1. Probabilistic Models

1.1 Learning Objectives

By the end of this lecture, you will be able to:

- Understand what constitutes a probabilistic model
- State and interpret Kolmogorov's three axioms
- Apply the discrete probability law and uniform probability law
- Derive basic properties from the axioms

1.2 Building a Probabilistic Model

A probabilistic model consists of two ingredients:

1.2 Building a Probabilistic Model

A probabilistic model consists of two ingredients:

1. A **sample space** S: the set of all possible outcomes

1.2 Building a Probabilistic Model

A probabilistic model consists of two ingredients:

1. A **sample space** S: the set of all possible outcomes
2. A **probability law** P: assigns probabilities to events

1.2 Building a Probabilistic Model

A probabilistic model consists of two ingredients:

1. A **sample space** S: the set of all possible outcomes
2. A **probability law** P: assigns probabilities to events

To be a proper probability model, the probability law must follow certain rules — the **axioms**.

1.3 Probability Laws

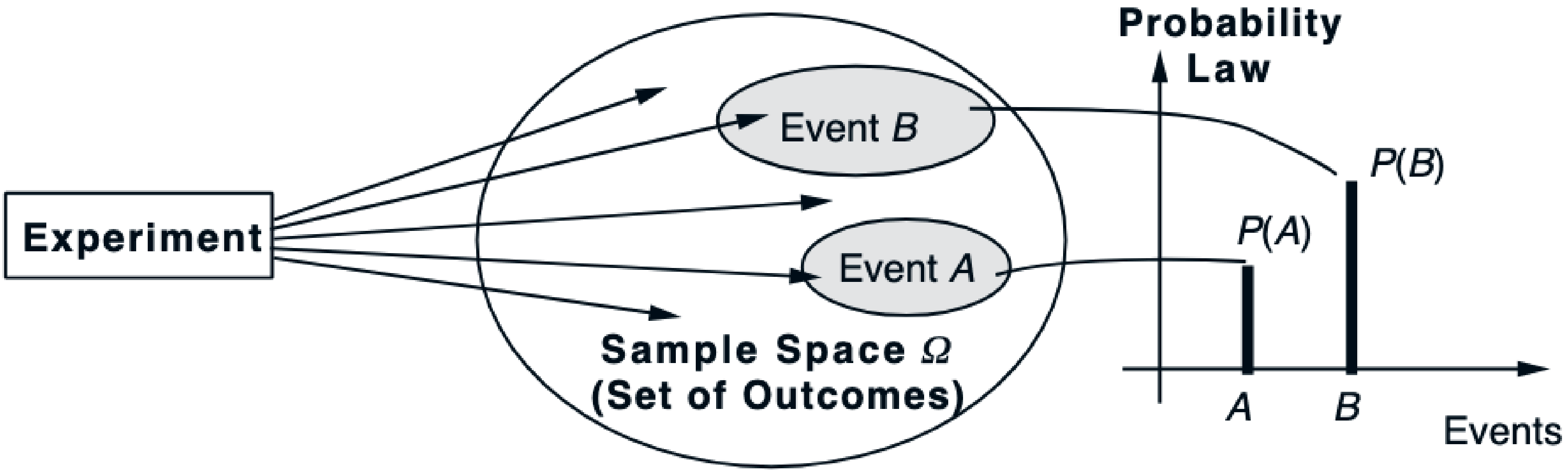


Figure 1.2: The main ingredients of a probabilistic model.

Bertsekas and Tsitsiklis

1.4 Probability Laws

We’ve already seen a certain kind of probability law…

Can you describe the law that governs the probability of outcomes for a fair die?

1.5 Kolmogorov’s Axioms

For a sample space S and events $A, B, \ldots$:

1.5 Kolmogorov's Axioms

For a sample space S and events $A, B, ...$:

Definition: Axiom 1 (Non-negativity)

$P(A) \geq 0$ for every event A

1.5 Kolmogorov's Axioms

For a sample space S and events $A, B, \ldots$:

Definition: Axiom 1 (Non-negativity)

$P(A) \geq 0$ for every event A

Definition: Axiom 2 (Normalization)

$P(S) = 1$

1.5 Kolmogorov's Axioms

For a sample space S and events $A, B, \ldots$:

Definition: Axiom 1 (Non-negativity)

$P(A) \geq 0$ for every event A

Definition: Axiom 2 (Normalization)

