

# Session 1.2: Bayes Theorem

Bayesian modelling for Spatial and Spatio-temporal data, Imperial College

# Learning objectives

After this lecture you should be able to

- Distinguish between conditional probability and likelihood
- Compute joint and conditional probabilities
- Use Bayes theorem to obtain posterior probabilities

The topics treated in this lecture are presented in Chapter 3 of the book Blangiardo and Cameletti (2015) and in Chapter 2 of the book Johnson, Ott, and Dogucu (2022).

# Outline

1. Conditional probability and likelihood
2. Normalising constant
3. Bayes Theorem

# Conditional probability and likelihood

## Example: COVID-19 test

- A COVID-19 test has shown to have 80% sensitivity and 99% specificity
- In England, COVID-19 prevalence is 6%

What is the chance that a patient testing positive actually does have COVID-19?

# Example: COVID-19 test

- A COVID-19 test has shown to have 80% sensitivity and 99% specificity
- In England, COVID-19 prevalence is 6%

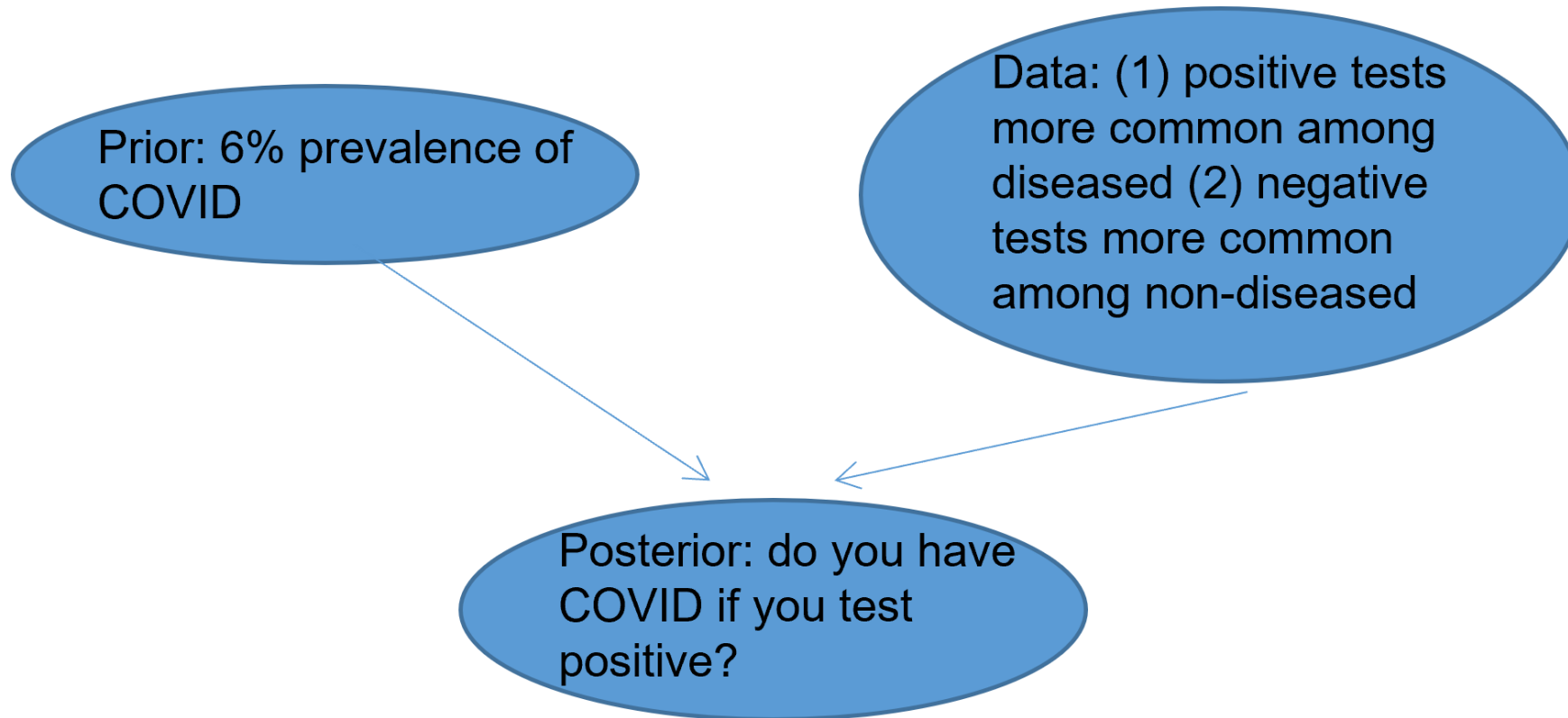
What is the chance that a patient testing positive actually does have COVID-19?

