Sets and Probabilistic Models

Berlin Chen
Department of Computer Science & Information Engineering
National Taiwan Normal University

Reference:

- D. P. Bertsekas, J. N. Tsitsiklis, *Introduction to Probability* , Sections 1.1-1.2

Sets (1/2)

- A **set** is a collection of objects which are the **elements** of the set
  - If x is an element of set S, denoted by $x \in S$
  - Otherwise denoted by $x \notin S$
- A set that has no elements is called **empty set** is denoted by Ø
- Set specification
  - Countably finite: $\{1,2,3,4,5,6\}$
  - Countably infinite: $\{0,2,-2,4,-4,\ldots\}$
  - With a certain property: $\{k \mid k/2 \text{ is integer}\}$ (countably infinite)

    $\{x \mid 0 \le x \le 1\}$ (uncountable)

    $\{x \mid x \text{ satisfies } P\}$

    such that

Sets (2/2)

- If every element of a set S is also an element of a set T, then S is a **subset** of T
  - Denoted by $S \subset T$ or $T \supset S$

- If $S \subset T$ and $T \subset S$, then the two sets are **equal**
  - Denoted by $S = T$

- The universal set, denoted by Ω, which contains all objects of interest in a particular context
  - After specifying the context in terms of universal set Ω, we only consider sets S that are subsets of Ω

Set Operations (1/3)

- Complement
  - The **complement** of a set S with respect to the universe Ω, is the set $\{x \in \Omega \mid x \notin S\}$,namely, the set of all elements that do not belong to S, denoted by S^c
  - The complement of the universe $\Omega^c = \varnothing$

- Union
  - The **union** of two sets S and T is the set of all elements that belong to S or T, denoted by $S \bigcup T$
    $S \bigcup T = \{x \mid x \in S \text{ or } x \in T\}$
- Intersection
  - The **intersection** of two sets S and T is the set of all elements that belong to both S and T, denoted by $S \bigcap T$
    $S \bigcap T = \{x \mid x \in S \text{ and } x \in T\}$

Set Operations (2/3)

- The union or the intersection of several (or even infinite many) sets

$$\bigcup_{n=1}^{\infty} S_n = S_1 \cup S_2 \cup \cdots = \{x \mid x \in S_n \text{ for some } n\}$$

$$\bigcap_{n=1}^{\infty} S_n = S_1 \cap S_2 \cap \cdots = \{x \mid x \in S_n \text{ for all } n\}$$

- Disjoint
  - Two sets are **disjoint** if their intersection is empty (e.g.,$S \cap T$ = Ø)

- Partition
  - A collection of sets is said to be a **partition** of a set S if the sets in the collection are disjoint and their union is S

Set Operations (3/3)