$P(S) = 1$

Definition: Axiom 3 (Additivity)

If A and B are disjoint events, then $P(A \cup B) = P(A) + P(B)$

1.5 Kolmogorov's Axioms

For a sample space S and events $A, B, \ldots$:

Definition: Axiom 1 (Non-negativity)

$P(A) \geq 0$ for every event A

Definition: Axiom 2 (Normalization)

$P(S) = 1$

Definition: Axiom 3 (Additivity)

If A and B are disjoint events, then $P(A \cup B) = P(A) + P(B)$ More generally, for disjoint $A_1, A_2, \ldots$:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

1.6 Interpreting the Axioms: The Mass Analogy

Think of probability as **spreading one unit of mass** over the sample space.

1.6 Interpreting the Axioms: The Mass Analogy

Think of probability as **spreading one unit of mass** over the sample space.

- Each outcome gets some mass (possibly zero)
- Total mass equals 1 (normalization)
- Mass of an event = sum of masses of its outcomes (additivity)
- All masses are non-negative

1.6 Interpreting the Axioms: The Mass Analogy

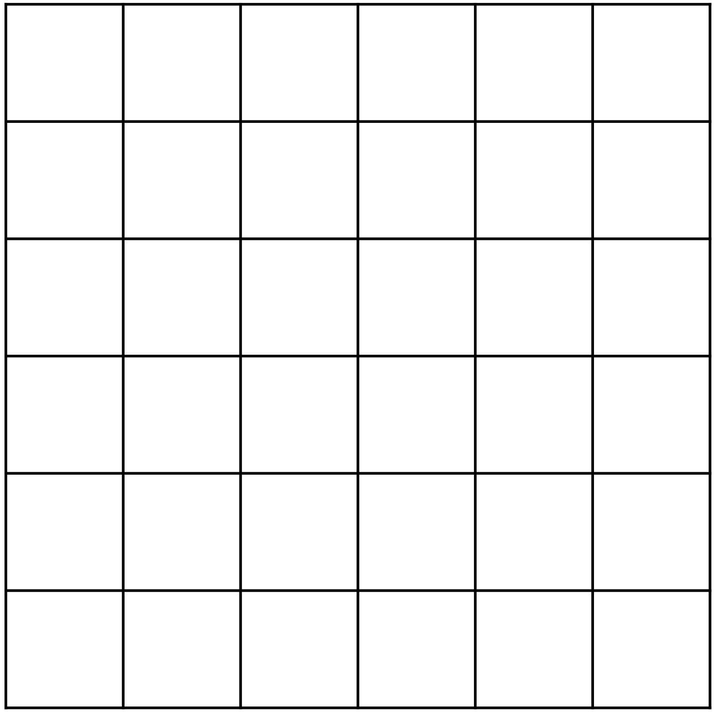
Think of probability as **spreading one unit of mass** over the sample space.

- Each outcome gets some mass (possibly zero)
- Total mass equals 1 (normalization)
- Mass of an event = sum of masses of its outcomes (additivity)
- All masses are non-negative

$P(A)$ = total mass assigned to outcomes in A

1.7 Mass Analogy

Draw pile of sand spread over sample space



1.8 Properties from the Axioms

From just three axioms, we can derive all probability rules.

1.8 Properties from the Axioms

From just three axioms, we can derive all probability rules.

Example: Empty Set Probability

We know: $S \cup \emptyset = S$ and they're disjoint

By additivity: $P(S) + P(\emptyset) = P(S)$

Therefore: $P(\emptyset) = 0$

1.9 The Complement Rule

For any event A:

$$P(A^c) = 1 - P(A)$$

1.9 The Complement Rule

For any event A:

$$P(A^c) = 1 - P(A)$$

Proof:

- A and A^c are disjoint

1.9 The Complement Rule

For any event A:

$$P(A^c) = 1 - P(A)$$

Proof:

- A and A^c are disjoint
- $A \cup A^c = S$

1.9 The Complement Rule

For any event A:

$$P(A^c) = 1 - P(A)$$

Proof:

- A and A^c are disjoint
- $A \cup A^c = S$
- By additivity: $P(A) + P(A^c) = P(S) = 1$

1.9 The Complement Rule

For any event A:

$$P(A^c) = 1 - P(A)$$

Proof:

- A and A^c are disjoint
- $A \cup A^c = S$
- By additivity: $P(A) + P(A^c) = P(S) = 1$
- Rearranging: $P(A^c) = 1 - P(A)$

1.9 The Complement Rule

For any event A:

$$P(A^c) = 1 - P(A)$$

Proof:

- A and A^c are disjoint
- $A \cup A^c = S$
- By additivity: $P(A) + P(A^c) = P(S) = 1$
- Rearranging: $P(A^c) = 1 - P(A)$

Note: This is often easier than computing $P(A)$ directly!

