Probability for Computer Science

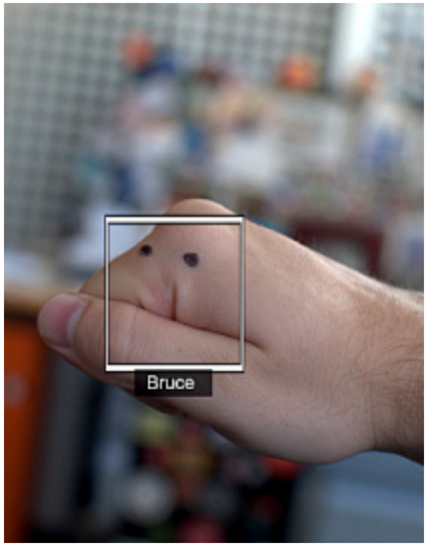
Spring 2021

Prof. Claire Monteleoni

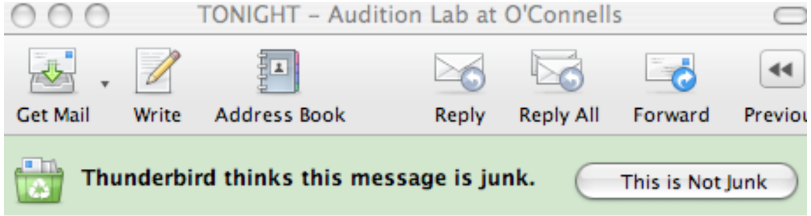
Boulder

- What is the probability that Siri gives you the wrong answer?

- What is the probability that a face detection algorithm makes a mistake?



- What is the probability that an email is spam?



  - If the email contains “!!!!!!!!!!!” is this probability higher?

Probability and Uncertainty in CS

- What is the expected number of text messages that will be delayed or lost by a mobile provider, each day?
- With what probability will a server fail?
  - If data is stored in the cloud (replicated across many servers), what is the probability that access to the data will fail (all servers fail simultaneously)?
- What is the expected running time of a randomized algorithm?
  - What is its probability of returning the correct output?
- What is the probability that a robot reaches its destination?
- With what probability do we expect a server to be hacked?
- What is the expected number of bugs in 1000 lines of code?

Probability and Uncertainty in CS

Probability is useful all over CS, e.g.

- Computer networks
- Operating systems
- Databases
- Security
- AI
- Vision
- Robotics
- Natural Language Processing
- Machine Learning, Data Mining

Getting set up in this course

- Make sure you have access to the textbook:
  Dimitri P. Bertsekas and John N. Tsitsiklis: Introduction to Probability (2nd Edition). Athena Scientific, 2008.
  - Do readings assigned on schedule, before each class.

- Canvas course page:
  - Read the Syllabus for details on the course
  - Check the Schedule often (linked from Canvas)
  - Fill out Intro Questionnaire (linked from Schedule) this week

- Course page on Piazza.com (linked from Canvas)
  (Free) Sign-up and join the Piazza course this week
  - Check Piazza often for announcements

- Join a team: find 1-3 teammates, talk to students at lecture or do a team search on Piazza. Meet regularly to discuss/practice homework, problems after first trying them on your own.

Course outline

- Introduction to Probability Theory, e.g.
  - Random variables: discrete, continuous
  - Conditioning, independence
  - Joint and conditional distributions
  - Problem solving
- Probability-based concepts and uses in CS, and Reasoning under uncertainty, e.g.
  - Information theory, coding
  - Inference, including Bayesian
  - Markov chains
  - Intro to randomized algorithms

Guidelines

Please read Syllabus for further details on course.

