Bayes' theorem

Chrysafis Vogiatzis

Department of Industrial and Enterprise Systems Engineering
University of Illinois at Urbana-Champaign

Lecture 4

ILLINOIS
ISE | Industrial & Enterprise Systems Engineering
GRAINGER COLLEGE OF ENGINEERING

Last time..

- **probability of union of two events**:

$$\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$$

  - $P(A \cup B) = P(A) + P(B)$, if A and B are mutually exclusive events.

- **conditional probability**:

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

- **independent events**: two events A and B are independent if $P(A|B) = P(A)$ and $P(B|A) = P(B)$.

- **multiplication rule for probabilities**:

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

  - $P(A \cap B) = P(A) \cdot P(B)$, if A and B are independent events.

ILLINOIS

Multiplication rule: example

Example

A student is registered for two graded and co-requisite courses, where the first course helps quite significantly with the second one. This means that an A in the first course, increases the chances of getting an A in the second one to 80%; not getting an A in the first class leaves their chances at the second class the same. Historically, the student has received an A with probability 30% in their classes.

Let A_1 be the event that the student gets an A in the first class and A_2 that they get an A in the second class.

- What is the probability that they get an A in the first class?

  $P(A_1) = 30\%.$

- What is the probability that they get an A in the second class given that they have gotten an A in the first one?

  $P(A_2|A_1) = 80\%.$

- What is the probability that they get an A in both classes?

Using the multiplication rule:

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2|A_1) = 24\%.$$

ILLINOIS

Multiplication rule: example

Example

A student is registered for two graded and co-requisite courses, where the first course helps quite significantly with the second one. This means that an A in the first course, increases the chances of getting an A in the second one to 80%; not getting an A in the first class leaves their chances at the second class the same. Historically, the student has received an A with probability 30% in their classes.

Let A_1 be the event that the student gets an A in the first class and A_2 that they get an A in the second class.

- What is the probability that they get an A in the first class?

  $P(A_1) = 30\%.$

- What is the probability that they get an A in the second class given that they have gotten an A in the first one?

  $P(A_2|A_1) = 80\%.$

- What is the probability that they get an A in both classes?

Using the multiplication rule:

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2|A_1) = 24\%.$$

Multiplication rule: example

Example

A student is registered for two graded and co-requisite courses, where the first course helps quite significantly with the second one. This means that an A in the first course, increases the chances of getting an A in the second one to 80%; not getting an A in the first class leaves their chances at the second class the same. Historically, the student has received an A with probability 30% in their classes.

Let A_1 be the event that the student gets an A in the first class and A_2 that they get an A in the second class.

- What is the probability that they get an A in the first class?

  $P(A_1) = 30\%.$

- What is the probability that they get an A in the second class given that they have gotten an A in the first one?

  $P(A_2|A_1) = 80\%.$

- What is the probability that they get an A in both classes?

Using the multiplication rule:

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2|A_1) = 24\%.$$

ILLINOIS

Multiplication rule: example

Example

A student is registered for two graded and co-requisite courses, where the first course helps quite significantly with the second one. This means that an A in the first course, increases the chances of getting an A in the second one to 80%; not getting an A in the first class leaves their chances at the second class the same. Historically, the student has received an A with probability 30% in their classes.

Let A_1 be the event that the student gets an A in the first class and A_2 that they get an A in the second class.

- What is the probability that they get an A in the first class?

  $P(A_1) = 30\%.$

- What is the probability that they get an A in the second class given that they have gotten an A in the first one?

  $P(A_2|A_1) = 80\%.$

- What is the probability that they get an A in both classes?

Using the multiplication rule:

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2|A_1) = 24\%.$$

ILLINOIS

Multiplication rule: example

Example

A student is registered for two graded and co-requisite courses, where the first course helps quite significantly with the second one. This means that an A in the first course, increases the chances of getting an A in the second one to 80%; not getting an A in the first class leaves their chances at the second class the same. Historically, the student has received an A with probability 30% in their classes.

Let A_1 be the event that the student gets an A in the first class and A_2 that they get an A in the second class.

- What is the probability that they get an A in the first class?

  $P(A_1) = 30\%.$

- What is the probability that they get an A in the second class given that they have gotten an A in the first one?

  $P(A_2|A_1) = 80\%.$

- What is the probability that they get an A in both classes?

Using the multiplication rule:

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2|A_1) = 24\%.$$

ILLINOIS

Multiplication rule: example

Example

A student is registered for two graded and co-requisite courses, where the first course helps quite significantly with the second one. This means that an A in the first course, increases the chances of getting an A in the second one to 80%; not getting an A in the first class leaves their chances at the second class the same. Historically, the student has received an A with probability 30% in their classes.