1.10 More Properties

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

1.10 More Properties

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

Proof: We can write $B = A \cup (B \cap A^c)$ where the sets are disjoint.

By additivity: $P(B) = P(A) + P(B \cap A^c) \geq P(A)$

(since $P(B \cap A^c) \geq 0$ by non-negativity)

1.10 More Properties

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

Proof: We can write $B = A \cup (B \cap A^c)$ where the sets are disjoint.

By additivity: $P(B) = P(A) + P(B \cap A^c) \geq P(A)$

(since $P(B \cap A^c) \geq 0$ by non-negativity)

1.11 The Union Bound

Union Bound: $P(A \cup B) \leq P(A) + P(B)$

(Equality when A and B are disjoint)

1.12 Inclusion-Exclusion Principle

Definition: Inclusion-Exclusion (Two Events)

For any events A and B:

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

Draw Venn diagram showing A, B, and $A \cap B$

1.12 Inclusion-Exclusion Principle

Definition: Inclusion-Exclusion (Two Events)

For any events A and B:

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

Draw Venn diagram showing A, B, and $A \cap B$

Why subtract $P(A \cap B)$? Because outcomes in both A and B get counted twice when we add $P(A) + P(B)$.

1.12 Inclusion-Exclusion Principle

Definition: Inclusion-Exclusion (Two Events)

For any events A and B:

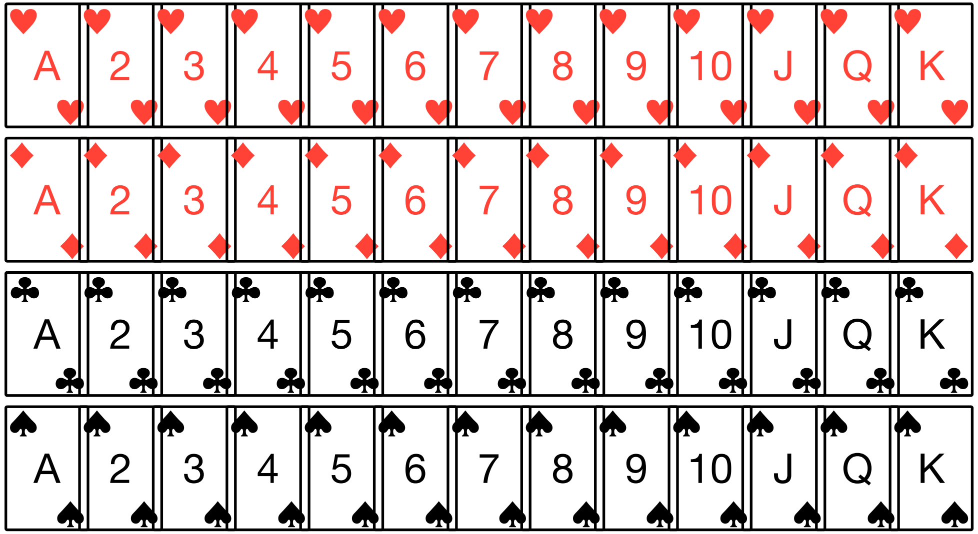
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Draw Venn diagram showing A, B, and $A \cap B$

Why subtract $P(A \cap B)$? Because outcomes in both A and B get counted twice when we add $P(A) + P(B)$.

