Basics of Probability

Jared K. Harpole

Department of Psychology
University of Kansas

August 13, 2015
About this Course

- The field of probability is vast and we are only going to scratch the surface (e.g. discrete)
- Michael Jordan did not become one of the best basketball players in history in a day.
- At times the material may be dense or unclear, please stop me with questions
What is Probability

This is a complicated question but we will cover the two most common interpretations.

To motivate the definitions let me ask you how you would determine the probability of getting heads from a single flip of a penny?

Classical/Frequentist (Long-term frequency)

We would flip the coin a large number of times and count the number of successes and divide by the number of flips.

Bayesian (Subjective belief)

Make some assumptions about the penny (e.g. assume the penny is a fair coin), flip the coin a couple of times, divide the number of successes (heads) by the number of tosses, then update your belief and repeat this process.

These two camps of thought fight a lot, but for the purposes of this presentation we will mainly focus on the Classical/frequentist interpretation as this is what is most commonly taught in practice.
What is Probability

- This is a complicated question but we will cover the two most common interpretations.
What is Probability

- This is a complicated question but we will cover the two most common interpretations.
- To motivate the definitions let me ask you how you would determine the probability of getting heads from a single flip of a penny?
What is Probability

- This is a complicated question but we will cover the two most common interpretations.
- To motivate the definitions let me ask you how you would determine the probability of getting heads from a single flip of a penny?
  - Classical/Frequentist (Long-term frequency)
What is Probability

- This is a complicated question but we will cover the two most common interpretations.
- To motivate the definitions let me ask you how you would determine the probability of getting heads from a single flip of a penny?
  - Classical/Frequentist (Long-term frequency)
    - We would flip the coin a large number of times and count the number of successes and divide by the number of flips.
What is Probability

- This is a complicated question but we will cover the two most common interpretations.
- To motivate the definitions let me ask you how you would determine the probability of getting heads from a single flip of a penny?
  - Classical/Frequentist (Long-term frequency)
    - We would flip the coin a large number of times and count the number of successes and divide by the number of flips.
  - Bayesian (Subjective belief)
What is Probability

- This is a complicated question but we will cover the two most common interpretations.
- To motivate the definitions let me ask you how you would determine the probability of getting heads from a single flip of a penny?
  - Classical/Frequentist (Long-term frequency)
    - We would flip the coin a large number of times and count the number of successes and divide by the number of flips.
  - Bayesian (Subjective belief)
    - Make some assumptions about the penny (e. g. assume the penny is a fair coin), flip the coin a couple of times, divide the number of successes (heads) by the number of tosses, then update your belief and repeat this process.
What is Probability

- This is a complicated question but we will cover the two most common interpretations.
- To motivate the definitions let me ask you how you would determine the probability of getting heads from a single flip of a penny?
  - Classical/Frequentist (Long-term frequency)
    - We would flip the coin a large number of times and count the number of successes and divide by the number of flips.
  - Bayesian (Subjective belief)
    - Make some assumptions about the penny (e. g. assume the penny is a fair coin), flip the coin a couple of times, divide the number of successes (heads) by the number of tosses, then update your belief and repeat this process.
- These two camps of thought fight a lot, but for the purposes of this presentation we will mainly focus on the Classical/frequentist interpretation as this is what is most commonly taught in practice.
Why should you care?

Sets

If we did not have sets none of us would be here.

Form the underpinnings of probability and statistics.

May be annoying but need to be appreciated.
Why should you care?

- Sets are an abstract concept
Why should you care?

- Sets are an abstract concept.
- If we did not have sets none of us would be here.
Why should you care?

- Sets are an abstract concept
- If we did not have sets none of us would be here
- Form the underpinnings of probability and statistics
Why should you care?

- Sets are an abstract concept.
- If we did not have sets none of us would be here.
- Form the underpinnings of probability and statistics.
- May be annoying but need to be appreciated.
A Gentle Introduction to Sets

- Probability makes use of sets and we briefly review them here.
A Gentle Introduction to Sets

- Probability makes use of sets and we briefly review them here.
- A set is a collection of objects which belong to the set.
A Gentle Introduction to Sets

- Probability makes use of sets and we briefly review them here.
- A set is a collection of objects which belong to the set
  - Ex. number of sides on a 6-sided dice \{1, 2, 3, 4, 5, 6\}
A Gentle Introduction to Sets

- Probability makes use of sets and we briefly review them here.
- A set is a collection of objects which belong to the set
  - Ex. number of sides on a 6-sided dice \(\{1, 2, 3, 4, 5, 6\}\)
  - Ex. even roll of 6-sided dice \(\{2, 4, 6\}\)
A Gentle Introduction to Sets

- Probability makes use of sets and we briefly review them here.
- A set is a collection of objects which belong to the set
  - Ex. number of sides on a 6-sided dice \( \{1, 2, 3, 4, 5, 6\} \)
  - Ex. even roll of 6-sided dice \( \{2, 4, 6\} \)
  - Ex. flipping a coin once \( \{H, T\} \)
Probability makes use of sets and we briefly review them here.

A set is a collection of objects which belong to the set

- Ex. number of sides on a 6-sided dice \{1, 2, 3, 4, 5, 6\}
- Ex. even roll of 6-sided dice \{2, 4, 6\}
- Ex. flipping a coin once \{H, T\}
- Ex. flipping a coin 2 times \{HH, HT, TH, TT\}
A Gentle Introduction to Sets

- Probability makes use of sets and we briefly review them here.
- A set is a collection of objects which belong to the set
  - Ex. number of sides on a 6-sided dice \( \{1, 2, 3, 4, 5, 6\} \)
  - Ex. even roll of 6-sided dice \( \{2, 4, 6\} \)
  - Ex. flipping a coin once \( \{H, T\} \)
  - Ex. flipping a coin 2 times \( \{HH, HT, TH, TT\} \)
- Special Events in Sets
A Gentle Introduction to Sets

- Probability makes use of sets and we briefly review them here.
- A set is a collection of objects which belong to the set
  - Ex. number of sides on a 6-sided dice \( \{1, 2, 3, 4, 5, 6\} \)
  - Ex. even roll of 6-sided dice \( \{2, 4, 6\} \)
  - Ex. flipping a coin once \( \{H, T\} \)
  - Ex. flipping a coin 2 times \( \{HH, HT, TH, TT\} \)
- Special Events in Sets
  - The null set/empty set (nothing happens) \( \emptyset = \{\} \)
A Gentle Introduction to Sets

- Probability makes use of sets and we briefly review them here.
- A set is a collection of objects which belong to the set
  - Ex. number of sides on a 6-sided dice $\{1, 2, 3, 4, 5, 6\}$
  - Ex. even roll of 6-sided dice $\{2, 4, 6\}$
  - Ex. flipping a coin once $\{H, T\}$
  - Ex. flipping a coin 2 times $\{HH, HT, TH, TT\}$

- Special Events in Sets
  - The null set/empty set (nothing happens) $\emptyset = \{\}$
  - The entire collection of events that can occur is the sample space denoted here by $\Omega$
A Gentle Introduction to Sets

- Probability makes use of sets and we briefly review them here.
- A set is a collection of objects which belong to the set
  - Ex. number of sides on a 6-sided dice \( \{1, 2, 3, 4, 5, 6\} \)
  - Ex. even roll of 6-sided dice \( \{2, 4, 6\} \)
  - Ex. flipping a coin once \( \{H, T\} \)
  - Ex. flipping a coin 2 times \( \{HH, HT, TH, TT\} \)
- Special Events in Sets
  - The null set/empty set (nothing happens) \( \emptyset = {} \)
  - The entire collection of events that can occur is the sample space denoted here by \( \Omega \)
    - Ex. sample space of 1 coin flip \( \Omega = \{H, T\} \)
A Gentle Introduction to Sets

• Probability makes use of sets and we briefly review them here.
• A set is a collection of objects which belong to the set
  • Ex. number of sides on a 6-sided dice \( \{1, 2, 3, 4, 5, 6\} \)
  • Ex. even roll of 6-sided dice \( \{2, 4, 6\} \)
  • Ex. flipping a coin once \( \{H, T\} \)
  • Ex. flipping a coin 2 times \( \{HH, HT, TH, TT\} \)
• Special Events in Sets
  • The null set/empty set (nothing happens) \( \emptyset = \{\} \)
  • The entire collection of events that can occur is the sample space denoted here by \( \Omega \)
    • Ex. sample space of 1 coin flip \( \Omega = \{H, T\} \)
    • Ex. sample space of 2 coin flips \( \Omega = \{HH, HT, TH, TT\} \)
A Gentle Introduction to Sets

- Probability makes use of sets and we briefly review them here.
- A set is a collection of objects which belong to the set
  - Ex. number of sides on a 6-sided dice \{1, 2, 3, 4, 5, 6\}
  - Ex. even roll of 6-sided dice \{2, 4, 6\}
  - Ex. flipping a coin once \{H, T\}
  - Ex. flipping a coin 2 times \{HH, HT, TH, TT\}

- Special Events in Sets
  - The null set/empty set (nothing happens) \(\emptyset = {}\)
  - The entire collection of events that can occur is the sample space denoted here by \(\Omega\)
    - Ex. sample space of 1 coin flip \(\Omega = \{H, T\}\)
    - Ex. sample space of 2 coin flips \(\Omega = \{HH, HT, TH, TT\}\)
    - Ex. sample space of 1 roll of 6-sided die \(\Omega = \{1, 2, 3, 4, 5, 6\}\)
Learning Check - Sets

- What are the two most common interpretations of probability?
- What is the set of the first three letters of the alphabet?
- What are the two special events common to all sets?
- What is the sample space of rolling an even number on a 6-sided dice?
A Gentle Introduction to Set Operations - Subset

An example of a set $A$ which is a subset of $\Omega$:

Set $A$ is a subset of $\Omega$ denoted $A \subseteq \Omega$. 

Ex. One roll of a 6-sided dice. 

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

Set $A$ is an even roll ($A = \{2, 4, 6\}$) is a set and is a subset of $\{1, 2, 3, 4, 5, 6\}$ 

We say $\{2, 4, 6\} \subseteq \{1, 2, 3, 4, 5, 6\}$.
A Gentle Introduction to Set Operations - Subset

- An example of a set $A$ which is a subset of $\Omega$:

\[ \Omega \]

\[ A \]

- Set $A$ is a subset of $\Omega$ denoted $A \subseteq \Omega$. 
An example of a set $A$ which is a subset of $\Omega$:

- Set $A$ is a subset of $\Omega$ denoted $A \subseteq \Omega$.
- Ex. One roll of a 6-sided dice.
A Gentle Introduction to Set Operations - Subset

An example of a set $A$ which is a subset of $\Omega$:

- Set $A$ is a subset of $\Omega$ denoted $A \subseteq \Omega$.
- Ex. One roll of a 6-sided dice.
  - $\Omega = \{1, 2, 3, 4, 5, 6\}$
A Gentle Introduction to Set Operations - Subset

- An example of a set $A$ which is a subset of $\Omega$:

  ![Set Diagram]

- Set $A$ is a subset of $\Omega$ denoted $A \subseteq \Omega$.
- Ex. One roll of a 6-sided dice.
  - $\Omega = \{1, 2, 3, 4, 5, 6\}$
  - Set $A$ is an even roll ($A = \{2, 4, 6\}$) is a set and is a subset of $\{1, 2, 3, 4, 5, 6\}$
A Gentle Introduction to Set Operations - Subset

- An example of a set $A$ which is a subset of $\Omega$:

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

  Set $A$ is an even roll ($A = \{2, 4, 6\}$) is a set and is a subset of $\{1, 2, 3, 4, 5, 6\}$

  We say $\{2, 4, 6\} \subseteq \{1, 2, 3, 4, 5, 6\}$
The complement of a set $A$ (denoted $A^c$) with respect to sample space $\Omega$ is illustrated as $A^c$ refers to all elements of $\Omega$ that are not in $A$.

Ex. One roll of 6-sided dice
Given $\Omega = \{1, 2, 3, 4, 5, 6\}$
Set of rolling a 1 $A = \{1\}$
The complement of rolling a 1 $A^c = \{2, 3, 4, 5, 6\}$
The complement of a set $A$ (denoted $A^c$) with respect to sample space $\Omega$ is illustrated as

And refers to all elements of $\Omega$ that are not in $A$
The complement of a set $A$ (denoted $A^c$) with respect to sample space $\Omega$ is illustrated as

And refers to all elements of $\Omega$ that are not in $A$

Ex. One roll of 6-sided dice
The complement of a set $A$ (denoted $A^c$) with respect to sample space $\Omega$ is illustrated as

And refers to all elements of $\Omega$ that are not in $A$

Ex. One roll of 6-sided dice

Given $\Omega = \{1, 2, 3, 4, 5, 6\}$
The complement of a set $A$ (denoted $A^c$) with respect to sample space $\Omega$ is illustrated as

And refers to all elements of $\Omega$ that are not in $A$

Ex. One roll of 6-sided dice

- Given $\Omega = \{1, 2, 3, 4, 5, 6\}$
- Set of rolling a 1 $A = \{1\}$
A Gentle Introduction to Set Operations - Complement

- The complement of a set $A$ (denoted $A^c$) with respect to sample space $\Omega$ is illustrated as

  ![Diagram of a set $\Omega$ with a subset $A$ and its complement $A^c$]

- And refers to all elements of $\Omega$ that are not in $A$

- Ex. One roll of 6-sided dice
  - Given $\Omega = \{1, 2, 3, 4, 5, 6\}$
  - Set of rolling a 1 $A = \{1\}$
  - The complement of rolling a 1 $A^c = \{2, 3, 4, 5, 6\}$
A Gentle Introduction to Set Operations - Union

Given two sets \( A \) and \( B \) that are subsets of the sample space \( \Omega \), the union of \( A \) and \( B \) denoted as \( A \cup B \) is illustrated as:

Refers to all elements that are in \( A \) or \( B \) or in both \( A \) and \( B \).

Example: One roll of a 6-sided dice.

Given \( \Omega = \{1, 2, 3, 4, 5, 6\} \),

Set of rolling a 1 is denoted as \( A = \{1\} \).

Set of not rolling a 1 is denoted as \( B = \{2, 3, 4, 5, 6\} \).

Then \( A \cup B = \{1\} \cup \{2, 3, 4, 5, 6\} = \{1, 2, 3, 4, 5, 6\} \).
A Gentle Introduction to Set Operations - Union

- Given two sets $A$ and $B$ that are subsets of the sample space $\Omega$, the union of $A$ and $B$ denoted $A \cup B$ is illustrated as

- Refers to all elements that are in $A$ or $B$ or in both $A$ and $B$
A Gentle Introduction to Set Operations - Union

- Given two sets $A$ and $B$ that are subsets of the sample space $\Omega$ the union of $A$ and $B$ denoted $A \cup B$ is illustrated as

- Refers to all elements that are in $A$ or $B$ or in both $A$ and $B$
- Ex. One roll of a 6-sided dice
Given two sets $A$ and $B$ that are subsets of the sample space $\Omega$ the union of $A$ and $B$ denoted $A \cup B$ is illustrated as

- Refers to all elements that are in $A$ or $B$ or in both $A$ and $B$
- Ex. One roll of a 6-sided dice
  - Given $\Omega = \{1, 2, 3, 4, 5, 6\}$
Given two sets $A$ and $B$ that are subsets of the sample space $\Omega$ the union of $A$ and $B$ denoted $A \cup B$ is illustrated as

- Refers to all elements that are in $A$ or $B$ or in both $A$ and $B$
- Ex. One roll of a 6-sided dice
  - Given $\Omega = \{1, 2, 3, 4, 5, 6\}$
  - Set of rolling a 1 is denoted as $A = \{1\}$
A Gentle Introduction to Set Operations - Union

- Given two sets \( A \) and \( B \) that are subsets of the sample space \( \Omega \) the union of \( A \) and \( B \) denoted \( A \cup B \) is illustrated as

![Venn Diagram](image)

- Refers to all elements that are in \( A \) or \( B \) or in both \( A \) and \( B \)

- Ex. One roll of a 6-sided dice
  - Given \( \Omega = \{1, 2, 3, 4, 5, 6\} \)
  - Set of rolling a 1 is denoted as \( A = \{1\} \)
  - Set of not rolling a 1 is denoted as \( B = \{2, 3, 4, 5, 6\} \)
A Gentle Introduction to Set Operations - Union

- Given two sets $A$ and $B$ that are subsets of the sample space $\Omega$ the union of $A$ and $B$ denoted $A \cup B$ is illustrated as

- Refers to all elements that are in $A$ or $B$ or in both $A$ and $B$
- Ex. One roll of a 6-sided dice
  - Given $\Omega = \{1, 2, 3, 4, 5, 6\}$
  - Set of rolling a 1 is denoted as $A = \{1\}$
  - Set of not rolling a 1 is denoted as $B = \{2, 3, 4, 5, 6\}$
  - Then $A \cup B = \{1\} \cup \{2, 3, 4, 5, 6\} = \{1, 2, 3, 4, 5, 6\}$
A Gentle Introduction to Set Operations - Intersection

Given sets \( A \) and \( B \) that are subsets of the sample space \( \Omega \), the intersection of \( A \) and \( B \) denoted \( A \cap B \) is illustrated by the elements that are in both \( A \) and \( B \).