Let A_1 be the event that the student gets an A in the first class and A_2 that they get an A in the second class.

- What is the probability that they get an A in the first class?

$P(A_1) = 30\%.$

- What is the probability that they get an A in the second class given that they have gotten an A in the first one?

$P(A_2|A_1) = 80\%.$

- What is the probability that they get an A in both classes?

Using the multiplication rule:

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2|A_1) = 24\%.$$

ILLINOIS

The law of total probability: motivation

We can take it one step further: **what is the probability the student gets an A in the second class?**

- We can think of this as a tree of parallel universes:

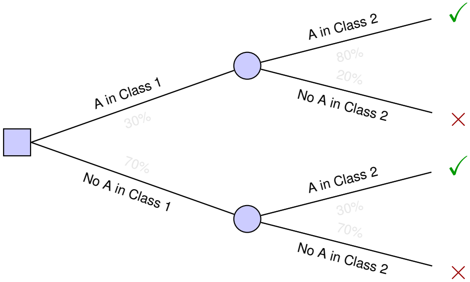
- We can also add the probabilities to the tree to help us.
- The multiplication rule states that we can multiply probabilities along the way to a branch of the tree.
- Adding the two positive outcomes, we have 45%.

ILLINOIS

The law of total probability: motivation

We can take it one step further: **what is the probability the student gets an A in the second class?**

- We can think of this as a tree of parallel universes:

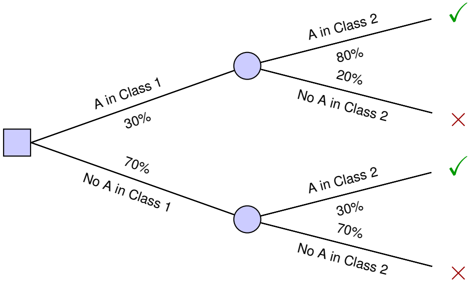


- We can also add the probabilities to the tree to help us.
- The multiplication rule states that we can multiply probabilities along the way to a branch of the tree.
- Adding the two positive outcomes, we have 45%.

The law of total probability: motivation

We can take it one step further: **what is the probability the student gets an A in the second class?**

- We can think of this as a tree of parallel universes:

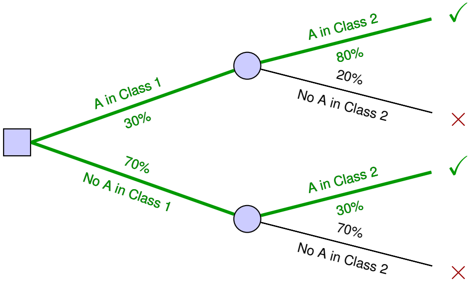


- We can also add the probabilities to the tree to help us.
- The multiplication rule states that we can multiply probabilities along the way to a branch of the tree.
- Adding the two positive outcomes, we have 45%.

The law of total probability: motivation

We can take it one step further: **what is the probability the student gets an A in the second class?**

- We can think of this as a tree of parallel universes:



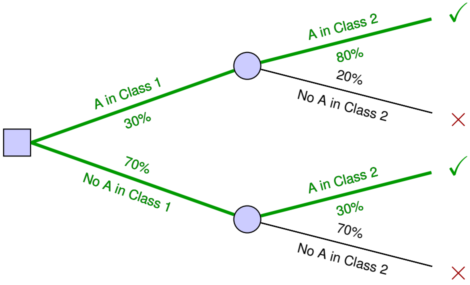
- We can also add the probabilities to the tree to help us.
- The multiplication rule states that we can multiply probabilities along the way to a branch of the tree.
- Adding the two positive outcomes, we have 45%.

ILLINOIS

The law of total probability: motivation

We can take it one step further: **what is the probability the student gets an A in the second class?**

- We can think of this as a tree of parallel universes:



- We can also add the probabilities to the tree to help us.
- The multiplication rule states that we can multiply probabilities along the way to a branch of the tree.
- Adding the two positive outcomes, we have 45%.

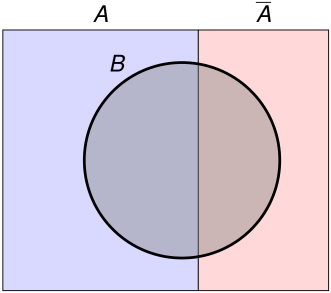
The law of total probability: derivations

- Mutually exclusive events: getting an A in Class 1 and not getting an A in Class 1 are mutually exclusive.
- Collectively exhaustive events: getting an A in Class 1 and not getting an A in Class 1 together describe everything that can happen in Class 1.
- The multiplication rule: we used that to calculate the probability of each possible "path".
- Added them up.

