THE UNIVERSITY OF TEXAS AT AUSTIN
Department of Statistics and Data Sciences
College of Natural Sciences

SDS 321: Introduction to Probability and Statistics

Lecture 1: Axioms of Probability

Purnamrita Sarkar
Department of Statistics and Data Science
The University of Texas at Austin

www.cs.cmu.edu/∼psarkar/teaching

Getting Started

Your instructor:	Prof. Purna Sarkar
email:	`purna.sarkar@austin.utexas.edu`
Office Hours:	Tuesdays 11:30-12:30, GDC 7.306
Your TA:	Krishna Teja Rekapalli
email:	krrish1729@gmail.com
TA Office Hours:	Wednesdays 5-7pm

Course Overview

- This course provides an introduction to probability and statistics.
- The first section will be on fundamentals of probability, including:
  - Discrete and continuous random variables
  - Combinatorics
  - Multiple random variables
  - Functions of random variables
  - Limit theorems
- The second section will be on statistics, including:
  - Parameter estimation
  - Hypothesis testing
- We will consider mainly classical statistics. If time permits we will discuss Bayesian Statistics.

Course materials

- Course syllabus, slides and homework assignments will be posted at `www.cs.cmu.edu/~psarkar/teaching`
- Grades will be posted at `canvas.utexas.edu`
- The course text books are
  1. Introduction to Probability, by Dimitri P. Bertsekas and John N. Tsitsiklis.
  2. A First Course in Probability, by Sheldon Ross

  - Another good book that covers similar material is
    - Introduction to Probability, by Charles M. Grinstead and Laurie J. Snell

Assessment

- 4 exams. 2 midterms, and the final will consist of 2 midterm length exams.
- I will take the best 3 out of 4.
- The final grade will be 25% Homework, 25% from all 3 exams.
- Homework will be assigned (approximately) weekly, with roughly 10 homeworks in total.
- Homeworks should be submitted via Canvas by 5pm one week after it is assigned.

What is probability?

"If I flip this coin, the probability of getting heads is 0.5"

- What does this mean?

What is probability?

"If I flip this coin, the probability of getting heads is 0.5"

- What does this mean?
- If I were to toss the coin 10 times, roughly 5 times I will see a head.
- A probability of 1 means it is certain, a probability of 0 means it is impossible.
- In general, you do an *experiment* many times, and you count how many times a particular *event* occurs. The proportion roughly gives you the probability of that particular event.

Experiments and events?

- Experiment: Tossing a coin twice

- Experiment: You throw two dice

Experiments and events?

- Experiment: Tossing a coin twice
  - Event: you get two heads
- Experiment: You throw two dice

Experiments and events?

- Experiment: Tossing a coin twice
  - Event: you get two heads
  - Event: you get two different outcomes
- Experiment: You throw two dice

Experiments and events?

- Experiment: Tossing a coin twice
  - Event: you get two heads
  - Event: you get two different outcomes
- Experiment: You throw two dice
  - Event: the sum of the rolls is six

Experiments and events?

- Experiment: Tossing a coin twice
  - Event: you get two heads
  - Event: you get two different outcomes
- Experiment: You throw two dice
  - Event: the sum of the rolls is six
  - Event: you get two odd faces.

Sample space

- The sample space Ω is the set of all possible outcomes of an experiment.

Sample space

- The sample space Ω is the set of all possible outcomes of an experiment.
- You rolled one die. What is Ω? .

Sample space

- The sample space Ω is the set of all possible outcomes of an experiment.
- You rolled one die. What is Ω? $\{1, 2, 3, 4, 5, 6\}$.

Sample space

- The sample space Ω is the set of all possible outcomes of an experiment.
- You rolled one die. What is Ω? $\{1, 2, 3, 4, 5, 6\}$.
- You tossed three coins together. What is Ω?

.

Sample space

- The sample space Ω is the set of all possible outcomes of an experiment.
- You rolled one die. What is Ω? $\{1, 2, 3, 4, 5, 6\}$.
- You tossed three coins together. What is Ω?
  $\{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$.
- The different elements of a sample space must be **mutually exclusive** and **collectively exhaustive**.

Sample space

- The sample space Ω is the set of all possible outcomes of an experiment.
- You rolled one die. What is Ω? $\{1, 2, 3, 4, 5, 6\}$.
- You tossed three coins together. What is Ω?
  $\{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$.
- The different elements of a sample space must be **mutually exclusive** and **collectively exhaustive**.
  - Ω for three coin tosses cannot be $\{\text{at least one head}, \text{at most one tail}\}$.