We have two pieces of information:

- 1 Our prior suggests that the COVID-19 prevalence in the country is low (6%)
- 2 Our data suggest that our diagnostic test is accurate

# Example: COVID-19 test

How can we balance these two pieces of information to answer the question about having the disease?



# Prior probability model

- Let's look at our prior: COVID-19 prevalence in the country is 6%. How can we formalise it?

Define A as the event: **a person has COVID-19 in England**

Then  $P(A) = 0.06$  and consequently  $P(A^C) = 0.94$ .

Remember that a valid probability model must:

- 1 account for all possible events (having or not having COVID-19);
- 2 assign prior probabilities to each event.

Also

- 1 each probability must be between 0 and 1;
- 2 these probabilities must sum to one.



# Conditional probability

Now we summarise the **data** that we get from the diagnostic tests:

- 80% sensitivity: if a person has COVID-19 they will test positive 80 out of 100 times
- 99% specificity: if a person does not have COVID-19 they will test negative 99 out of 100 times

# Conditional probability

Now we summarise the **data** that we get from the diagnostic tests:

- 80% sensitivity: if a person has COVID-19 they will test positive 80 out of 100 times
- 99% specificity: if a person does not have COVID-19 they will test negative 99 out of 100 times

These are **conditional probabilities** and defining B the event: a person tests positive for COVID-19, we can summarise the above information in

$$P(B \mid A) = 0.8 \text{ and } P(B \mid A^C) = 0.01$$

# Some rules of conditional probabilities

In general, comparing the conditional vs unconditional probabilities  $P(B \mid A)$  vs  $P(B)$ , reveals the extent to which information about  $B$  changes in light of  $A$

In some cases, the certainty of an event  $B$  might increase in light of new data  $A$ :

- if you eat hamburgers and chips every day, your probability of having high cholesterol is higher than in the general population

$$P(B \mid A) > P(B)$$

# Some rules of conditional probabilities

In general, comparing the conditional vs unconditional probabilities  $P(B \mid A)$  vs  $P(B)$ , reveals the extent to which information about  $B$  changes in light of  $A$

In some cases, the certainty of an event  $B$  might increase in light of new data  $A$ :

- if you eat hamburgers and chips every day, your probability of having high cholesterol is higher than in the general population

$$P(B \mid A) > P(B)$$

In some cases, the certainty of an event  $B$  might decrease in light of new data  $A$ :

- if you are vaccinated against flu, your probability of getting into hospital with serious flu complications decreases

$$P(B \mid A) < P(B)$$

# Some rules of conditional probabilities

In general, comparing the conditional vs unconditional probabilities  $P(B \mid A)$  vs  $P(A)$ , reveals the extent to which information about  $B$  changes in light of  $A$

In some cases, the certainty of an event  $B$  might increase in light of new data  $A$ :

- if you eat hamburgers and chips every day, your probability of having high cholesterol is higher than in the general population

$$P(B \mid A) > P(B)$$

In some cases, the certainty of an event  $B$  might decrease in light of new data  $A$ :

- if you are vaccinated against flu, your probability of getting into hospital with serious flu complications decreases

$$P(B \mid A) < P(B)$$

The order of conditioning is also important, as generally  $P(B \mid A) \neq P(A \mid B)$ : for instance in India the probability of getting bitten by a snake after a week of torrential rain is  $P(B \mid A) = 0.4$ ; but this does not mean that there is a 0.4 probability of a week of torrential rain after someone is bitten by a snake  $P(A \mid B)$ .