The law of total probability: visually

Consider the events $B \cap A$ and $B \cap \overline{A}$.

- Are they mutually exclusive?
- Is $B = (B \cap A) \cup (B \cap \overline{A})$?
- Based on that: $P(B) = P(B \cap A) + P(B \cap \overline{A})$.
- Finally, from the multiplication rule:
  - $P(B \cap A) = P(A) \cdot P(B|A)$
  - $P\left(B \cap \overline{A}\right) = P(\overline{A}) \cdot P(B|\overline{A})$



The law of total probability for two events:

$$P(B) = P(A) \cdot P(B|A) + P(\overline{A}) \cdot P(B|\overline{A}).$$

Or, in general, for m mutually exclusive and collectively exhaustive events $A_1, A_2, \ldots, A_m$, we have:

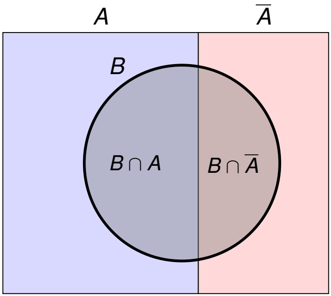
The law of total probability:

$$P(B) = \sum_{i=1}^{m} P(A_i) \cdot P(B|A_i).$$

The law of total probability: visually

Consider the events $B \cap A$ and $B \cap \overline{A}$.

- Are they mutually exclusive?
- Is $B = (B \cap A) \cup \left(B \cap \overline{A}\right)$?
- Based on that: $P(B) = P(B \cap A) + P\left(B \cap \overline{A}\right)$.
- Finally, from the multiplication rule:
  - $P(B \cap A) = P(A) \cdot P(B|A)$
  - $P\left(B \cap \overline{A}\right) = P(\overline{A}) \cdot P(B|\overline{A})$



The law of total probability for two events:

$$P(B) = P(A) \cdot P(B|A) + P(\overline{A}) \cdot P(B|\overline{A}).$$

Or, in general, for m mutually exclusive and collectively exhaustive events $A_1, A_2, \ldots, A_m$, we have:

The law of total probability:

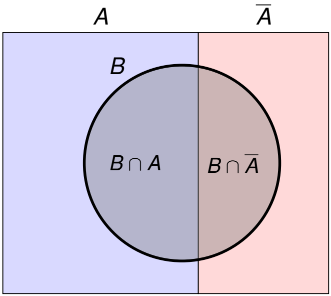
$$P(B) = \sum_{i=1}^{m} P(A_i) \cdot P(B|A_i).$$

ILLINOIS

The law of total probability: visually

Consider the events $B \cap A$ and $B \cap \overline{A}$.

- Are they mutually exclusive?
- Is $B = (B \cap A) \cup \left(B \cap \overline{A}\right)$?
- Based on that:
  $P(B) = P(B \cap A) + P\left(B \cap \overline{A}\right)$.
- Finally, from the multiplication rule:
  - $P(B \cap A) = P(A) \cdot P(B|A)$
  - $P\left(B \cap \overline{A}\right) = P(\overline{A}) \cdot P(B|\overline{A})$



The law of total probability for two events:

$$P(B) = P(A) \cdot P(B|A) + P(\overline{A}) \cdot P(B|\overline{A}).$$

Or, in general, for m mutually exclusive and collectively exhaustive events $A_1, A_2, \ldots, A_m$, we have:

The law of total probability:

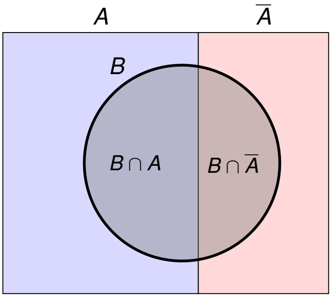
$$P(B) = \sum_{i=1}^{m} P(A_i) \cdot P(B|A_i).$$

ILLINOIS

The law of total probability: visually

Consider the events $B \cap A$ and $B \cap \overline{A}$.