Ex. Two sets

\[
A = \{1, 2, 3\} \quad \text{and} \quad B = \{2, 3, 4\}
\]

The intersection of \( A \) and \( B \) is
\[
\{1, 2, 3\} \cap \{2, 3, 4\} = \{2, 3\}
\]

Ex. Two sets

\[
C = \{1, 2, 3\} \quad \text{and} \quad D = \{4, 5, 6\}
\]

The intersection of \( C \) and \( D \) is
\[
\{1, 2, 3\} \cap \{4, 5, 6\} = \emptyset
\]
Given sets $A$ and $B$ that are subsets of the sample space $\Omega$, the intersection of $A$ and $B$ denoted $A \cap B$ is illustrated by

Refers to the elements that are in both $A$ and $B$
A Gentle Introduction to Set Operations - Intersection

- Given sets $A$ and $B$ that are subsets of the sample space $\Omega$, the intersection of $A$ and $B$ denoted $A \cap B$ is illustrated by

  \[ \Omega \]

  \[ A \cap B \]

- Refers to the elements that are in both $A$ and $B$
- Ex. Two sets $A = \{1, 2, 3\}$ and $B = \{2, 3, 4\}$
Given sets $A$ and $B$ that are subsets of the sample space $\Omega$, the intersection of $A$ and $B$ denoted $A \cap B$ is illustrated by

- Refers to the elements that are in both $A$ and $B$
- Ex. Two sets $A = \{1, 2, 3\}$ and $B = \{2, 3, 4\}$
  - The intersection of $A$ and $B$ $\{1, 2, 3\} \cap \{2, 3, 4\} = \{2, 3\}$
Given sets $A$ and $B$ that are subsets of the sample space $\Omega$, the intersection of $A$ and $B$ denoted $A \cap B$ is illustrated by

- Refers to the elements that are in both $A$ and $B$
- Ex. Two sets $A = \{1, 2, 3\}$ and $B = \{2, 3, 4\}$
  - The intersection of $A$ and $B$ \( \{1, 2, 3\} \cap \{2, 3, 4\} = \{2, 3\} \)
- Ex. Two sets $C = \{1, 2, 3\}$ and $D = \{4, 5, 6\}$
A Gentle Introduction to Set Operations - Intersection

- Given sets $A$ and $B$ that are subsets of the sample space $\Omega$, the intersection of $A$ and $B$ denoted $A \cap B$ is illustrated by

- Refers to the elements that are in both $A$ and $B$
- Ex. Two sets $A = \{1, 2, 3\}$ and $B = \{2, 3, 4\}$
  - The intersection of $A$ and $B$ $\{1, 2, 3\} \cap \{2, 3, 4\} = \{2, 3\}$
- Ex. Two sets $C = \{1, 2, 3\}$ and $D = \{4, 5, 6\}$
  - The intersection of $C$ and $D = \{1, 2, 3\} \cap \{4, 5, 6\} = \emptyset$
A Gentle Introduction to Set Operations- Disjoint

Two sets are said to be disjoint if their intersection is the empty set. It is clear from the picture that the intersection of $A$ and $B$ is empty or $A \cap B = \emptyset$.

Ex. Given two sets $A = \{1, 2, 3, 4\}$ and $B = \{5, 6, 7, 8\}$, these two sets are disjoint because they don't share any common elements.
Two sets are said to be disjoint if their intersection is the empty set. It is clear from the picture that the intersection of $A$ and $B$ is empty or $A \cap B = \emptyset$. 
Two sets are said to be disjoint if their intersection is the empty set.

It is clear from the picture that the intersection of $A$ and $B$ is empty or $A \cap B = \emptyset$.

Ex. Given two sets $A = \{1, 2, 3, 4\}$ and $B = \{5, 6, 7, 8\}$
Two sets are said to be disjoint if their intersection is the empty set.

It is clear from the picture that the intersection of $A$ and $B$ is empty or $A \cap B = \emptyset$.

Ex. Given two sets $A = \{1, 2, 3, 4\}$ and $B = \{5, 6, 7, 8\}$

These two sets are disjoint because they don’t share any common elements.
Two sets are said to be disjoint if their intersection is the empty set.

It is clear from the picture that the intersection of \( A \) and \( B \) is empty or \( A \cap B = \emptyset \).

Ex. Given two sets \( A = \{1, 2, 3, 4\} \) and \( B = \{5, 6, 7, 8\} \)
- These two sets are disjoint because they don’t share any common elements.
- \( \{1, 2, 3, 4\} \cap \{5, 6, 7, 8\} = \emptyset \)
A Gentle Introduction to Set Operations - Partition

A collection of sets is said to be a partition of a sample space \( \Omega \) if the sets in the collection are disjoint and the union of the collection of sets equals the sample space.

For example, rolling one 6-sided dice \( \Omega = \{1, 2, 3, 4, 5, 6\} \)

The two sets \( A = \{2, 4, 6\} \) (even number) and \( B = \{1, 3, 5\} \) (odd number) form a partition of the sample space.

\[ \{2, 4, 6\} \cup \{1, 3, 5\} = \{1, 2, 3, 4, 5, 6\} = \Omega \]
A Gentle Introduction to Set Operations - Partition

- A collection of sets is said to be a partition of a sample space $\Omega$ if the sets in the collection are disjoint and the union of the collection of sets equals the sample space.

- Here $A_1, A_2, A_3, A_4$ for a partition of the sample space $\Omega$
A Gentle Introduction to Set Operations - Partition

- A collection of sets is said to be a partition of a sample space \( \Omega \) if the sets in the collection are disjoint and the union of the collection of sets equals the sample space.

Here \( A_1, A_2, A_3, A_4 \) for a partition of the sample space \( \Omega \)

Ex. Rolling one 6-sided dice \( \Omega = \{1, 2, 3, 4, 5, 6\} \)
A Gentle Introduction to Set Operations - Partition

- A collection of sets is said to be a partition of a sample space \( \Omega \) if the sets in the collection are disjoint and the union of the collection of sets equals the sample space.

Here \( A_1, A_2, A_3, A_4 \) for a partition of the sample space \( \Omega \)

Ex. Rolling one 6-sided dice \( \Omega = \{1, 2, 3, 4, 5, 6\} \)
  - The two sets \( A = \{2, 4, 6\} \) (even number) and \( B = \{1, 3, 5\} \) (odd number) form a partition of the sample space.
A Gentle Introduction to Set Operations - Partition

- A collection of sets is said to be a partition of a sample space \( \Omega \) if the sets in the collection are disjoint and the union of the collection of sets equals the sample space.

Here \( A_1, A_2, A_3, A_4 \) for a partition of the sample space \( \Omega \)
- Ex. Rolling one 6-sided dice \( \Omega = \{1, 2, 3, 4, 5, 6\} \)
  - The two sets \( A = \{2, 4, 6\} \) (even number) and \( B = \{1, 3, 5\} \) (odd number) form a partition of the sample space.
  - \( \{2, 4, 6\} \cup \{1, 3, 5\} = \{1, 2, 3, 4, 5, 6\} = \Omega \)
**A Gentle Introduction to Set Operations - Illustrative Summary**

- Picture of what we just learned\(^1\)

![Illustrative diagrams of set operations](image)

**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 \(\Omega\).

\(^1\)Image from Bertsekas and Tsitsiklis (2008, p. 5)
Learning Check - Set Operations

- What is the sample space of a single dice roll? What is a subset of that sample space? (There are many right answers)
- Give a union and intersection of two sets \( A = \{1, 2, 3\} \) and \( B = \{3, 4, 5\} \) which are subsets of the sample space \( \Omega = \{1, 2, 3, 4, 5, 6, 7, 8\} \)
- Explain in words or with a picture what a partition is...Use your own words.
- What are the two most common interpretations of probability?
Probability Models - Overview

Elements of a Probability Model:
- Experiment is an underlying process (e.g. three coin tosses, experimental manipulation in psychology, etc.)
- Sample space (Ω) aka set of all possible outcomes of the experiment
- Subset of the sample space is an event here
- Probability law \( P(\cdot) \) which assigns a number that must satisfy certain properties

\(^1\)Image from Bertsekas and Tsitsiklis (2008, p. 6)
Probability Models - Overview

- Elements of a Probability Model:

1 Image from Bertsekas and Tsitsiklis (2008, p. 6)
Probability Models - Overview

- **Elements of a Probability Model:**
  - Experiment is an underlying process (e.g. three coin tosses an experimental manipulation in psychology, etc.)

\[1\] Image from Bertsekas and Tsitsiklis (2008, p. 6)
Probability Models - Overview

- Elements of a Probability Model:
  - Experiment is an underlying process (e.g. three coin tosses an experimental manipulation in psychology, etc.)
  - Sample space ($\Omega$) aka set of all possible outcomes of the experiment

\(^1\)Image from Bertsekas and Tsitsiklis (2008, p. 6)
Probability Models - Overview

- **Elements of a Probability Model:**
  - Experiment is an underlying process (e.g. three coin tosses an experimental manipulation in psychology, etc.)
  - Sample space ($\Omega$) aka set of all possible outcomes of the experiment
  - Subset of the sample space is an event here $A$ and $B$
Probability Models - Overview

- Elements of a Probability Model:
  - Experiment is an underlying process (e.g. three coin tosses an experimental manipulation in psychology, etc.)
  - Sample space (\( \Omega \)) aka set of all possible outcomes of the experiment
  - Subset of the sample space is an event here \( A \) and \( B \)
  - Probability law \( P(\cdot) \) which assigns a number that must satisfy certain properties

\(^1\)Image from Bertsekas and Tsitsiklis (2008, p. 6)
Probability Models - Overview

- Elements of a Probability Model:
  - Experiment is an underlying process (e.g. three coin tosses an experimental manipulation in psychology, etc.)
  - Sample space ($\Omega$) aka set of all possible outcomes of the experiment
  - Subset of the sample space is an event here $A$ and $B$
  - Probability law $P(\cdot)$ which assigns a number that must satisfy certain properties

---

1 Image from Bertsekas and Tsitsiklis (2008, p. 6)
Given an event $A$, which is a subset of sample space $\Omega$, a number $P(A)$, denoted the probability of $A$, must satisfy

1. **Non-negativity**
   
   $P(A) \geq 0$ for every event $A$ that is a member of $\Omega$

2. **Normalization**
   
   $P(\Omega) = 1$, the probability of the sample space equals 1

3. **Additivity**
   
   If events $A$ and $B$ are disjoint (e.g. $A \cap B = \emptyset$) then the probability of their union is the sum of the individual events:
   
   $P(A \cup B) = P(A) + P(B)$
   
   This generalizes to more than two events such that if $A_1, A_2, \ldots, A_n$ is a sequence of disjoint events then
   
   $P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n)$
Given an event $A$, which is a subset of sample space $\Omega$, a number $P(A)$, denoted the probability of $A$, must satisfy

1. **Non-negativity**
   
   $P(A) \geq 0$ for every event $A$ that is a member of $\Omega$.

2. **Normalization**
   
   $P(\Omega) = 1$, the probability of the sample space equals 1.

3. **Additivity**
   
   If events $A$ and $B$ are disjoint (e.g. $A \cap B = \emptyset$) then the probability of their union is the sum of the individual events:
   
   $$P(A \cup B) = P(A) + P(B).$$
   
   This generalizes to more than two events such that if $A_1, \ldots, A_n$ is a sequence of disjoint events then
   
   $$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n).$$
Probability Models - Probability Laws/Axioms

- Given an event \( A \), which is a subset of sample space \( \Omega \), a number \( P(A) \), denoted the probability of \( A \), must satisfy

1. (Non-negativity) \( P(A) \geq 0 \) for every event \( A \) that is a member of \( \Omega \).
Probability Models - Probability Laws/Axioms

- Given an event $A$, which is a subset of sample space $\Omega$, a number $P(A)$, denoted the probability of $A$, must satisfy

1. (Non-negativity) $P(A) \geq 0$ for every event $A$ that is a member of $\Omega$

2. (Normalization) $P(\Omega) = 1$, the probability of the sample space equals 1
Given an event $A$, which is a subset of sample space $\Omega$, a number $P(A)$, denoted the probability of $A$, must satisfy

1. (Non-negativity) $P(A) \geq 0$ for every event $A$ that is a member of $\Omega$
2. (Normalization) $P(\Omega) = 1$, the probability of the sample space equals 1
3. (Additivity) If events $A$ and $B$ are disjoint (e.g. $A \cap B = \emptyset$) then the probability of their union is the sum of the individual events: $P(A) \cup P(B) = P(A) + P(B)$. 
Given an event $A$, which is a subset of sample space $\Omega$, a number $P(A)$, denoted the probability of $A$, must satisfy

1. (Non-negativity) $P(A) \geq 0$ for every event $A$ that is a member of $\Omega$
2. (Normalization) $P(\Omega) = 1$, the probability of the sample space equals 1
3. (Additivity) If events $A$ and $B$ are disjoint (e.g. $A \cap B = \emptyset$) then the probability of their union is the sum of the individual events: $P(A) \cup P(B) = P(A) + P(B)$.
   - This generalizes to more than two events such that if $A_1, \ldots, A_n$ is a sequence of disjoint events then $P(A_1) \cup P(A_2) \cup \cdots \cup P(A_n) = P(A_1) + P(A_2) + \cdots + P(A_n)$
Probability Models - Example 1

From the previous slide the 3 Probability Laws:

1. \( P(A) \geq 0 \) for all events \( A \)
2. \( P(\Omega) = 1 \)
3. If \( A \cap B = \emptyset \) then \( P(A \cup B) = P(A) + P(B) \)

Let's take a real world example: A single toss of a coin with outcomes heads (\( H \)) and (\( T \)) tails. The sample space \( \Omega = \{ H, T \} \).

Events \( \{ H, T \} \) (\( \Omega \)); \( \{ H \} \) (heads); \( \{ T \} \) (tails); \( \{ \emptyset \} \) (nothing).

Assume coin is fair which means \( P(H) = P(T) = 0.5 \).

Events \( H \) and \( T \) are disjoint so then \( P(H \cup T) = P(H) + P(T) = 1 = \Omega \).

\( P(\emptyset) = 0 \) (Note this is a provable statement but I won't do it here).
Probability Models - Example 1

- From the previous slide the 3 Probability Laws:
  1. \( P(A) \geq 0 \) for all events \( A \)
  2. \( P(\Omega) = 1 \)
  3. If \( A \cap B = \emptyset \) then \( P(A \cup B) = P(A) + P(B) \)
Probability Models - Example 1

- From the previous slide the 3 Probability Laws:
  1. $P(A) \geq 0$ for all events $A$
  2. $P(\Omega) = 1$
  3. If $A \cap B = \emptyset$ then $P(A \cup B) = P(A) + P(B)$

- Let’s take a real world example: A single toss of a coin with outcomes heads ($H$) and ($T$) tails. The sample space $\Omega = \{H, T\}$. 
Probability Models - Example 1

- From the previous slide the 3 Probability Laws:
  1. $P(A) \geq 0$ for all events $A$
  2. $P(\Omega) = 1$
  3. If $A \cap B = \emptyset$ then $P(A \cup B) = P(A) + P(B)$

- Let’s take a real world example: A single toss of a coin with outcomes heads $(H)$ and $(T)$ tails. The sample space $\Omega = \{H, T\}$.
  - Events $\{H, T\}$ (\Omega); $\{H\}$ (heads); $\{T\}$ (tails); $\{\emptyset\}$ (nothing)
From the previous slide the 3 Probability Laws:

1. \( P(A) \geq 0 \) for all events \( A \)
2. \( P(\Omega) = 1 \)
3. If \( A \cap B = \emptyset \) then \( P(A \cup B) = P(A) + P(B) \)

Let’s take a real world example: A single toss of a coin with outcomes heads (\( H \)) and (\( T \)) tails. The sample space \( \Omega = \{H, T\} \).

- Events \( \{H, T\} \) (\( \Omega \)); \( \{H\} \) (\textit{heads}); \( \{T\} \) (\textit{tails}); \( \{\emptyset\} \) (\textit{nothing})
- Assume coin is fair which means \( P(H) = P(T) = 0.5 \)
Probability Models - Example 1

- From the previous slide the 3 Probability Laws:
  1. $P(A) \geq 0$ for all events $A$
  2. $P(\Omega) = 1$
  3. If $A \cap B = \emptyset$ then $P(A \cup B) = P(A) + P(B)$

- Let’s take a real world example: A single toss of a coin with outcomes heads ($H$) and ($T$) tails. The sample space $\Omega = \{H, T\}$.
  - Events $\{H, T\}$ ($\Omega$); $\{H\}$ (heads); $\{T\}$ (tails); $\{\emptyset\}$ (nothing)
  - Assume coin is fair which means $P(H) = P(T) = 0.5$

- Events $H$ and $T$ are disjoint so then $P(H \cup T) = P(H) + P(T) = 1 = \Omega$
From the previous slide the 3 Probability Laws:

1. $P(A) \geq 0$ for all events $A$
2. $P(\Omega) = 1$
3. If $A \cap B = \emptyset$ then $P(A \cup B) = P(A) + P(B)$

Let’s take a real world example: A single toss of a coin with outcomes heads ($H$) and (T) tails. The sample space $\Omega = \{H, T\}$.

- Events $\{H, T\}$ ($\Omega$); $\{H\}$ (heads); $\{T\}$ (tails); $\{\emptyset\}$ (nothing)
- Assume coin is fair which means $P(H) = P(T) = 0.5$

Events $H$ and $T$ are disjoint so then

$P(H \cup T) = P(H) + P(T) = 1 = \Omega$

$P(\emptyset) = 0$ (Note this is a provable statement but I won’t do it here)
Consider an experiment with three coin tosses. The sample space \( \Omega = \{ \text{HHH}, \text{HHT}, \text{HTH}, \text{THH}, \text{HTT}, \text{THT}, \text{TTH}, \text{TTT} \} \). Assume each event has probability of \( \frac{1}{8} \). What is the probability that exactly 2 tails occur?

