Statistics Practical Task 2

Bronwyn Bowles-King

1. You are given the values of P(B) and P(A and B) in the following scenarios. You are welcome to calculate P(A|B) if you choose to do so:

1.1 You work for a risk analysis insurer. You have read that this year, out of all drivers on the road, 5% have had accidents under the age of 25. You have also read that 10% of all drivers are under the age of 25. A new client approaches you and states that their age is 22. You want to calculate the chance that this driver has had an accident this year based on their age.

- Event A: Driver had an accident this year.
- Event B: Driver is under 25 years old.

Facts given:

- Out of all drivers, 5% have had accidents and are under the age of 25.
- 10% of all drivers are under this age.

P(A and B) is 0.05 because 5% of all drivers are both under 25 and had an accident.

P(B) is 0.10 because 10% of all drivers are under 25 years of age. B is true for the potential client.

> What is the probability of A given B is true in this case?

This relationship of "A given B" is expressed as A|B. P(A|B) is the probability that A is true when B is also true (assuming some cause between them). P(A|B) is calculated based on P(B) and P(A and B) as follows:

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

$$= \frac{5}{10}$$

$$= 50\%$$

The likelihood that the client has had an accident in the past year is 50%.

1.2 Your friend told you that they would buy you lunch if you can flip a coin and have it land on heads twice. You flip it the first time, and it lands on heads. What are your chances now of it landing on heads again?

- Event A: Second coin flip is heads.
- Event B: First coin flip is heads.

Facts given:

- First coin flip is heads.
- Coin has two sides and thus a 50% of either being the result.
- Both A and B are heads.

> What is the probability of A given B is true in this case? The situation is simpler here because the events are actually independent.

$$\mathrm{P(A|B) = P(A)}$$

$$= 50\%$$

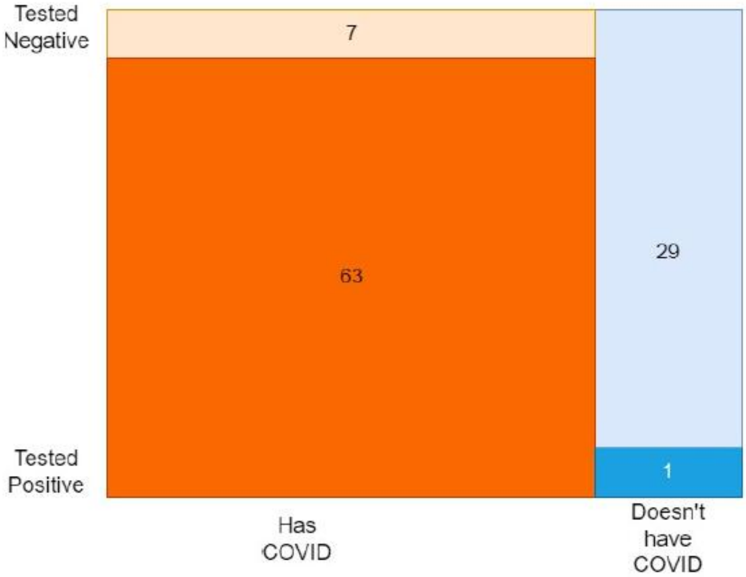
1.3 You were always told that knowing Maths helps you to achieve 80% in Computer Science. You read some statistics showing that 30% of all Computer Science graduates took Maths and achieved 80%. Overall, 60% of all Computer Science graduates took Maths. Considering you took Maths, what are your chances of achieving 80%?

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

$$\mathrm{P(A|B)} = \frac{30}{60} = \frac{1}{2} = 50\%$$

2. Consider a mock study about COVID diagnosis with a total of 100 participants. A visual representation of participants and their COVID and diagnostic test result status is given below.

The two orange areas show the total number of people who have COVID, and the two blue areas show the total number of people who do not have COVID. The two darkly-coloured areas at the bottom show the people who tested positive for COVID. The two lightly-coloured areas show the people who tested negative for COVID.



2.1 Using this diagram, state the following:

- H: our hypothesis
- E: our evidence

What the hypothesis should be is not exactly clear to me from the information given, but it could be that given a person is tested for COVID-19, they are more likely to have it. Or, it could be simply that a person has COVID-19. I will go with this simpler hypothesis. The evidence we have is a set of COVID-19 tests.

In summary:

H: Person has COVID-19.
E: A positive test for COVID-19.