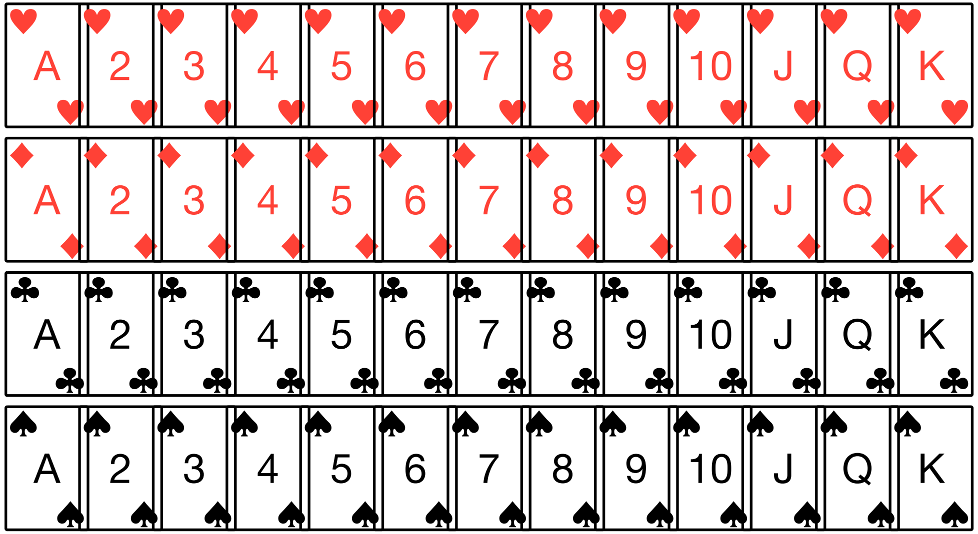
Note: We'll explore inclusion-exclusion for more than two events in a later lecture.

1.13 Inclusion-Exclusion Principle Example



$P(\text{hearts or face card})?$

1.13 Inclusion-Exclusion Principle Example



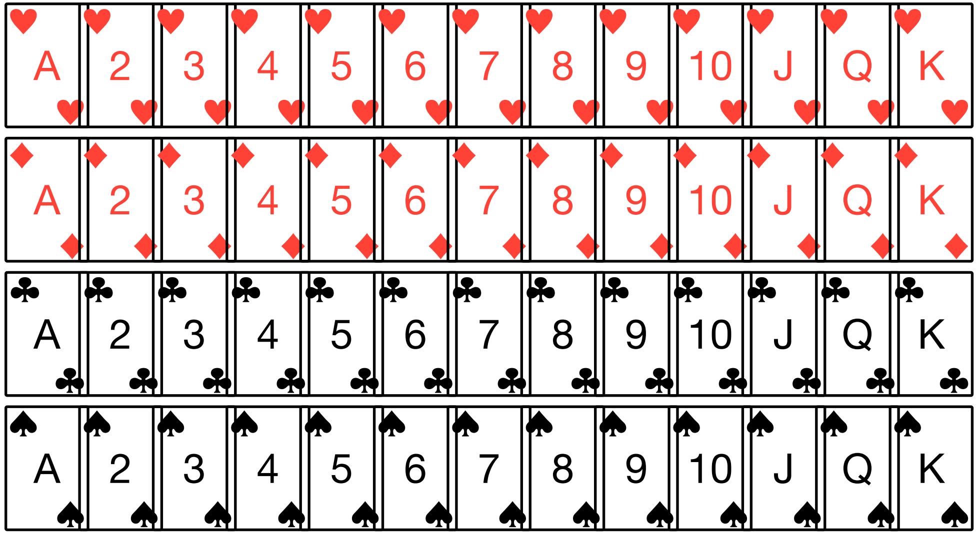
$P(\text{hearts or face card})$?

$P(\text{hearts or face card}) =$

$P(\text{hearts}) + P(\text{face card}) -$

$P(\text{hearts and face card})$

1.13 Inclusion-Exclusion Principle Example



$P(\text{hearts or face card})$?

$P(\text{hearts or face card}) =$

$P(\text{hearts}) + P(\text{face card}) -$

$P(\text{hearts and face card})$

$P(\text{hearts or face card}) = \frac{13}{52} + \frac{12}{52} - \frac{3}{52} = \frac{22}{52} = \frac{11}{26}$

1.14 Inclusion-Exclusion Principle (Three Events)

For any events A, B, and C:

$$P(A \cup B \cup C) =$$

$$P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

1.15 Inclusion-Exclusion Principle (General Case)

For any events $A_1, A_2, ..., A_n$:

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) - ... + (-1)^{n+1} P(A_1 \cap A_2 \cap ... \cap A_n)$$

1.16 Another Useful Property

For any events A, B, and C:

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

1.16 Another Useful Property

For any events A, B, and C:

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

This partitions $A \cup B \cup C$ into three **disjoint** pieces:

- Outcomes in A
- Outcomes in B but not A
- Outcomes in C but not A or B

1.16 Another Useful Property

For any events A, B, and C:

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

This partitions $A \cup B \cup C$ into three **disjoint** pieces:

- Outcomes in A
- Outcomes in B but not A
- Outcomes in C but not A or B

Note: This generalizes: any union can be written as a disjoint union by “removing” previous sets.