Sample space

- The sample space Ω is the set of all possible outcomes of an experiment.
- You rolled one die. What is Ω? $\{1, 2, 3, 4, 5, 6\}$.
- You tossed three coins together. What is Ω? $\{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$.
- The different elements of a sample space must be **mutually exclusive** and **collectively exhaustive**.
  - Ω for three coin tosses cannot be $\{\text{at least one head}, \text{at most one tail}\}$.
- An event is a collection of possible outcomes.

Simple and compound events

- Simple event:
  - Your two coin tosses came up HH.
  - Your rolled die shows a 6
- Compound event: can be decomposed into simple events
  - Your two coin tosses give two different outcomes
  - The sum of the two rolled dice is six
  - You got two odd faces from rolling two dice.

Simple and compound events

- Simple event:
  - Your two coin tosses came up HH.
  - Your rolled die shows a 6
- Compound event: can be decomposed into simple events
  - Your two coin tosses give two different outcomes
    - You got HT or TH.
  - The sum of the two rolled dice is six
  - You got two odd faces from rolling two dice.

Simple and compound events

- Simple event:
  - Your two coin tosses came up HH.
  - Your rolled die shows a 6
- Compound event: can be decomposed into simple events
  - Your two coin tosses give two different outcomes
    - You got HT or TH.
  - The sum of the two rolled dice is six
    - You got $(1,5)$ or $(2,4)$ or $(3,3)$ or $(4,2)$ or $(5,1)$.
  - You got two odd faces from rolling two dice.

Simple and compound events

- Simple event:
  - Your two coin tosses came up HH.
  - Your rolled die shows a 6
- Compound event: can be decomposed into simple events
  - Your two coin tosses give two different outcomes
    - You got HT or TH.
  - The sum of the two rolled dice is six
    - You got $(1, 5)$ or $(2, 4)$ or $(3, 3)$ or $(4, 2)$ or $(5, 1)$.
  - You got two odd faces from rolling two dice.
    - You got $(1, 1)$, or $(1, 3)$ or $\ldots$.

Sets and sample spaces

We need to introduce some mathematical concepts to define probability more concretely:

- A **set** is a collection of objects, which are called **elements**
  - The natural numbers are a set, where the elements are individual numbers.
  - This class is the set, where the elements are the professor, the TA and the students.
- If an element x is in a set S, we write $x \in S$.
- If a set contains no elements, we call it the **empty set**, $\emptyset$.
- If a set contains every possible element, we call it the **universal set**, Ω.

Sets

- A set can be **finite** (e.g. the set of people in this class) or **infinite** (e.g. the set of real numbers).
  - Set of primary colors = {red, blue, yellow}.

Sets

- A set can be **finite** (e.g. the set of people in this class) or **infinite** (e.g. the set of real numbers).
  - Set of primary colors = {red, blue, yellow}.
- If we can enumerate the elements of an infinite set, i.e. arrange the elements in a list, we say it is countable.
  - Set of positive integers $= \{1, 2, \dots\}$

Sets

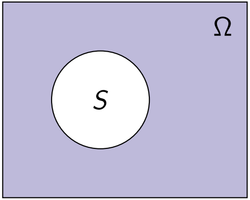
- A set can be **finite** (e.g. the set of people in this class) or **infinite** (e.g. the set of real numbers).
  - Set of primary colors = {red, blue, yellow}.
- If we can enumerate the elements of an infinite set, i.e. arrange the elements in a list, we say it is countable.
  - Set of positive integers = $\{1, 2, \dots\}$
- If we cannot enumerate the elements, we say it is uncountable.
  - the real numbers
  - the set of all subsets of natural numbers, aka the power set

Sets

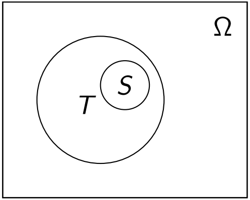
- A set can be **finite** (e.g. the set of people in this class) or **infinite** (e.g. the set of real numbers).
  - Set of primary colors = {red, blue, yellow}.
- If we can enumerate the elements of an infinite set, i.e. arrange the elements in a list, we say it is countable.
  - Set of positive integers $= \{1, 2, \dots\}$
- If we cannot enumerate the elements, we say it is uncountable.
  - the real numbers
  - the set of all subsets of natural numbers, aka the power set
- We can use curly brackets to describe a set in terms of its elements:
  - Sample space of a die roll: $S = \{1, 2, 3, 4, 5, 6\}$
  - Arbitrary set where all the elements meet some criterion C:
    $S = \{x | x \text{ satisfies } C\}$

Operations on sets

- Let the **universal set** Ω be the set of all objects we might possibly be interested in.
- The **complement**, S^C, of a set S, w.r.t. Ω, is the set of all elements that are in Ω but not in S. So $\Omega^C = \emptyset$.
- We say $S \subseteq T$, if every element in S is also in T.
- $S \subseteq T$ and $T \subseteq S$ if and only if $S = T$.



