Probability for Computer Science

Spring 2021

Lecture 4

Prof. Claire Monteleoni

Boulder

Today

- Probability Laws (continued)
- Conditional Probability
- Multiplication Rule
- The Total Probability Theorem
  - Divide and Conquer method
- Bayes' Rule
- If time: Independence

Probability Axioms

1. **(Nonnegativity)** $\mathbf{P}(A) \geq 0$, for every event A.
2. **(Additivity)** If A and B are two disjoint events, then the probability of their union satisfies

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B).$$

   More generally, if the sample space has an infinite number of elements and $A_1, A_2, \ldots$ is a sequence of disjoint events, then the probability of their union satisfies

$$\mathbf{P}(A_1 \cup A_2 \cup \cdots) = \mathbf{P}(A_1) + \mathbf{P}(A_2) + \cdots.$$

3. **(Normalization)** The probability of the entire sample space Ω is equal to 1, that is, $\mathbf{P}(\Omega) = 1$.

Credit: Bertsekas & Tsitsiklis, 2008

Some Properties of Probability Laws

Consider a probability law, and let A, B, and C be events.

(a) If $A \subset B$, then $\mathbf{P}(A) \leq \mathbf{P}(B)$.

(b) $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$.

(c) $\mathbf{P}(A \cup B) \leq \mathbf{P}(A) + \mathbf{P}(B)$.

Credit: Bertsekas & Tsitsiklis, 2008

Conditional Probability

A technique to reason about the outcome of an experiment, given partial information.

E.g.

How likely is it that a person has a particular disease, given that the medical test for it turned out negative?

If a word starts with the letter t, what is the probability that its second letter is h?

Properties of Conditional Probability

- The conditional probability of an event A, given an event B with $\mathbf{P}(B) > 0$, is defined by

$$\mathbf{P}(A \mid B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)},$$

  and specifies a new (conditional) probability law on the same sample space Ω. In particular, all properties of probability laws remain valid for conditional probability laws.

- Conditional probabilities can also be viewed as a probability law on a new universe B, because all of the conditional probability is concentrated on B.

- If the possible outcomes are finitely many and equally likely, then

$$\mathbf{P}(A \mid B) = \frac{\text{number of elements of } A \cap B}{\text{number of elements of } B}.$$

Credit: Bertsekas & Tsitsiklis, 2008

Multiplication Rule

Assuming that all of the conditioning events have positive probability, we have

$$\mathbf{P}\left(\cap_{i=1}^{n} A_i\right) = \mathbf{P}(A_1)\mathbf{P}(A_2 \mid A_1)\mathbf{P}(A_3 \mid A_1 \cap A_2) \cdots \mathbf{P}\left(A_n \mid \cap_{i=1}^{n-1} A_i\right).$$

The multiplication rule can be verified by writing

$$\mathbf{P}\left(\cap_{i=1}^{n} A_i\right) = \mathbf{P}(A_1) \cdot \frac{\mathbf{P}(A_1 \cap A_2)}{\mathbf{P}(A_1)} \cdot \frac{\mathbf{P}(A_1 \cap A_2 \cap A_3)}{\mathbf{P}(A_1 \cap A_2)} \cdots \frac{\mathbf{P}\left(\cap_{i=1}^{n} A_i\right)}{\mathbf{P}\left(\cap_{i=1}^{n-1} A_i\right)}$$

