Probability for Computer Science

Spring 2021

Lecture 2

Prof. Claire Monteleoni

Boulder

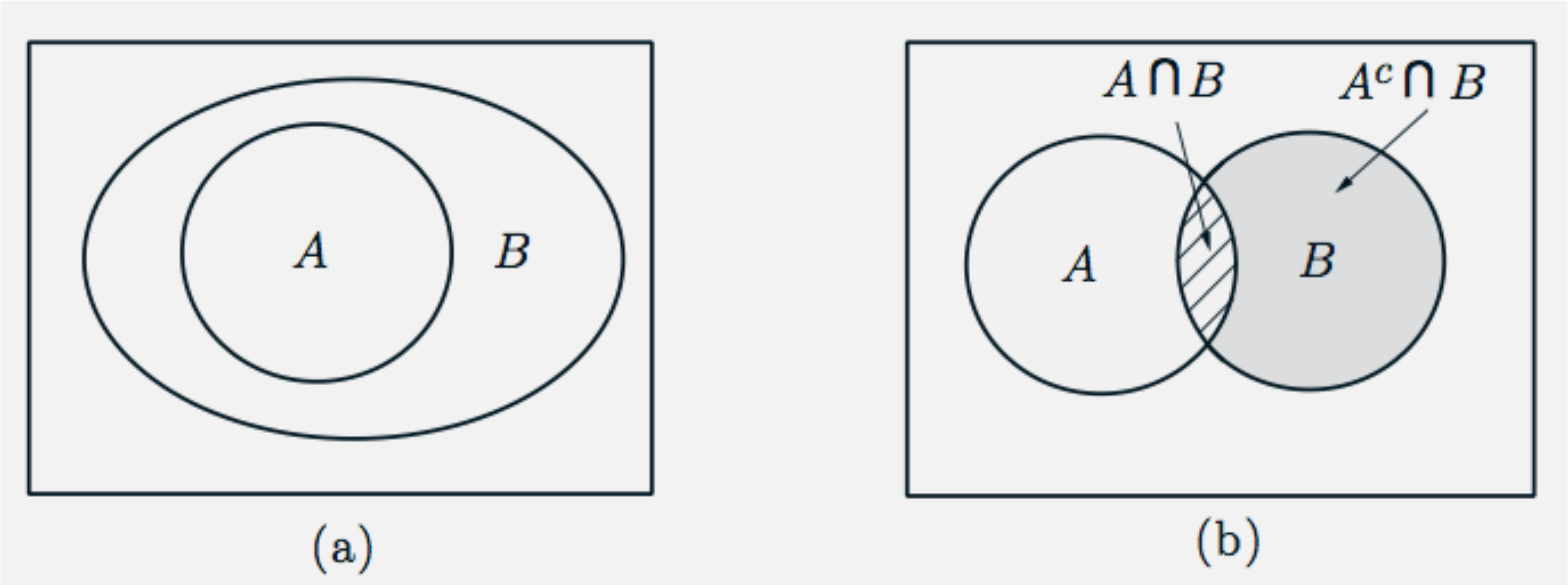
Today

- Intro. Set Theory (continued)
- Probabilistic Models
- Probability Laws, and the Axioms of Probability

If time:

