P}(\mathcal{S}) = 0.001.$$

Next, note that

$$\mathbb{P}(\mathcal{S}^c) = 1 - \mathbb{P}(\mathcal{S}) = 1 - 0.001 = 0.999.$$

and

$$\mathbb{P}(\mathcal{T}|\mathcal{S}^c) = 1 - \mathbb{P}(\mathcal{T}^c|\mathcal{S}^c) = 1 - 0.99 = 0.01.$$

At this point, we can plug everything into the Bayes' rule formula:

$$\begin{aligned}\mathbb{P}(\mathcal{S}|\mathcal{T}) &= \frac{\mathbb{P}(\mathcal{T}|\mathcal{S})\mathbb{P}(\mathcal{S})}{\mathbb{P}(\mathcal{T}|\mathcal{S})\mathbb{P}(\mathcal{S}) + \mathbb{P}(\mathcal{T}|\mathcal{S}^c)\mathbb{P}(\mathcal{S}^c)} \\ &= \frac{0.99 \cdot 0.001}{0.99 \cdot 0.001 + 0.01 \cdot 0.999} \\ &\approx 0.09016393443.\end{aligned}$$

This means only about 9% of the diagnosed patients will actually have the infection. Why does this test generate so many false positives? As you can see from the arithmetic above, the problem is $\mathbb{P}(\mathcal{T}|\mathcal{S}^c)$, the probability that the bacteria are falsely found in a healthy patient. Even though $\mathbb{P}(\mathcal{T}|\mathcal{S}^c)$ is only 1%, this means 1% of all the healthy patients (who represent 99.9% of the population) are being incorrectly diagnosed, overwhelming the total number of sick patients.

In order to avoid overdiagnosis, the test must be much more accurate. For example, if the error rate in the test when administered to a healthy person (i.e., the so-called “false-positive rate") were reduced by an order of magnitude (i.e., $\mathbb{P}(\mathcal{T}|\mathcal{S}^c) = 0.001$), the answer would change to

$$\mathbb{P}(\mathcal{S}|\mathcal{T}) = \frac{0.99 \cdot 0.001}{0.99 \cdot 0.001 + 0.001 \cdot 0.999} \approx 0.5,$$

i.e., the fraction of people with the infection among all those who test positive increases to 50%. A further reduction in the false-positive error rate by another order of magnitude (i.e., $\mathbb{P}(\mathcal{T}|\mathcal{S}^c) = 0.0001$) would produce

$$\mathbb{P}(\mathcal{S}|\mathcal{T}) = \frac{0.99 \cdot 0.001}{0.99 \cdot 0.001 + 0.0001 \cdot 0.999} \approx 0.9,$$

so now 90% of patients who test positive would be actually have the infection. This exercise emphasizes the importance of designing the observation model (medical test in this example) to match the prior frequency of the event we are aiming to detect.

Submit You have used 0 of 5 attempts

Discussion

Topic: Conditioning on Events / Exercise: Bayes' Theorem and Total Probability

Hide Discussion

Add a Post

limit of 5 submissions

discussion posted 2 months ago by **zafar-hussain**

this limit of 5 submission is counter productive, it must be removed to made to some high value

This post is visible to everyone.

+ Expand discussion

another very poor section with no background given before one does this hard exercise

discussion posted 2 months ago by anonymous

very less instruction here as well

This post is visible to everyone.

+ Expand discussion

What is difference between marginalization and law of total probability?

question posted 2 months ago by anonymous

Hi. I was revisiting this...

This post is visible to everyone.

+ Expand discussion

About $\mathbb {P}(\mathcal{T} | \mathcal{S}^ c)$

discussion posted 2 months ago by **alfonsho**

I worked out $\mathbb{P}(\mathcal{T}|\mathcal{S}^c)$ using the conditional probability definition and got

```
In [137]: 0.01/0.99
```

...

This post is visible to everyone.

+ Expand discussion

Conditioning rules vs formal logic

discussion posted 2 months ago by **deep-one**

I completed the task in two ways - using both the conditioning rules and the decision tree/formal logic. Are those two approaches of equal worth...

This post is visible to everyone.

+ Expand discussion

unicode symbol resources

discussion posted 2 months ago by **hedges333**

Here are some useful unicode resources for typing symbols in comments:

https://en.wikipedia.org/wiki/Mathematical_operators_and_symbols_in_Unicode

https://en.wikipedia.org/wiki/Unicode_subscripts_and_superscripts

https://en.wikipedia.org/wiki/List_of_logic_symbols

https://en.wikipedia.org/wiki/Greek_alphabet

PLMGTFY...

This post is visible to everyone.

+ Expand discussion

Superscript "C"

discussion posted 2 months ago by anonymous

While I was able to figure this out - was the meaning of the "Superscript C" ever covered here? Additionally, while I'm at it - was the meaning...

This post is visible to everyone.

+ Expand discussion

So many conditioning probability

discussion posted 2 months ago by **Belter**

There are 8 different conditioning probability, according to my own understanding:

- A: following conditioning probability can be used to judge...

This post is visible to everyone.

+ Expand discussion

what is the meaning of the last sentence of the solution part?

discussion posted 2 months ago by **ggiannak17**

What it means "to match the prior frequency of the event we are aiming to detect. "?

This post is visible to everyone.

+ Expand discussion

More than intuition

discussion posted 2 months ago by **washington314**

I love this example because without much analysis people consistently thinks incorrect and makes a poor decision.

This post is visible to everyone.

+ Expand discussion

Help with Exercise: Bayes' Theorem and Total Probability

discussion posted 2 months ago by **ThatIsNoK**

Hello,

unfortunately I am running out of attempts. How I am supposed to give the input? Is 99% = 0.99 or should I use 99 for 99%? Thanks for...

This post is visible to everyone.

+ Expand discussion

Interesting question

discussion posted 2 months ago by **HaroldHenson**

I found this question quite illuminating. It is no wonder that the Doctors recommend a second test for a rare disease.

This post is visible to everyone.

+ Expand discussion

1