Exactly 2 tails occur = \{ HTT, THT, TTH \}

These three events are disjoint so using the additivity law:

\[
P(\{ HTT, THT, TTH \}) = P(\{ TTH \}) + P(\{ THT \}) + P(\{ HTT \}) = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8}\]
Consider an experiment with three coin tosses. The sample space
\[ \Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\} \]
Probability Models - Example 2

- Consider an experiment with three coin tosses. The sample space
  \( \Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\} \)
- Assume each event has probability of \(1/8\)
Consider an experiment with three coin tosses. The sample space
\[ \Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\} \]

Assume each event has probability of 1/8

What is the probability that exactly 2 tails occur?
Consider an experiment with three coin tosses. The sample space
\[ \Omega = \{ HHH, HHT, HTH, THH, HTT, THT, TTH, TTT \} \]
Assume each event has probability of 1/8
What is the probability that exactly 2 tails occur?
- \{ exactly 2 tails occur \} = \{ HTT, THT, TTH \}
Probability Models - Example 2

- Consider an experiment with three coin tosses. The sample space
  \[ \Omega = \{ HHH, HHT, HTH, THH, HTT, THT, TTH, TTT \} \]
- Assume each event has probability of 1/8
- What is the probability that exactly 2 tails occur?
  - \{exactly 2 tails occur\} = \{HTT, THT, TTH\}
  - These three events are disjoint so using the additivity law
Probability Models - Example 2

- Consider an experiment with three coin tosses. The sample space
  \[ \Omega = \{ HHH, HHT, HTH, THH, HTT, THT, TTH, TTT \} \]
- Assume each event has probability of 1/8
- What is the probability that exactly 2 tails occur?
  - \{exactly 2 tails occur\} = \{HTT, THT, TTH\}
  - These three events are disjoint so using the additivity law
  - \[ P(\{HTT, THT, TTH\}) = P(\{TTH\}) + P(\{THT\}) + P(\{HTT\}) = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8} \]
Learning Check - Probability Models

1. What 4 things are part of a probability model?

2. What are the 3 probability laws? Write out the mathematical symbols and meanings.

3. Given an experiment with three coin tosses, with sample space \( \Omega = \{ HHH, HHT, HTH, THH, HTT, THT, TTH, TTT \} \), and probability of each event being 1/8; what is the probability of rolling three heads or rolling three tails?

4. Give an example of 3 sets that form a partition of a single dice roll.
Conditional Probability - Overview

Conditional probability allows us to reason probabilistically when we have partial information about our problem. Some examples of conditional probability questions are:

- What is the probability I will get an A in this class given I have a B in the class at this point?
- What is the probability I have cancer given that I tested positive?
- What is the probability a student matriculates given their scores on measures of achievement?

More formally, given an experiment, sample space, and probability law, then if we know some event $B$ has occurred, we can use this in reasoning about the probability of event $A$ occurring.
Conditional Probability - Overview

- Conditional probability allows us to reason probabilistically when we have partial information about our problem.
Conditional Probability - Overview

- Conditional probability allows us to reason probabilistically when we have partial information about our problem.
- Some examples of conditional probability questions are:
Conditional probability allows us to reason probabilistically when we have partial information about our problem.

Some examples of conditional probability questions are:

- What is the probability I will get an A in this class given I have a B in the class at this point?
- What is the probability I have cancer given that I tested positive?
- What is the probability a student matriculates given their scores on measures of achievement?
Conditional Probability - Overview

- Conditional probability allows us to reason probabilistically when we have partial information about our problem.
- Some examples of conditional probability questions are:
  - What is the probability I will get an A in this class given I have a B in the class at this point?
  - What is the probability I have cancer given that I tested positive?
Conditional Probability - Overview

- Conditional probability allows us to reason probabilistically when we have partial information about our problem.
- Some examples of conditional probability questions are:
  - What is the probability I will get an A in this class given I have a B in the class at this point?
  - What is the probability I have cancer given that I tested positive?
  - What is the probability a student matriculates given their scores on measures of achievement?
Conditional Probability - Overview

- Conditional probability allows us to reason probabilistically when we have partial information about our problem.
- Some examples of conditional probability questions are:
  - What is the probability I will get an A in this class given I have a B in the class at this point?
  - What is the probability I have cancer given that I tested positive?
  - What is the probability a student matriculates given their scores on measures of achievement?
- More formally given an experiment, sample space, and probability law then if we know some event $B$ has occurred then we can use this in reasoning about the probability of event $A$ occurring
Conditional Probability - Overview

The new probability law specifies that the probability of event $A$ conditional on event $B$ is given by:

$$P(A | B) = \frac{P(A \cap B)}{P(B)}$$

We assume that $P(B) > 0$ otherwise this is undefined.

Let's take an example of a single roll of a 6-sided fair dice. What is the probability that we roll a 6 given that we know $X$ is even?

$$P(X = 6 | X \text{ is even}) = \frac{P(X = 6 \cap X \text{ is even})}{P(X \text{ is even})} = \frac{1}{6} \div \frac{1}{2} = \frac{1}{3}$$
Conditional Probability - Overview

- The new probability law specifies that the probability of event $A$ conditional on event $B$ is given by:

$$P(A | B) = \frac{P(A \cap B)}{P(B)}$$

We assume that $P(B) > 0$ otherwise this is undefined.

Let's take an example of a single roll of a 6-sided fair dice. What is the probability that we roll a 6 given that we know $X$ is even?

$$P(X = 6 | X \text{ is even}) = \frac{P(X = 6 \cap X \text{ is even})}{P(X \text{ is even})} = \frac{1}{6} \cdot \frac{1}{2} = \frac{1}{12}$$
Conditional Probability - Overview

- The new probability law specifies that the probability of event $A$ conditional on event $B$
- Mathematically this is written as:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- We assume that $P(B) > 0$ otherwise this is undefined.
The new probability law specifies that the probability of event $A$ conditional on event $B$

Mathematically this is written as:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

We assume that $P(B) > 0$ otherwise this is undefined.

Let’s take an example of a single roll of a 6-sided fair dice.
Conditional Probability - Overview

- The new probability law specifies that the probability of event $A$ conditional on event $B$
- Mathematically this is written as:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- We assume that $P(B) > 0$ otherwise this is undefined
- Let’s take an example of a single roll of a 6-sided fair dice.
- What is the probability that we roll a 6 given that we know $X$ is even?

$$P(X=6|X \text{ even}) = \frac{P(X=6 \cap X \text{ even})}{P(X \text{ even})} = \frac{1}{6} \times \frac{1}{2} = \frac{1}{12}$$
Conditional Probability - Overview

- The new probability law specifies that the probability of event $A$ conditional on event $B$
- Mathematically this is written as:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- We assume that $P(B) > 0$ otherwise this is undefined
- Let’s take an example of a single roll of a 6-sided fair dice.
- What is the probability that we roll a 6 given that we know $X$ is even?
  - $P(X = 6 | X \text{ is even}) = \frac{P(X=6 \cap X \text{ is even})}{P(X \text{ is even})}$
Conditional Probability - Overview

- The new probability law specifies that the probability of event $A$ conditional on event $B$
- Mathematically this is written as:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- We assume that $P(B) > 0$ otherwise this is undefined
- Let’s take an example of a single roll of a 6-sided fair dice.
- What is the probability that we roll a 6 given that we know $X$ is even?
  - $P(X = 6|X \text{ is even}) = \frac{P(X=6 \cap X \text{ is even})}{P(X \text{ is even})}$
  - $P(X = 6|X \text{ is even}) = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{2}{6} = \frac{1}{3}$
Conditional Probability - Example 1

Assume that we toss a fair coin 3 times and define the events $A = \{\text{more heads than tails occur}\}$ and $B = \{\text{1st toss is a head}\}$. We wish to find $P(A|B)$ when

The sample space $\Omega = \{\text{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT}\}$

What is $P(B)$? Hint... Count the number of times more heads than tails occur and divide by size of $\Omega$

$P(B) = \frac{1}{2}$

What is $P(A \cap B)$?

$P(A \cap B) = \frac{3}{8}$

$A \cap B = \{\text{HHH, HHT, HTH, THH}\} \cap \{\text{HHH, HHT, HTH, THH}\} = \{\text{HHH, HHT, HTH, THH}\} / 8$

What is $P(A|B)$? Hint... Plug and Chug

$P(A|B) = \frac{3}{8} \cdot \frac{1}{2} = \frac{3}{8} \cdot \frac{1}{2} = \frac{3}{4}$
Assume that we toss a fair coin 3 times and define the events
\( A = \{ \text{more heads than tails occur} \} \) and
\( B = \{ \text{1st toss is a head} \} \). We wish to find \( P(A|B) \) when
Conditional Probability - Example 1

- Assume that we toss a fair coin 3 times and define the events
  \(A = \{\text{more heads than tails occur}\}\) and
  \(B = \{1\text{st toss is a head}\}\). We wish to find \(P(A|B)\) when
- The sample space
  \(\Omega = \{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT\}\)
Conditional Probability - Example 1

- Assume that we toss a fair coin 3 times and define the events
  \(A = \{\text{more heads than tails occur}\}\) and
  \(B = \{\text{1st toss is a head}\}\). We wish to find \(P(A|B)\) when
- The sample space
  \(\Omega = \{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT\}\)
- What is \(P(B)\) Hint...Count the number of times more heads than tails occur and divide by size of \(\Omega\)
Conditional Probability - Example 1

Assume that we toss a fair coin 3 times and define the events

\[ A = \{ \text{more heads than tails occur} \} \] and

\[ B = \{ \text{1st toss is a head} \}. \]

We wish to find \( P(A|B) \) when

The sample space

\[ \Omega = \{ HHH, HHT, HTH, THH, TTH, THT, HTT, TTT \} \]

What is \( P(B) \)? Hint...Count the number of times more heads than tails occur and divide by size of \( \Omega \)

\[ P(B) = \frac{1}{2} \]
Conditional Probability - Example 1

- Assume that we toss a fair coin 3 times and define the events
  \( A = \{ \text{more heads than tails occur} \} \) and
  \( B = \{ 1\text{st toss is a head} \} \). We wish to find \( P(A|B) \) when
- The sample space
  \( \Omega = \{ HHH, HHT, HTH, THH, TTH, THT, HTT, TTT \} \)
- What is \( P(B) \) Hint...Count the number of times more heads than tails occur and divide by size of \( \Omega \)
  - \( P(B) = \frac{1}{2} \)
- What is \( P(A \cap B) \)?
Conditional Probability - Example 1

- Assume that we toss a fair coin 3 times and define the events
  \( A = \{ \text{more heads than tails occur} \} \) and
  \( B = \{ \text{1st toss is a head} \} \). We wish to find \( P(A|B) \) when
- The sample space
  \( \Omega = \{ \text{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT} \} \)
- What is \( P(B) \)? Hint...Count the number of times more heads than tails occur and divide by size of \( \Omega \)
  - \( P(B) = \frac{1}{2} \)
- What is \( P(A \cap B) \)?
  - \( P(A \cap B) = \frac{3}{8} \)
Conditional Probability - Example 1

- Assume that we toss a fair coin 3 times and define the events 
  \( A = \{ \text{more heads than tails occur} \} \) and 
  \( B = \{ \text{1st toss is a head} \} \). We wish to find \( P(A|B) \) when 
- The sample space 
  \( \Omega = \{ HHH, HHT, HTH, THH, TTH, THT, HTT, TTT \} \) 
- What is \( P(B) \)? Hint...Count the number of times more heads than tails occur and divide by size of \( \Omega \) 
  - \( P(B) = \frac{1}{2} \) 
- What is \( P(A \cap B) \)? 
  - \( P(A \cap B) = \frac{3}{8} \) 
  - \( A \cap B = \{ HHH, HHT, HTH, THH \} \cap \{ HHH, HHT, HTH \} = \{ HHH, HHT, HTH \} / 8 \)
Assume that we toss a fair coin 3 times and define the events
\( A = \{ \text{more heads than tails occur} \} \) and
\( B = \{ \text{1st toss is a head} \} \). We wish to find \( P(A|B) \) when

The sample space
\( \Omega = \{ HHH, HHT, HTH, THH, TTH, THT, HTT, TTT \} \)

What is \( P(B) \)?
Hint...Count the number of times more heads than tails occur and divide by size of \( \Omega \)

\[ P(B) = \frac{1}{2} \]

What is \( P(A \cap B) \)?

\[ P(A \cap B) = \frac{3}{8} \]

\( A \cap B = \{ HHH, HHT, HTH, THH \} \cap \{ HHH, HHT, HTH \} = \{ HHH, HHT, HTH \}/8 \)

What is \( P(A|B) \)?
Hint...Plug and Chug
Conditional Probability - Example 1

- Assume that we toss a fair coin 3 times and define the events
  \(A = \{\text{more heads than tails occur}\}\) and
  \(B = \{\text{1st toss is a head}\}\). We wish to find \(P(A|B)\) when
- The sample space
  \(\Omega = \{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT\}\)
- What is \(P(B)\)? Hint... Count the number of times more heads than tails occur and divide by size of \(\Omega\)
  \(P(B) = \frac{1}{2}\)
- What is \(P(A \cap B)\)?
  \(P(A \cap B) = \frac{3}{8}\)
  \(A \cap B = \{HHH, HHT, HTH, THH\} \cap \{HHH, HHT, HTH\} = \{HHH, HHT, HTH\} / 8\)
- What is \(P(A|B)\)? Hint... Plug and Chug
  \(P(A|B) = \frac{\frac{3}{8}}{\frac{1}{2}} = \frac{3}{8} \times 2 = \frac{6}{8} = \frac{3}{4}\)
Let's take another example with a contingency table of Jared's Demerits for the 2012-2013 Academic School Year:

<table>
<thead>
<tr>
<th></th>
<th>Minor (M)</th>
<th>Severe (S)</th>
<th>Row Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carol (C)</td>
<td>5</td>
<td>11</td>
<td>16</td>
</tr>
<tr>
<td>Billy (B)</td>
<td>5</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>Column Totals</td>
<td>10</td>
<td>15</td>
<td>25</td>
</tr>
</tbody>
</table>

What is the probability Carol gave the demerit given that the demerit was severe ($P(C|S)$)?

$$P(C|S) = \frac{P(C \cap S)}{P(S)} = \frac{11}{25}$$

$P(C \cap S)$ is the number where row is $C$ and column is $S$ divided by total $P(S)$ is the column total of $S$ divided by total.
Let’s take another example with a contingency table of Jared’s Demerits for the 2012-2013 Academic School Year.

<table>
<thead>
<tr>
<th></th>
<th>Minor (M)</th>
<th>Severe (S)</th>
<th>Row Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carol (C)</td>
<td>5</td>
<td>11</td>
<td>16</td>
</tr>
<tr>
<td>Billy (B)</td>
<td>5</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>Column Totals</td>
<td>10</td>
<td>15</td>
<td>25</td>
</tr>
</tbody>
</table>

What is the probability Carol gave the demerit given that the demerit was severe ($P(C|S)$)?

$$P(C|S) = \frac{11}{25}$$

$P(C \cap S)$ is number where row is C and column is S divided by total.

$P(S)$ is the column total of S divided by total.
Conditional Probability - Example 2

Let’s take another example with a contingency table of Jared’s Demerits for the 2012-2013 Academic School Year.

<table>
<thead>
<tr>
<th></th>
<th>Minor (M)</th>
<th>Severe (S)</th>
<th>Row Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carol (C)</td>
<td>5</td>
<td>11</td>
<td>16</td>
</tr>
<tr>
<td>Billy (B)</td>
<td>5</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>Column Totals</td>
<td>10</td>
<td>15</td>
<td>25</td>
</tr>
</tbody>
</table>

What is the probability Carol gave the demerit given that the demerit was severe \( P(C|S) \)?
Conditional Probability - Example 2

Let’s take another example with a contingency table of Jared’s Demerits for the 2012-2013 Academic School Year

<table>
<thead>
<tr>
<th></th>
<th>Minor (M)</th>
<th>Severe (S)</th>
<th>Row Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carol (C)</td>
<td>5</td>
<td>11</td>
<td>16</td>
</tr>
<tr>
<td>Billy (B)</td>
<td>5</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>Column Totals</td>
<td>10</td>
<td>15</td>
<td>25</td>
</tr>
</tbody>
</table>

What is the probability Carol gave the demerit given that the demerit was severe ($P(C|S)$)?

$$P(C|S) = \frac{11}{25} = \frac{11}{15}$$
Let’s take another example with a contingency table of Jared’s Demerits for the 2012-2013 Academic School Year:

<table>
<thead>
<tr>
<th></th>
<th>Minor (M)</th>
<th>Severe (S)</th>
<th>Row Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carol (C)</td>
<td>5</td>
<td>11</td>
<td>16</td>
</tr>
<tr>
<td>Billy (B)</td>
<td>5</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>Column Totals</td>
<td>10</td>
<td>15</td>
<td>25</td>
</tr>
</tbody>
</table>

What is the probability Carol gave the demerit given that the demerit was severe ($P(C|S)$)?

$$P(C|S) = \frac{11}{25} \div \frac{15}{25} = \frac{11}{15}$$

$P(C \cap S)$ is number where row is C and column is S divided by total.
Conditional Probability - Example 2

Let’s take another example with a contingency table of Jared’s Demerits for the 2012-2013 Academic School Year.

<table>
<thead>
<tr>
<th></th>
<th>Minor (M)</th>
<th>Severe (S)</th>
<th>Row Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carol (C)</td>
<td>5</td>
<td>11</td>
<td>16</td>
</tr>
<tr>
<td>Billy (B)</td>
<td>5</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>Column Totals</td>
<td>10</td>
<td>15</td>
<td>25</td>
</tr>
</tbody>
</table>

What is the probability Carol gave the demerit given that the demerit was severe ($P(C|S)$)?

