Algoritmos y Programación III

Andrés A. Aristizábal P. aaaristizabal@icesi.edu.co
Angela Villota apvillota@icesi.edu.co

Departamento de Computación y Sistemas Inteligentes

2024-2

Agenda

1. **Reasoning with uncertainty**
   - Probability review

Probability review

Which are the three basic components in **probability models**?

Probability review

Which are the three basic components in **probability models**?

- The set of all possible outcomes, denoted Ω.
- The collection of all sets of outcomes (events), denoted $\mathcal{A}$
- And a probability measure $\mathbb{P}$
- Specification of the triple $(\Omega, \mathcal{A}, \mathbb{P})$ defines the probability space which models a real-world measurement or experimental process.

Probability review

1. Sample Space (Ω)
The set of all possible outcomes, denoted Ω.

- *Example*: Consider flipping a fair coin. The sample space is:

$$\Omega = \{\text{Heads}, \text{Tails}\}$$

2. Event Collection ($\mathcal{A}$)
The collection of all sets of outcomes (events), denoted $\mathcal{A}$.

- *Example*: The event of flipping a head or a tail is represented as:

$$\mathcal{A} = \{\emptyset, \{\text{Heads}\}, \{\text{Tails}\}, \{\text{Heads}, \text{Tails}\}\}$$

  where $\emptyset$ is the empty event (no outcome).

3. Probability Measure ($\mathbb{P}$)
A probability measure $\mathbb{P}$ assigns a probability to each event.

- *Example*: For a fair coin:

$$\mathbb{P}(\text{Heads}) = 0{,}5, \quad \mathbb{P}(\text{Tails}) = 0{,}5$$

4. Probability Space

The triple $(\Omega, \mathcal{A}, \mathbb{P})$ defines the probability space.

- *Example*: For a fair coin toss, the probability space is:

$$(\Omega, \mathcal{A}, \mathbb{P}) = (\{\text{Heads, Tails}\}, \{\emptyset, \{\text{Heads}\}, \{\text{Tails}\}, \{\text{Heads, Tails}\}\}, \mathbb{P})$$

Probability review

What are elementary events?

Probability review

What are elementary events?

- Events that mutually exclude each other, but cover all possible outcomes of the attempt.
- An elementary event is an event that contains exactly one outcome from the sample space.
- It cannot be broken down further into simpler events.

Probability review

In the coin flip example, the sample space is:

$$\Omega = \{\text{Heads}, \text{Tails}\}$$

The elementary events are:

- $\{\text{Heads}\}$: The event where the outcome is "Heads"
- $\{\text{Tails}\}$: The event where the outcome is "Tails"

Which are the impossible and the certain events?

Probability review

Which are the impossible and the certain events?

- An impossible event is an event that cannot happen under any circumstances. It is represented by the empty set. Since it never occurs, its probability is always zero.
- A certain event is an event that is guaranteed to occur. Since it is guaranteed to happen, its probability is always one.
- Ω itself is denoted the certain event.
- $\emptyset$ is the impossible event.

Probability review

How do we assign a probability to an event?

Probability review

How do we assign a probability to an event?

- Let Ω be a sample space $\{\omega_1, ..., \omega_n\}$ and for $A \subseteq \Omega$, let $|A|$ denote the number of elements in A.
- Then the probability associated with equally likely events

$$P(A) = \frac{|A|}{|\Omega|}$$

  reports the fraction of outcomes in Ω that are also in A.

What follows immediately?

What follows immediately?

- Every elementary event has the probability $\frac{1}{|\Omega|}$.
- The requirement that elementary events have equal probability is called the *Laplace assumption* and the probabilities calculated thereby are called Laplace probabilities.

What do we get from this?

Probability review

What do we get from this?

- $0 \leq P(A) \leq 1$
- If $A \cap B = \emptyset$, then $P(A \cup B) = P(A) + P(B)$.
- $P(\Omega) = 1$

Probability review

What do we get from this?

Probability review

What do we get from this?

1. $0 \leq P(A) \leq 1$
The probability of any event A must lie between 0 and 1.

- *Example*: For rolling a six-sided die, $P(\text{rolling a 3}) = \frac{1}{6}$ satisfies $0 \leq \frac{1}{6} \leq 1$.

2. If $A \cap B = \emptyset$**, then** $P(A \cup B) = P(A) + P(B)$
If two events are disjoint, their combined probability is the sum of their individual probabilities.

- *Example*: Consider rolling a die and flipping a coin.

$$P(\text{roll a 2 or flip Heads}) = \frac{1}{6} + \frac{1}{2} = \frac{2}{3}$$

3. $P(\Omega) = 1$
The probability of the entire sample space is always 1.

- *Example*: For rolling a die, one of the numbers 1 through 6 must occur, so:

$$P(\{1, 2, 3, 4, 5, 6\}) = 1$$

Probability review

What else can we derive?