# Some rules of conditional probabilities

In general, comparing the conditional vs unconditional probabilities  $P(B \mid A)$  vs  $P(B)$ , reveals the extent to which information about  $B$  changes in light of  $A$

In some cases, the certainty of an event  $B$  might increase in light of new data  $A$ :

- if you eat hamburgers and chips every day, your probability of having high cholesterol is higher than in the general population

$$P(B \mid A) > P(B)$$

In some cases, the certainty of an event  $B$  might decrease in light of new data  $A$ :

- if you are vaccinated against flu, your probability of getting into hospital with serious flu complications decreases

$$P(B \mid A) < P(B)$$

The order of conditioning is also important, as generally  $P(B \mid A) \neq P(A \mid B)$ : for instance in India the probability of getting bitten by a snake after a week of torrential rain is  $P(B \mid A) = 0.4$ ; but this does not mean that there is a 0.4 probability of a week of torrential rain after someone is bitten by a snake  $P(A \mid B)$ .

Finally information about  $A$  does not always change our understanding of  $B$ : then the two events are **independent** and  $P(B \mid A) = P(B)$

# Some rules of conditional probabilities

In general, comparing the conditional vs unconditional probabilities  $P(B \mid A)$  vs  $P(B)$ , reveals the extent to which information about  $B$  changes in light of  $A$

In some cases, the certainty of an event  $B$  might increase in light of new data  $A$ :

- if you eat hamburgers and chips every day, your probability of having high cholesterol is higher than in the general population

$$P(B \mid A) > P(B)$$

In some cases, the certainty of an event  $B$  might decrease in light of new data  $A$ :

- if you are vaccinated against flu, your probability of getting into hospital with serious flu complications decreases

$$P(B \mid A) < P(B)$$

The order of conditioning is also important, as generally  $P(B \mid A) \neq P(A \mid B)$ : for instance in India the probability of getting bitten by a snake after a week of torrential rain is  $P(B \mid A) = 0.4$ ; but this does not mean that there is a 0.4 probability of a week of torrential rain after someone is bitten by a snake  $P(A \mid B)$ .

Finally information about  $A$  does not always change our understanding of  $B$ : then the two events are **independent** and  $P(B \mid A) = P(B)$

See recording 2 for an additional recap on Probability

# Some rules of conditional probabilities

- Provable from probability axioms

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

- If  $A_i$  is a set of mutually exclusive and exhaustive events (i.e.  $A_i \cap A_j = \emptyset$ ,  $P(\bigcup_i A_i) = \sum_i P(A_i) = 1$ ), then

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_j P(B|A_j)P(A_j)}$$



# Likelihood

Let's re-examine the COVID-19 example: we know that if someone has the disease their probability of testing positive is much higher than if they do not (0.8 vs 0.01), so what we think is that the probability of having the disease if testing positive must be high.

We are moving unconsciously from conditional probability to likelihood.

# Likelihood

Let's re-examine the COVID-19 example: we know that if someone has the disease their probability of testing positive is much higher than if they do not (0.8 vs 0.01), so what we think is that the probability of having the disease if testing positive must be high.

We are moving unconsciously from conditional probability to likelihood.

When  $A$  is known, the conditional probability function  $P(. | A)$  allows to compute the probabilities of an unknown event  $B$  or  $B^C$ :

$$P(B | A) \text{ vs } P(B^C | A)$$

When  $B$  is known (i.e. observed), the likelihood function  $L(. | B)$  allows to evaluate the relative compatibility of the data  $B$  with the event  $A$  or  $A^C$ :

$$L(A | B) = P(B | A) \text{ vs } L(A^C | B) = P(B | A^C)$$

# Likelihood

Let's re-examine the COVID-19 example: we know that if someone has the disease their probability of testing positive is much higher than if they do not (0.8 vs 0.01), so what we think is that the probability of having the disease if testing positive must be high.