- $P(C|S) = \frac{11}{25} = \frac{11}{15}$
- $P(C \cap S)$ is number where row is C and column is S divided by total
- $P(S)$ is the column total of S divided by total
Conditional Probability - Learning Check

Given the following contingency table of Jared’s Demerits during the 2013-2014 School Year

<table>
<thead>
<tr>
<th></th>
<th>Minor (M)</th>
<th>Severe (S)</th>
<th>Row Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carol (C)</td>
<td>8</td>
<td>6</td>
<td>14</td>
</tr>
<tr>
<td>Billy (B)</td>
<td>5</td>
<td>12</td>
<td>17</td>
</tr>
<tr>
<td>Column Totals</td>
<td>?</td>
<td>18</td>
<td><strong>31</strong></td>
</tr>
</tbody>
</table>

- What is $P(M)$? Hint you will need to find the column total of $M$ to get this.
- What is $P(B|M)$?
- What is $P(B|S)$?
The multiplication rule (chain rule or product rule) allows us to compute the probability of an intersection of events. From the definition of conditional probability:

\[ P(A | B) = \frac{P(A \cap B)}{P(B)} \]

Then

\[ P(A \cap B) = P(A | B) P(B) \]

(This is the product rule for events \( A \) and \( B \)).

But wait there is more!!! More generally for events \( A_1, \ldots, A_n \):

\[ P(\bigcap_{i=1}^{n} A_i) = P(A_1) P(A_2 | A_1) P(A_3 | A_1 \cap A_2) \cdots P(A_n | \bigcap_{i=1}^{n-1} A_i) \]

Let's take an example so you understand what is happening.
The multiplication rule (chain rule or product rule) allows us to compute the probability of an intersection of events.
Multiplication Rule - Overview

- The multiplication rule (chain rule or product rule) allows us to compute the probability of an intersection of events.
- From Def. of conditional probability:
The multiplication rule (chain rule or product rule) allows us to compute the probability of an intersection of events. From the definition of conditional probability:

\[ P(A|B) = \frac{P(A \cap B)}{P(B)} \]
The multiplication rule (chain rule or product rule) allows us to compute the probability of an intersection of events.

From Def. of conditional probability:

- \( P(A|B) = \frac{P(A \cap B)}{P(B)} \)
- Then \( P(A \cap B) = P(A|B)P(B) \) (This is the product rule for events A and B)
Multiplication Rule - Overview

- The multiplication rule (chain rule or product rule) allows us to compute the probability of an intersection of events
- From Def. of conditional probability:
  - \( P(A|B) = \frac{P(A \cap B)}{P(B)} \)
  - Then \( P(A \cap B) = P(A|B)P(B) \) (This is the product rule for events \( A \) and \( B \))
- But wait there is more!!! More generally for events \( A_1, \ldots, A_n \)
  \[
P(\cap_{i=1}^{n} A_i) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|\cap_{i=1}^{n-1} A_i)
\]
The multiplication rule (chain rule or product rule) allows us to compute the probability of an intersection of events. From the definition of conditional probability:

\[ P(A|B) = \frac{P(A \cap B)}{P(B)} \]

Then \( P(A \cap B) = P(A|B)P(B) \) (This is the product rule for events \( A \) and \( B \)).

But wait there is more!!! More generally for events \( A_1, \ldots, A_n \):

\[ P(\bigcap_{i=1}^{n} A_i) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|\bigcap_{i=1}^{n-1} A_i) \]

Let’s take an example so you understand what is happening.
Why Should you care?

- The multiplication rule is what allows for logistic regressions, SEMs, and other modeling tools to be used super fast.
- I should say the multiplication rule provides a framework from which we can build software to solve problems.
Multiplication Rule - Ex.

Assume we want to know the probability of the intersection of $A_1, A_2, A_3, A_4$ (here $n = 4$).

$$P(A_1 \cap A_2 \cap A_3 \cap A_4) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)P(A_4|A_1 \cap A_2 \cap A_3)$$

Toss a fair coin 3 times. Let $A = A_1 \cap A_2 \cap A_3$.

$$P(A) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)$$

$$P(A) = P(A_1)P(A_2)P(A_3) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$$

Note this example assumes independence of the tosses which we will talk about shortly...stay tuned.
Multiplication Rule - Ex.

- From previous slide, given events $A_1, \ldots, A_n$

\[ P(\cap_{i=1}^{n} A_i) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|\cap_{i=1}^{n-1} A_i) \]

Toss a fair coin 3 times. Let $A = A_1 \cap A_2 \cap A_3$

\[ P(A) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) = \left(\frac{1}{2}\right) \left(\frac{1}{2}\right) \left(\frac{1}{2}\right) = \frac{1}{8} \]
Multiplication Rule - Ex.

- From previous slide.. given events $A_1, \ldots, A_n$

\[ P(\bigcap_{i=1}^{n} A_i) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|\bigcap_{i=1}^{n-1} A_i) \]

- Assume we want to know the probability of the intersection of $A_1, A_2, A_3, A_4$ (here $n = 4$)

Note this example assumes independence of the tosses which we will talk about shortly... stay tuned.
Multiplication Rule - Ex.

- From previous slide..given events $A_1, \ldots, A_n$

\[ P(\cap_{i=1}^{n} A_i) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|\cap_{i=1}^{n-1} A_i) \]

- Assume we want to know the probability of the intersection of $A_1, A_2, A_3, A_4$ (here $n = 4$)
  - Want this $P(A_1 \cap A_2 \cap A_3 \cap A_4) = P(\cap_{i=1}^{4} A_i)$
Multiplication Rule - Ex.

• From previous slide..given events $A_1, \ldots, A_n$

\[ P(\cap_{i=1}^{n} A_i) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|\cap_{i=1}^{n-1} A_i) \]

• Assume we want to know the probability of the intersection of $A_1, A_2, A_3, A_4$ (here $n = 4$)
  • Want this $P(A_1 \cap A_2 \cap A_3 \cap A_4) = P(\cap_{i=1}^{4} A_i)$
  • $P(\cap_{i=1}^{4} A_i) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)P(A_4|A_1 \cap A_2 \cap A_3)$

Toss a fair coin 3 times. Let $A = A_1 \cap A_2 \cap A_3$

\[ P(A) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \]

\[ = P(A_1)P(A_2)P(A_3|A_1 \cap A_2) \]

\[ = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8} \]
Multiplication Rule - Ex.

- From previous slide, given events $A_1, \ldots, A_n$

  \[ P(\cap_{i=1}^{n} A_i) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|\cap_{i=1}^{n-1} A_i) \]

- Assume we want to know the probability of the intersection of $A_1, A_2, A_3, A_4$ (here $n = 4$)
  - Want this $P(A_1 \cap A_2 \cap A_3 \cap A_4) = P(\cap_{i=1}^{4} A_i)$
  - $P(\cap_{i=1}^{4} A_i) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)P(A_4|A_1 \cap A_2 \cap A_3)$

- Toss a fair coin 3 times. Let $A = A_1 \cap A_2 \cap A_3$
Multiplication Rule - Ex.

- From previous slide..given events $A_1, \ldots, A_n$

$$P(\bigcap_{i=1}^{n} A_i) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|\bigcap_{i=1}^{n-1} A_i)$$

- Assume we want to know the probability of the intersection of $A_1, A_2, A_3, A_4$ (here $n = 4$)
  - Want this $P(A_1 \cap A_2 \cap A_3 \cap A_4) = P(\bigcap_{i=1}^{4} A_i)$
  - $P(\bigcap_{i=1}^{4} A_i) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)P(A_4|A_1 \cap A_2 \cap A_3)$

- Toss a fair coin 3 times. Let $A = A_1 \cap A_2 \cap A_3$
  - $P(A) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)$
**Multiplication Rule - Ex.**

- From previous slide, given events $A_1, \ldots, A_n$

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

- Assume we want to know the probability of the intersection of $A_1, A_2, A_3, A_4$ (here $n=4$)
  - Want this $P(A_1 \cap A_2 \cap A_3 \cap A_4) = P(\cap_{i=1}^4 A_i)$
  - $P(\cap_{i=1}^4 A_i) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)P(A_4|A_1 \cap A_2 \cap A_3)$

- Toss a fair coin 3 times. Let $A = A_1 \cap A_2 \cap A_3$
  - $P(A) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)$
  - $P(A) = P(A_1)P(A_2)P(A_3) = 1/2 \times 1/2 \times 1/2 = 1/8$
Multiplication Rule - Ex.

- From previous slide..given events $A_1, \ldots, A_n$

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

- Assume we want to know the probability of the intersection of $A_1, A_2, A_3, A_4$ (here $n = 4$)
  - Want this $P(A_1 \cap A_2 \cap A_3 \cap A_4) = P(\cap_{i=1}^{4} A_i)$
  - $P(\cap_{i=1}^{4} A_i) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)P(A_4|A_1 \cap A_2 \cap A_3)$

- Toss a fair coin 3 times. Let $A = A_1 \cap A_2 \cap A_3$

$$P(A) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)$$

$$P(A) = P(A_1)P(A_2)P(A_3) = 1/2 * 1/2 * 1/2 = 1/8$$

- Note this example assumes independence of the tosses which we will talk about shortly...stay tuned.
Total Probability Theorem - Overview
Total Probability Theorem - Overview

- If we are given events $A_1, A_2, A_3$ which partition the sample space then the probability of $B$ [$P(B)$] can be visualized as follows:

![Venn diagram showing the partition of the sample space by $A_1$, $A_2$, and $A_3$, and the event $B$ as a union of intersections with these events.](image)
If we are given events $A_1, A_2, A_3$ which partition the sample space then the probability of $B$ [$P(B)$] can be visualized as follows:

The events $A_1, A_2, A_3$ form a partition
If we are given events $A_1, A_2, A_3$ which partition the sample space then the probability of $B$ [$P(B)$] can be visualized as follows:

- The events $A_1, A_2, A_3$ form a partition
- The event $B$ is a disjoint union of intersections with the events $A_1, A_2, A_3$
If we are given events $A_1, A_2, A_3$ which partition the sample space then the probability of $B$ [$P(B)$] can be visualized as follows:

- The events $A_1, A_2, A_3$ form a partition
- The event $B$ is a disjoint union of intersections with the events $A_1, A_2, A_3$
- $B = (B \cap A_1) \cup (B \cap A_2) \cup (B \cap A_3)$
Total Probability Theorem - Overview
Total Probability Theorem - Overview

- The probability of $B$ [$P(B)$] can be obtained by the additivity axiom since events $A_1$, $A_2$, and $A_3$ form a partition.
Total Probability Theorem - Overview

- The probability of B \([P(B)]\) can be obtained by the additivity axiom since events \(A_1, A_2, \text{ and } A_3\) form a partition
- \(P(B) = P(B \cap A_1) + P(B \cap A_2) + P(B \cap A_3)\)
The probability of B \( P(B) \) can be obtained by the additivity axiom since events \( A_1, A_2, \) and \( A_3 \) form a partition:

\[
P(B) = P(B \cap A_1) + P(B \cap A_2) + P(B \cap A_3)
\]

- \( P(B|A) = \frac{P(A \cap B)}{P(A)} \implies P(B|A)P(A) = P(B \cap A) \)
The probability of B \( P(B) \) can be obtained by the additivity axiom since events \( A_1, A_2, \) and \( A_3 \) form a partition.

\[
P(B) = P(B \cap A_1) + P(B \cap A_2) + P(B \cap A_3)
\]

\[
P(B|A) = \frac{P(A \cap B)}{P(A)} \quad \implies \quad P(B|A)P(A) = P(B \cap A)
\]

\[
P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3)
\]
Total Probability Theorem - A Closer Look
Total Probability Theorem - A Closer Look

\[ P(B) = P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B) \]
Total Probability Theorem - A Closer Look

- \( P(B) = P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B) \)
- \( P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3) \)
Total Probability Theorem - A Closer Look

- \( P(B) = P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B) \)
- \( P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3) \)

Why should you care?
Total Probability Theorem - Example

Assume you enter a competition and the probability of winning is given by $P(B)$. If there are 3 types of players ($A_1, A_2, A_3$) with a probability of playing against a given player given as $P(A_1) = 0.5$, $P(A_2) = 0.25$, $P(A_3) = 0.25$. The conditional probabilities of winning given you are playing against the three types of players are $P(B | A_1) = 0.3$, $P(B | A_2) = 0.4$, $P(B | A_3) = 0.5$. What is the probability of winning $P(B)$?
Total Probability Theorem - Example

Assume you enter a competition and the probability of winning is given by $P(B)$. 
Total Probability Theorem - Example

- Assume you enter a competition and the probability of winning is given by $P(B)$.
- If there are 3 types of players ($A_1, A_2, A_3$) with a probability of playing against a given player given as $P(A_1) = .5$, $P(A_2) = .25$, $P(A_3) = .25$. 

What is the probability of winning $[P(B)]$?
Total Probability Theorem - Example

- Assume you enter a competition and the probability of winning is given by $P(B)$.
- If there are 3 types of players ($A_1, A_2, A_3$) with a probability of playing against a given player given as $P(A_1) = .5, P(A_2) = .25, P(A_3) = .25$.
- The conditional probabilities of winning given you are playing against the three types of players are $P(B|A_1) = .3, P(B|A_2) = .4, P(B|A_3) = .5$.
Total Probability Theorem - Example

- Assume you enter a competition and the probability of winning is given by \( P(B) \).
- If there are 3 types of players \((A_1, A_2, A_3)\) with a probability of playing against a given player given as \( P(A_1) = .5, P(A_2) = .25, P(A_3) = .25 \).
- The conditional probabilities of winning given you are playing against the three types of players are \( P(B|A_1) = .3, P(B|A_2) = .4, P(B|A_3) = .5 \).
- What is the probability of winning \( [P(B)] \)?
Assume you enter a competition and the probability of winning is given by $P(B)$.

If there are 3 types of players ($A_1, A_2, A_3$) with a probability of playing against a given player given as $P(A_1) = .5$, $P(A_2) = .25$, $P(A_3) = .25$.

The conditional probabilities of winning given you are playing against the three types of players are $P(B|A_1) = .3$, $P(B|A_2) = .4$, $P(B|A_3) = .5$.

What is the probability of winning [$P(B)$]?

Go ahead and try
Bayes Theorem - Overview

Let \( A \) and \( B \) be two events and assume \( P(A) > 0 \). Bayes' Theorem relates conditional probabilities and is defined as follows:

\[
P(B | A) = \frac{P(A | B) P(B)}{P(A)}
\]

We can also make inferences about \( P(A | B) \) as follows:

\[
P(A | B) = \frac{P(B | A) P(A)}{P(B)}
\]

I won't prove it but Bayes' Theorem follows from the definition of conditional probability.
Let $A$ and $B$ be two events and assume $P(A) > 0$. Bayes’ Theorem relates conditional probabilities and is defined as follows:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$
Let $A$ and $B$ be two events and assume $P(A) > 0$. Bayes’ Theorem relates conditional probabilities and is defined as follows:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

We can also make inferences about $P(A|B)$ as follows:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$
Bayes Theorem - Overview

Let $A$ and $B$ be two events and assume $P(A) > 0$. Bayes’ Theorem relates conditional probabilities and is defined as follows:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

We can also make inferences about $P(A|B)$ as follows:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

I won’t prove it but Bayes’ Theorem follows from the definition of conditional probability.
Bayes’ Theorem - Overview Contin...

Now for a more general case, let $A_1, A_2, \ldots, A_n$ be disjoint events that form a partition of the sample space and $P(A_i) > 0$ for all $i$. Then given an event $B$ such that $P(B) > 0$, we have

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)}$$

Note $P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_n)P(A_n)$.
Now for a more general case, let $A_1, A_2, \ldots, A_n$ be disjoint events that form a partition of the sample space and $P(A_i) > 0$ for all $i$. Then given an event $B$ such that $P(B) > 0$.

\[
P(A_i \mid B) = \frac{P(B \mid A_i) P(A_i)}{P(B)}
\]
Now for a more general case, let $A_1, A_2, \ldots, A_n$ be disjoint events that form a partition of the sample space and $P(A_i) > 0$ for all $i$. Then given an event $B$ such that $P(B) > 0$, we have

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)}$$

Note

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_n)P(A_n)$$
Now for a more general case, let $A_1, A_2, \ldots, A_n$ be disjoint events that form a partition of the sample space and $P(A_i) > 0$ for all $i$. Then given an event $B$ such that $P(B) > 0$.

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)}$$

Note

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_n)P(A_n)$$
Bayes’ Theorem - Ex.

If an aircraft is present in a certain area, a radar detects it and generates an alarm signal with probability 0.99. If an aircraft is not present the radar generates a (false) alarm, with probability 0.10. We assume that an aircraft is present with probability 0.05. What is the probability of an aircraft presence given the radar generates an alarm?

First we need to define the events $A$ = \{aircraft present\} and $B$ = \{radar generates alarm\}.

The problem gives us the following:

$P(A) = 0.05$, $P(B|A) = 0.99$, $P(B|A^c) = 0.10$

What is $P(A^c)$?

$P(A^c) = 0.95$

Problem from Bertsekas and Tsitsiklis (2008, p. 33)
Bayes’ Theorem - Ex.

- If an aircraft is present in a certain area, a radar detects it and generates an alarm signal with probability 0.99. If an aircraft is not present the radar generates a (false) alarm, with probability 0.10. We assume that an aircraft is present with probability 0.05. What is the probability of an aircraft presence given the radar generates an alarm?\(^1\)

\(^1\)Problem from Bertsekas and Tsitsiklis (2008, p. 33)
Bayes’ Theorem - Ex.

- If an aircraft is present in a certain area, a radar detects it and generates an alarm signal with probability 0.99. If an aircraft is not present the radar generates a (false) alarm, with probability 0.10. We assume that an aircraft is present with probability 0.05. What is the probability of an aircraft presence given the radar generates an alarm?\(^1\)

- First we need to define the events

\(^1\)Problem from Bertsekas and Tsitsiklis (2008, p. 33)
Bayes’ Theorem - Ex.