2. Discrete Probability Laws

2.1 Sample Space Depends on the Question

Example: Two Network Games

Consider monitoring 10 packets on a network, each may be **corrupted** (C) or **intact** (I):

2.1 Sample Space Depends on the Question

Example: Two Network Games

Consider monitoring 10 packets on a network, each may be **corrupted** (C) or **intact** (I):

Game 1: You win $1 for each intact packet.

Game 2: You win $1 per packet until the first corruption, then $2 per packet until the second corruption, etc.

2.1 Sample Space Depends on the Question

Example: Two Network Games

Consider monitoring 10 packets on a network, each may be **corrupted** (C) or **intact** (I):

Game 1: You win \$1 for each intact packet.

Game 2: You win \$1 per packet until the first corruption, then \$2 per packet until the second corruption, etc.

For Game 1: $S = \{0, 1, 2, ..., 10\}$ (total count of intact packets)

For Game 2: $S = \{\text{all } \ 2^{10} \text{ sequences of C and I}\}$ (order matters!)

2.1 Sample Space Depends on the Question

Example: Two Network Games

Consider monitoring 10 packets on a network, each may be **corrupted** (C) or **intact** (I):

Game 1: You win $1 for each intact packet.

Game 2: You win $1 per packet until the first corruption, then $2 per packet until the second corruption, etc.

For Game 1: $S = \{0, 1, 2, ..., 10\}$ (total count of intact packets)

For Game 2: $S = \{\text{all } \ 2^{10} \text{ sequences of C and I}\}$ (order matters!)

Note: The sample space should capture exactly what matters for the problem.

2.2 Discrete Sample Spaces

When the sample space is **finite** or **countably infinite**:

$$S = \{s_1, s_2, ..., s_n\}$$

2.2 Discrete Sample Spaces

When the sample space is **finite** or **countably infinite**:

$$S = \{s_1, s_2, ..., s_n\}$$

We can specify the probability law by assigning a probability p_i to each outcome s_i.

2.2 Discrete Sample Spaces

When the sample space is **finite** or **countably infinite**:

$$S = \{s_1, s_2, ..., s_n\}$$

We can specify the probability law by assigning a probability p_i to each outcome s_i.

Requirements (from axioms):

- Each $p_i \geq 0$ (non-negativity)
- $\sum_{i=1}^{n} p_i = 1$ (normalization)

2.3 Discrete Probability Law

Definition: Discrete Probability Law

For any event A:

$$P(A) = \sum_{s_i \in A} P(\{s_i\})$$

The probability of an event is the sum of the probabilities of its elements.

2.3 Discrete Probability Law

Definition: Discrete Probability Law

For any event A:

$$P(A) = \sum_{s_i \in A} P(\{s_i\})$$

The probability of an event is the sum of the probabilities of its elements.

This follows directly from the additivity axiom!

2.4 Special Case: Equally Likely Outcomes

Definition: Discrete Uniform Probability Law

If all n outcomes are equally likely:

$$P(\{s_i\}) = \frac{1}{n} \quad \text{for all } i$$

Then for any event A:

$$P(A) = \frac{|A|}{|S|} = \frac{\text{number of favorable outcomes}}{\text{total number of outcomes}}$$

2.4 Special Case: Equally Likely Outcomes

Definition: Discrete Uniform Probability Law

If all n outcomes are equally likely:

$$P(\{s_i\}) = \frac{1}{n} \quad \text{for all } i$$

Then for any event A:

$$P(A) = \frac{|A|}{|S|} = \frac{\text{number of favorable outcomes}}{\text{total number of outcomes}}$$

This is exactly our **naive definition** from week 1!

2.5 Example: Rolling Two Dice

Roll two fair 4-sided dice. Sample space: all pairs (i, j) where $i, j \in \{1, 2, 3, 4\}$.

2.5 Example: Rolling Two Dice

Roll two fair 4-sided dice. Sample space: all pairs (i, j) where $i, j \in \{1, 2, 3, 4\}$.

Total outcomes: $4 \times 4 = 16$, each with probability $\frac{1}{16}$.

