Bayesian Inference and Hypothesis Testing

Christos Dimitrakakis

October 28, 2024

Bayes theorem

- Recall the definition of Conditional probability:

$$P(A|B) = P(A \cap B)/P(B)$$

i.e. the probability of A happening if B happens.

Bayes theorem

- Recall the definition of Conditional probability:

$$P(A|B) = P(A \cap B)/P(B)$$

i.e. the probability of A happening if B happens.

- It is also true that:

$$P(B|A) = P(A \cap B)/P(A)$$

Bayes theorem

- Recall the definition of Conditional probability:

$$P(A|B) = P(A \cap B)/P(B)$$

i.e. the probability of A happening if B happens.

- It is also true that:

$$P(B|A) = P(A \cap B)/P(A)$$

- Combining the two equations, reverse the conditioning:

$$P(A|B) = P(B|A)P(A)/P(B)$$

Bayes theorem

- Recall the definition of Conditional probability:

$$P(A|B) = P(A \cap B)/P(B)$$

  i.e. the probability of A happening if B happens.
- It is also true that:

$$P(B|A) = P(A \cap B)/P(A)$$

- Combining the two equations, reverse the conditioning:

$$P(A|B) = P(B|A)P(A)/P(B)$$

- So we can reverse the order of conditioning, i.e. relate to the probability of A given B to that of B given A.

The cards problem

1. Print out a number of cards, with either [A|A], [A|B] or [B|B] on their sides.
2. If you have an A, what is the probability of an A on the other side?
3. Have the students perform the experiment with:
    - 3.1 Draw a random card.
    - 3.2 Count the number of people with A.
    - 3.3 What is the probability that somebody with an A on one side will have an A on the other?
    - 3.4 Half of the people should have an A?

The cards problem

1. Print out a number of cards, with either [A|A], [A|B] or [B|B] on their sides.
2. If you have an A, what is the probability of an A on the other side?
3. Have the students perform the experiment with:
    3.1 Draw a random card.
    3.2 Count the number of people with A.
    3.3 What is the probability that somebody with an A on one side will have an A on the other?
    3.4 Half of the people should have an A?

The prior and posterior probabilities

A	A	2/6	A observed	2/3
A	B	1/6	A observed	1/3
B	A	1/6		
B	B	2/6		

The murder problem

- A murder occurred in a house over Christmas. There were n people inside, plus the victim. Person X, the victim's son, is accused of a murder.

The murder problem

- A murder occurred in a house over Christmas. There were n people inside, plus the victim. Person X, the victim's son, is accused of a murder.
- There are two possibilities:

The murder problem

- A murder occurred in a house over Christmas. There were n people inside, plus the victim. Person X, the victim's son, is accused of a murder.
- There are two possibilities:
  - H_0: They are innocent.

The murder problem

- A murder occurred in a house over Christmas. There were n people inside, plus the victim. Person X, the victim's son, is accused of a murder.
- There are two possibilities:
  - H_0: They are innocent.
  - H_1: They are guilty.

The murder problem

- A murder occurred in a house over Christmas. There were n people inside, plus the victim. Person X, the victim's son, is accused of a murder.
- There are two possibilities:
  - H_0: They are innocent.
  - H_1: They are guilty.

The murder problem

- A murder occurred in a house over Christmas. There were n people inside, plus the victim. Person X, the victim's son, is accused of a murder.
- There are two possibilities:
  - H_0: They are innocent.
  - H_1: They are guilty.

What is your belief that they have committed the crime?

The murder problem

- A murder occurred in a house over Christmas. There were n people inside, plus the victim. Person X, the victim's son, is accused of a murder.
- There are two possibilities:
  - H_0: They are innocent.
  - H_1: They are guilty.

What is your belief that they have committed the crime?

Prior elicitation

- All those that think the accused is guilty, raise their hand.

The murder problem

- A murder occurred in a house over Christmas. There were n people inside, plus the victim. Person X, the victim's son, is accused of a murder.
- There are two possibilities:
  - H_0: They are innocent.
  - H_1: They are guilty.