- If an aircraft is present in a certain area, a radar detects it and generates an alarm signal with probability 0.99. If an aircraft is not present the radar generates a (false) alarm, with probability 0.10. We assume that an aircraft is present with probability 0.05. What is the probability of an aircraft presence given the radar generates an alarm?¹

- First we need to define the events
  - \( A = \{ \text{aircraft present} \} \) and \( B = \{ \text{radar generates alarm} \} \)

¹Problem from Bertsekas and Tsitsiklis (2008, p. 33)
Bayes’ Theorem - Ex.

- If an aircraft is present in a certain area, a radar detects it and generates an alarm signal with probability 0.99. If an aircraft is not present the radar generates a (false) alarm, with probability 0.10. We assume that an aircraft is present with probability 0.05. What is the probability of an aircraft presence given the radar generates an alarm?\(^1\)

- First we need to define the events
  - \(A = \{\text{aircraft present}\}\) and \(B = \{\text{radar generates alarm}\}\)

- The problem gives us the following:
  \[P(A) = 0.05, \ P(B|A) = 0.99, \ P(B|A^c) = 0.10\]

\(^1\)Problem from Bertsekas and Tsitsiklis (2008, p. 33)
Bayes’ Theorem - Ex.

- If an aircraft is present in a certain area, a radar detects it and generates an alarm signal with probability 0.99. If an aircraft is not present the radar generates a (false) alarm, with probability 0.10. We assume that an aircraft is present with probability 0.05. What is the probability of an aircraft presence given the radar generates an alarm?\(^1\)

- First we need to define the events
  - \(A = \{\text{aircraft present}\}\) and \(B = \{\text{radar generates alarm}\}\)

- The problem gives us the following:
  - \(P(A) = 0.05,\ P(B|A) = 0.99,\ P(B|A^c) = 0.10\)

- What is \(P(A^c)\)?

\(^1\)Problem from Bertsekas and Tsitsiklis (2008, p. 33)
Bayes’ Theorem - Ex.

- If an aircraft is present in a certain area, a radar detects it and generates an alarm signal with probability 0.99. If an aircraft is not present the radar generates a (false) alarm, with probability 0.10. We assume that an aircraft is present with probability 0.05. What is the probability of an aircraft presence given the radar generates an alarm?\(^1\)
- First we need to define the events
  - \(A = \{\text{aircraft present}\}\) and \(B = \{\text{radar generates alarm}\}\)
- The problem gives us the following:
  - \(P(A) = 0.05,\ P(B|A) = 0.99,\ P(B|A^c) = 0.10\)
  - What is \(P(A^c)\)?

\(^1\)Problem from Bertsekas and Tsitsiklis (2008, p. 33)
If an aircraft is present in a certain area, a radar detects it and generates an alarm signal with probability 0.99. If an aircraft is not present the radar generates a (false) alarm, with probability 0.10. We assume that an aircraft is present with probability 0.05. What is the probability of an aircraft presence given the radar generates an alarm?\(^1\)

First we need to define the events

- \(A = \{\text{aircraft present}\}\) and \(B = \{\text{radar generates alarm}\}\)

The problem gives us the following:

- \(P(A) = 0.05, \ P(B|A) = 0.99, \ P(B|A^c) = 0.10\)
- What is \(P(A^c)\)?
  - \(P(A^c) = 0.95\)

\(^1\)Problem from Bertsekas and Tsitsiklis (2008, p. 33)
Bayes’ Theorem Ex. Contin...

From previous slide

\( A = \{ \text{aircraft present} \} \) and \( B = \{ \text{radar generates alarm} \} \)

The problem gives us the following:

\[
P(A) = 0.05, \quad P(B | A) = 0.99, \quad P(B | A^c) = 0.10, \quad P(A^c) = 0.95
\]

We want \( P(A | B) \)

Using Bayes' Theorem

\[
P(A | B) = \frac{P(B | A) P(A)}{P(B)}
\]

What is \( P(B) \)?

\[
P(B) = P(B | A) P(A) + P(B | A^c) P(A^c) = 0.99 \times 0.05 + 0.10 \times 0.95 = 0.3426
\]

Now plug and chug

What is \( P(A | B) \)?

\[
P(A | B) = \frac{0.99 \times 0.05}{0.3426} \approx 0.0565
\]
Bayes’ Theorem Ex. Contin...

- From previous slide

\[ P(A | B) = \frac{P(B | A) P(A)}{P(B)} \]

Where:
- \( P(A | B) \) is the probability of \( A \) given \( B \)
- \( P(B | A) \) is the probability of \( B \) given \( A \)
- \( P(A) \) is the prior probability of \( A \)
- \( P(B) \) is the marginal probability of \( B \)

Let's calculate \( P(A | B) \):

\[ P(A | B) = \frac{0.99 \times 0.05}{0.99 \times 0.05 + 0.10 \times 0.95} \approx 0.3426 \]
Bayes’ Theorem Ex. Contin...

- From previous slide
  - \( A = \{ \text{aircraft present} \} \) and \( B = \{ \text{radar generates alarm} \} \)
Bayes’ Theorem Ex. Contin...

- From previous slide
  - $A = \{\text{aircraft present}\}$ and $B = \{\text{radar generates alarm}\}$
  - The problem gives us the following:
    \[ P(A) = 0.05, \quad P(B|A) = 0.99, \quad P(B|A^c) = 0.10, \quad P(A^c) = 0.95 \]
Bayes’ Theorem Ex. Contin...

- From previous slide
  - \( A = \{\text{aircraft present}\} \) and \( B = \{\text{radar generates alarm}\} \)
- The problem gives us the following:
  - \( P(A) = 0.05, \ P(B|A) = 0.99, \ P(B|A^c) = 0.10, \ P(A^c) = 0.95 \)
- We want \( P(A|B) \)

From previous slide

- $A = \{\text{aircraft present}\}$ and $B = \{\text{radar generates alarm}\}$

The problem gives us the following:

- $P(A) = 0.05$, $P(B|A) = 0.99$, $P(B|A^c) = 0.10$, $P(A^c) = 0.95$

We want $P(A|B)$

Using Bayes’ Theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$$P(A|B) = \frac{0.99 \times 0.05}{0.99 \times 0.05 + 0.10 \times 0.95} \approx 0.3426$$
Bayes’ Theorem Ex. Contin...

- From previous slide
  - $A = \{\text{aircraft present}\}$ and $B = \{\text{radar generates alarm}\}$
- The problem gives us the following:
  - $P(A) = 0.05$, $P(B|A) = 0.99$, $P(B|A^c) = 0.10$, $P(A^c) = 0.95$
- We want $P(A|B)$
- Using Bayes’ Theorem $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$
- What is $P(B)$?

\[ P(B) = P(B|A)P(A) + P(B|A^c)P(A^c) = 0.99 \times 0.05 + 0.10 \times 0.95 \approx 0.3426 \]
Bayes’ Theorem Ex. Contin...

- From previous slide
  - \( A = \{ \text{aircraft present} \} \) and \( B = \{ \text{radar generates alarm} \} \)
- The problem gives us the following:
  \[ P(A) = 0.05, \ P(B|A) = 0.99, \ P(B|A^c) = 0.10, \ P(A^c) = 0.95 \]
- We want \( P(A|B) \)
- Using Bayes’ Theorem \( P(A|B) = \frac{P(B|A)P(A)}{P(B)} \)
- What is \( P(B) \)?
Bayes’ Theorem Ex. Contin...

- From previous slide
  - \( A = \{\text{aircraft present}\} \) and \( B = \{\text{radar generates alarm}\} \)
- The problem gives us the following:
  \[ P(A) = 0.05, \ P(B|A) = 0.99, \ P(B|A^c) = 0.10, \ P(A^c) = 0.95 \]
- We want \( P(A|B) \)
- Using Bayes’ Theorem \( P(A|B) = \frac{P(B|A)P(A)}{P(B)} \)
- What is \( P(B) \)?
  \[ P(B) = P(B|A)P(A) + P(B|A^c)P(A^c) \]
Bayes’ Theorem Ex. Contin...

- From previous slide
  - \( A = \{ \text{aircraft present} \} \) and \( B = \{ \text{radar generates alarm} \} \)

- The problem gives us the following:
  \[ P(A) = 0.05, \quad P(B|A) = 0.99, \quad P(B|A^c) = 0.10, \quad P(A^c) = 0.95 \]

- We want \( P(A|B) \)

- Using Bayes’ Theorem
  \[ P(A|B) = \frac{P(B|A)P(A)}{P(B)} \]

- What is \( P(B) \)?
  - \( P(B) = P(B|A)P(A) + P(B|A^c)P(A^c) \)

- Now plug and chug
Bayes’ Theorem Ex. Contin...

- From previous slide
  - $A = \{\text{aircraft present}\}$ and $B = \{\text{radar generates alarm}\}$
- The problem gives us the following:
  - $P(A) = 0.05$, $P(B|A) = 0.99$, $P(B|A^c) = 0.10$, $P(A^c) = 0.95$
- We want $P(A|B)$
- Using Bayes’ Theorem
  - $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$
- What is $P(B)$?
  - $P(B) = P(B|A)P(A) + P(B|A^c)P(A^c)$
- Now plug and chug
- What is $P(A|B)$?
Bayes’ Theorem Ex. Contin...

- From previous slide
  - \( A = \{ \text{aircraft present} \} \) and \( B = \{ \text{radar generates alarm} \} \)
- The problem gives us the following:
  \[ P(A) = 0.05, \ P(B|A) = 0.99, \ P(B|A^c) = 0.10, \ P(A^c) = 0.95 \]
- We want \( P(A|B) \)
- Using Bayes’ Theorem \( P(A|B) = \frac{P(B|A)P(A)}{P(B)} \)
- What is \( P(B) \)?
  - \( P(B) = P(B|A)P(A) + P(B|A^c)P(A^c) \)
- Now plug and chug
- What is \( P(A|B) \)?
  - \( P(A|B) = \frac{0.99 \times 0.05}{0.99 \times 0.05 + 0.10 \times 0.95} \approx 0.3426 \)
Bayes’ Theorem - Learning Check

- Remember $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$

- You are given the following information about a medical test related to a disease. The false negative rate of the test is 16% ($P(\neg|sick)$), false positive rate of the test is 19% ($P(\neg|not\ sick)$), and the probability of someone having the disease (being sick) is 8% ($P(sick)$).

- What is the $P(sick|\neg)$ the probability of being sick given the test was positive?
Independence - Overview

Independence is a very important property in probability theory. With conditional probability $P(A | B)$ representing the partial information event $B$ provides about $A$. What if knowing event $B$ provides us with no information about the probability of $A$? This is precisely the definition of independence. More formally, two events $A$ and $B$ are independent if $P(A | B) = P(A)$ $P(A \cap B) = P(A) \times P(B)$. A simple example is flipping coins; does the fact that you got heads tell you any additional information about the probability you will get tails and vice versa? No!
Independence is a very important property in probability theory.
Independence - Overview

- Independence is a very important property in probability theory.
- With conditional probability $P(A|B)$ represented the partial information event $B$ provides about $A$. 

Independence - Overview

- Independence is a very important property in probability theory.
- With conditional probability $P(A|B)$ represented the partial information event $B$ provides about $A$.
- What if knowing event $B$ provides us with no information about the probability of $A$ [$P(A)$]?
Independence - Overview

- Independence is a very important property in probability theory.
- With conditional probability $P(A|B)$ represented the partial information event $B$ provides about $A$.
- What if knowing event $B$ provides us with no information about the probability of $A$ [$P(A)$]?
- This is precisely the definition of independence.
Independence - Overview

- Independence is a very important property in probability theory.
- With conditional probability $P(A|B)$ represented the partial information event $B$ provides about $A$.
- What if knowing event $B$ provides us with no information about the probability of $A$ [$P(A)$]?
- This is precisely the definition of independence.
- More formally two events $A$ and $B$ are independent if
Independence - Overview

- Independence is a very important property in probability theory.
- With conditional probability $P(A|B)$ represented the partial information event $B$ provides about $A$.
- What if knowing event $B$ provides us with no information about the probability of $A$ [$P(A)$]?
- This is precisely the definition of independence.
- More formally two events $A$ and $B$ are independent if
  - $P(A|B) = P(A)$
Independence - Overview

- Independence is a very important property in probability theory.
- With conditional probability $P(A|B)$ represented the partial information event $B$ provides about $A$.
- What if knowing event $B$ provides us with no information about the probability of $A$ [$P(A)$]?
- This is precisely the definition of independence.
- More formally, two events $A$ and $B$ are independent if:
  - $P(A|B) = P(A)$
  - $P(A \cap B) = P(A) \times P(B)$

A simple example is flipping coins: does the fact that you got heads tell you any additional information about the probability you will get tails and vice versa. No!
Independence - Overview

- Independence is a very important property in probability theory.
- With conditional probability $P(A|B)$ represented the partial information event $B$ provides about $A$.
- What if knowing event $B$ provides us with no information about the probability of $A$ [$P(A)$]?
- This is precisely the definition of independence.
- More formally, two events $A$ and $B$ are independent if
  - $P(A|B) = P(A)$
  - $P(A \cap B) = P(A) \times P(B)$
- A simple example is flipping coins, does the fact that you got heads tell you any additional information about the probability you will get tails and vice versa. No!
Independence - Multiplication Rule

Remember this example from the Multiplication Rule:

Given:

\[ P(A_1 \cap \ldots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \ldots P(A_n|A_1 \cap A_2 \cap \ldots \cap A_{n-1}) \]

Toss a fair coin 3 times. Let:

\[ A = A_1 \cap A_2 \cap A_3 \]

Then:

\[ P(A) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \]

\[ P(A) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8} \]

The assumption of independence simplifies this problem.
Independence - Multiplication Rule

- Remember this example from the Multiplication Rule

Given $A_1, \ldots, A_n$

$$P(A_1 \cap \cdots \cap A_n) = P(A_1)P(A_2 | A_1)P(A_3 | A_1 \cap A_2)\cdots P(A_n | A_1 \cap \cdots \cap A_{n-1})$$

Toss a fair coin 3 times. Let $A = A_1 \cap A_2 \cap A_3$

$$P(A) = P(A_1)P(A_2 | A_1)P(A_3 | A_1 \cap A_2) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$$

The assumption of independence simplifies this problem.
Independence - Multiplication Rule

- Remember this example from the Multiplication Rule
- Given $A_1, \ldots, A_n$

\[ P(\cap_{i=1}^n A_i) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|\cap_{i=1}^{n-1} A_i) \]
Independence - Multiplication Rule

- Remember this example from the Multiplication Rule
  - Given $A_1, \ldots, A_n$

  \[
P(\cap_{i=1}^{n} A_i) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|\cap_{i=1}^{n-1} A_i)
\]

- Toss a fair coin 3 times. Let $A = A_1 \cap A_2 \cap A_3$
Independence - Multiplication Rule

- Remember this example from the Multiplication Rule
  - Given $A_1, \ldots, A_n$
    \[ P(\cap_{i=1}^n A_i) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|\cap_{i=1}^{n-1} A_i) \]
  - Toss a fair coin 3 times. Let $A = A_1 \cap A_2 \cap A_3$
  - $P(A) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)$
Independence - Multiplication Rule

- Remember this example from the Multiplication Rule
  - Given $A_1, \ldots, A_n$
    \[
P(\cap_{i=1}^n A_i) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|\cap_{i=1}^{n-1} A_i)
    \]
  - Toss a fair coin 3 times. Let $A = A_1 \cap A_2 \cap A_3$
    - $P(A) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)$
    - $P(A) = P(A_1)P(A_2)P(A_3) = 1/2 \times 1/2 \times 1/2 = 1/8$
Independence - Multiplication Rule

- Remember this example from the Multiplication Rule
  - Given $A_1, \ldots, A_n$

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

- Toss a fair coin 3 times. Let $A = A_1 \cap A_2 \cap A_3$
  - $P(A) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)$
  - $P(A) = P(A_1)P(A_2)P(A_3) = 1/2 \times 1/2 \times 1/2 = 1/8$
  - The assumption of independence simplifies this problem.
Independence - Example 1

Given the following 2x2 table below with coffee user (Y/N) and passing probability theory (Y/N)

<table>
<thead>
<tr>
<th></th>
<th>Pass (P)</th>
<th>Fail (F)</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drinker (D)</td>
<td>0.51</td>
<td>0.09</td>
<td>0.60</td>
</tr>
<tr>
<td>Non-drinker (ND)</td>
<td>0.34</td>
<td>0.06</td>
<td>0.40</td>
</tr>
<tr>
<td>Column Total</td>
<td>0.85</td>
<td>0.15</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Now using conditional probability we can test independence
Are Passing and Coffee Drinker independent?

Check if \( P(D \cap P) = P(D) \times P(P) = 0.85 \times 0.60 = 0.51 \) Yes!

Check if \( P(ND \cap P) = P(ND) \times P(P) = 0.85 \times 0.40 = 0.34 \) Yes!

Check if \( P(D \cap F) = P(D) \times P(F) = 0.60 \times 0.15 = 0.09 \) Yes!

Check if \( P(ND \cap F) = P(ND) \times P(F) = 0.40 \times 0.15 = 0.06 \) Yes!

Indeed Passing and Coffee Drinker are independent!
Independence - Example 1

- Given the following 2x2 table below with coffee user (Y/N) and passing probability theory (Y/N)
### Independence - Example 1

Given the following 2x2 table below with coffee user (Y/N) and passing probability theory (Y/N)

<table>
<thead>
<tr>
<th></th>
<th>Pass (P)</th>
<th>Fail (F)</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drinker (D)</td>
<td>.51</td>
<td>.09</td>
<td>.60</td>
</tr>
<tr>
<td>Non-drinker (ND)</td>
<td>.34</td>
<td>.06</td>
<td>.40</td>
</tr>
<tr>
<td>Column Total</td>
<td>.85</td>
<td>.15</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Now using conditional probability we can test independence:

- Check if $P(D \cap P) = P(D) \times P(P) = .85 \times .60 = .51$ Yes!
- Check if $P(ND \cap P) = P(ND) \times P(P) = .85 \times .40 = .34$ Yes!
- Check if $P(D \cap F) = P(D) \times P(F) = .60 \times .15 = .09$ Yes!
- Check if $P(ND \cap F) = P(ND) \times P(F) = .40 \times .15 = .06$ Yes!