- Conditional Probability

Venn Diagrams



Credit: Bertsekas & Tsitsiklis, 2008

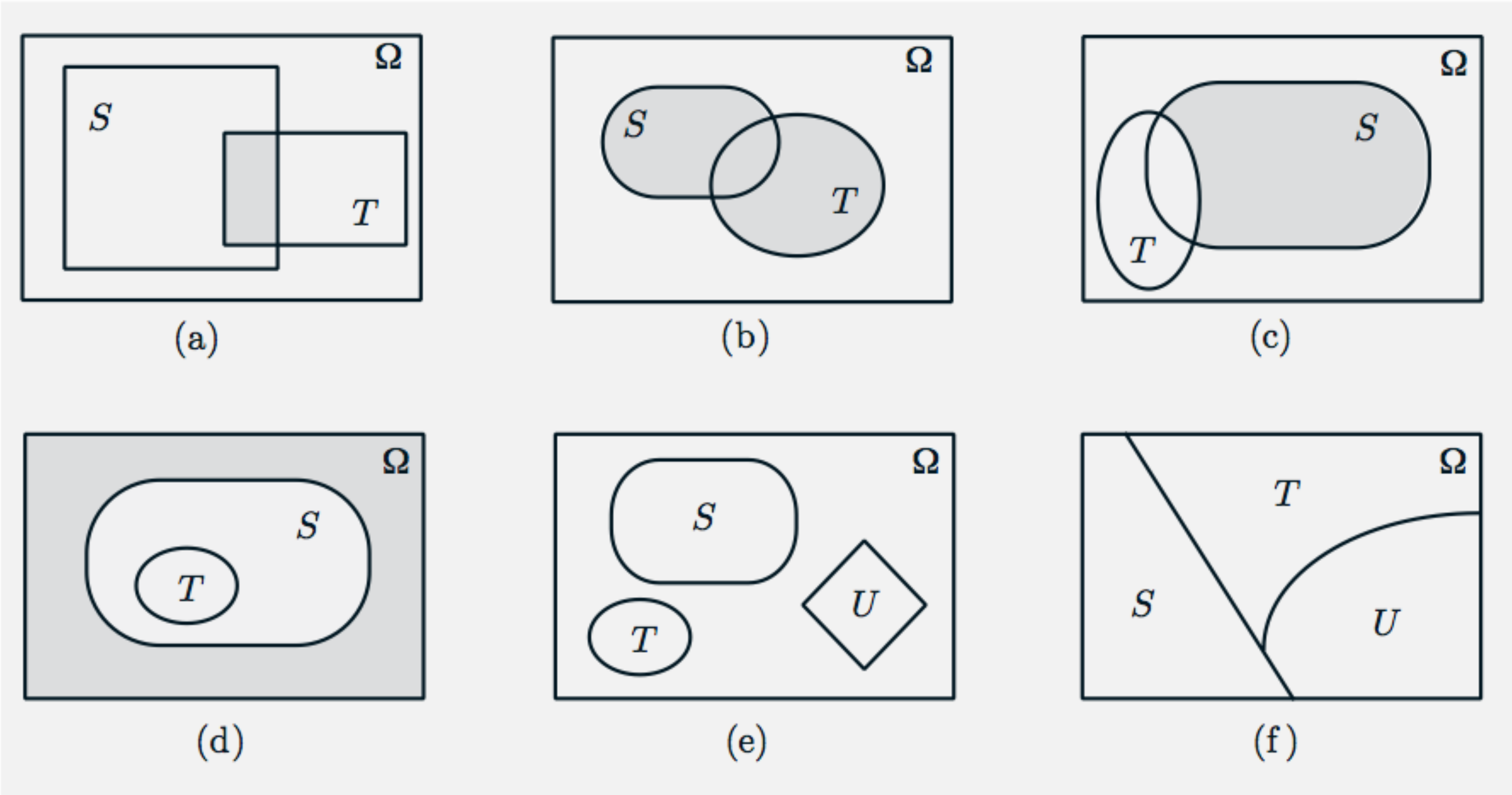


Figure 1.1: Examples of Venn diagrams. (a) The shaded region is $S \cap T$. (b) The shaded region is $S \cup T$. (c) The shaded region is $S \cap T^c$. (d) Here, $T \subset S$. The shaded region is the complement of S. (e) The sets S, T, and U are disjoint. (f) The sets S, T, and U form a partition of the set Ω.

Credit: Bertsekas & Tsitsiklis, 2008

Set Operations

$$S \cup T = T \cup S,$$
$$S \cap (T \cup U) = (S \cap T) \cup (S \cap U),$$
$$(S^c)^c = S,$$
$$S \cup \Omega = \Omega,$$

$$S \cup (T \cup U) = (S \cup T) \cup U,$$
$$S \cup (T \cap U) = (S \cup T) \cap (S \cup U),$$
$$S \cap S^c = \emptyset,$$
$$S \cap \Omega = S.$$

DeMorgan's Laws:

$$\left(\bigcup_n S_n\right)^c = \bigcap_n S_n^c, \qquad \left(\bigcap_n S_n\right)^c = \bigcup_n S_n^c.$$

Review Chapter 1, especially p. 5, if these are unfamiliar.

Exercise: Verify these identities, using Venn diagrams to gain intuition. Then prove them.

Credit: Bertsekas & Tsitsiklis, 2008

Defining a probabilistic model

An experiment is any activity that results in **exactly one outcome.** E.g.

- Flipping a coin
- Flipping two coins in a row

A probabilistic model is defined as follows:

1. A sample space: the set of all possible outcomes of an experiment

2. A probability law: assigns probabilities to events. An event is a **set of possible outcomes**.

The Sample Space, Ω

The set of all possible outcomes of an experiment.

Elements of this set must be:

- Mutually exclusive: running the experiment results in one unique outcome.
  - Every possible outcome of the experiment is a **unique** element of Ω, not contained in multiple elements.
- Collectively exhaustive: the experiment always results in an outcome from the sample space.
  - Ω contains **all possible outcomes** of the experiment.