We are moving unconsciously from conditional probability to likelihood.

When  $A$  is known, the conditional probability function  $P(. | A)$  allows to compute the probabilities of an unknown event  $B$  or  $B^C$ :

$$P(B | A) \text{ vs } P(B^C | A)$$

When  $B$  is known (i.e. observed), the likelihood function  $L(. | B)$  allows to evaluate the relative compatibility of the data  $B$  with the event  $A$  or  $A^C$ :

$$L(A | B) = P(B | A) \text{ vs } L(A^C | B) = P(B | A^C)$$

So far we have (i) the prior evidence of getting COVID-19 and (ii) the likelihood which tells that a positive test is more likely among diseased people:

Table: Prior probability and likelihood

Event	A	$A^C$	Total
Prior	0.06	0.96	1.00
Likelihood	0.80	0.01	0.81

# Normalising constant

# Normalising constant

The marginal probability of testing positive  $P(B)$  provides an important point of comparison. This is the last bit of information we need. Let's try to fill in the table below:

Event	B	B <sup>C</sup>	Total
A			0.06
A <sup>C</sup>			0.94
Total			1

First let's look at the row A: there are those who tests positive AND have the disease and those who do not test positive AND have the disease. To get these probabilities remember that  $P(A) = 0.06$  and that  $P(B \mid A) = 0.8$ , so that

- 1  $P(A \cap B) = 0.06 \times 0.8 = 0.048$  Then using a similar rationale we can get:
- 2  $P(A \cap B^C) = P(B^C \mid A) \times P(A) = (1 - 0.8) \times 0.06 = 0.012$
- 3  $P(A^C \cap B) = P(B \mid A^C) \times P(A^C) = 0.01 \times 0.94 = 0.0094$
- 4  $P(A^C \cap B^C) = P(B^C \mid A^C) \times P(A^C) = 0.99 \times 0.94 = 1 - 0.048 - 0.012 - 0.0094 = 0.931$

And finally the  $P(B) = 0.048 + 0.0094 = 0.057$

# Normalising constant

The marginal probability of testing positive  $P(B)$  provides an important point of comparison. This is the last bit of information we need. Let's try to fill in the table below:

Event	B	B <sup>C</sup>	Total
A	0.048	0.012	0.06
A <sup>C</sup>	0.009	0.931	0.94
Total	0.057	0.943	1

First let's look at the row A: there are those who tests positive AND have the disease and those who do not test positive AND have the disease. To get these probabilities remember that  $P(A) = 0.06$  and that  $P(B \mid A) = 0.8$ , so that

- 1  $P(A \cap B) = 0.06 \times 0.8 = 0.048$  Then using a similar rationale we can get:
- 2  $P(A \cap B^C) = P(B^C \mid A) \times P(A) = (1 - 0.8) \times 0.06 = 0.012$
- 3  $P(A^C \cap B) = P(B \mid A^C) \times P(A^C) = 0.01 \times 0.94 = 0.009$
- 4  $P(A^C \cap B^C) = P(B^C \mid A^C) \times P(A^C) = 0.99 \times 0.94 = 1 - 0.048 - 0.012 - 0.0094 = 0.931$

And finally the  $P(B) = 0.048 + 0.0094 = 0.057$

# Bayes Theorem

# Now we put all together...

We are now ready to answer the question:

What is the chance that a patient testing positive actually does have COVID-19?

Going back to the table we can zoom in into the people testing positive

Event	B	B <sup>c</sup>	Total
A	0.048	0.012	0.06
A <sup>c</sup>	0.009	0.931	0.94
Total	0.057	0.943	1