Probability review

What else can we derive?

- If $A_1, ..., A_k$ are pairwise disjoint or mutually exclusive, ($A_i \cap A_j = \emptyset$ *if* $\neq j$.) then $P(A_1 \cup A_2 \cup ... \cup A_k) = P(A_1) + P(A_2) + ... + P(A_k)$.
- For any two events A and B, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
- If $A \subseteq B$ then $P(A) \leq P(B)$.
- Letting $\neg B$ denote the complement of A, then $P(\neg B) = 1 - P(A)$.
- Therefore the range of the function we call the probability is a subset of the interval [0,1].
- If $A_1, .., A_n$ are the elementary events, then $\sum_{i=1}^{n} P(A_i) = 1$ (normalization condition).

$$P\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P(A_j).$$

Probability review

Let us look at an example

Probability review

Let us look at an example

Example

$$
\begin{aligned}
\Omega &= \textit{\{all outcomes of a roll of a die\}} \\
&= \{1,2,3,4,5,6\} \\
\mathcal{A} &= \textit{\{all possible set of outcomes\}} \\
&= \{\{1\}, ..., \{6\}, \{1,2\}, ..., \{5,6\}, ..., \{1,2,3,4,5,6\}\} \\
\mathbb{P} &= \textit{probability of all sets / events}
\end{aligned}
$$

- *What is the probability of a given* $\omega \in \Omega$ *say* $\omega = 3$*?*
- *What is the probability of the event* $\omega \in \{1,2,3\}$*?*

Probability review

Let us look at an example

Example

$$
\begin{aligned}
\Omega &= \textit{\{all outcomes of a roll of a die\}} \\
&= \{1, 2, 3, 4, 5, 6\} \\
\mathcal{A} &= \textit{\{all possible set of outcomes\}} \\
&= \{\{1\}, ..., \{6\}, \{1, 2\}, ..., \{5, 6\}, ..., \{1, 2, 3, 4, 5, 6\}\} \\
\mathbb{P} &= \textit{probability of all sets / events}
\end{aligned}
$$

- *What is the probability of a given* $\omega \in \Omega$ *say* $\omega = 3$*?*
- *What is the probability of the event* $\omega \in \{1, 2, 3\}$*? The event includes multiple outcomes:* $\omega \in \{1, 2, 3\}$*.*

$$P(\omega \in \{1, 2, 3\}) = P(\omega = 1) + P(\omega = 2) + P(\omega = 3)$$

Since each outcome has the same probability:

$$P(\omega \in \{1, 2, 3\}) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2}$$

Probability review

What is $P(A \cap B)$?

Probability review

What is $P(A \cap B)$?

- The expression $P(A \cap B)$ or equivalently P(A, B) stands for the probability of the events $A \cap B$. represents the joint probability of two events A and B occurring at the same time.
- We are often interested in the probabilities of all elementary events, that is, of all combinations of all values of the variables A and B.
- For the binary variables A and B these are $P(A, B), P(A, \neg B), P(\neg A, B), P(\neg A, \neg B)$.
- The **joint probability distribution of the variables A and B** is the **vector** $(P(A, B), P(A, \neg B), P(\neg A, B), P(\neg A, \neg B))$ consisting of the four values a distribution for the previous example.

Probability review

So what is the joint probability distribution?

So what is the joint probability distribution?

- **Joint Probability Distribution:** A function that gives the probability of two or more random variables taking specific values simultaneously.
- **Describes Relationships:** It captures how the outcomes of different variables interact with one another.
- **Notation:** $P(X = x, Y = y)$ specifies the joint probability for specific values of the random variables.

Probability review

So what is the joint probability distribution?

- **Joint Probability Distribution:** A function that gives the probability of two or more random variables taking specific values simultaneously.
- **Describes Relationships:** It captures how the outcomes of different variables interact with one another.
- **Notation:** $P(X = x, Y = y)$ specifies the joint probability for specific values of the random variables.

Example

Consider two binary random variables X (Raining: Yes/No) and Y (Carrying Umbrella: Yes/No):

	$Y =$ Yes	$Y =$ No
$X =$ Yes	0,3	0,2
$X =$ No	0,1	0,4

Probability review

What about conditional probability?

Probability review

What about conditional probability?