Events

An event, A, is a collection of possible outcomes, a subset of the sample space. $A \subseteq \Omega$

E.g. let the experiment be rolling a single die. Then

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

We can define the following events:

- "The number is even." $A = \{2, 4, 6\}$
- "The number is less than 4." $B = \{1, 2, 3\}$
- "The number is 6." $C = \{6\}$

Events

An event, A, is **a set of outcomes**, i.e., a subset of the sample space. $A \subseteq \Omega$

Since events are sets, we can do set operations on them.

E.g. given $A = \{2, 4, 6\}, B = \{1, 2, 3\}$

- What is $A \cup B$?
- What is $A \cap B$?

Probabilistic Model

To define a probabilistic model, one must define:

1. A sample space, Ω: the set of all possible outcomes of the experiment.

2. A probability law. A probability law assigns a probability, P(A), to every event, A.
   – Must satisfy the Axioms of Probability.

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

Discrete Probability Law

If the sample space consists of a finite number of possible outcomes, then the probability law is specified by the probabilities of the events that consist of a single element. In particular, the probability of any event $\{s_1, s_2, \ldots, s_n\}$ is the sum of the probabilities of its elements:

$$\mathbf{P}\big(\{s_1, s_2, \ldots, s_n\}\big) = \mathbf{P}(s_1) + \mathbf{P}(s_2) + \cdots + \mathbf{P}(s_n).$$

Credit: Bertsekas & Tsitsiklis, 2008

Discrete Uniform Probability Law

If the sample space consists of n possible outcomes which are equally likely (i.e., all single-element events have the same probability), then the probability of any event A is given by