and using the conditional probability rules we get

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = 0.048/0.057 = 0.84$$

**This is building Bayes Theorem from scratch.**

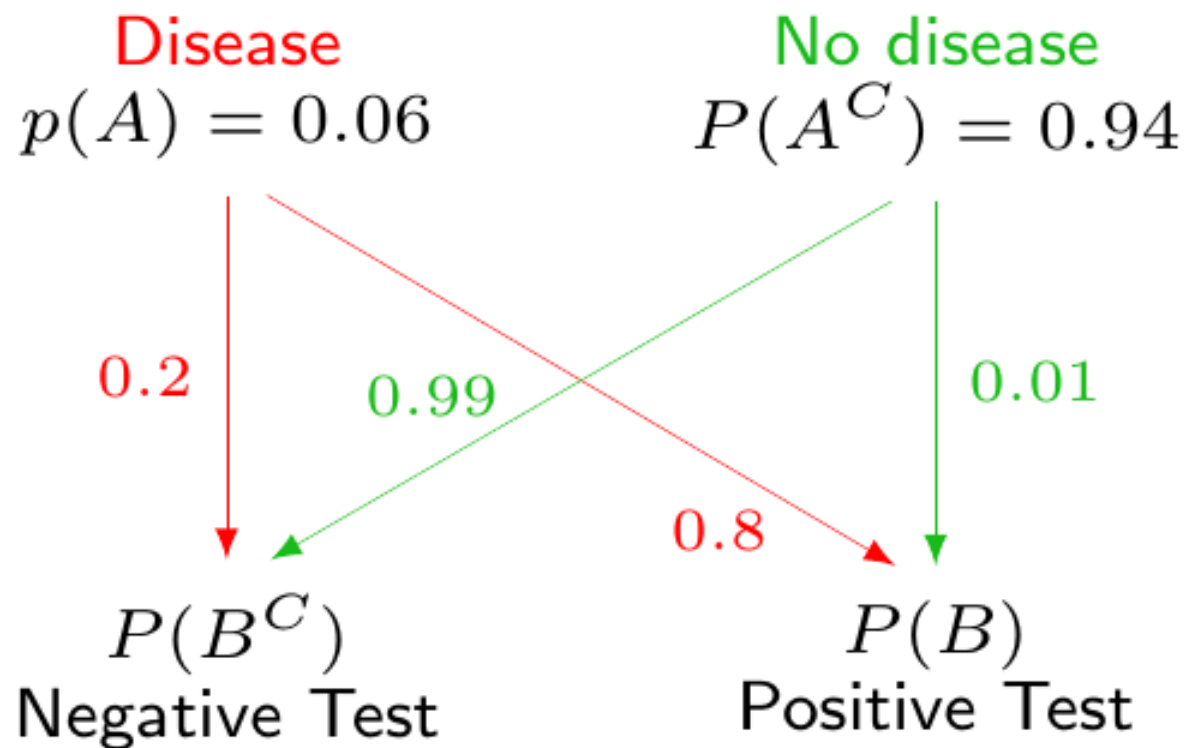
Now remember that

$$P(A \cap B) = P(B \mid A)P(A)$$

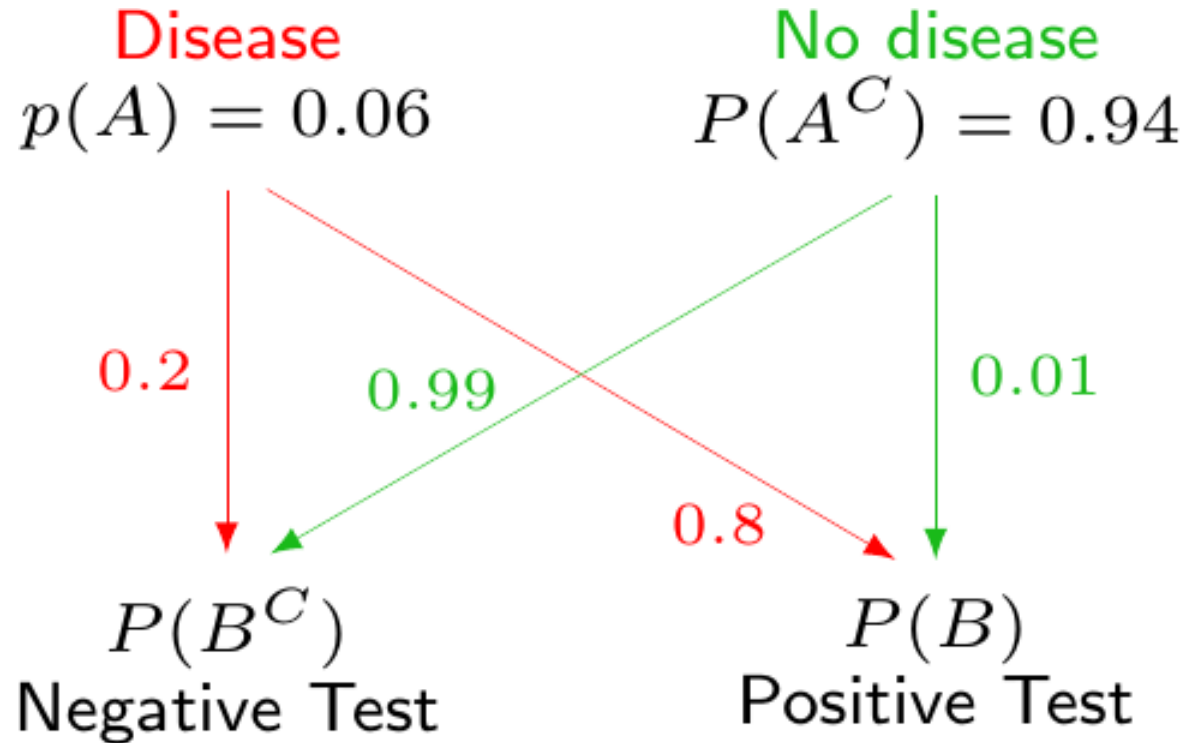
then Bayes' theorem will calculate  $P(A \mid B)$  combining