- Visualization of set operations with Venn diagrams

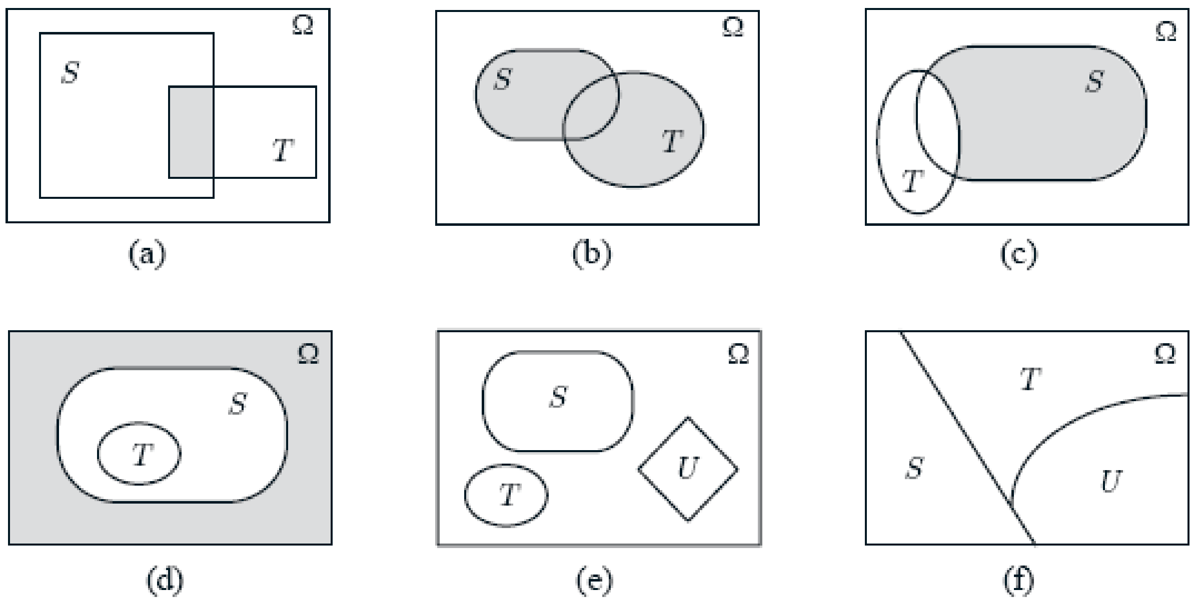


Figure 1.1: Examples of Venn diagrams. (a) The shaded region is $S \cap T$. (b) The shaded region is $S \cup T$. (c) The shaded region is $S \cap T^c$. (d) Here, $T \subset S$. The shaded region is the complement of S. (e) The sets S, T, and U are disjoint. (f) The sets S, T, and U form a partition of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

The Algebra of Sets

- The following equations are the elementary consequences of the set definitions and operations

commutative

$$S \cup T = T \cup S,$$

associative

$$S \cup (T \cup U) = (S \cup T) \cup U$$

distributive

$$S \cap (T \cup U) = (S \cap T) \cup (S \cap U),$$

distributive

$$S \cup (T \cap U) = (S \cup T) \cap (S \cup U),$$

$$(S^c)^c = S,$$

$$S \cap S^c = \varnothing$$

$$S \cup \Omega = \Omega,$$

$$S \cap \Omega = S.$$

- De Morgan's law

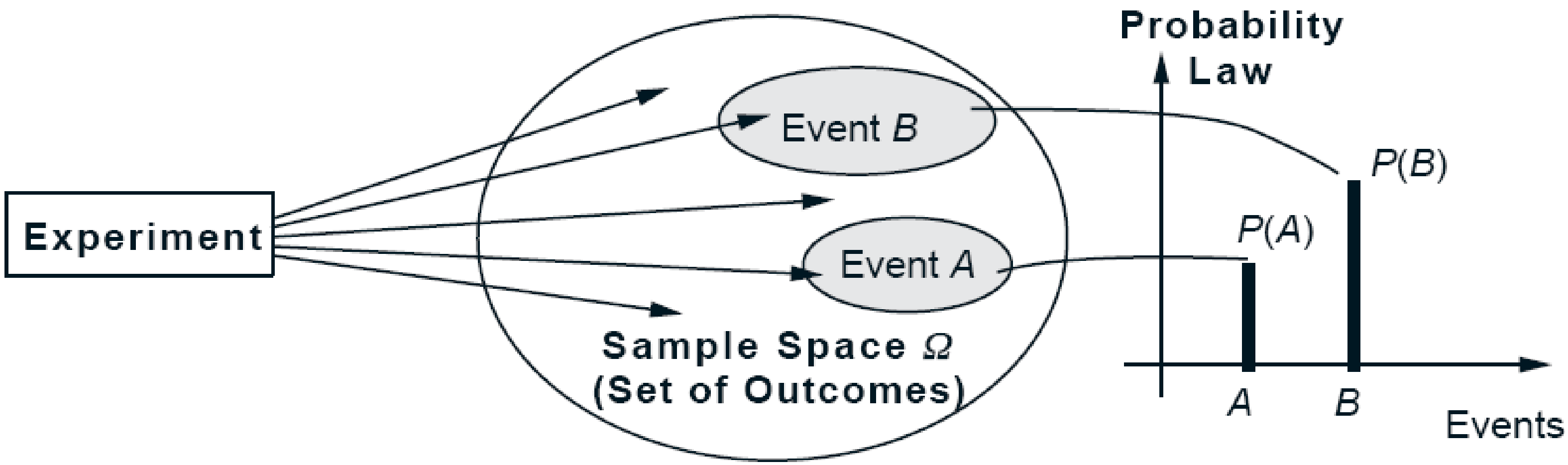
$$\left(\bigcup_n S_n\right)^c = \bigcap_n S_n^c \qquad \left(\bigcap_n S_n\right)^c = \bigcup_n S_n^c$$