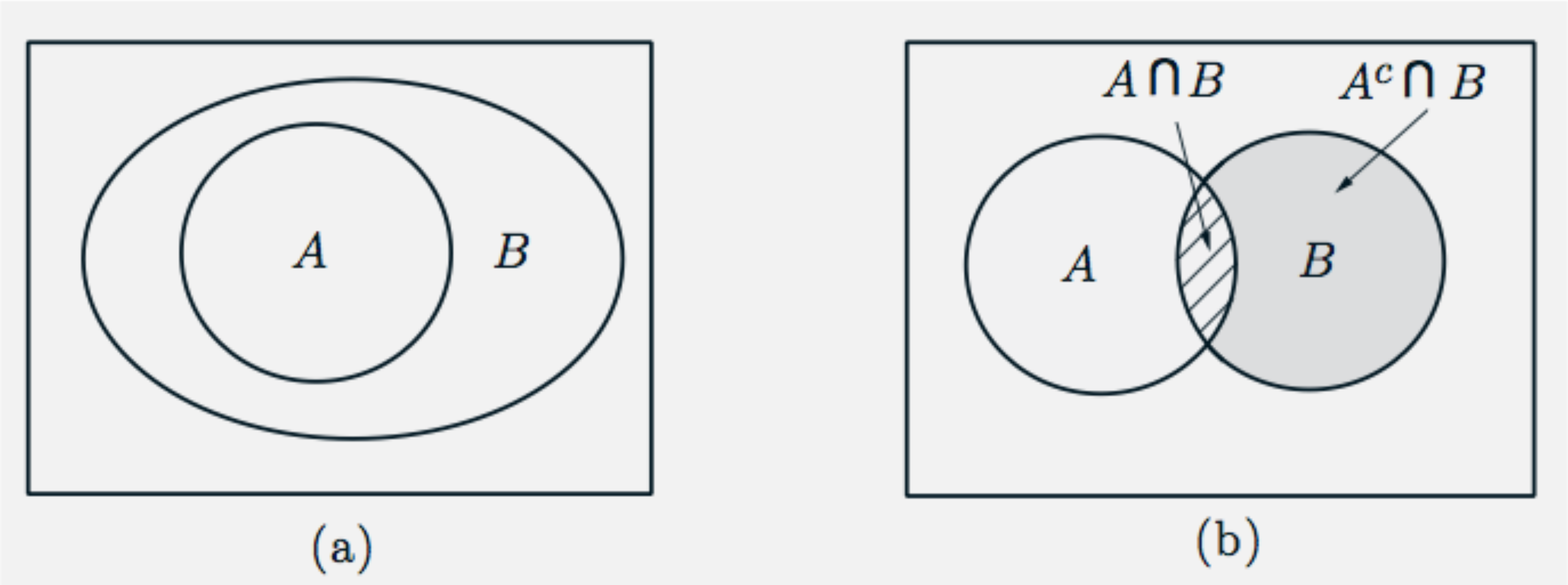
Coursework & Grading

- 7 Homeworks. These are not to be turned in, however they will help you prepare for the quizzes and exams.
- 6 quizzes; your lowest quiz grade is dropped. Your best 5 quiz grades contribute 30% to the course grade.
- Midterm Exam on Thursday February 25th, worth 25% of the course grade.
- Final Exam on WEDNESDAY May 5th, worth 35% of the course grade.
- 10% In-class exercises & participation. There will be some pencil-and-paper exercises given in lecture. Sometimes you will be asked to complete these exercises in groups, sometimes on your own. These exercises will be graded based on effort. It's ok to get the wrong answer, but you need to show that you tried. Missed exercises cannot be made up, but your lowest exercise grade will be dropped. Class participation also contributes to your grade, including on Piazza (e.g. asking questions and answering other students' questions).

Advice for success in this course

- Complete the assigned readings before each lecture. Readings are listed on the schedule; check the schedule often for changes/updates.
- Start the homeworks early. You are highly encouraged to form study groups of 2-4 students, and discuss the homework problems together once you have attempted them on your own. You will need to master the skills to solve the homework problems, in order to prepare for the quizzes and exams.
- Ask and answer questions on Piazza. Helping to answer questions can help you solidify your own understanding.
- Attend office hours to clarify remaining questions.