S^C is the shaded region



$S \subset T \subset \Omega$

Operations on sets: Union, Intersection, Difference

- The **union**, $S \cup T$, of two sets S and T is the set of elements that are in either S or T (or both): $S \cup T = \{x | x \in S \text{ or } x \in T\}$.

Operations on sets: Union, Intersection, Difference

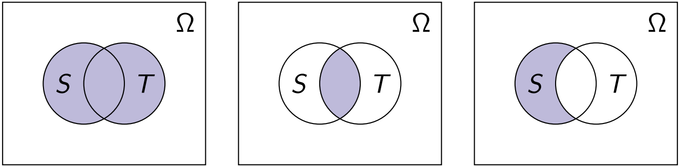
- The **union**, $S \cup T$, of two sets S and T is the set of elements that are in either S or T (or both): $S \cup T = \{x | x \in S \text{ or } x \in T\}$.
- The **intersection**, $S \cap T$, of two sets S and T is the set of elements that are in both S and T: $S \cap T = \{x | x \in S \text{ and } x \in T\}$

Operations on sets: Union, Intersection, Difference

- The **union**, $S \cup T$, of two sets S and T is the set of elements that are in either S or T (or both): $S \cup T = \{x | x \in S \text{ or } x \in T\}$.
- The **intersection**, $S \cap T$, of two sets S and T is the set of elements that are in both S and T: $S \cap T = \{x | x \in S \text{ and } x \in T\}$
- The **difference**, $S \setminus T$, of two sets S and T is the set of elements that are in S, but not in T: $S \setminus T = \{x | x \in S \text{ and } x \notin T\}$

Operations on sets: Union, Intersection, Difference

- The **union**, $S \cup T$, of two sets S and T is the set of elements that are in either S or T (or both): $S \cup T = \{x | x \in S \text{ or } x \in T\}$.
- The **intersection**, $S \cap T$, of two sets S and T is the set of elements that are in both S and T: $S \cap T = \{x | x \in S \text{ and } x \in T\}$
- The **difference**, $S \setminus T$, of two sets S and T is the set of elements that are in S, but not in T: $S \setminus T = \{x | x \in S \text{ and } x \notin T\}$



$S \cup T$ $\quad$ $S \cap T$ $\quad$ $S \setminus T = S \cap T^c$

Operations on sets

- We can extend the notions of union and intersection to multiple (even infinitely many!) sets:

$$\bigcup_{i=1}^{n} S_n = S_1 \cup S_2 \cup \cdots \cup S_n = \{x | x \in S_n \text{ for some } 1 \leq i \leq n\}$$

$$\bigcap_{i=1}^{n} S_n = S_1 \cap S_2 \cap \cdots \cap S_n = \{x | x \in S_n \text{ for all } 1 \leq i \leq n\}$$

- We say two sets are **disjoint** if their intersection is empty.
- We say a collection of sets are disjoint if no two sets have any common elements.
- If a collection of disjoint sets have union S, we call them a **partition** of S.

Probability laws

- The probability law assigns to an event E a non-negative number $P(E)$ which encodes our belief/knowledge about the "likelihood" of the event E.
- Axioms of probability:

Probability laws

- The probability law assigns to an event E a non-negative number $P(E)$ which encodes our belief/knowledge about the "likelihood" of the event E.
- Axioms of probability:
  - **Nonnegativity**: $P(A) \geq 0$, for every event A.
  - **Additivity**: If A and B are two disjoint events, then the probability of their union satisfies $P(A \cup B) = P(A) + P(B)$.
    This extends to the union of infinitely many disjoint events:
    $$P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots$$
  - **Normalization**: The probability of the entire sample space Ω is equal to 1, i.e. $P(\Omega) = 1$

Examples

You tossed two fair dice together. What is the probability of the event $E = \{\text{sum of the rolls} = 6\}$?

Examples

You tossed two fair dice together. What is the probability of the event $E = \{\text{sum of the rolls} = 6\}$?