Indeed Passing and Coffee Drinker are independent!
Independence - Example 1

- Given the following 2x2 table below with coffee user (Y/N) and passing probability theory (Y/N)

<table>
<thead>
<tr>
<th></th>
<th>Pass (P)</th>
<th>Fail (F)</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drinker (D)</td>
<td>.51</td>
<td>.09</td>
<td>.60</td>
</tr>
<tr>
<td>Non-drinker (ND)</td>
<td>.34</td>
<td>.06</td>
<td>.40</td>
</tr>
<tr>
<td>Column Total</td>
<td>.85</td>
<td>.15</td>
<td>1.00</td>
</tr>
</tbody>
</table>

- Now using conditional probability we can test independence
Independence - Example 1

- Given the following 2x2 table below with coffee user (Y/N) and passing probability theory (Y/N)

<table>
<thead>
<tr>
<th></th>
<th>Pass (P)</th>
<th>Fail (F)</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drinker (D)</td>
<td>.51</td>
<td>.09</td>
<td>.60</td>
</tr>
<tr>
<td>Non-drinker (ND)</td>
<td>.34</td>
<td>.06</td>
<td>.40</td>
</tr>
<tr>
<td>Column Total</td>
<td><strong>.85</strong></td>
<td><strong>.15</strong></td>
<td><strong>1.00</strong></td>
</tr>
</tbody>
</table>

- Now using conditional probability we can test independence

- Are Passing and Coffee Drinker independent?
Independence - Example 1

Given the following 2x2 table below with coffee user (Y/N) and passing probability theory (Y/N)

<table>
<thead>
<tr>
<th></th>
<th>Pass (P)</th>
<th>Fail (F)</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drinker (D)</td>
<td>.51</td>
<td>.09</td>
<td>.60</td>
</tr>
<tr>
<td>Non-drinker (ND)</td>
<td>.34</td>
<td>.06</td>
<td>.40</td>
</tr>
<tr>
<td>Column Total</td>
<td>.85</td>
<td>.15</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Now using conditional probability we can test independence.
Are Passing and Coffee Drinker independent?
- Check if $P(D \cap P) = P(D) \times P(P) = .85 \times .60 = .51$ Yes!
Independence - Example 1

Given the following 2x2 table below with coffee user (Y/N) and passing probability theory (Y/N)

<table>
<thead>
<tr>
<th></th>
<th>Pass (P)</th>
<th>Fail (F)</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drinker (D)</td>
<td>.51</td>
<td>.09</td>
<td>.60</td>
</tr>
<tr>
<td>Non-drinker (ND)</td>
<td>.34</td>
<td>.06</td>
<td>.40</td>
</tr>
<tr>
<td>Column Total</td>
<td>.85</td>
<td>.15</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Now using conditional probability we can test independence

Are Passing and Coffee Drinker independent?

- Check if \( P(D \cap P) = P(D) \times P(P) = .85 \times .60 = .51 \) Yes!
- Check if \( P(ND \cap P) = P(ND) \times P(P) = .85 \times .40 = .34 \) Yes!
Independence - Example 1

Given the following 2x2 table below with coffee user (Y/N) and passing probability theory (Y/N)

<table>
<thead>
<tr>
<th></th>
<th>Pass (P)</th>
<th>Fail (F)</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drinker (D)</td>
<td>.51</td>
<td>.09</td>
<td>.60</td>
</tr>
<tr>
<td>Non-drinker (ND)</td>
<td>.34</td>
<td>.06</td>
<td>.40</td>
</tr>
<tr>
<td>Column Total</td>
<td>.85</td>
<td>.15</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Now using conditional probability we can test independence

Are Passing and Coffee Drinker independent?

- Check if $P(D \cap P) = P(D) \times P(P) = .85 \times .60 = .51$ Yes!
- Check if $P(ND \cap P) = P(ND) \times P(P) = .85 \times .40 = .34$ Yes!
- Check if $P(D \cap F) = P(D) \times P(F) = .60 \times .15 = .09$ Yes!
Independence - Example 1

- Given the following 2x2 table below with coffee user (Y/N) and passing probability theory (Y/N)

<table>
<thead>
<tr>
<th></th>
<th>Pass (P)</th>
<th>Fail (F)</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drinker (D)</td>
<td>.51</td>
<td>.09</td>
<td>.60</td>
</tr>
<tr>
<td>Non-drinker (ND)</td>
<td>.34</td>
<td>.06</td>
<td>.40</td>
</tr>
<tr>
<td>Column Total</td>
<td>.85</td>
<td>.15</td>
<td>1.00</td>
</tr>
</tbody>
</table>

- Now using conditional probability we can test independence.
- Are Passing and Coffee Drinker independent?
  - Check if \( P(D \cap P) = P(D) \times P(P) = 0.85 \times 0.60 = 0.51 \) Yes!
  - Check if \( P(ND \cap P) = P(ND) \times P(P) = 0.85 \times 0.40 = 0.34 \) Yes!
  - Check if \( P(D \cap F) = P(D) \times P(F) = 0.60 \times 0.15 = 0.09 \) Yes!
  - Check if \( P(ND \cap F) = P(ND) \times P(F) = 0.40 \times 0.15 = 0.06 \) Yes!
Independence - Example 1

- Given the following 2x2 table below with coffee user (Y/N) and passing probability theory (Y/N)

<table>
<thead>
<tr>
<th></th>
<th>Pass (P)</th>
<th>Fail (F)</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drinker (D)</td>
<td>.51</td>
<td>.09</td>
<td>.60</td>
</tr>
<tr>
<td>Non-drinker (ND)</td>
<td>.34</td>
<td>.06</td>
<td>.40</td>
</tr>
<tr>
<td>Column Total</td>
<td>.85</td>
<td>.15</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Now using conditional probability we can test independence.

Are Passing and Coffee Drinker independent?
- Check if $P(D \cap P) = P(D) \times P(P) = .85 \times .60 = .51$ Yes!
- Check if $P(ND \cap P) = P(ND) \times P(P) = .85 \times .40 = .34$ Yes!
- Check if $P(D \cap F) = P(D) \times P(F) = .60 \times .15 = .09$ Yes!
- Check if $P(ND \cap F) = P(ND) \times P(F) = .40 \times .15 = .06$ Yes!

Indeed Passing and Coffee Drinker are independent!
Independence - Example 2

Given the following 2x2 table below with coffee user (Y/N) and passing probability theory (Y/N)

<table>
<thead>
<tr>
<th></th>
<th>Pass (P)</th>
<th>Fail (F)</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drinker (D)</td>
<td>.48</td>
<td>.12</td>
<td>.60</td>
</tr>
<tr>
<td>Non-drinker (ND)</td>
<td>.34</td>
<td>.06</td>
<td>.40</td>
</tr>
<tr>
<td>Column Total</td>
<td>.82</td>
<td>.18</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Now using conditional probability we can test independence

Are Passing and Coffee Drinker independent?

Check if $P(D \cap P) = P(D) \times P(P) = .49 \neq .48$ No!

Check if $P(ND \cap P) = P(ND) \times P(P) = .33 \neq .34$ No!

I don't need to compute anymore it is clear in this case there is dependence!
Independence - Example 2

Given the following 2x2 table below with coffee user (Y/N) and passing probability theory (Y/N):

<table>
<thead>
<tr>
<th></th>
<th>Pass (P)</th>
<th>Fail (F)</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drinker (D)</td>
<td>.48</td>
<td>.12</td>
<td>.60</td>
</tr>
<tr>
<td>Non-drinker (ND)</td>
<td>.34</td>
<td>.06</td>
<td>.40</td>
</tr>
</tbody>
</table>

Column Total: .82 (Pass) + .18 (Fail) = 1.00

Now using conditional probability we can test independence:

Are Passing and Coffee Drinker independent?

Check if

\[ P(D \cap P) = P(D) \times P(P) = .49 \]
\[ \neq .48 \]

No!

Check if

\[ P(ND \cap P) = P(ND) \times P(P) = .33 \]
\[ \neq .34 \]

No!

I don't need to compute anymore it is clear in this case there is dependence!
Independence - Example 2

- Given the following 2x2 table below with coffee user (Y/N) and passing probability theory (Y/N)

<table>
<thead>
<tr>
<th></th>
<th>Pass (P)</th>
<th>Fail (F)</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drinker (D)</td>
<td>.48</td>
<td>.12</td>
<td>.60</td>
</tr>
<tr>
<td>Non-drinker (ND)</td>
<td>.34</td>
<td>.06</td>
<td>.40</td>
</tr>
<tr>
<td>Column Total</td>
<td>.82</td>
<td>.18</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Now using conditional probability we can test independence.

- Are Passing and Coffee Drinker independent?

Check if $P(D \cap P) = P(D) \times P(P) = .49 \neq .48$ No!

Check if $P(ND \cap P) = P(ND) \times P(P) = .33 \neq .34$ No!

I don't need to compute anymore it is clear in this case there is dependence!
Independence - Example 2

- Given the following 2x2 table below with coffee user (Y/N) and passing probability theory (Y/N)

<table>
<thead>
<tr>
<th></th>
<th>Pass (P)</th>
<th>Fail (F)</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drinker (D)</td>
<td>.48</td>
<td>.12</td>
<td>.60</td>
</tr>
<tr>
<td>Non-drinker (ND)</td>
<td>.34</td>
<td>.06</td>
<td>.40</td>
</tr>
<tr>
<td>Column Total</td>
<td>.82</td>
<td>.18</td>
<td>1.00</td>
</tr>
</tbody>
</table>

- Now using conditional probability we can test independence
Independence - Example 2

- Given the following 2x2 table below with coffee user (Y/N) and passing probability theory (Y/N)

<table>
<thead>
<tr>
<th></th>
<th>Pass (P)</th>
<th>Fail (F)</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drinker (D)</td>
<td>.48</td>
<td>.12</td>
<td>.60</td>
</tr>
<tr>
<td>Non-drinker (ND)</td>
<td>.34</td>
<td>.06</td>
<td>.40</td>
</tr>
<tr>
<td>Column Total</td>
<td>.82</td>
<td>.18</td>
<td>1.00</td>
</tr>
</tbody>
</table>

- Now using conditional probability we can test independence
- Are Passing and Coffee Drinker independent?
### Independence - Example 2

Given the following 2x2 table below with coffee user (Y/N) and passing probability theory (Y/N)

<table>
<thead>
<tr>
<th></th>
<th>Pass (P)</th>
<th>Fail (F)</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drinker (D)</td>
<td>.48</td>
<td>.12</td>
<td>.60</td>
</tr>
<tr>
<td>Non-drinker (ND)</td>
<td>.34</td>
<td>.06</td>
<td>.40</td>
</tr>
<tr>
<td>Column Total</td>
<td>.82</td>
<td>.18</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Now using conditional probability we can test independence

Are Passing and Coffee Drinker independent?

- Check if \( P(D \cap P) = P(D) \times P(P) = .82 \times .60 = .49 \neq .48 \) No!
Independence - Example 2

Given the following 2x2 table below with coffee user (Y/N) and passing probability theory (Y/N)

<table>
<thead>
<tr>
<th></th>
<th>Pass (P)</th>
<th>Fail (F)</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drinker (D)</td>
<td>.48</td>
<td>.12</td>
<td>.60</td>
</tr>
<tr>
<td>Non-drinker (ND)</td>
<td>.34</td>
<td>.06</td>
<td>.40</td>
</tr>
<tr>
<td>Column Total</td>
<td>.82</td>
<td>.18</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Now using conditional probability we can test independence

Are Passing and Coffee Drinker independent?

Check if \( P(D \cap P) = P(D) \times P(P) = .82 \times .60 = .49 \neq .48 \) No!

Check if \( P(ND \cap P) = P(ND) \times P(P) = .82 \times .40 = .33 \neq .34 \) No!
Independence - Example 2

- Given the following 2x2 table below with coffee user (Y/N) and passing probability theory (Y/N)

<table>
<thead>
<tr>
<th></th>
<th>Pass (P)</th>
<th>Fail (F)</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drinker (D)</td>
<td>.48</td>
<td>.12</td>
<td>.60</td>
</tr>
<tr>
<td>Non-drinker (ND)</td>
<td>.34</td>
<td>.06</td>
<td>.40</td>
</tr>
<tr>
<td>Column Total</td>
<td>.82</td>
<td>.18</td>
<td>1.00</td>
</tr>
</tbody>
</table>

- Now using conditional probability we can test independence
- Are Passing and Coffee Drinker independent?
  - Check if $P(D \cap P) = P(D) \times P(P) = .82 \times .60 = .49 \neq .48$ No!
  - Check if $P(ND \cap P) = P(ND) \times P(P) = .82 \times .40 = .33 \neq .34$ No!

- I don’t need to compute anymore it is clear in this case there is dependence!
Conditional Independence - Overview

What if we observe some event $C$. Does knowing something about $C$ influence the joint probability of events $A$ and $B$ [$P(A \cap B)$]?

If the answer is no then $P(A \cap B | C) = P(A | C) \times P(B | C)$.

Also if $A$ and $B$ are conditionally independent given $C$ then $P(A | B \cap C) = P(A | C)$.

Let's assume that we have two students John and Sally who live in different suburbs of New York City. Sally drives to school and John takes the train. It is reasonable to assume that $J = \{\text{John is late}\}$ and $S = \{\text{Sally is late}\}$ are independent. If both of them take the train then it is reasonable to assume the events are independent if no train strike occurs.
Conditional Independence - Overview

- What if we observe some event $C$. Does knowing something about $C$ influence the joint probability of events $A$ and $B$ [$P(A \cap B)$]?
Conditional Independence - Overview

- What if we observe some event \( C \). Does knowing something about \( C \) influence the joint probability of events \( A \) and \( B \) \([P(A \cap B)]\)?
- If the answer is no then \( P(A \cap B | C) = P(A | C) \times P(B | C) \)

Let's assume that we have two students John and Sally who live in different suburbs of New York City. Sally drives to school and John takes the train. It is reasonable to assume that \( J = \{\text{John is late}\} \) and \( S = \{\text{Sally is late}\} \) are independent. If both of them take the train then it is reasonable to assume the events are independent if no train strike occurs.
Conditional Independence - Overview

- What if we observe some event $C$. Does knowing something about $C$ influence the joint probability of events $A$ and $B$ [$P(A \cap B)$]?  

- If the answer is no then $P(A \cap B | C) = P(A | C) \times P(B | C)$

- Also if $A$ and $B$ are conditionally independent given $C$ then $P(A | B \cap C) = P(A | C)$
Conditional Independence - Overview

- What if we observe some event $C$. Does knowing something about $C$ influence the joint probability of events $A$ and $B$ $[P(A \cap B)]$?
- If the answer is no then $P(A \cap B | C) = P(A | C) \ast P(B | C)$
- Also if $A$ and $B$ are conditionally independent given $C$ then $P(A | B \cap C) = P(A | C)$
- Let’s assume that we have two students John and Sally who live in different suburbs of New York City. Sally drives to school and John takes the train.
Conditional Independence - Overview

- What if we observe some event $C$. Does knowing something about $C$ influence the joint probability of events $A$ and $B$ $[P(A \cap B)]$?
- If the answer is no then $P(A \cap B|C) = P(A|C) \times P(B|C)$
- Also if $A$ and $B$ are conditionally independent given $C$ then $P(A|B \cap C) = P(A|C)$
- Let’s assume that we have two students John and Sally who live in different suburbs of New York City. Sally drives to school and John takes the train.
- It is reasonable to assume that $J = \{John \ is \ late\}$ and $S = \{Sally \ is \ late\}$ are independent
Conditional Independence - Overview

- What if we observe some event \( C \). Does knowing something about \( C \) influence the joint probability of events \( A \) and \( B \) \( [P(A \cap B)] \)?
- If the answer is no then \( P(A \cap B | C) = P(A | C) \times P(B | C) \)
- Also if \( A \) and \( B \) are conditionally independent given \( C \) then \( P(A | B \cap C) = P(A | C) \)
- Let's assume that we have two students John and Sally who live in different suburbs of New York City. Sally drives to school and John takes the train.
- It is reasonable to assume that \( J = \{John \ is \ late\} \) and \( S = \{Sally \ is \ late\} \) are independent
- If both of them take the train then it is reasonable to assume the events are independent if no train strike occurs.
Suppose we want to reason about a student being accepted to MIT or Harvard for graduate studies. In most reasonable distributions these two events would not be independent. If I know you were admitted to MIT then it means you probably have a decent shot at being admitted to Harvard and vice versa.

Suppose the two schools base their admissions on the students GPA and suppose our student has a 4.0 GPA. Then we could reason and argue that once we know a students GPA this will tell us all the information of being admitted to MIT. Knowing the student was admitted to Harvard does not tell us anything more about being admitted to MIT.

Formally:

\[ P(\text{MIT} | \text{GPA} \cap \text{Harvard}) = P(\text{MIT} | \text{GPA}) \]

Example based on Koller and Friedman (2008, pp. 23-24)
Conditional Independence - Example

Suppose we want to reason about a student being accepted to MIT or Harvard for graduate studies.

\[ P(\text{MIT} | \text{GPA} \cap \text{Harvard}) = P(\text{MIT} | \text{GPA}) \]

1Example based on Koller and Friedman (2008, pp. 23-24)
Suppose we want to reason about a student being accepted to MIT or Harvard for graduate studies. In most reasonable distributions these two events would not be independent. Suppose the two schools base their admissions on the students GPA and suppose our student has a 4.0 GPA. Then we could reason and argue that once we know a student's GPA this will tell us all the information of being admitted to MIT. Knowing the student was admitted to Harvard does not tell us anything more about being admitted to MIT. Formally $P(\text{MIT} | \text{GPA} \cap \text{Harvard}) = P(\text{MIT} | \text{GPA})$.