information from the prior  $P(A)$  and the likelihood of observing  $A$  with the event  $B$ , given by  $P(B \mid A)$



# Does it really work?



# Does it really work?

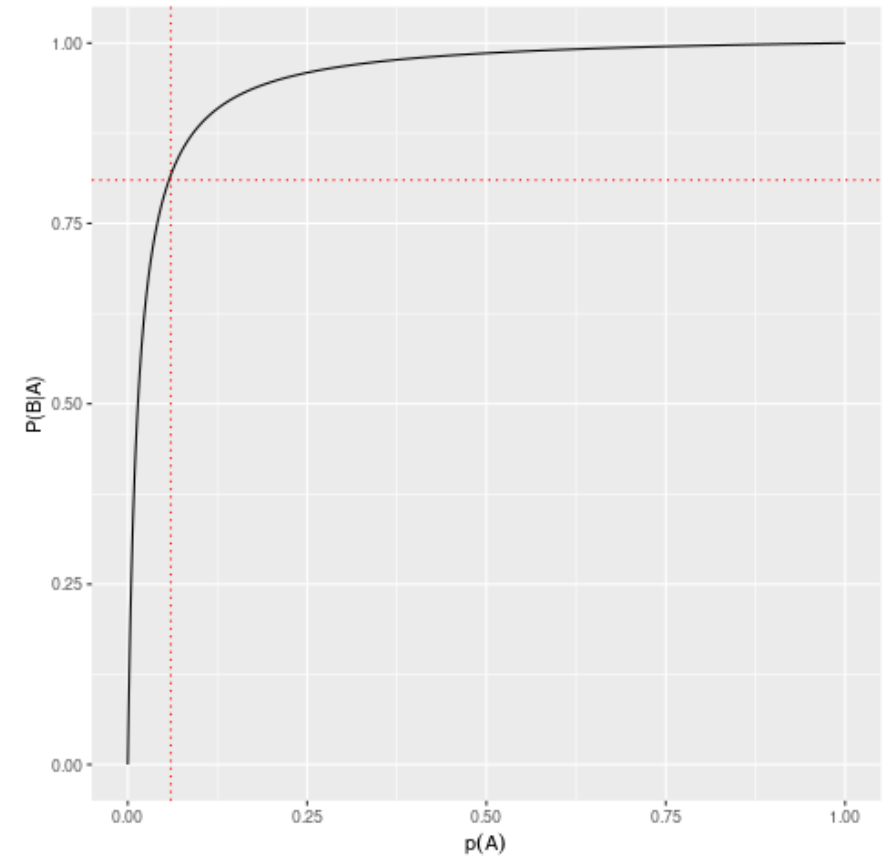


Using Bayes' theorem we get

$$p(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^C)P(A^C)} = \frac{0.8 \times 0.06}{0.8 \times 0.06 + 0.01 \times 0.94} = 0.84$$