Venn Diagrams



Credit: Bertsekas & Tsitsiklis, 2008

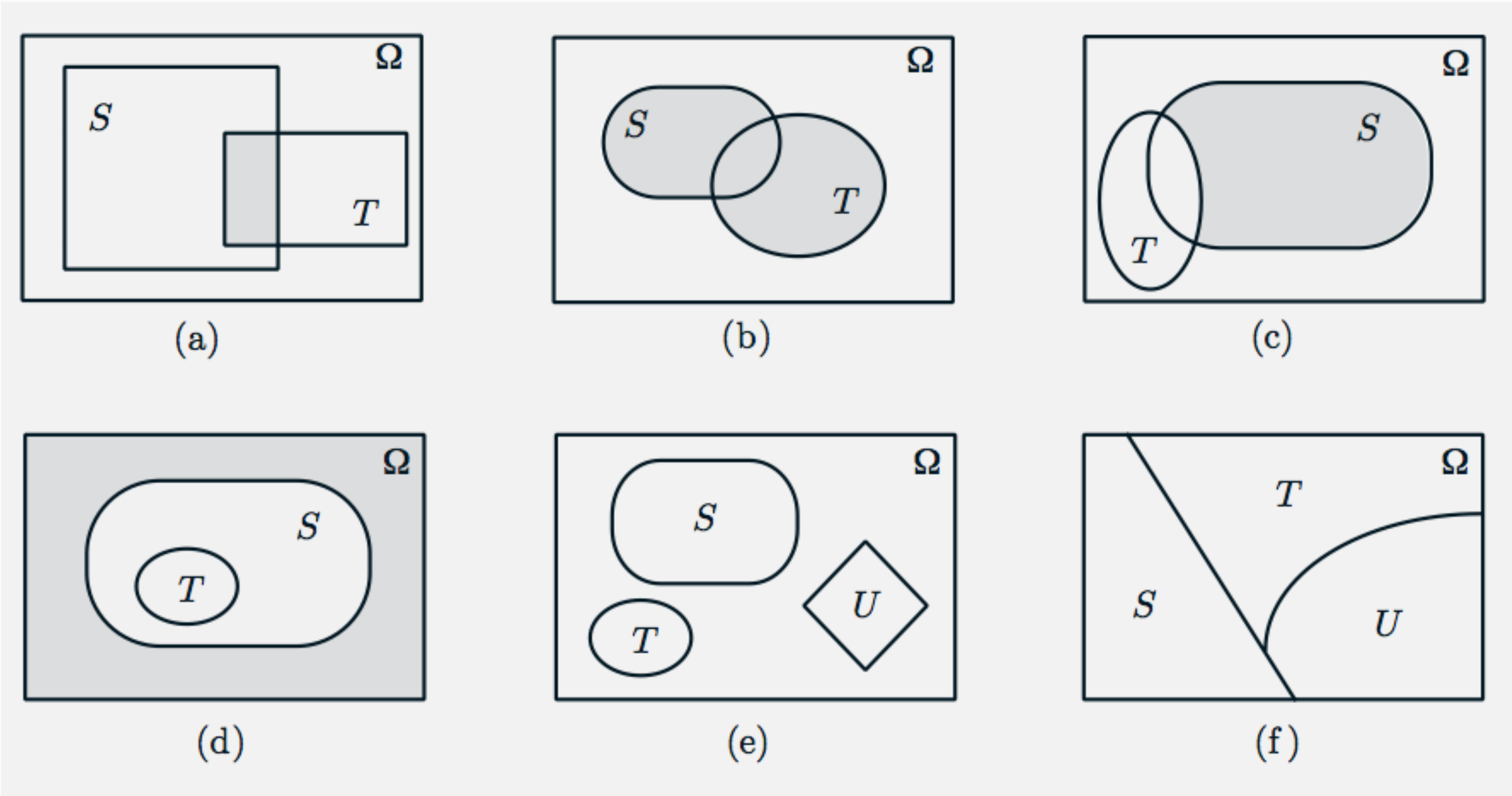


Figure 1.1: Examples of Venn diagrams. (a) The shaded region is $S \cap T$. (b) The shaded region is $S \cup T$. (c) The shaded region is $S \cap T^c$. (d) Here, $T \subset S$. The shaded region is the complement of S. (e) The sets S, T, and U are disjoint. (f) The sets S, T, and U form a partition of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

Credit: Bertsekas & Tsitsiklis, 2008

Set Operations

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

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