---

1 Example based on Koller and Friedman (2008, pp. 23-24)
Conditional Independence - Example

- Suppose we want to reason about a student being accepted to MIT or Harvard for graduate studies.
- In most reasonable distributions these two events would not be independent
  - If I know you were admitted to MIT then it means you probably have a decent shot at being admitted to Harvard and vice versa.

---

Example based on Koller and Friedman (2008, pp. 23-24)
Suppose we want to reason about a student being accepted to MIT or Harvard for graduate studies.

In most reasonable distributions these two events would not be independent
- If I know you were admitted to MIT then it means you probably have a decent shot at being admitted to Harvard and vice versa.

Suppose the two schools base their admissions on the students GPA and suppose our student has a 4.0 GPA.

---

1 Example based on Koller and Friedman (2008, pp. 23-24)
Suppose we want to reason about a student being accepted to MIT or Harvard for graduate studies.

In most reasonable distributions these two events would not be independent

If I know you were admitted to MIT then it means you probably have a decent shot at being admitted to Harvard and vice versa.

Suppose the two schools base their admissions on the students GPA and suppose our student has a 4.0 GPA.

Then we could reason and argue that once we know a students GPA this will tell us all the information of being admitted to MIT. Knowing the student was admitted to Harvard does not tell us anything more about being admitted to MIT.

---

Example based on Koller and Friedman (2008, pp. 23-24)
Conditional Indendence - Example

Suppose we want to reason about a student being accepted to MIT or Harvard for graduate studies.

In most reasonable distributions these two events would not be independent
- If I know you were admitted to MIT then it means you probably have a decent shot at being admitted to Harvard and vice versa.

Suppose the two schools base their admissions on the students GPA and suppose our student has a 4.0 GPA.

Then we could reason and argue that once we know a students GPA this will tell us all the information of being admitted to MIT. Knowing the student was admitted to Harvard does not tell us anything more about being admitted to MIT.

Formally \( P(MIT | GPA \cap Harvard) = P(MIT | GPA) \)

\(^1\)Example based on Koller and Friedman (2008, pp. 23-24)
Pairwise Independence and Independence

It is important to note that pairwise independence does not imply independence. For independence to hold the following must hold for three events $A_1, A_2, A_3$:

\[
P(A_1 \cap A_2) = P(A_1) \cdot P(A_2)
\]

\[
P(A_1 \cap A_3) = P(A_1) \cdot P(A_3)
\]

\[
P(A_2 \cap A_3) = P(A_2) \cdot P(A_3)
\]

If one of these conditions fails then independence fails.
Pairwise Independence and Independence

- It is important to note that pairwise independence does not imply independence.
Pairwise Independence and Independence

- It is important to note that pairwise independence does not imply independence.
- For independence to hold the following must hold for three events $A_1, A_2, A_3$:  
  
  \[
  P(A_1 \cap A_2) = P(A_1) \times P(A_2),
  \]
  
  \[
  P(A_1 \cap A_3) = P(A_1) \times P(A_3),
  \]
  
  \[
  P(A_2 \cap A_3) = P(A_2) \times P(A_3),
  \]
  
  \[
  P(A_1 \cap A_2 \cap A_3) = P(A_1) \times P(A_2) \times P(A_3).
  \]
Pairwise Independence and Independence

- It is important to note that pairwise independence does not imply independence.
- For independence to hold the following must hold for three events $A_1, A_2, A_3$:
  - $P(A_1 \cap A_2) = P(A_1) \times P(A_2)$
Pairwise Independence and Independence

- It is important to note that pairwise independence does not imply independence.
- For independence to hold the following must hold for three events $A_1, A_2, A_3$:
  - $P(A_1 \cap A_2) = P(A_1) \times P(A_2)$
  - $P(A_1 \cap A_3) = P(A_1) \times P(A_3)$
  - $P(A_2 \cap A_3) = P(A_2) \times P(A_3)$
  - $P(A_1 \cap A_2 \cap A_3) = P(A_1) \times P(A_2) \times P(A_3)$

If one of these conditions fails then independence fails.
Pairwise Independence and Independence

- It is important to note that pairwise independence does not imply independence.
- For independence to hold the following must hold for three events $A_1, A_2, A_3$:
  - $P(A_1 \cap A_2) = P(A_1) \times P(A_2)$
  - $P(A_1 \cap A_3) = P(A_1) \times P(A_3)$
  - $P(A_2 \cap A_3) = P(A_2) \times P(A_3)$
  - $P(A_1 \cap A_2 \cap A_3) = P(A_1) \times P(A_2) \times P(A_3)$

If one of these conditions fails then independence fails.
It is important to note that pairwise independence does not imply independence.

For independence to hold the following must hold for three events $A_1, A_2, A_3$:

- $P(A_1 \cap A_2) = P(A_1) \times P(A_2)$
- $P(A_1 \cap A_3) = P(A_1) \times P(A_3)$
- $P(A_2 \cap A_3) = P(A_2) \times P(A_3)$
- $P(A_1 \cap A_2 \cap A_3) = P(A_1) \times P(A_2) \times P(A_3)$
Pairwise Independence and Independence

- It is important to note that pairwise independence does not imply independence.
- For independence to hold the following must hold for three events $A_1, A_2, A_3$:
  - $P(A_1 \cap A_2) = P(A_1) \times P(A_2)$
  - $P(A_1 \cap A_3) = P(A_1) \times P(A_3)$
  - $P(A_2 \cap A_3) = P(A_2) \times P(A_3)$
  - $P(A_1 \cap A_2 \cap A_3) = P(A_1) \times P(A_2) \times P(A_3)$
- If one of these conditions fails then independence fails
Learning Check - Independence

- Why is independence important in statistics and probability?
- Given two events $A$ and $B$ if $A$ and $B$ are independent then what is the probability of their intersection ($P(A \cap B)$)?
- If I have three events $B_1, B_2, B_3$ and I want to know if the three events are independent what must hold for this to be true?
Counting Techniques - Overview

Determining probabilities often involves counting the number of outcomes for various events. Counting techniques help with this process. Given a sample space $\Omega$ with a finite number of equally likely outcomes, if we want to determine the probability of an event $A$, then:

$$P(A) = \frac{\text{# of elements in } A}{\text{# of elements in } \Omega}$$

Here I am going to cover only three techniques:

- Fundamental counting principle
- Permutations
- Combinations

Why do we care?
Counting Techniques - Overview

- Determining probabilities often involves counting the number of outcomes for various events
Counting Techniques - Overview

- Determining probabilities often involves counting the number of outcomes for various events
- Counting techniques help with this process
Counting Techniques - Overview

- Determining probabilities often involves counting the number of outcomes for various events.
- Counting techniques help with this process.
- Given a sample space $\Omega$ with a finite number equally likely outcomes if we want to determine the probability of an event $A$ then:

\[
P(A) = \frac{\# \text{ of elements in } A}{\# \text{ of elements in } \Omega}
\]
Counting Techniques - Overview

- Determining probabilities often involves counting the number of outcomes for various events.
- Counting techniques help with this process.
- Given a sample space \( \Omega \) with a finite number equally likely outcomes if we want to determine the probability of an event \( A \) then:

\[
P(A) = \frac{\# \text{ of elements in } A}{\# \text{ of elements in } \Omega}
\]

- Here I am going to cover only three techniques.
Counting Techniques - Overview

- Determining probabilities often involves counting the number of outcomes for various events
- Counting techniques help with this process
- Given a sample space $\Omega$ with a finite number equally likely outcomes if we want to determine the probability of an event $A$ then:

$$P(A) = \frac{\# \text{ of elements in } A}{\# \text{ of elements in } \Omega}$$

- Here I am going to cover only three techniques
  - Fundamental counting principle
Counting Techniques - Overview

- Determining probabilities often involves counting the number of outcomes for various events.
- Counting techniques help with this process.
- Given a sample space $\Omega$ with a finite number equally likely outcomes if we want to determine the probability of an event $A$ then:

$$ P(A) = \frac{\# \text{ of elements in } A}{\# \text{ of elements in } \Omega} $$

- Here I am going to cover only three techniques:
  - Fundamental counting principle
  - permutations
Counting Techniques - Overview

- Determining probabilities often involves counting the number of outcomes for various events.
- Counting techniques help with this process.
- Given a sample space $\Omega$ with a finite number equally likely outcomes if we want to determine the probability of an event $A$ then:

$$P(A) = \frac{\# \text{ of elements in } A}{\# \text{ of elements in } \Omega}$$

- Here I am going to cover only three techniques:
  - Fundamental counting principle
  - permutations
  - combinations
- Why do we care?
Fundamental Counting Principle

Consider a process that has \( r \) stages. For the first stage, there are \( n_1 \) possible outcomes. For every possible outcome at the first stage, there are \( n_2 \) outcomes at the second stage. This generalizes to an arbitrary number of stages where the total number of outcomes of the \( r \)-stage process is \( n_1 \times n_2 \times \cdots \times n_r \).

For example, assume we are flipping a fair coin 3 times. Here \( r = 3 \), \( n_1 = 2 \), \( n_2 = 2 \), and \( n_3 = 2 \). Using the fundamental counting principle, the total number of outcomes is \( 2 \times 2 \times 2 = 8 \).

If you don't believe me, here is the set:

\[
\{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT\}
\]
Consider a process that has \( r \) stages
Fundamental Counting Principle

- Consider a process that has \( r \) stages.
- For the first stage there are \( n_1 \) possible outcomes.
Fundamental Counting Principle

- Consider a process that has \( r \) stages
- For the first stage there are \( n_1 \) possible outcomes
- For every possible outcome at \( n_1 \) there are \( n_2 \) outcomes at the second stage
Consider a process that has $r$ stages
- For the first stage there are $n_1$ possible outcomes
- For every possible outcome at $n_1$ there are $n_2$ outcomes at the second stage
- This generalizes to an arbitrary number of stages where the total number of outcomes of the $r$-stage process is $n_1 \times n_2 \times \cdots \times n_r$
Consider a process that has \( r \) stages

For the first stage there are \( n_1 \) possible outcomes

For every possible outcome at \( n_1 \) there are \( n_2 \) outcomes at the second stage

This generalizes to an arbitrary number of stages where the total number of outcomes of the \( r \)-stage process is \( n_1 \times n_2 \times \cdots \times n_r \)

For example assume we are flipping a fair coin 3 times
Fundamental Counting Principle

- Consider a process that has \( r \) stages
- For the first stage there are \( n_1 \) possible outcomes
- For every possible outcome at \( n_1 \) there are \( n_2 \) outcomes at the second stage
- This generalizes to an arbitrary number of stages where the total number of outcomes of the \( r \)-stage process is \( n_1 \times n_2 \times \cdots \times n_r \)
- For example assume we are flipping a fair coin 3 times
- Here \( r = 3, \ n_1 = 2, \ n_2 = 2, \ n_3 = 2 \)
Fundamental Counting Principle

- Consider a process that has $r$ stages
- For the first stage there are $n_1$ possible outcomes
- For every possible outcome at $n_1$ there are $n_2$ outcomes at the second stage
- This generalizes to an arbitrary number of stages where the total number of outcomes of the $r$-stage process is $n_1 \times n_2 \times \cdots \times n_r$
- For example assume we are flipping a fair coin 3 times
- Here $r = 3$, $n_1 = 2$, $n_2 = 2$, $n_3 = 2$
- Using the fundamental counting principle the total number of outcomes is $2 \times 2 \times 2 = 8$
Consider a process that has $r$ stages
For the first stage there are $n_1$ possible outcomes
For every possible outcome at $n_1$ there are $n_2$ outcomes at the second stage
This generalizes to an arbitrary number of stages where the total number of outcomes of the $r$-stage process is $n_1 \times n_2 \times \cdots \times n_r$
For example assume we are flipping a fair coin 3 times
Here $r = 3$, $n_1 = 2$, $n_2 = 2$, $n_3 = 2$
Using the fundamental counting principle the total number of outcomes is $2 \times 2 \times 2 = 8$
If you don’t believe me here is the set:
$\{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT\}$
Fundamental Counting Principle - Illustration

From before assuming a 3 stage process with two outcomes:

{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}
Fundamental Counting Principle - Illustration

- From before assuming a 3 stage process with two outcomes
Fundamental Counting Principle - Illustration

- From before assuming a 3 stage process with two outcomes

From previous slide:

\{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}
Fundamental Counting Principle - Illustration

From before assuming a 3 stage process with two outcomes:

{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}

From previous slide:
Permutations - Overview

Permutations involve the use of factorials the $n!$ operator:

$$n! = n \times (n-1) \times \cdots \times 1$$

For example, $3! = 3 \times 2 \times 1 = 6$. Note $0! = 1$.

If we have $n$ distinct objects and we want to count the number of ways we can pick out $k$ out of $n$ objects ($k \leq n$) and arrange them in distinct $k$-object sequences, the formula for $k$-permutations is:

$$\frac{n!}{(n-k)!}$$

Let's take a simple example... What if I want to determine the number of two person pairs from Bob, Alice, and Eve assuming order matters.

$n = 3$, $k = 2$, $3! (3-2)! = 3 \times 2 \times 1 = 6$

$\{ (B, A); (A, B); (B, E); (E, B); (A, E); (E, A) \}$
Permutations - Overview

- Permutations involve the use of factorials the ! operator

\[ n! = n \times (n-1) \times \cdots \times 1 \]

For example, to find the number of two-person pairs from Bob, Alice, and Eve assuming order matters:

\[ n = 3, \quad k = 2, \quad 3! \cdot (3-2)! = 3 \times 2 \times 1 = 6 \]

\{ (B, A), (A, B), (B, E), (E, B), (A, E), (E, A) \}
Permutations - Overview

- Permutations involve the use of factorials the ! operator
- \( n! = n \times (n - 1) \times \cdots \times 1 \) or \( 3! = 3 \times 2 \times 1 = 6 \). Note \( 0! = 1 \)
Permutations - Overview

- Permutations involve the use of factorials the ! operator
- $n! = n \cdot (n-1) \cdot \cdots \cdot 1$ or $3! = 3 \cdot 2 \cdot 1 = 6$. Note $0! = 1$
- If we $n$ distinct objects and we want to count the number of ways we can pick out $k$ out of $n$ objects ($k \leq n$) and arrange them in distinct $k$-object sequences.

\[
\text{The formula for } k\text{-permutations is } \frac{n!}{(n-k)!}.
\]

Let's take a simple example... What if I want to determine the number of two person pairs from Bob, Alice, and Eve assuming order matters.

$n = 3$, $k = 2$, $3! \cdot (3-2)! = 3 \cdot 2 \cdot 1 = 6$

\[
\text{\{(B,A); (A,B); (B,E); (E,B); (A,E); (E,A)\}}
\]
Permutations - Overview

- Permutations involve the use of factorials the ! operator
- $n! = n \times (n - 1) \times \cdots \times 1$ or $3! = 3 \times 2 \times 1 = 6$. Note $0! = 1$
- If we $n$ distinct objects and we want to count the number of ways we can pick out $k$ out of $n$ objects ($k \leq n$) and arrange them in distinct $k$-object sequences.
- The formula for $k$-permutations is

$$\frac{n!}{(n - k)!}$$
Permutations - Overview

- Permutations involve the use of factorials the ! operator
- \( n! = n \times (n - 1) \times \cdots \times 1 \) or \( 3! = 3 \times 2 \times 1 = 6 \). Note \( 0! = 1 \)
- If we \( n \) distinct objects and we want to count the number of ways we can pick out \( k \) out of \( n \) objects \( (k \leq n) \) and arrange them in distinct \( k \)-object sequences.
- The formula for \( k \)-permutations is
  \[
  \frac{n!}{(n - k)!}
  \]
- Let’s take a simple example...What if I want to determine the number of two person pairs from Bob, Alice, and Eve assuming order matters.
Permutations - Overview

- Permutations involve the use of factorials the ! operator
- \( n! = n \times (n - 1) \times \cdots \times 1 \) or \( 3! = 3 \times 2 \times 1 = 6 \). Note \( 0! = 1 \)
- If we \( n \) distinct objects and we want to count the number of ways we can pick out \( k \) out of \( n \) objects \((k \leq n)\) and arrange them in distinct \( k \)-object sequences.
- The formula for \( k \)-permutations is
  \[
  \frac{n!}{(n-k)!}
  \]
- Let’s take a simple example...What if I want to determine the number of two person pairs from Bob, Alice, and Eve assuming order matters.
  - \( n = 3, \ k = 2, \ \frac{3!}{(3-2)!} = \frac{3 \times 2 \times 1}{1} = 6 \)
Permutations - Overview

- Permutations involve the use of factorials the ! operator
- $n! = n \times (n - 1) \times \cdots \times 1$ or $3! = 3 \times 2 \times 1 = 6$. Note $0! = 1$
- If we $n$ distinct objects and we want to count the number of ways we can pick out $k$ out of $n$ objects ($k \leq n$) and arrange them in distinct $k$-object sequences.
- The formula for $k$-permutations is
  \[
  \frac{n!}{(n - k)!}
  \]
- Let’s take a simple example...What if I want to determine the number of two person pairs from Bob, Alice, and Eve assuming order matters.
  - $n = 3, k = 2, \frac{3!}{(3-2)!} = \frac{3 \times 2 \times 1}{1} = 6$
  - $\{(B, A); (A, B); (B, E); (E, B); (A, E); (E, A)\}$
Combinations - Overview

From the previous slide you might have thought I had some screws loose in my head by assuming that order matters. Combinations are similar to permutations with one big exception: Order Does Not Matter!