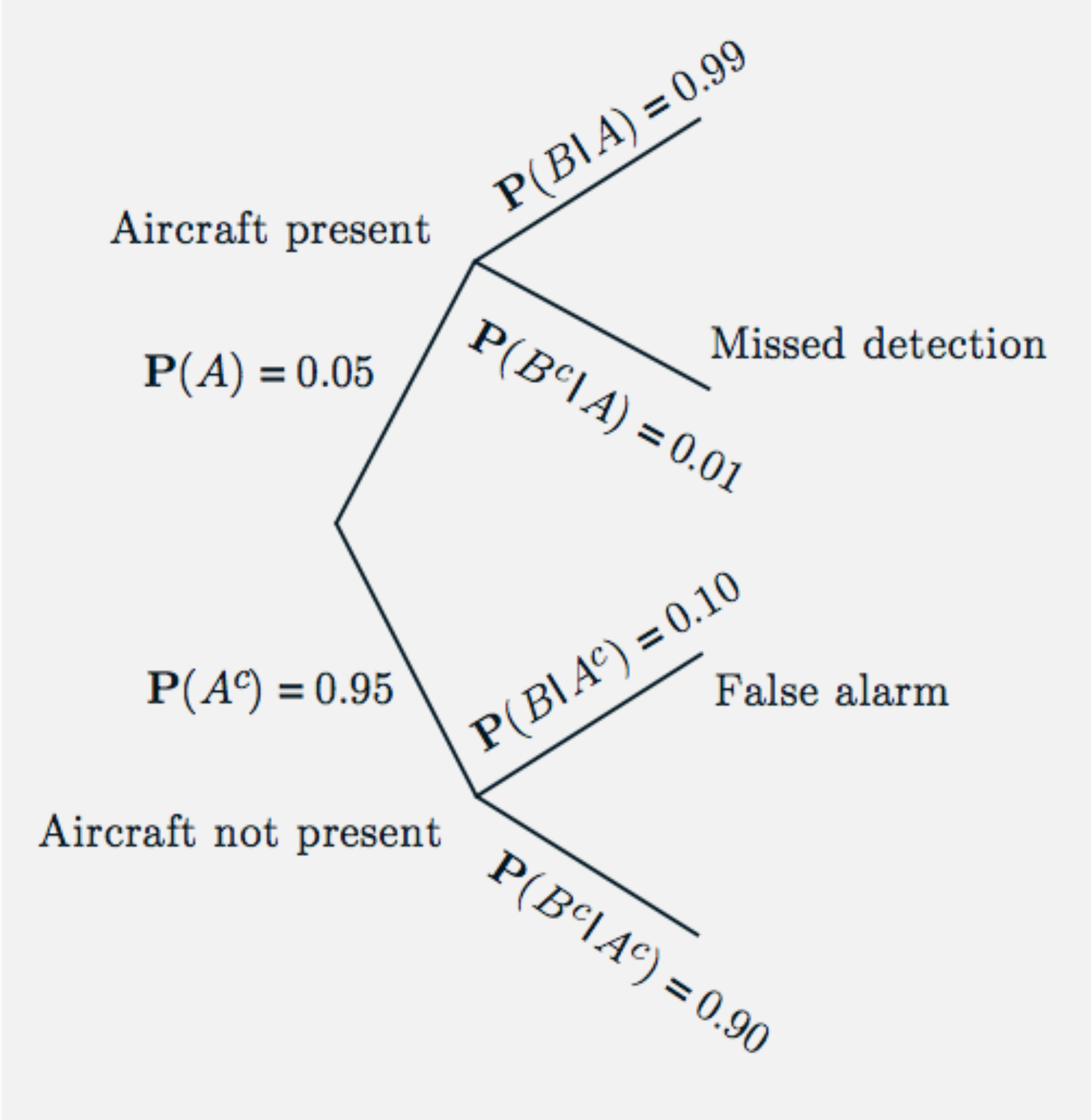
Credit: Bertsekas & Tsitsiklis, 2008

Example

- If there’s an aircraft in a certain region, the radar system generates an alarm with probability 0.99.
- If there’s no aircraft in the region, the radar system generates an alarm with probability 0.1.
- The probability of an aircraft being in the region is 0.05.

Q: What is the probability that an aircraft is in the region, and the radar system does not generate an alarm?

Credit: Bertsekas & Tsitsiklis, 2008



Credit: Bertsekas & Tsitsiklis, 2008

The Sequential Method

(a) We set up the tree so that an event of interest is associated with a leaf. We view the occurrence of the event as a sequence of steps, namely, the traversals of the branches along the path from the root to the leaf.

(b) We record the conditional probabilities associated with the branches of the tree.

(c) We obtain the probability of a leaf by multiplying the probabilities recorded along the corresponding path of the tree.

Given enough practice, you can apply the Multiplication Rule without drawing a tree.