- It is a measure of the likelihood of an event occurring given that another event has already occurred.
- It answers questions like, **What is the probability of event A happening given that event B has occurred?**

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

Probability review

Let us look at an example

Probability review

Let us look at an example

Example

$$\begin{aligned} \Omega &= \{1, 2, 3, 4, 5, 6\} \\ A &= \{1, 2\} \\ B &= \{2, 3\} \end{aligned}$$

- *We say that A occurs if the outcome is either 1 or 2.*
- *Now suppose you are told that B occurs.*
- *Then it seems more probable that A has also occurred.*
- *The probability that A occurs, without knowledge of whether B has occurred, is 1/3.*
- *That is, P(A) = 1/3.*

Probability review

Let us look at an example

Example

$$\begin{aligned} \Omega &= \{1,2,3,4,5,6\} \\ A &= \{1,2\} \\ B &= \{2,3\} \end{aligned}$$

- *We say that A occurs if the outcome is either 1 or 2.*
- *Now suppose you are told that B occurs.*
- *Then it seems more probable that A has also occurred.*
- *The probability that A occurs, without knowledge of whether B has occurred, is 1/3.*
- *That is, P(A) = 1/3.*

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\{2\})}{P(\{2,3\})} = \frac{1/6}{1/3} = \frac{1}{2}$$

- We see that knowledge of B provides important information about A

When do we call two events independent?

When do we call two events independent?

Definition

- *Two events A and B are said to be independent if P(A|B) =* $\mathbb{P}(A)$.
- *In other words, B provides no information about whether A has occurred.*
- $P(A \cap B) = P(A) \cdot P(B)$

Probability review

Let us look at an example

Let us look at an example

Example

Suppose we have two dice.

$$\begin{aligned} \Omega &= \{\text{all pairs of outcomes of the roll of two dice}\} \\ &= \{\{1,1\},\{1,2\},...,\{6,6\}\} \end{aligned}$$

- *Let A = {1st die is 1} and B = {2nd die is 1}.*
- $P(A|B) = \dfrac{P(A=\{1\},B=\{1\})}{P(B=\{1\})} = \dfrac{1/36}{1/6} = \dfrac{1}{6}$
- *As we know, the value of one die does not influence the other.*
- *The two outcomes are independent.*

Example

- *For a roll of two dice, the probability of rolling two sixes is* $\dfrac{1}{36}$ *if the two dice are independent.*
- $P(A=\{6\},B=\{6\}) = P(A=\{6\}) \cdot P(B=\{6\}) = \dfrac{1}{6} \cdot \dfrac{1}{6} = \dfrac{1}{36}$

Probability review

What is the chain rule?

Probability review

What is the chain rule?

The **chain rule** allows us to decompose the probability of the intersection of several events into a product of conditional probabilities.

- Consider the definition of conditional probability for $P(A \cap B)$ results in the so-called product rule

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

- This rule can be generalized for the case of n variables by repeated application of the above rule we obtain the chain rule.
- Because the chain rule holds for all values of the variables $X_1, ..., X_n$ it has been formulated for the distribution using the symbol **P.**

$$\begin{aligned}
&\boldsymbol{P}(X_1, \dots, X_n) \\
&\quad = \boldsymbol{P}(X_n | X_1, \dots, X_{n-1}) \cdot \boldsymbol{P}(X_1, \dots, X_{n-1}) \\
&\quad = \boldsymbol{P}(X_n | X_1, \dots, X_{n-1}) \cdot \boldsymbol{P}(X_{n-1} | X_1, \dots, X_{n-2}) \cdot \boldsymbol{P}(X_1, \dots, X_{n-2}) \\
&\quad = \boldsymbol{P}(X_n | X_1, \dots, X_{n-1}) \cdot \boldsymbol{P}(X_{n-1} | X_1, \dots, X_{n-2}) \cdot \ldots \cdot \boldsymbol{P}(X_n | X_1) \cdot \boldsymbol{P}(X_1) \\
&\quad = \prod_{i=1}^{n} \boldsymbol{P}(X_n | X_1, \dots, X_{i-1}),
\end{aligned}$$

Let's take a closer look

Let's take a closer look

The chain rule allows us to express the joint probability of events as a product of conditional probabilities.

- For two events A and B:

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

- For three events A, B, and C:

$$P(A \cap B \cap C) = P(A) \cdot P(B|A) \cdot P(C|A \cap B)$$

- General chain rule for n events:

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdot \cdots \cdot P(A_n|A_1 \cap \cdots \cap A_{n-1})$$

Probability review

What is marginalization?

Probability review

What is marginalization?

Marginalization is a process where we sum over the possible values of a random variable to eliminate it from a joint probability distribution.

- We know that $A = (A \cap B) \cup (A \cap \neg B)$ is true for binary variables A and B. This means that event A can be split into two mutually exclusive cases:
  - A occurs together with B, Or
  - A occurs together with $\neg B$(the complement of B).
- To find the probability of A, we apply this decomposition to express P(A) as a sum of two terms,

$$P(A) = P((A \cap B) \cup (A \cap \neg B)) = P(A \cap B) + P(A \cap \neg B)$$

- By summation over the two values of B, the variable B is eliminated.
- For more than two variables, marginalization can be applied in a similar way. Suppose we have $X_1, ..., X_d$, a variable, for example X_d, can be eliminated by summation over all of its values:

$$P(X_1 = x_1, ..., X_{d-1} = x_{d-1}) = \sum_{X_d} P(X_1 = x_1, ..., X_{d-1} = x_{d-1}, X_d = x_d)$$

- The application of this formula is called marginalization.