What is your belief that they have committed the crime?

Prior elicitation

- All those that think the accused is guilty, raise their hand.
- Divide by the number of people in class

The murder problem

- A murder occurred in a house over Christmas. There were n people inside, plus the victim. Person X, the victim's son, is accused of a murder.
- There are two possibilities:
  - H_0: They are innocent.
  - H_1: They are guilty.

What is your belief that they have committed the crime?

Prior elicitation

- All those that think the accused is guilty, raise their hand.
- Divide by the number of people in class
- Let us call this $P(H_1)$.

The murder problem

- A murder occurred in a house over Christmas. There were n people inside, plus the victim. Person X, the victim's son, is accused of a murder.
- There are two possibilities:
  - H_0: They are innocent.
  - H_1: They are guilty.

What is your belief that they have committed the crime?

Prior elicitation

- All those that think the accused is guilty, raise their hand.
- Divide by the number of people in class
- Let us call this $P(H_1)$.
- This is a purely subjective measure!

DNA test

- Let us now do a DNA test on the suspect

DNA test

- Let us now do a DNA test on the suspect

DNA test properties

- D: Test is positive

DNA test

- Let us now do a DNA test on the suspect

DNA test properties

- D: Test is positive
- $P(D|H_0) = 10\%$: False positive rate

DNA test

- Let us now do a DNA test on the suspect

DNA test properties

- D: Test is positive
- $P(D|H_0) = 10\%$: False positive rate
- $P(D|H_1) = 100\%$: True positive rate

DNA test

- Let us now do a DNA test on the suspect

DNA test properties

- D: Test is positive
- $P(D|H_0) = 10\%$: False positive rate
- $P(D|H_1) = 100\%$: True positive rate

DNA test

- Let us now do a DNA test on the suspect

DNA test properties

- D: Test is positive
- $P(D|H_0) = 10\%$: False positive rate
- $P(D|H_1) = 100\%$: True positive rate

Run the test

- The result is either positive or negative ($\neg D$).

DNA test

- Let us now do a DNA test on the suspect

DNA test properties

- D: Test is positive
- $P(D|H_0) = 10\%$: False positive rate
- $P(D|H_1) = 100\%$: True positive rate

Run the test

- The result is either positive or negative ($\neg D$).
- What is your belief now that the suspect is guilty?

Everybody is a suspect

- Run a DNA test on everybody in the house.

Everybody is a suspect

- Run a DNA test on everybody in the house.
- What is different from before?

Everybody is a suspect

- Run a DNA test on everybody in the house.
- What is different from before?
- Who has a positive test?

Everybody is a suspect

- Run a DNA test on everybody in the house.
- What is different from before?
- Who has a positive test?
- What is your belief that the people with the positive test are guilty?

Explanation

- Prior: $P(H_i)$.

$$
\begin{aligned}
P(D) &= P(D \cap H_0) + P(D \cap H_1) && (1)\\
&= P(D|H_0)P(H_0) + P(D|H_1)P(H_1) && (2)
\end{aligned}
$$

- Posterior: $P(H_0|D) = \frac{P(D|H_0)P(H_0)}{P(D|H_0)P(H_0)+P(D|H_1)P(H_1)}$
- Assuming $P(D|H_1) = 1$, and setting $P(H_0) = q$, this gives

$$
P(H_0|D) = \frac{0.1q}{0.1q + 1 - q} = \frac{q}{10 - 9q}
$$

- The posterior can always be updated with more data!

Explanation

- Prior: $P(H_i)$.
- Likelihood $P(D|H_i)$.

$$
\begin{align}
P(D) &= P(D \cap H_0) + P(D \cap H_1) \tag{1}\\
&= P(D|H_0)P(H_0) + P(D|H_1)P(H_1) \tag{2}
\end{align}
$$

- Posterior: $P(H_0|D) = \frac{P(D|H_0)P(H_0)}{P(D|H_0)P(H_0)+P(D|H_1)P(H_1)}$
- Assuming $P(D|H_1) = 1$, and setting $P(H_0) = q$, this gives

$$
P(H_0|D) = \frac{0.1q}{0.1q + 1 - q} = \frac{q}{10 - 9q}
$$

- The posterior can always be updated with more data!