2.5 Example: Rolling Two Dice

Roll two fair 4-sided dice. Sample space: all pairs (i, j) where $i, j \in \{1, 2, 3, 4\}$.

Total outcomes: $4 \times 4 = 16$, each with probability $\frac{1}{16}$.

What is $P(\text{sum is even})$?

2.5 Example: Rolling Two Dice

Roll two fair 4-sided dice. Sample space: all pairs (i, j) where $i, j \in \{1, 2, 3, 4\}$.

Total outcomes: $4 \times 4 = 16$, each with probability $\frac{1}{16}$.

What is $P(\text{sum is even})$?

Count: $(1, 1), (1, 3), (2, 2), (2, 4), (3, 1), (3, 3), (4, 2), (4, 4) \longrightarrow$ 8 outcomes

$$P(\text{sum is even}) = \frac{8}{16} = \frac{1}{2}$$

2.5 Example: Rolling Two Dice

Roll two fair 4-sided dice. Sample space: all pairs (i, j) where $i, j \in \{1, 2, 3, 4\}$.

Total outcomes: $4 \times 4 = 16$, each with probability $\frac{1}{16}$.

What is $P(\text{sum is even})$?

Count: $(1, 1), (1, 3), (2, 2), (2, 4), (3, 1), (3, 3), (4, 2), (4, 4) \longrightarrow$ 8 outcomes

$$P(\text{sum is even}) = \frac{8}{16} = \frac{1}{2}$$

What is $P(\text{at least one die shows } 4)$?

2.5 Example: Rolling Two Dice

Roll two fair 4-sided dice. Sample space: all pairs (i, j) where $i, j \in \{1, 2, 3, 4\}$.

Total outcomes: $4 \times 4 = 16$, each with probability $\frac{1}{16}$.

What is $P(\text{sum is even})$?

Count: $(1, 1), (1, 3), (2, 2), (2, 4), (3, 1), (3, 3), (4, 2), (4, 4) \longrightarrow$ 8 outcomes

$$P(\text{sum is even}) = \frac{8}{16} = \frac{1}{2}$$

What is $P(\text{at least one die shows } 4)$?

$$P(\text{at least one } 4) = \frac{7}{16}$$

2.6 Example: Three Coin Tosses

Experiment: Toss a fair coin three times.

2.6 Example: Three Coin Tosses

Experiment: Toss a fair coin three times.

Sample space:

$$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

2.6 Example: Three Coin Tosses

Experiment: Toss a fair coin three times.

Sample space:

$$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

With fair coin: each outcome has probability $\frac{1}{8}$

2.6 Example: Three Coin Tosses

Experiment: Toss a fair coin three times.

Sample space:

$$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

With fair coin: each outcome has probability $\frac{1}{8}$

Let A = “exactly 2 heads” = $\{HHT, HTH, THH\}$

2.6 Example: Three Coin Tosses

Experiment: Toss a fair coin three times.

Sample space:

$$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

With fair coin: each outcome has probability $\frac{1}{8}$

Let A = “exactly 2 heads” = $\{HHT, HTH, THH\}$

$$P(A) = P(\{HHT\}) + P(\{HTH\}) + P(\{THH\}) = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8}$$

2.7 Example: Biased Coin

Now suppose the coin has $P(\text{heads}) = 0.6$

2.7 Example: Biased Coin

Now suppose the coin has $P(\mathrm{heads}) = 0.6$

The tosses are independent, so:

- $P(\{HHH\}) = 0.6 \times 0.6 \times 0.6 = 0.216$
- $P(\{HHT\}) = 0.6 \times 0.6 \times 0.4 = 0.144$
- $P(\{TTT\}) = 0.4 \times 0.4 \times 0.4 = 0.064$
- …

2.7 Example: Biased Coin

Now suppose the coin has $P(\mathrm{heads}) = 0.6$

The tosses are independent, so:

- $P(\{HHH\}) = 0.6 \times 0.6 \times 0.6 = 0.216$
- $P(\{HHT\}) = 0.6 \times 0.6 \times 0.4 = 0.144$
- $P(\{TTT\}) = 0.4 \times 0.4 \times 0.4 = 0.064$
- …

3. Continuous Probability Laws

3.1 Continuous Sample Spaces

What if outcomes are **continuous** (e.g., real numbers)?

3.1 Continuous Sample Spaces

What if outcomes are **continuous** (e.g., real numbers)?