Probabilistic Models (1/2)

- A probabilistic model is a mathematical description of an uncertainty situation
  - It has to be in accordance with a fundamental framework to be discussed shortly
- Elements of a probabilistic model
  - The **sample space**
    - The set of all possible outcomes of an experiment
  - The **probability law**
    - Assign to a set A of possible outcomes (also called an **event**) a nonnegative number $\mathbf{P}(A)$ (called the **probability** of A) that encodes our knowledge or belief about the collective "likelihood" of the elements of A

Probabilistic Models (2/2)

- The main ingredients of a probabilistic model



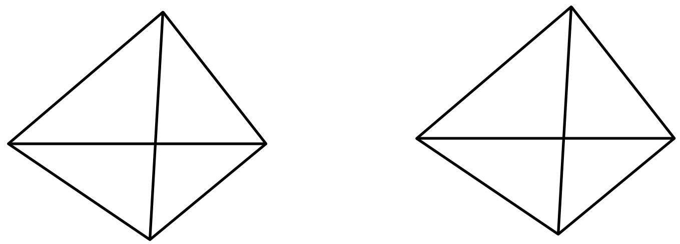
Sample Spaces and Events (1/2)

- Each probabilistic model involves an underlying process, called the **experiment**
  - That produces exactly one out of several possible **outcomes**
  - The set of all possible outcomes is called the **sample space** of the experiment, denoted by Ω
  - A subset of the sample space (a collection of possible outcomes) is called an **event**

- Examples of the **experiment**
  - A single toss of a coin (finite outcomes)
  - Three tosses of two dice (finite outcomes)
  - An infinite sequences of tosses of a coin (infinite outcomes)
  - Throwing a dart on a square (infinite outcomes), etc.

Sample Spaces and Events (2/2)

- Properties of the sample space
  - Elements of the sample space must be **mutually exclusive**
  - The sample space must be **collectively exhaustive**
  - The sample space should be at the “right” granularity (avoiding irrelevant details)



Granularity of the Sample Space

- Example 1.1. Consider two alternative games, both involving ten successive coin tosses:
  - Game 1: We receive $1 each time a head comes up
  - Game 2: We receive $1 for every coin toss, up to and including the first time a head comes up. Then, we receive $2 for every coin toss, up to the second time a head comes up. More generally, the dollar amount per toss is doubled each time a head comes up

>> Game 1 consists of 11 (0,1,..,10) possible outcomes (of money received)

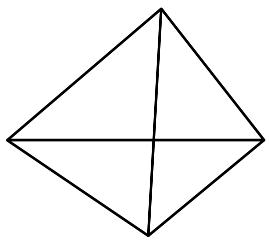
>> Game 2 consists of ?? possible outcomes (of money received)

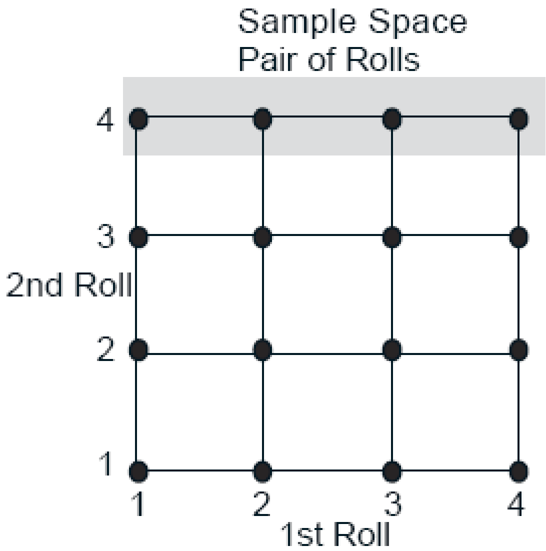
- A finer description is needed 1, 1, 1, 2, 2, 2, 4, 4, 4, 8
- E.g., each outcome corresponds to a possible ten-long sequence of heads and tails (will each sequence have a distinct outcome?)