What about that?

What about that?

- This summation can continue with the variables $X_1, ..., X_{d-1}$ until just one variable is left.
- Marginalization can also be applied to the distribution $P(X_1, ..., X_d)$.
- The resulting distribution $P(X_1, ..., X_{d-1})$ is called the marginal distribution.

Probability review

An example to understand this notion

Probability review

An example to understand this notion

Example

We observe the set of all patients who come to the doctor with acute stomach pain. For each patient the leukocyte value is measured, which is a metric for the relative abundance of white blood cells in the blood. We define the variable Leuko, which is true if and only if the leukocyte value is greater than 10,000. This indicates an infection in the body. Otherwise we define the variable App, which tells us whether the patient has appendicitis, that is, an infected appendix. The distribution P(App, Leuko) of these two variables is given in the following table:

$\mathbf{P}(App, Leuko)$	App	$\neg App$	Total
$Leuko$	0.23	0.31	0.54
$\neg Leuko$	0.05	0.41	0.46
Total	0.28	0.72	1

In this example, we are marginalizing over the variable App (appendicitis) to find the probability of Leuko (high leukocyte count).

In the last row the sum over the rows is given, and in the last column the sum of the columns is given.

- To find $P(\text{Leuko})$ (the probability of a high leukocyte count), we sum the joint probabilities where Leuko $=$ True for both values of App (True and False):

  $$P(\text{Leuko}) = P(\text{App} = \text{True}, \text{Leuko} = \text{True}) + P(\text{App} = \text{False}, \text{Leuko} = \text{True})$$

  Using the values from the table:

  $$P(\text{Leuko}) = 0{,}36 + 0{,}18 = 0{,}54$$

  This is the marginal probability that a patient has a leukocyte count higher than 10,000.
- The conditional probability of a high leukocyte count given that the patient has appendicitis is $P(Leuko|App) = 0{,}82$
- Use the conditional probability formula to verify this.

What is the Bayes' Theorem?

What is the Bayes' Theorem?

Bayes'Theorem gives a mathematical rule for inverting conditional probabilities, allowing us to find the probability of a cause given its effect.

- Having $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$
- Having $P(B|A) = \dfrac{P(B \cap A)}{P(A)}$
- By solving both equations for $P(A \cap B)$ and equating them we obtain Bayes' theorem

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

Using Bayes' Theorem we can get $P(App|Leuko)$

Example

$$P(App|Leuko) = \frac{P(Leuko|App) \cdot P(App)}{P(Leuko)} = \frac{0.82 \cdot 0.28}{0.54} = 0.43.$$

Probability review

Exercise

Suppose a medical test is used to detect a certain disease. This test is known to be 95 % accurate, meaning that it correctly identifies 95 % of the people who have the disease, and it correctly identifies 95 % of the people who do not have the disease. The prevalence of the disease in the population is 1 %.

- *Now, if a person tests positive for the disease using this test, what is the probability that they actually have the disease?*
- *Let event A be the event that a person has the disease.*
- *Let event B be the event that a person tests positive for the disease.*

Bayes' Theorem Application

From the problem statement:

- A medical test detects a certain disease with 95 % accuracy.
- Prevalence of the disease in the population: 1 %.
- Let A: the event that a person has the disease.
- Let B: the event that a person tests positive for the disease.

Given Information

- True Positive Rate (Sensitivity):

$$P(B|A) = 0{,}95$$

- True Negative Rate (Specificity):

$$P(\neg B|\neg B) = 0{,}95 \quad \Rightarrow \quad P(B|\neg A) = 1 - P(\neg B|\neg A) = 0{,}05$$

- Prevalence of the disease:

$$P(A) = 0{,}01$$

- Probability of not having the disease:

$$P(\neg B) = 1 - P(A) = 0{,}99$$

Finding Marginal Probability

Marginal Probability of Testing Positive:

$$\begin{aligned} P(B) &= P(B|A) \cdot P(A) + P(B|\neg B) \cdot P(\neg B) \\ &= (0{,}95 \cdot 0{,}01) + (0{,}05 \cdot 0{,}99) \\ &= 0{,}0095 + 0{,}0495 \\ &= 0{,}059 \end{aligned}$$

Using Bayes'Theorem:

$$\begin{aligned} P(A|B) &= \frac{P(B|A) \cdot P(A)}{P(B)} \\ &= \frac{0{,}95 \cdot 0{,}01}{0{,}059} \\ &= \frac{0{,}0095}{0{,}059} \\ &\approx 0{,}161 \end{aligned}$$