Credit: Bertsekas & Tsitsiklis, 2008

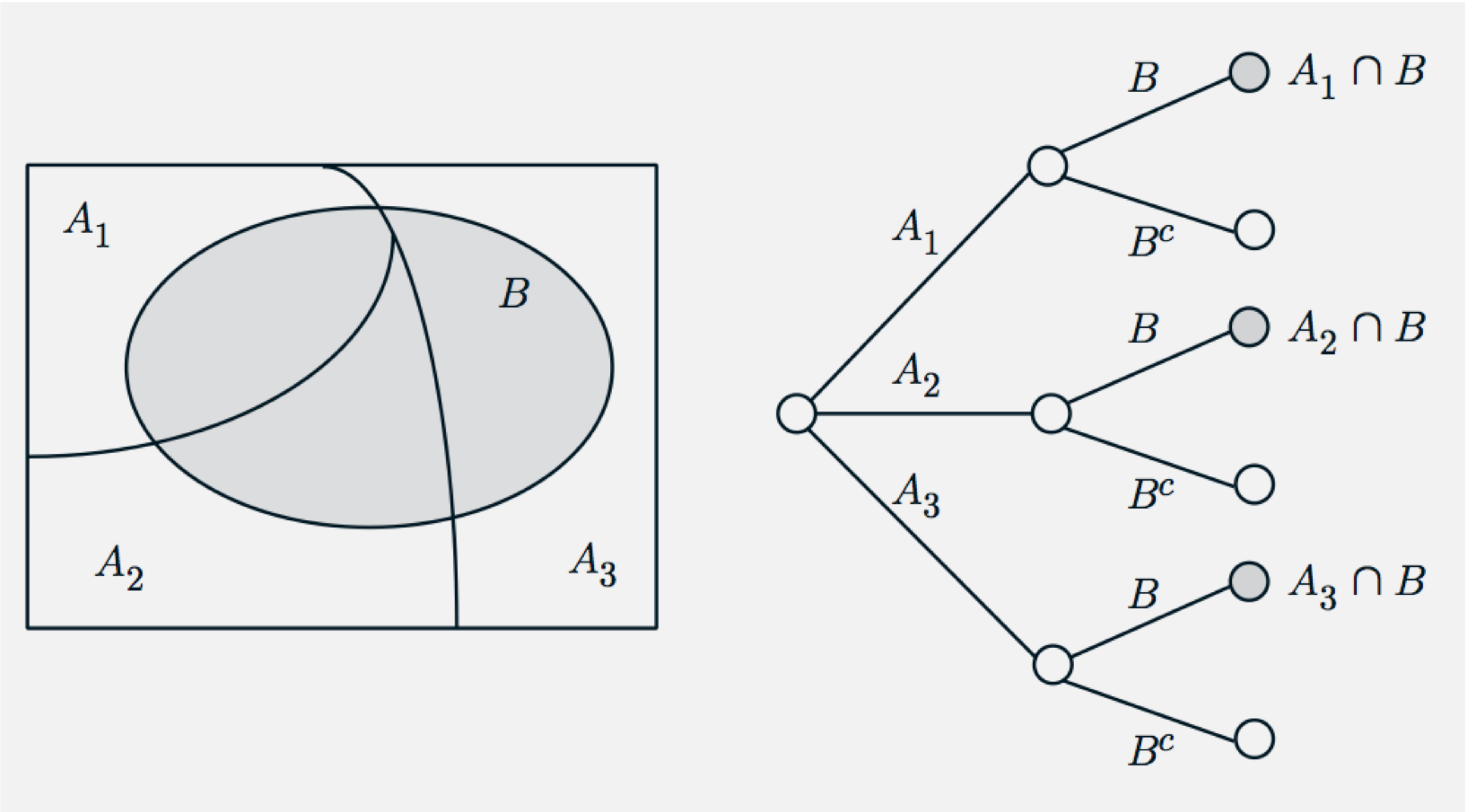
Total Probability Theorem

Let $A_1, \ldots, A_n$ be disjoint events that form a partition of the sample space (each possible outcome is included in exactly one of the events $A_1, \ldots, A_n$) and assume that $\mathbf{P}(A_i) > 0$, for all i. Then, for any event B, we have