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

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

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

The multiplication rule can be verified by writing

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

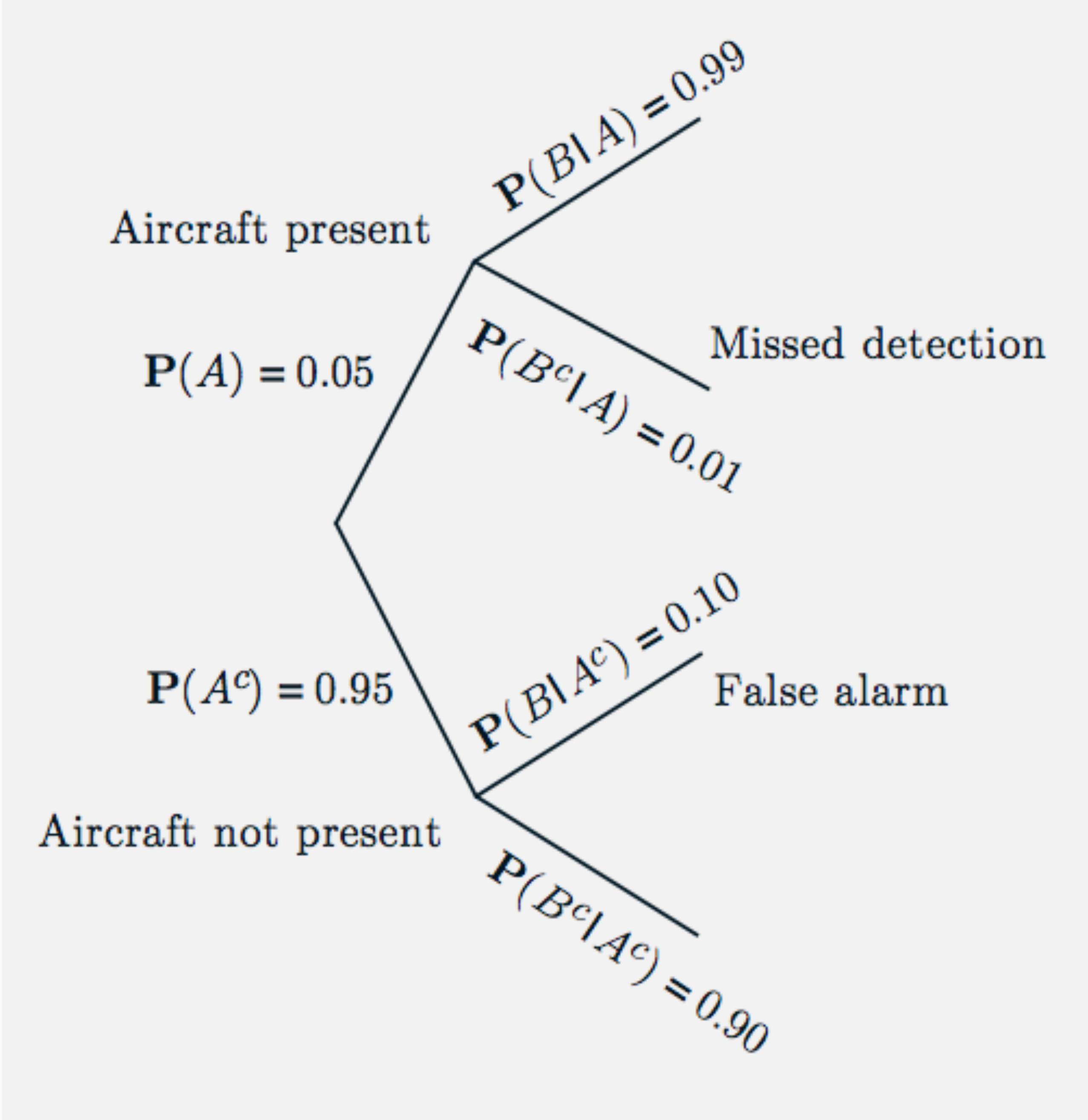
Credit: Bertsekas & Tsitsiklis, 2008

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

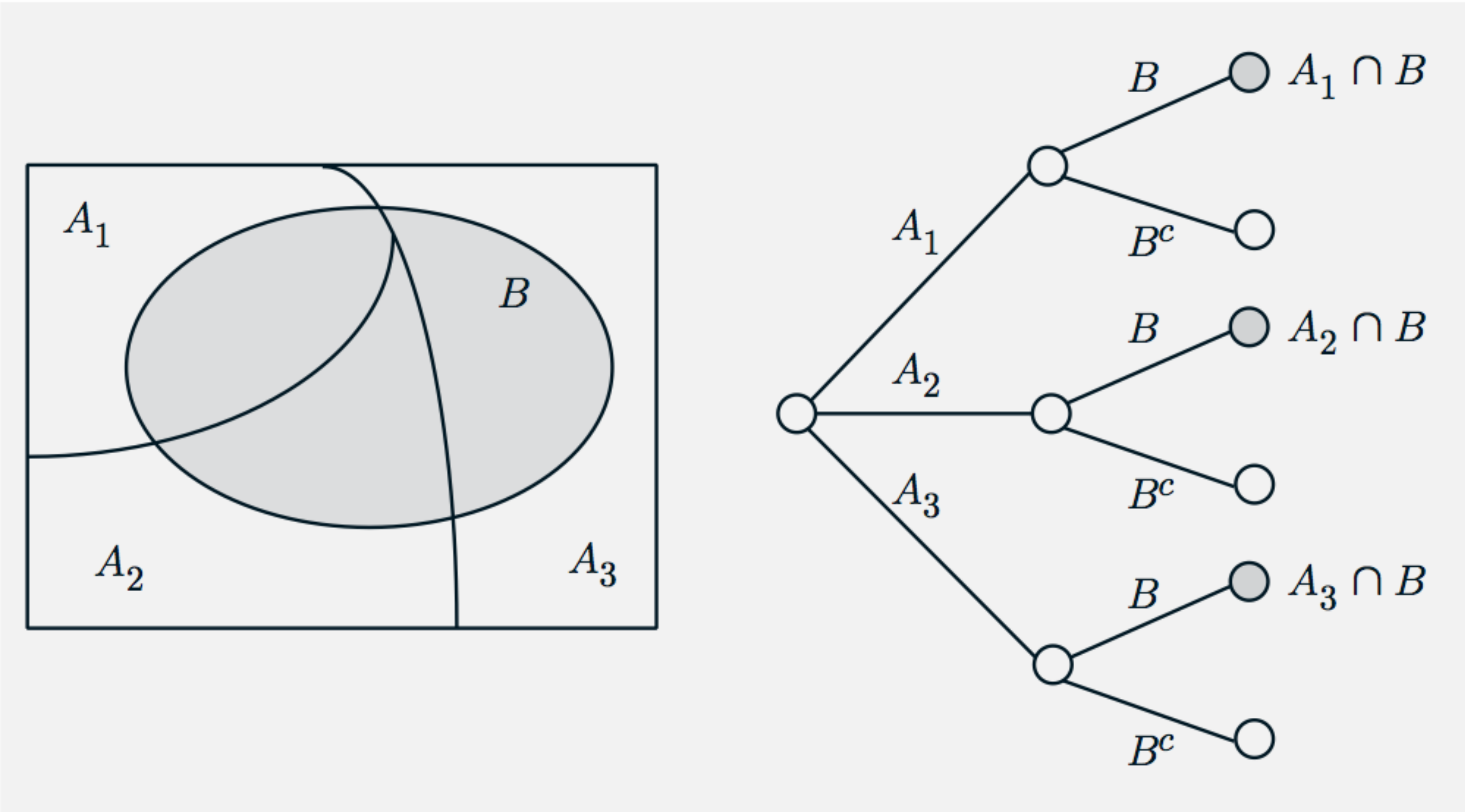
Total Probability Theorem

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:

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