Example: Response time T of a server, where $T \in [0, \infty)$

3.1 Continuous Sample Spaces

What if outcomes are **continuous** (e.g., real numbers)?

Example: Response time T of a server, where $T \in [0, \infty)$

Problem: We can't sum over uncountably many outcomes!

3.1 Continuous Sample Spaces

What if outcomes are **continuous** (e.g., real numbers)?

Example: Response time T of a server, where $T \in [0, \infty)$

Problem: We can't sum over uncountably many outcomes!

Solution: Use a **probability density function** (PDF) $f(x)$ and integrate:

$$P(A) = \int_A f(x)\,\mathrm{d}x$$

3.2 Single Points Have Zero Probability

In continuous models, $P(\{x\}) = 0$ for any single point x!

3.2 Single Points Have Zero Probability

In continuous models, $P(\{x\}) = 0$ for any single point x!

Why? If $P(\{x\}) > 0$ for any point, then by additivity, intervals would have infinite probability.

3.2 Single Points Have Zero Probability

In continuous models, $P(\{x\}) = 0$ for any single point x!

Why? If $P(\{x\}) > 0$ for any point, then by additivity, intervals would have infinite probability.

This seems strange but is consistent:

- $P(T = 3.0000\ldots \text{ exactly}) = 0$
- $P(2.9 < T < 3.1) > 0$

3.2 Single Points Have Zero Probability

In continuous models, $P(\{x\}) = 0$ for any single point x!

Why? If $P(\{x\}) > 0$ for any point, then by additivity, intervals would have infinite probability.

This seems strange but is consistent:

- $P(T = 3.0000\ldots \text{ exactly}) = 0$
- $P(2.9 < T < 3.1) > 0$

Note: We can only assign positive probability to **intervals** (or unions of intervals).

3.3 Continuous Uniform Distribution

Definition: Continuous Uniform Law

If outcomes are “equally likely” over interval $[a, b]$:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a, b] \\ 0 & \text{otherwise} \end{cases}$$

For any subinterval $[c, d] \subseteq [a, b]$:

$$P(c \leq X \leq d) = \frac{d-c}{b-a} = \frac{\text{length of event}}{\text{length of sample space}}$$

3.4 Example: Random Number Generator

A random number generator produces values uniformly in $[0, 1]$.

3.4 Example: Random Number Generator

A random number generator produces values uniformly in $[0, 1]$.

$$P(0.3 \leq X \leq 0.7) = 0.7 - 0.3 = 0.4$$

3.4 Example: Random Number Generator

A random number generator produces values uniformly in $[0, 1]$.

$$P(0.3 \leq X \leq 0.7) = 0.7 - 0.3 = 0.4$$

$$P(X = 0.5) = 0$$

3.4 Example: Random Number Generator

A random number generator produces values uniformly in $[0, 1]$.

$$P(0.3 \leq X \leq 0.7) = 0.7 - 0.3 = 0.4$$

$$P(X = 0.5) = 0$$

$$P(X \leq 0.25 \text{ or } X \geq 0.75) = 0.25 + 0.25 = 0.5$$

3.5 Example: Meeting Times

Alice and Bob agree to meet between 12:00 and 1:00 PM.

Each arrives at a uniformly random time. Whoever arrives first waits 15 minutes.

3.5 Example: Meeting Times

Alice and Bob agree to meet between 12:00 and 1:00 PM.

Each arrives at a uniformly random time. Whoever arrives first waits 15 minutes.

What is P(they meet)?

3.5 Example: Meeting Times

Alice and Bob agree to meet between 12:00 and 1:00 PM.

Each arrives at a uniformly random time. Whoever arrives first waits 15 minutes.

What is P(they meet)?

Sample space: $S = [0, 1] \times [0, 1]$ (arrival times in hours)

They meet if $|T_A - T_B| \leq 0.25$

3.5 Example: Meeting Times

Alice and Bob agree to meet between 12:00 and 1:00 PM.

Each arrives at a uniformly random time. Whoever arrives first waits 15 minutes.

What is P(they meet)?

Sample space: $S = [0, 1] \times [0, 1]$ (arrival times in hours)

They meet if $|T_A - T_B| \leq 0.25$

$$P(\text{meet}) = 1 - 2 \cdot \frac{(0.75)^2}{2} = 1 - \frac{9}{16} = \frac{7}{16}$$

Draw unit square with shaded meeting region