- Are they mutually exclusive?
- Is $B = (B \cap A) \cup \left(B \cap \overline{A}\right)$?
- Based on that:
  $P(B) = P(B \cap A) + P\left(B \cap \overline{A}\right)$.
- Finally, from the multiplication rule:
  - $P(B \cap A) = P(A) \cdot P(B|A)$
  - $P\left(B \cap \overline{A}\right) = P(\overline{A}) \cdot P(B|\overline{A})$



The law of total probability for two events:

$$P(B) = P(A) \cdot P(B|A) + P(\overline{A}) \cdot P(B|\overline{A}).$$

Or, in general, for m mutually exclusive and collectively exhaustive events $A_1, A_2, \ldots, A_m$, we have:

The law of total probability:

$$P(B) = \sum_{i=1}^{m} P(A_i) \cdot P(B|A_i).$$

ILLINOIS

Bayes' theorem: motivating example

Example

Mantoux test:

- *diagnostic tool for TB.*
- *pretty good success rate:*
  - *a person with TB will test positive 99% of the time.*
  - *a person without TB will test negative 99% of the time.*
- *Only 0.05% of the children world population has TB.*

A kid just tested positive in the Mantoux test:

what is the probability they actually do have TB?

ILLINOIS

Bayes' theorem: setup

What we have:

- n states/hypotheses $S_1, S_2, \ldots, S_n$, each with their own probability $P(S_i), i = 1, \ldots, n$.
  - States need to be mutually exclusive.
  - States need to be collectively exhaustive.
  - $P(S_i)$ are also called "prior probabilities".
  - Often comes from our beliefs/biases about the world.
- m test outcomes/observations $O_1, O_2, \ldots, O_m$, each with their own probability of appearing for each state $P(O_j|S_i)$.
  - The test does not imply the state!
  - $P(O_j|S_i)$ are also called "likelihood probabilities".
  - Often come from historical data and previous observations.

What we need:

- $P(S_i|O_j)$: the probability of state/hypothesis S_i to be true, given the observation/evidence O_j.

ILLINOIS

Bayes' theorem: statement

The Bayes' theorem formally states that:

$$P(S_i|O_j) = \frac{P(S_i) \cdot P(O_j|S_i)}{\sum_{k=1}^{n} P(S_k) \cdot P(O_j|S_k)}$$

Let's focus on each term:

- numerator: from the multiplication rule, this is equal to $P(S_i \cap O_j)$.
  - $P(S_i \cap O_j)$ is also termed the joint probability.
- denominator: from the law of total probability, this is equal to $P(O_j)$.
  - $P(O_j)$ is also referred to as the marginal probability.

ILLINOIS

Bayes' theorem: statement

The Bayes' theorem formally states that:

$$P(S_i|O_j) = \frac{P(S_i) \cdot P(O_j|S_i)}{\sum_{k=1}^{n} P(S_k) \cdot P(O_j|S_k)}$$

Let's focus on each term:

- numerator: from the multiplication rule, this is equal to $P(S_i \cap O_j)$.
  - $P(S_i \cap O_j)$ is also termed the joint probability.
- denominator: from the law of total probability, this is equal to $P(O_j)$.
  - $P(O_j)$ is also referred to as the marginal probability.

ILLINOIS

Bayes' theorem: the Mantoux test example

Back to the Mantoux test.

States of the world?

- S_1 : kid has TB;
- S_2 : kid does not have TB.

Possible experiment outcomes?

- O_1 : positive Mantoux result;
- O_2 : negative Mantoux result.

Prior probabilities?

- $P(S_1) = 0.0005$;
- $P(S_2) = 0.9995$.

Likelihood probabilities?

- $P(O_1|S_1) = 0.99$, $P(O_2|S_1) = 0.01$,
- $P(O_1|S_2) = 0.01$, $P(O_2|S_2) = 0.99$,

Bayes' theorem: the Mantoux test example

Back to the Mantoux test.

States of the world?

- S_1 : kid has TB;
- S_2 : kid does not have TB.

Possible experiment outcomes?

- O_1 : positive Mantoux result;
- O_2 : negative Mantoux result.

Prior probabilities?

- $P(S_1) = 0.0005$;
- $P(S_2) = 0.9995$.

Likelihood probabilities?

- $P(O_1|S_1) = 0.99$, $P(O_2|S_1) = 0.01$,
- $P(O_1|S_2) = 0.01$, $P(O_2|S_2) = 0.99$,

Bayes' theorem: the Mantoux test example

Back to the Mantoux test.

States of the world?

- S_1 : kid has TB;
- S_2 : kid does not have TB.

Possible experiment outcomes?

- O_1 : positive Mantoux result;
- O_2 : negative Mantoux result.

Prior probabilities?

- $P(S_1) = 0.0005$;
- $P(S_2) = 0.9995$.

Likelihood probabilities?

- $P(O_1|S_1) = 0.99$, $P(O_2|S_1) = 0.01$,
- $P(O_1|S_2) = 0.01$, $P(O_2|S_2) = 0.99$,

ILLINOIS

Bayes' theorem: the Mantoux test example

Back to the Mantoux test.

States of the world?