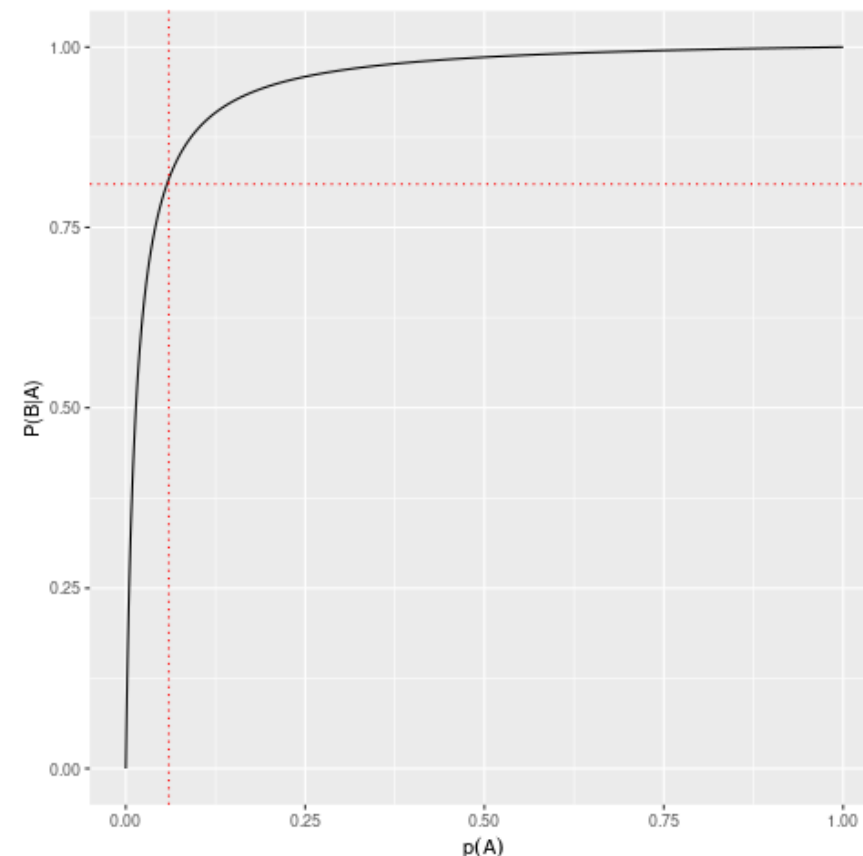
# Comments

- The disease prevalence can be thought of as a *prior* probability ( $p = 0.06$ )
- Observing a positive result causes us to modify this probability to  $p = 0.84$ . This is our *posterior* probability that patient is COVID-19 positive.



# Comments

- The disease prevalence can be thought of as a *prior* probability ( $p = 0.06$ )
- Observing a positive result causes us to modify this probability to  $p = 0.84$ . This is our *posterior* probability that patient is COVID-19 positive.



- Bayes theorem applied to *observables* (as in diagnostic testing) is uncontroversial and established
- More controversial in general statistical analyses: *parameters* are unknown quantities, and prior distributions need to be specified → **Bayesian inference**
- Stay tuned, we are going to dive into that next week!

# References

- Blangiardo, M. and M. Cameletti (2015). *Spatial and spatio-temporal Bayesian models with R-INLA*. John Wiley & Sons.
- Johnson, A. A., M. Q. Ott, and M. Dogucu (2022). *Bayes Rules!: An Introduction to Applied Bayesian Modeling*. CRC Press.