Explanation

- Prior: $P(H_i)$.
- Likelihood $P(D|H_i)$.
- Posterior: $P(H_i|D) = P(D \cap H_i)/P(D) = P(D|H_i)P(H_i)/P(D)$

$$P(D) = P(D \cap H_0) + P(D \cap H_1) \tag{1}$$
$$= P(D|H_0)P(H_0) + P(D|H_1)P(H_1) \tag{2}$$

- Posterior: $P(H_0|D) = \frac{P(D|H_0)P(H_0)}{P(D|H_0)P(H_0)+P(D|H_1)P(H_1)}$
- Assuming $P(D|H_1) = 1$, and setting $P(H_0) = q$, this gives

$$P(H_0|D) = \frac{0.1q}{0.1q + 1 - q} = \frac{q}{10 - 9q}$$

- The posterior can always be updated with more data!

Explanation

- Prior: $P(H_i)$.
- Likelihood $P(D|H_i)$.
- Posterior: $P(H_i|D) = P(D \cap H_i)/P(D) = P(D|H_i)P(H_i)/P(D)$
- Marginal probability:

$$\begin{aligned} P(D) &= P(D \cap H_0) + P(D \cap H_1) \qquad (1) \\ &= P(D|H_0)P(H_0) + P(D|H_1)P(H_1) \qquad (2) \end{aligned}$$

- Posterior: $P(H_0|D) = \frac{P(D|H_0)P(H_0)}{P(D|H_0)P(H_0)+P(D|H_1)P(H_1)}$
- Assuming $P(D|H_1) = 1$, and setting $P(H_0) = q$, this gives

$$P(H_0|D) = \frac{0.1q}{0.1q + 1 - q} = \frac{q}{10 - 9q}$$

- The posterior can always be updated with more data!

Python example

```
def get_posterior(prior, data, likelihood):
    marginal = prior * likelihood[data][0] + (1 - prior) * likelihoo
    posterior = prior * likelihood[data][0] / marginal
    return posterior

import numpy as np
prior = 0.9 # Pr(H1)
likelihood = np.zeros([2, 2])
likelihood[0][0] = 0.9 # Pr(F|H0)
likelihood[1][0] = 0.1 # Pr(T|H0)
likelihood[0][1] = 0 # Pr(F|H1)
likelihood[1][1] = 1 # Pr(T|H1)
data = 1
return get_posterior(prior, data, likelihood)
```

Types of hypothesis testing problems

Simple Hypothesis Test

Example: DNA evidence, Covid tests

- Two hypothesese H_0, H_1
- $P(D|H_i)$ is defined for all i

Multiple Hypotheses Test

Example: Model selection

- H_i: One of many mutually exclusive models
- $P(D|H_i)$ is defined for all i

Null Hypothesis Test

Example: Are men's and women's heights the same?

- H_0: The 'null' hypothesis
- $P(D|H_0)$ is defined
- The alternative is undefined

Pitfalls

Problem definition

- Defining the models $P(D|H_i)$ incorrectly.

Pitfalls

Problem definition

- Defining the models $P(D|H_i)$ incorrectly.
- Using an "unreasonable" prior $P(H_i)$

Pitfalls

Problem definition

- Defining the models $P(D|H_i)$ incorrectly.
- Using an "unreasonable" prior $P(H_i)$

Pitfalls

Problem definition

- Defining the models $P(D|H_i)$ incorrectly.
- Using an "unreasonable" prior $P(H_i)$

The garden of many paths

- Having a huge hypothesis space

Pitfalls

Problem definition

- Defining the models $P(D|H_i)$ incorrectly.
- Using an "unreasonable" prior $P(H_i)$

The garden of many paths

- Having a huge hypothesis space
- Selecting the relevant hypothesis after seeing the data