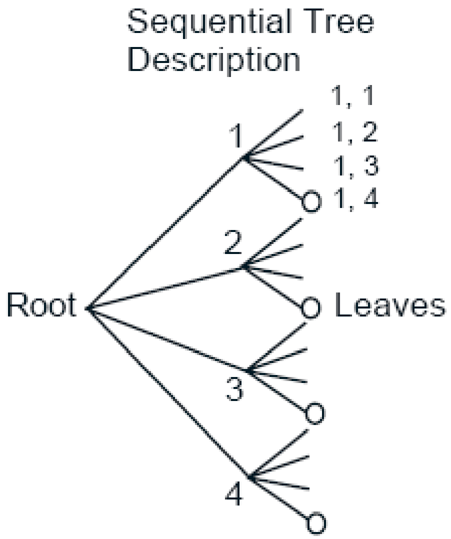
Sequential Probabilistic Models

- Many experiments have an inherent sequential character
  - Tossing a coin three times
  - Observing the value of stock on five successive days
  - Receiving eight successive digits at a communication receiver

>> They can be described by means of a **tree-based sequential description**







Probability Laws

- Given the sample space associated with an experiment is settled on, a probability law
  - Specify the **likelihood** of any outcome, and/or of any set of possible outcomes (an event)
  - Or alternatively, assign to every event A, a number $\mathbf{P}(A)$, called the **probability** of A, satisfying the following axioms:

Probability Axioms

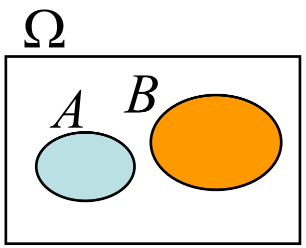
1. **(Nonnegativity)** $\mathbf{P}(A) \geq 0$, for every event A.
2. **(Additivity)** If A and B are two disjoint events, then the probability of their union satisfies

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

   Furthermore, 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, $\mathbf{P}(\Omega) = 1$.



$\mathbf{P}(\Omega) = 1$

$\mathbf{P}(\varnothing) = 0$

Probability Laws for Discrete Models

- 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:

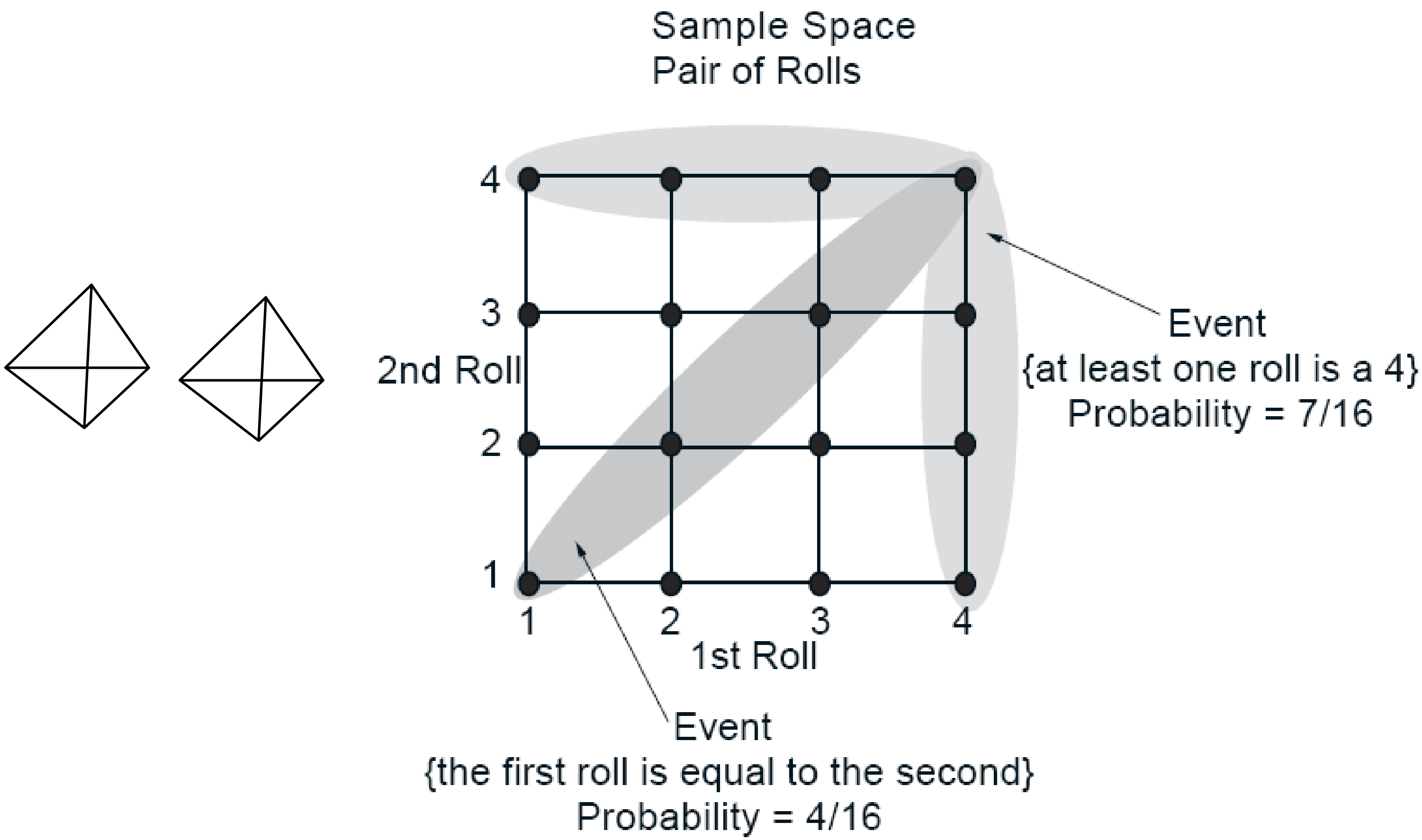
$$\mathbf{P}(\{s_1, s_2, \ldots, s_n\}) = \mathbf{P}(\{s_1\}) + \mathbf{P}(\{s_2\}) + \cdots + \mathbf{P}(\{s_n\})$$
$$= \mathbf{P}(s_1) + \mathbf{P}(s_2) + \cdots + \mathbf{P}(s_n)$$

- 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

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

An Example of Sample Space and Probability Law

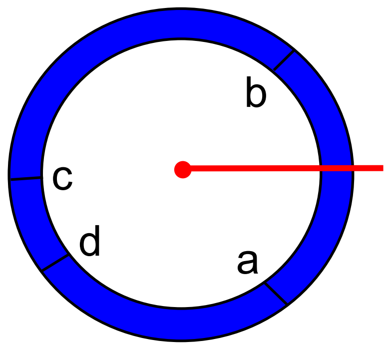
- The experiment of rolling a pair of 4-sided dice



Continuous Models

- Probabilistic models with continuous sample spaces
  - It is inappropriate to assign probability to each single-element event (?)
  - Instead, it makes sense to assign probability to any interval (one-dimensional) or area (two-dimensional) of the sample space

- Example: a wheel of fortune

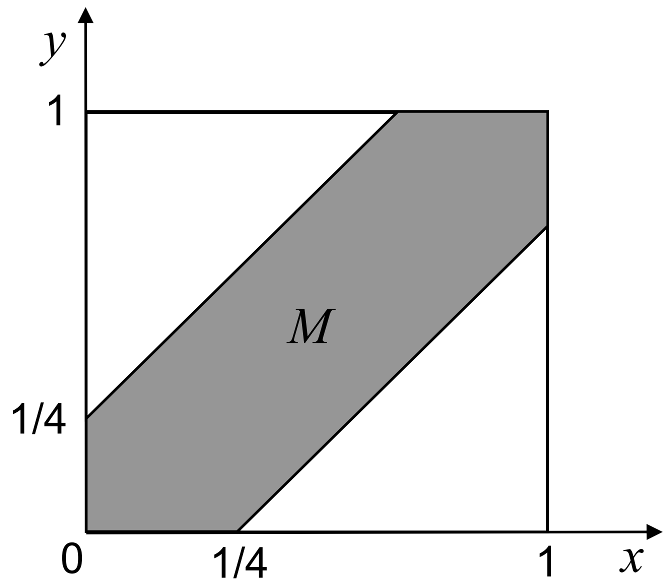


$$\mathbf{P}(\{0.3\}) = ?$$
$$\mathbf{P}(\{0.33\}) = ?$$
$$\mathbf{P}(\{0.333\}) = ?$$
$$\ldots$$