- The sample space is $\{(i, j) | 1 \leq i, j \leq 6\}$. There are a total of 36 outcomes.

Examples

You tossed two fair dice together. What is the probability of the event $E = \{\text{sum of the rolls} = 6\}$?

- The sample space is $\{(i,j) | 1 \leq i,j \leq 6\}$. There are a total of 36 outcomes.
- Since the dice are fair, each outcome is **equally likely**.
  - This means every outcome has probability $1/36$.

Examples

You tossed two fair dice together. What is the probability of the event $E = \{\text{sum of the rolls} = 6\}$?

- The sample space is $\{(i, j) | 1 \leq i, j \leq 6\}$. There are a total of 36 outcomes.
- Since the dice are fair, each outcome is **equally likely**.
  - This means every outcome has probability $1/36$.
- First decompose into simple events. We get a sum of 6 if we get $(1, 5)$ or $(2, 4)$ or $(3, 3)$ or $(4, 2)$ or $(5, 1)$.

Examples

You tossed two fair dice together. What is the probability of the event $E = \{\text{sum of the rolls} = 6\}$?

- The sample space is $\{(i, j) | 1 \leq i, j \leq 6\}$. There are a total of 36 outcomes.
- Since the dice are fair, each outcome is **equally likely**.
  - This means every outcome has probability $1/36$.
- First decompose into simple events. We get a sum of 6 if we get $(1, 5)$ or $(2, 4)$ or $(3, 3)$ or $(4, 2)$ or $(5, 1)$.
- Using the additivity law we have $P(E) = 5/36$.

Examples

You tossed two fair dice together. What is the probability of the event $E = \{\text{sum of the rolls} = 6\}$?

- The sample space is $\{(i,j)|1 \leq i,j \leq 6\}$. There are a total of 36 outcomes.
- Since the dice are fair, each outcome is **equally likely**.
  - This means every outcome has probability $1/36$.
- First decompose into simple events. We get a sum of 6 if we get $(1,5)$ or $(2,4)$ or $(3,3)$ or $(4,2)$ or $(5,1)$.
- Using the additivity law we have $P(E) = 5/36$.

This is an example of an **uniform distribution**, where all outcomes are equally likely.

Properties of probability laws

All the following can be proven by decomposing a set into disjoint partitions and using the **additivity** and **non-negativity** rules.

- If $A \subseteq B$, then $P(A) \leq P(B)$.

Properties of probability laws

All the following can be proven by decomposing a set into disjoint partitions and using the **additivity** and **non-negativity** rules.

- If $A \subseteq B$, then $P(A) \leq P(B)$.
  - $B = A \cup (A \setminus B)$, where A and $B \setminus A$ are disjoint. So $P(B) = P(A) + P(B \setminus A) \geq P(A)$.

Properties of probability laws

All the following can be proven by decomposing a set into disjoint partitions and using the **additivity** and **non-negativity** rules.

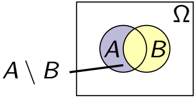
- If $A \subseteq B$, then $P(A) \leq P(B)$.
  - $B = A \cup (A \setminus B)$, where A and $B \setminus A$ are disjoint. So $P(B) = P(A) + P(B \setminus A) \geq P(A)$.

- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Properties of probability laws

All the following can be proven by decomposing a set into disjoint partitions and using the **additivity** and **non-negativity** rules.

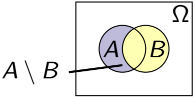
- If $A \subseteq B$, then $P(A) \leq P(B)$.
  - $B = A \cup (A \setminus B)$, where A and $B \setminus A$ are disjoint. So $P(B) = P(A) + P(B \setminus A) \geq P(A)$.
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
  - $A \cup B = (A \setminus B) \cup B$. So the additivity rule gives $P(A \cup B) = P(A \setminus B) + P(B)$. Can you finish the proof?



Properties of probability laws

All the following can be proven by decomposing a set into disjoint partitions and using the **additivity** and **non-negativity** rules.

- If $A \subseteq B$, then $P(A) \leq P(B)$.
  - $B = A \cup (A \setminus B)$, where A and $B \setminus A$ are disjoint. So $P(B) = P(A) + P(B \setminus A) \geq P(A)$.
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
  - $A \cup B = (A \setminus B) \cup B$. So the additivity rule gives $P(A \cup B) = P(A \setminus B) + P(B)$. Can you finish the proof?



- $P(A \cup B) \leq P(A) + P(B)$.
- $P(A \cup B \cup C) = P(A) + P(A^c \cap B) + P(A^c \cap B^c \cap C)$.