- S_1 : kid has TB;
- S_2 : kid does not have TB.

Possible experiment outcomes?

- O_1 : positive Mantoux result;
- O_2 : negative Mantoux result.

Prior probabilities?

- $P(S_1) = 0.0005$;
- $P(S_2) = 0.9995$.

Likelihood probabilities?

- $P(O_1|S_1) = 0.99$, $P(O_2|S_1) = 0.01$,
- $P(O_1|S_2) = 0.01$, $P(O_2|S_2) = 0.99$,

ILLINOIS

Bayes' theorem: the Mantoux test example

Back to the Mantoux test.

States of the world?

- S_1 : kid has TB;
- S_2 : kid does not have TB.

Possible experiment outcomes?

- O_1 : positive Mantoux result;
- O_2 : negative Mantoux result.

Prior probabilities?

- $P(S_1) = 0.0005$;
- $P(S_2) = 0.9995$.

Likelihood probabilities?

- $P(O_1|S_1) = 0.99$, $P(O_2|S_1) = 0.01$,
- $P(O_1|S_2) = 0.01$, $P(O_2|S_2) = 0.99$,

Bayes' theorem: the Mantoux test example

Back to the Mantoux test.

States of the world?

- S_1 : kid has TB;
- S_2 : kid does not have TB.

Possible experiment outcomes?

- O_1 : positive Mantoux result;
- O_2 : negative Mantoux result.

Prior probabilities?

- $P(S_1) = 0.0005$;
- $P(S_2) = 0.9995$.

Likelihood probabilities?

- $P(O_1|S_1) = 0.99$, $P(O_2|S_1) = 0.01$,
- $P(O_1|S_2) = 0.01$, $P(O_2|S_2) = 0.99$,

From Bayes' theorem:

$$P(S_1|O_1)$$

ILLINOIS

Bayes' theorem: the Mantoux test example

Back to the Mantoux test.

States of the world?

- S_1 : kid has TB;
- S_2 : kid does not have TB.

Possible experiment outcomes?

- O_1 : positive Mantoux result;
- O_2 : negative Mantoux result.

Prior probabilities?

- $P(S_1) = 0.0005$;
- $P(S_2) = 0.9995$.

Likelihood probabilities?

- $P(O_1|S_1) = 0.99$, $P(O_2|S_1) = 0.01$,
- $P(O_1|S_2) = 0.01$, $P(O_2|S_2) = 0.99$,

From Bayes' theorem:

$$P(S_1|O_1) = \frac{P(S_1) \cdot P(O_1|S_1)}{P(S_1) \cdot P(O_1|S_1) + P(S_2) \cdot P(O_1|S_2)}$$

ILLINOIS

Bayes' theorem: the Mantoux test example

Back to the Mantoux test.

States of the world?

- S_1 : kid has TB;
- S_2 : kid does not have TB.

Possible experiment outcomes?

- O_1 : positive Mantoux result;
- O_2 : negative Mantoux result.

Prior probabilities?

- $P(S_1) = 0.0005$;
- $P(S_2) = 0.9995$.

Likelihood probabilities?

- $P(O_1|S_1) = 0.99$, $P(O_2|S_1) = 0.01$,
- $P(O_1|S_2) = 0.01$, $P(O_2|S_2) = 0.99$,

From Bayes' theorem:

$$P(S_1|O_1) = \frac{P(S_1) \cdot P(O_1|S_1)}{P(S_1) \cdot P(O_1|S_1) + P(S_2) \cdot P(O_1|S_2)} =$$

$$= \frac{0.0005 \cdot 0.99}{0.0005 \cdot 0.99 + 0.9995 \cdot 0.01}$$

Bayes' theorem: the Mantoux test example

Back to the Mantoux test.

States of the world?

- S_1 : kid has TB;
- S_2 : kid does not have TB.

Possible experiment outcomes?

- O_1 : positive Mantoux result;
- O_2 : negative Mantoux result.

Prior probabilities?

- $P(S_1) = 0.0005$;
- $P(S_2) = 0.9995$.

Likelihood probabilities?

- $P(O_1|S_1) = 0.99$, $P(O_2|S_1) = 0.01$,
- $P(O_1|S_2) = 0.01$, $P(O_2|S_2) = 0.99$,

From Bayes' theorem:

$$P(S_1|O_1) = \frac{P(S_1) \cdot P(O_1|S_1)}{P(S_1) \cdot P(O_1|S_1) + P(S_2) \cdot P(O_1|S_2)} = \frac{0.0005 \cdot 0.99}{0.0005 \cdot 0.99 + 0.9995 \cdot 0.01} = 0.0472.$$