2.2 Then, give the values for the following:

- P(H)

$$\mathrm{P(H)} = 63 + 7 = 70$$

Explanation
P(H) is the probability that a person (going for a test) has COVID-19, whether they test positive (63 people) or the result is a false negative (7 people). So, the probability of the hypothesis (that a person in the sample has COVID-19) is 70%.

- P(E|H)

$$\mathrm{P(E|H)} = \frac{63}{70} \times 100 = 90\%$$

Explanation
P(E|H) is the probability of the "evidence" (a positive test) given the "hypothesis" (a person has COVID-19). It is the probability a person with COVID-19 tests positive when they are indeed infected. 63 people infected tested positive, so these are true positives. Seven people infected were missed and so these are false negatives.

P(E|H) answers the question: *How likely is the test accurate when the result is negative*? It answers this question by *contrasting true positives with false negatives*.

The probability of a correct positive test for COVID-19 in this sample is 90% (true positives), so 10% of cases are being missed (false negatives). This gives us an idea of how sensitive the test is or the ***extent to which a negative result is reliable***.

This can also help answer a critical medical question about the test: ***Should we retest someone when the result is negative?*** Missing 10% of COVID-19 cases is a strong enough reason to suggest that people who test negative but have symptoms still should be cautious. They should probably quarantine at home and wait to see how their symptoms develop. If they feel unwell, they should be retested.

As another example, in the case of HIV, medical staff recommend that people continue to test every year even if they previously had negative results. This is partly due to the fact that false negative results for HIV tests sometimes occur.

- P(E)

$$\mathrm{P(E)} = 63 + 1 = 64$$

Explanation
P(E) is the probability that a person in the sample tests positive for COVID-19, whether the result is accurate (63 people) or a false positive (1 person). So, the probability based on the test evidence (that a person in the sample has COVID-19) alone is 64%.

- P(H|E)

$$\mathrm{P(H|E)} = \frac{63}{64} \times 100 = 98\%$$

Explanation
P(H|E) means the probability of the “hypothesis” (someone has COVID-19) given the “evidence” (a positive test). Of the 64 people who tested positive, 63 actually have COVID-19 (true positive). One does not have COVID-19, so this was a false positive.

P(H|E) is 98% here, and so in 2% of cases, the test will show a false positive. This equation thus shows the ***extent to which a positive result is reliable***.

P(E|H) answers the question: *How likely is the test accurate when the result is positive*? It answers this question by *contrasting true positives with false positives*.

Knowing the probability that a person who tests positive actually has COVID-19 answers a critical medical question about the test: ***Should we retest someone when the result is positive?*** The low chance of a test being a false positive (2%) probably means doctors will not generally recommend a retest for a positive COVID-19 result. However, if there are circumstances that call for it, such as an absence of symptoms for a long period or the person really doubts the results, asking for a second test as confirmation is likely justified.

Medical staff need to consider how often a retest after a positive result is necessary based on the potential for test failure. As another example, this is often done in the case of an HIV-positive result so that the person is not needlessly put on treatment.

References

Ellinor, A., Williams, C., Strandberg, A. et al. (n.d.). Bayes' Theorem and Conditional Probability. Brilliant. https://brilliant.org/wiki/bayes-theorem

HyperionDev. (2025). Statistics. Course materials. Private repository, GitHub.

Joyce, J. (2021). Bayes' Theorem. *The Stanford Encyclopedia of Philosophy.* https://plato.stanford.edu/archives/fall2021/entries/bayes-theorem

Mathematics Exchange. (2015). Independence of Coin Tosses. https://math.stackexchange.com/questions/1431656/independence-of-coin-tosses

Schnipper, J. & Sax, P. (2020). Covid-19 test accuracy supplement: The math of Bayes' Theorem. Stat News. https://www.statnews.com/2020/08/20/covid-19-test-accuracy-supplement-the-math-of-bayes-theorem

Scott, N. (2021). Varsity Explains: Bayes' Theorem and COVID-19 testing. https://www.varsity.co.uk/science/21149

Notes on Assistance

I struggled with this task quite a bit and thus added additional explanations that I needed to make sense of the material. I asked a friend who studied actuarial science for assistance. She helped me see that for question 1, I was making things too complicated and so getting confused. For question 2, I did not initially understand the point of Bayesian statistics in terms of medical testing.

I also used Perplexity AI (https://www.perplexity.ai) to understand the real difference between P(E|H) and P(H|E), which I was confused by. It led me to the articles referenced above that helped. Writing descriptions of the meaning of each part of the equations and examples in my own words then helped me understand better.