$$\mathbf{P}(\{x \mid a \le x \le b\}) = ?$$

Another Example for Continuous Models

- **Example 1.5**: Romeo and Juliet have a date at a given time, and each will arrive at the meeting place with a delay between 0 and 1 hour, with all pairs of delays being equally likely. The first to arrive will wait for 15 minutes and will leave if the other has not yet arrived. What is the probability that they will meet?



x : arriving time for Romeo

y : arriving time for Juliet

M : the event that Romeo and Juliet will meet

$$M = \{(x, y) \mid |x - y| \le 1/4,\ 0 \le x \le 1,\ 0 \le y \le 1\}$$

$$1 - \left(\tfrac{3}{4}\right)\left(\tfrac{3}{4}\right) = \tfrac{7}{16}$$

Properties of Probability Laws (1/2)

- Probability laws have a number of properties, which can be deduced from the axioms. Some of them are summarized below

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

(d) $\mathbf{P}(A \cup B \cup C) = \mathbf{P}(A) + \mathbf{P}(A^c \cap B) + \mathbf{P}(A^c \cap B^c \cap C)$.

Properties of Probability Laws (2/2)

- Visualization and verification using Venn diagrams

Figure 1.6: Visualization and verification of various properties of probability laws using Venn diagrams. If $A \subset B$, then B is the union of the two disjoint events A and $A^c \cap B$; see diagram (a). Therefore, by the additivity axiom, we have

$$\mathbf{P}(B) = \mathbf{P}(A) + \mathbf{P}(A^c \cap B) \geq \mathbf{P}(A),$$

where the inequality follows from the nonnegativity axiom, and verifies property (a).

From diagram (b), we can express the events $A \cup B$ and B as unions of disjoint events:

$$A \cup B = A \cup (A^c \cap B), \qquad B = (A \cap B) \cup (A^c \cap B).$$

The additivity axiom yields

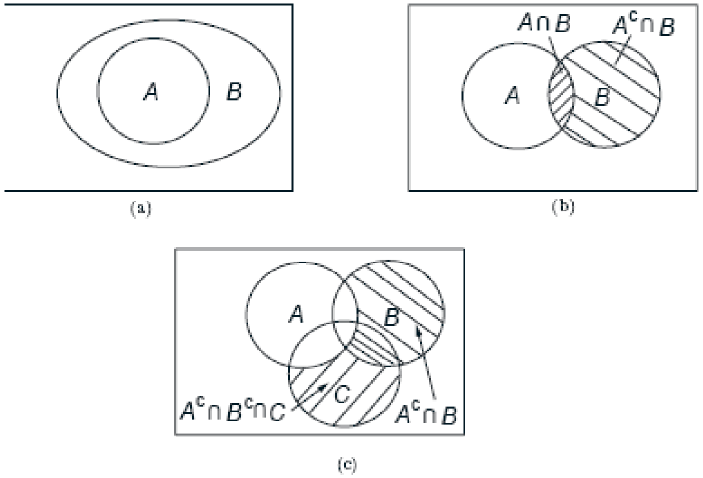
$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(A^c \cap B), \qquad \mathbf{P}(B) = \mathbf{P}(A \cap B) + \mathbf{P}(A^c \cap B).$$

Subtracting the second equality from the first and rearranging terms, we obtain $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$, verifying property (b). Using also the fact $\mathbf{P}(A \cap B) \geq 0$ (the nonnegativity axiom), we obtain $\mathbf{P}(A \cup B) \leq \mathbf{P}(A) + \mathbf{P}(B)$, verifying property (c)

From diagram (c), we see that the event $A \cup B \cup C$ can be expressed as a union of three disjoint events:

$$A \cup B \cup C = A \cup (A^c \cap B) \cup (A^c \cap B^c \cap C),$$

so property (d) follows as a consequence of the additivity axiom.