$$\begin{aligned}\mathbf{P}(B) &= \mathbf{P}(A_1 \cap B) + \cdots + \mathbf{P}(A_n \cap B) \\ &= \mathbf{P}(A_1)\mathbf{P}(B \mid A_1) + \cdots + \mathbf{P}(A_n)\mathbf{P}(B \mid A_n).\end{aligned}$$

$$P(B) = \sum_{i=1}^{n} P(A_i)P(B|A_i)$$

Credit: Bertsekas & Tsitsiklis, 2008



Credit: Bertsekas & Tsitsiklis, 2008

Divide and Conquer Method

- Choose a set of events A_1 , … , A_n that partition Ω and have known probabilities, $P(A_i)$.
- Compute $P(B \mid A_i)$ for each $i : 1 \leq i \leq n$.
- Solve for P(B) using Total Probability Theorem:

$$P(B) = \sum_{i=1}^{n} P(A_i) P(B|A_i)$$

Bayes' Rule

Bayes' Rule

Let $A_1, A_2, \ldots, A_n$ be disjoint events that form a partition of the sample space, and assume that $\mathbf{P}(A_i) > 0$, for all i. Then, for any event B such that $\mathbf{P}(B) > 0$, we have

$$\begin{aligned}\mathbf{P}(A_i \mid B) &= \frac{\mathbf{P}(A_i)\mathbf{P}(B \mid A_i)}{\mathbf{P}(B)} \\ &= \frac{\mathbf{P}(A_i)\mathbf{P}(B \mid A_i)}{\mathbf{P}(A_1)\mathbf{P}(B \mid A_1) + \cdots + \mathbf{P}(A_n)\mathbf{P}(B \mid A_n)}.\end{aligned}$$

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{j=1}^{n} P(A_j)P(B|A_j)}$$

Credit: Bertsekas & Tsitsiklis, 2008

Independence

- Two events A and B are said to be **independent** if

$$\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B).$$

If in addition, $\mathbf{P}(B) > 0$, independence is equivalent to the condition

$$\mathbf{P}(A \mid B) = \mathbf{P}(A).$$

- If A and B are independent, so are A and B^c.

Definition of Independence of Several Events

We say that the events $A_1, A_2, \ldots, A_n$ are **independent** if

$$\mathbf{P}\left(\bigcap_{i \in S} A_i\right) = \prod_{i \in S} \mathbf{P}(A_i), \qquad \text{for every subset } S \text{ of } \{1, 2, \ldots, n\}.$$

Credit: Bertsekas & Tsitsiklis, 2008

For the case of three events, A_1, A_2, and A_3, independence amounts to satisfying the four conditions

$$\begin{aligned}
\mathbf{P}(A_1 \cap A_2) &= \mathbf{P}(A_1)\,\mathbf{P}(A_2),\\
\mathbf{P}(A_1 \cap A_3) &= \mathbf{P}(A_1)\,\mathbf{P}(A_3),\\
\mathbf{P}(A_2 \cap A_3) &= \mathbf{P}(A_2)\,\mathbf{P}(A_3),\\
\mathbf{P}(A_1 \cap A_2 \cap A_3) &= \mathbf{P}(A_1)\,\mathbf{P}(A_2)\,\mathbf{P}(A_3).
\end{aligned}$$

Credit: Bertsekas & Tsitsiklis, 2008

Example

Experiment: toss a fair coin two times, independently.

H_1 = {The first toss is a head}

H_2 = {The second toss is a head}

D = {The two tosses have different outcomes}

Independence

- Two events A and B are said to be **independent** if

$$\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B).$$

If in addition, $\mathbf{P}(B) > 0$, independence is equivalent to the condition

$$\mathbf{P}(A \mid B) = \mathbf{P}(A).$$

- If A and B are independent, so are A and B^c.
- Two events A and B are said to be **conditionally independent**, given another event C with $\mathbf{P}(C) > 0$, if

$$\mathbf{P}(A \cap B \mid C) = \mathbf{P}(A \mid C)\mathbf{P}(B \mid C).$$

If in addition, $\mathbf{P}(B \cap C) > 0$, conditional independence is equivalent to the condition

$$\mathbf{P}(A \mid B \cap C) = \mathbf{P}(A \mid C).$$

- Independence does not imply conditional independence, and vice versa.

Credit: Bertsekas & Tsitsiklis, 2008