Model and Reality (1/2)

- Using the framework of probability theory to analyze uncertainty in a wide variety of contexts involves two distinct stages
  - In the first stage, we construct a probabilistic model, by specifying a probability law on a suitably defined sample space.
    - An open-ended task !
  - In the second stage, we work within a fully specified probabilistic model and derive the probabilities of certain events, or deduce some interesting properties
    - Tightly regulated by rules of ordinary logic and the axioms of probability.

Model and Reality (2/2)

- Example "Bertrand's Paradox"

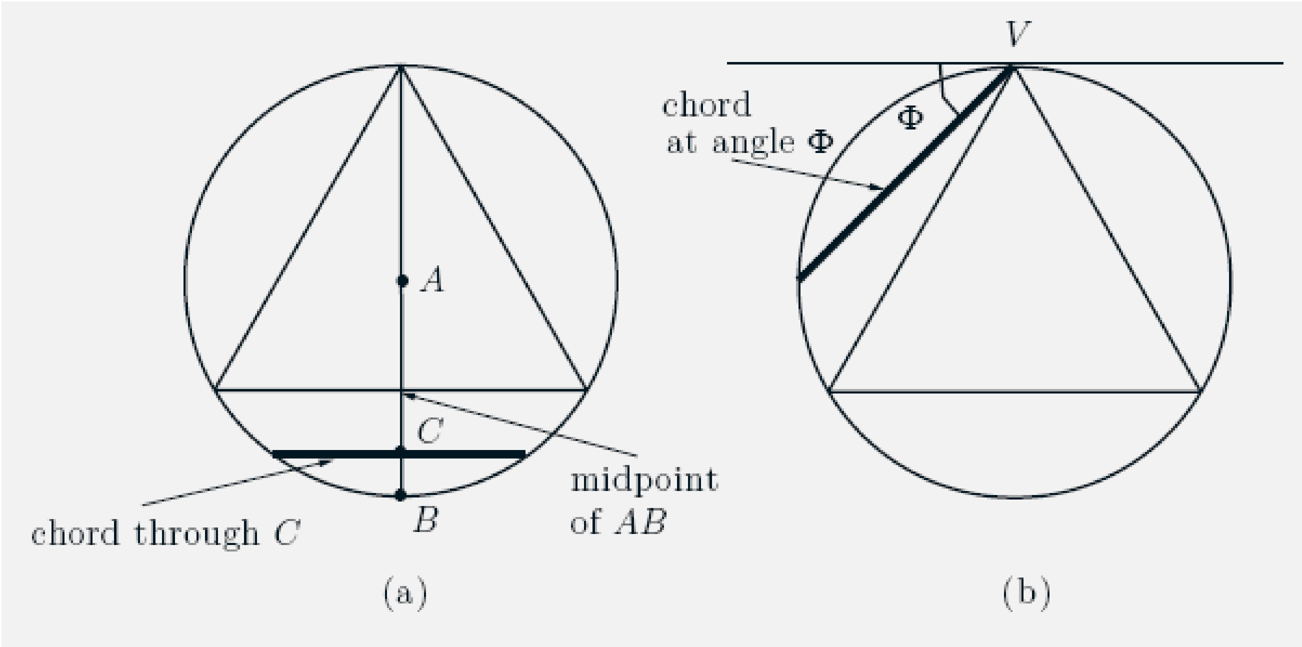


Figure 1.7: This example, presented by L. F. Bertrand in 1889, illustrates the need to specify unambiguously a probabilistic model. Consider a circle and an equilateral triangle inscribed in the circle. What is the probability that the length of a randomly chosen chord of the circle is greater than the side of the triangle? The answer here depends on the precise meaning of "randomly chosen." The two methods illustrated in parts (a) and (b) of the figure lead to contradictory results.

Recitation

- SECTION 1.1 Set
  - Problem 3
- SECTION 1.2 Probabilistic Models
  - Problems 5, 8 and 9