Let's take the example from the previous slide forming two person pairs from Bob, Alice, and Eve. Assuming order does not matter in this simple case there are three combinations:

\{(B, A); (B, E); (A, E)\}

The formula for combinations is sometimes called the binomial coefficient and given as:

\[^nC_k = \frac{n!}{k! \cdot (n-k)!}\]

For the above example:

\[3! \cdot 2! \cdot (3-2)! = 3 \cdot 2 \cdot 1 \cdot 2 \cdot 1 \cdot 1 = 3\]
Combinations - Overview

- From the previous slide you might have thought I had some screws loose in my head by assuming that order matters
Combinations - Overview

- From the previous slide you might have thought I had some screws loose in my head by assuming that order matters.
- Combinations are similar to permutations with one big exception Order Does Not Matter!

The formula for combinations is sometimes called the binomial coefficient and given as:

\[ \binom{n}{k} = \frac{n!}{k! \cdot (n-k)!} \]

For the above example:

\[ \frac{3!}{2! \cdot (3-2)!} = \frac{3 \cdot 2 \cdot 1}{2 \cdot 1 \cdot 1} = 3 \]
Combinations - Overview

- From the previous slide you might have thought I had some screws loose in my head by assuming that order matters.
- Combinations are similar to permutations with one big exception Order Does Not Matter!
- Let’s take the example from the previous slide forming two person pairs from Bob, Alice, and Eve. Assuming order does not matter in this simple case there are three combinations \{ (B, A); (B, E); (A, E) \}
Combinations - Overview

- From the previous slide you might have thought I had some screws loose in my head by assuming that order matters.
- Combinations are similar to permutations with one big exception: Order Does Not Matter!
- Let’s take the example from the previous slide forming two person pairs from Bob, Alice, and Eve. Assuming order does not matter in this simple case there are three combinations \[\{(B, A); (B, E); (A, E)\}\]
- The formula for combinations is sometimes called the binomial coefficient and given as:

\[
\binom{n}{k} = \frac{n!}{k! \times (n-k)!}
\]
Combinations - Overview

- From the previous slide you might have thought I had some screws loose in my head by assuming that order matters.
- Combinations are similar to permutations with one big exception: Order Does Not Matter!
- Let’s take the example from the previous slide forming two person pairs from Bob, Alice, and Eve. Assuming order does not matter in this simple case there are three combinations \{ (B, A); (B, E); (A, E) \}
- The formula for combinations is sometimes called the binomial coefficient and given as:

\[
\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}
\]

- For the above example, \( \frac{3!}{2! \cdot (3-2)!} = \frac{3 \cdot 2 \cdot 1}{2 \cdot 1 \cdot 1} = 3 \)
Learning Check - Counting

- If I have 12 letters from the alphabet what is the total number of ways to form triplets of letters from the 12 letters? (Hint: $n = 12$ and $k = 3$, order matters)
- If I have 5 numbered chairs and want to know the number of combinations of 2 chairs how do I find this? (Hint: $n = 5$, $k = 2$, order does not matter)
New Section
What is a Random Variable

When dealing with probabilistic models, the outcomes are numerical or can be associated with numerical values. Examples include:

- Stock prices (numerical)
- Pass/Fail or Success/Failure (could be coded as 1 for pass and 0 for fail)

When dealing with these numerical outcomes, it is often useful to assign probabilities to them. This is done through the notion of a random variable. The idea of a random variable is abstract:

A random variable is a function that maps every possible event from the sample space (\( \Omega \)) to the real line:

\[ X : \Omega \rightarrow \mathbb{R} \]

For example, if we toss a fair coin two times, then the sample space \( \Omega = \{ HH, HT, TH, TT \} \). We can define a random variable as:

\[ X(\text{HH}) = 0, \quad X(\text{HT}) = X(\text{TH}) = 1, \quad X(\text{HH}) = 2 \]
What is a Random Variable

- When dealing with probabilistic models the outcomes are numerical or can be associated with numerical values
What is a Random Variable

- When dealing with probabilistic models the outcomes are numerical or can be associated with numerical values
  - Stock prices (numerical)
What is a Random Variable

When dealing with probabilistic models the outcomes are numerical or can be associated with numerical values:
- Stock prices (numerical)
- Pass/Fail or Success/Failure (could be coded as 1 for pass and 0 for fail)
What is a Random Variable

- When dealing with probabilistic models the outcomes are numerical or can be associated with numerical values
  - Stock prices (numerical)
  - Pass/Fail or Success/Failure (could be coded as 1 for pass and 0 for fail)
- When dealing with these numerical outcomes it is often useful to assign probabilities to them
What is a Random Variable

- When dealing with probabilistic models the outcomes are numerical or can be associated with numerical values
  - Stock prices (numerical)
  - Pass/Fail or Success/Failure (could be coded as 1 for pass and 0 for fail)
- When dealing with these numerical outcomes it is often useful to assign probabilities to them
- This is done through the notion of a random variable. The idea of a random variable is abstract
What is a Random Variable

- When dealing with probabilistic models the outcomes are numerical or can be associated with numerical values
  - Stock prices (numerical)
  - Pass/Fail or Success/Failure (could be coded as 1 for pass and 0 for fail)
- When dealing with these numerical outcomes it is often useful to assign probabilities to them
- This is done through the notion of a random variable. The idea of a random variable is abstract
- A random variable is a function that maps every possible event from the sample space ($\Omega$) to the real line $X : \Omega \rightarrow R$
What is a Random Variable

- When dealing with probabilistic models the outcomes are numerical or can be associated with numerical values
  - Stock prices (numerical)
  - Pass/Fail or Success/Failure (could be coded as 1 for pass and 0 for fail)
- When dealing with these numerical outcomes it is often useful to assign probabilities to them
- This is done through the notion of a random variable. The idea of a random variable is abstract
- A random variable is a function that maps every possible event from the sample space (\( \Omega \)) to the real line \( X : \Omega \rightarrow \mathbb{R} \)
- For example if we toss a fair coin two times then the sample space \( \Omega = \{HH, HT, TH, TT\} \)
What is a Random Variable

- When dealing with probabilistic models the outcomes are numerical or can be associated with numerical values
  - Stock prices (numerical)
  - Pass/Fail or Success/Failure (could be coded as 1 for pass and 0 for fail)
- When dealing with these numerical outcomes it is often useful to assign probabilities to them
- This is done through the notion of a random variable. The idea of a random variable is abstract
- A random variable is a function that maps every possible event from the sample space (\(\Omega\)) to the real line \(X : \Omega \to \mathbb{R}\)
- For example if we toss a fair coin two times then the sample space \(\Omega = \{HH, HT, TH, TT\}\)
- We can define a random variable as \(X(HH) = 0, X(HT) = X(TH) = 1, X(HH) = 2\)
What is a Random Variable Cont...
Let’s take another Example Rolling a 4-sided fair die. Sample space (Ω) is \{1, 2, 3, 4\}
What is a Random Variable Contin...

- Let’s take another Example Rolling a 4-sided fair die. Sample space ($\Omega$) is {1, 2, 3, 4}
- We can define a random variable $X(j) = j$ for $j = 1, 2, 3, 4$

---

1Problem from Bertsekas and Tsitsiklis (2008, p. 72)
Let’s take another Example Rolling a 4-sided fair die. Sample space (Ω) is {1, 2, 3, 4}
We can define a random variable \( X(j) = j \) for \( j = 1, 2, 3, 4 \)
Here is an illustration of the idea of a random variable\(^1\):

\(^1\)Problem from Bertsekas and Tsitsiklis (2008, p. 72)
Let’s take another Example Rolling a 4-sided fair die. Sample space ($\Omega$) is \{1, 2, 3, 4\}
We can define a random variable $X(j) = j$ for $j = 1, 2, 3, 4$
Here is an illustration of the idea of a random variable:

1Problem from Bertsekas and Tsitsiklis (2008, p. 72)
Why should you care?

- Random variables take into account uncertainty
- If we know the process then we have a deterministic function
- It is often unreasonable to assume that you would know why something behaves a certain way all the time
Random variable is a real valued function of the outcome of the experiment.
A function of a random variable defines another random variable.
We can associate with each random variable certain averages of interest (e.g., Mean and Variance).
A random variable can be conditioned upon an event or another random variable.
The notion of independence of a random variable from an event or another random variable applies.
Random variables can be continuous or discrete. Here we only talk about the discrete case in detail.

Number of heads in a sequence of 5 coin tosses
Sum of two rolls of a 6-sided dice
Properties of Random Variables

- Random variable is a real valued function of the outcome of the experiment.
Properties of Random Variables

- Random variable is a real valued function of the outcome of the experiment
- A function of a random variable defines another random variable
Properties of Random Variables

- Random variable is a real valued function of the outcome of the experiment
- A function of a random variable defines another random variable
- We can associate with each random variable certain averages of interest (e.g. Mean and Variance)
Properties of Random Variables

- Random variable is a real valued function of the outcome of the experiment
- A function of a random variable defines another random variable
- We can associate with each random variable certain averages of interest (e.g. Mean and Variance)
- A random variable can be conditioned upon an event or another random variable
Properties of Random Variables

- Random variable is a real valued function of the outcome of the experiment
- A function of a random variable defines another random variable
- We can associate with each random variable certain averages of interest (e.g. Mean and Variance)
- A random variable can be conditioned upon an event or another random variable
- The notion of independence of a random variable from an event or another random variable applies
Properties of Random Variables

- Random variable is a real valued function of the outcome of the experiment
- A function of a random variable defines another random variable
- We can associate with each random variable certain averages of interest (e.g. Mean and Variance)
- A random variable can be conditioned upon an event or another random variable
- The notion of independence of a random variable from an event or another random variable applies
- Random variables can be continuous or discrete. Here we only talk about the discrete case in detail
Properties of Random Variables

- Random variable is a real valued function of the outcome of the experiment.
- A function of a random variable defines another random variable.
- We can associate with each random variable certain averages of interest (e.g. Mean and Variance).
- A random variable can be conditioned upon an event or another random variable.
- The notion of independence of a random variable from an event or another random variable applies.
- Random variables can be continuous or discrete. Here we only talk about the discrete case in detail.
  - Number of heads in a sequence of 5 coin tosses.
  - Sum of two rolls of a 6-sided dice.
Properties of Random Variables

- Random variable is a real valued function of the outcome of the experiment
- A function of a random variable defines another random variable
- We can associate with each random variable certain averages of interest (e.g. Mean and Variance)
- A random variable can be conditioned upon an event or another random variable
- The notion of independence of a random variable from an event or another random variable applies
- Random variables can be continuous or discrete. Here we only talk about the discrete case in detail
  - Number of heads in a sequence of 5 coin tosses
  - Sum of two rolls of a 6-sided dice
Discrete Random Variables - Overview

Discrete random variables are real valued functions of the outcome of the experiment that can take a finite or countably infinite number of values.

Discrete random variables have an associated probability mass function (PMF) which gives the probability that the random variable can take.

Discrete random variables have a cumulative mass function (CMF) which gives us the accumulated probability of a value up to the value of the random variable.

A function of a discrete random variable defines another random variable whose PMF can be obtained from the original random variable.
Discrete Random Variables - Overview

- Discrete random variables are real valued functions of the outcome of the experiment that can take a finite or countably infinite number of values.
Discrete Random Variables - Overview

- Discrete random variables are real valued functions of the outcome of the experiment that can take a finite or countably infinite number of values.
- Discrete random variables have an associated probability mass function (PMF) which gives the probability that the random variable can take.
Discrete random variables are real valued functions of the outcome of the experiment that can take a finite or countably infinite number of values.

Discrete random variables have an associated probability mass function (PMF) which gives the probability that the random variable can take.

Discrete random variables have a cumulative mass function (CMF) which gives us the accumulated probability of a value up to the value of the random variable $X$. 

A function of a discrete random variable defines another random variable whose PMF can be obtained from the original random variable.
Discrete Random Variables - Overview

- Discrete random variables are real valued functions of the outcome of the experiment that can take a finite or countably infinite number of values
- Discrete random variables have an associated probability mass function (PMF) which gives the probability that the random variable can take
- Discrete random variables have a cumulative mass function (CMF) which gives us the accumulated probability of a value up to the value of the random variable $X$
- A function of a discrete random variable defines another random variable whose PMF can be obtained from the original random variable
Probability Mass Functions

PMF is a function that tells us the probability of $x$, which is an observation of $X$. Given by $P(X = x)$.

Let's take an example... Assume an experiment of two tosses of a fair coin and let $X$ be the number of heads. Sample space $\Omega = \{HH, TH, HT, TT\}$

Mathematically this can be given as:

$$P(X = x) = \begin{cases} 
\frac{1}{4} & \text{if } x = 0 \text{ or } x = 2 \\
\frac{1}{2} & \text{if } x = 1 \\
0 & \text{otherwise}
\end{cases}$$

Note that:

$$\sum_{x=0}^{2} P(X = x) = 1$$

$$P(X > 0) = P(X = 1) + P(X = 2) = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}$$
Probability Mass Functions

- PMF is a function that tells us the probability of \( x \) which is an observation of \( X \). Given by \( P(X = x) \)
Probability Mass Functions

- PMF is a function that tells us the probability of \( x \) which is an observation of \( X \). Given by \( P(X = x) \)
- Let’s take an example...Assume an experiment of two tosses of a fair coin and let \( X \) be the number of heads.
Probability Mass Functions

- PMF is a function that tells us the probability of \( x \) which is an observation of \( X \). Given by \( P(X = x) \)
- Let’s take an example...Assume an experiment of two tosses of a fair coin and let \( X \) be the number of heads.
  - Sample space \( \Omega = \{HH, TH, HT, TT\} \)
Probability Mass Functions

- PMF is a function that tells us the probability of \( x \) which is an observation of \( X \). Given by \( P(X = x) \)
- Let's take an example...Assume an experiment of two tosses of a fair coin and let \( X \) be the number of heads.
  - Sample space \( \Omega = \{HH, TH, HT, TT\} \)
- Mathematically this can be given as:
  
  \[
  P(X = x) = \begin{cases} 
  1/4 & \text{if } x = 0 \text{ or } x = 2 \\
  1/2 & \text{if } x = 1 \\
  0 & \text{0 otherwise}
  \end{cases}
  \]
**Probability Mass Functions**

- PMF is a function that tells us the probability of $x$ which is an observation of $X$. Given by $P(X = x)$
- Let’s take an example...Assume an experiment of two tosses of a fair coin and let $X$ be the number of heads.
  - Sample space $\Omega = \{HH, TH, HT, TT\}$
- Mathematically this can be given as:
  $$P(X = x) = \begin{cases} 
1/4 & \text{if } x = 0 \text{ or } x = 2 \\
1/2 & \text{if } x = 1 \\
0 & \text{0 otherwise}
\end{cases}$$
- Note that $\sum_{x=0}^{2} P(X = x) = 1$
Probability Mass Functions

- PMF is a function that tells us the probability of $x$ which is an observation of $X$. Given by $P(X = x)$
- Let’s take an example...Assume an experiment of two tosses of a fair coin and let $X$ be the number of heads.
  - Sample space $\Omega = \{HH, TH, HT, TT\}$
- Mathematically this can be given as:
  
  $$P(X = x) = \begin{cases} 
  1/4 & \text{if } x = 0 \text{ or } x = 2 \\
  1/2 & \text{if } x = 1 \\
  0 & \text{otherwise}
  \end{cases}$$

- Note that $\sum_{x=0}^{2} P(X = x) = 1$
  
  $$P(X = 0) + P(X = 1) + P(X = 2) = 1/4 + 1/2 + 1/4 = 1$$
PMF is a function that tells us the probability of $x$ which is an observation of $X$. Given by $P(X = x)$

Let’s take an example...Assume an experiment of two tosses of a fair coin and let $X$ be the number of heads.

- Sample space $\Omega = \{HH, TH, HT, TT\}$
- Mathematically this can be given as:

$$P(X = x) = \begin{cases} 
1/4 & \text{if } x = 0 \text{ or } x = 2 \\
1/2 & \text{if } x = 1 \\
0 & \text{0 otherwise}
\end{cases}$$

- Note that $\sum_{x=0}^{2} P(X = x) = 1$
  - $P(X = 0) + P(X = 1) + P(X = 2) = 1/4 + 1/2 + 1/4 = 1$
  - $P(X > 0) = P(X = 1) + P(X = 2) = 1/2 + 1/4 = 3/4$
Probability Mass Function Example

From the previous slide:

\[ P(X = x) = \begin{cases} 1/4 & \text{if } x = 0 \text{ or } x = 2 \\ 1/2 & \text{if } x = 1 \\ 0 & \text{otherwise} \end{cases} \]
Probability Mass Function Example

- From the previous slide:

\[ P(X = x) = \begin{cases} 
1/4 & \text{if } x = 0 \text{ or } x = 2 \\
1/2 & \text{if } x = 1 \\
0 & \text{0 otherwise}
\end{cases} \]
Probability Mass Function Example

- From the previous slide:

\[ P(X = x) = \begin{cases} 
1/4 & \text{if } x = 0 \text{ or } x = 2 \\
1/2 & \text{if } x = 1 \\
0 & \text{otherwise}
\end{cases} \]

- Can be illustrated as:
From the previous slide:

\[ P(X = x) = \begin{cases} 
  1/4 & \text{if } x = 0 \text{ or } x = 2 \\
  1/2 & \text{if } x = 1 \\
  0 & \text{0 otherwise} 
\end{cases} \]

Can be illustrated as:
A cumulative mass function (CMF) accumulates probability up to the value of $x$. Take the previous example: Assume an experiment of two tosses of a fair coin and let $X$ be the number of heads. The sample space $\Omega = \{HH, TH, HT, TT\}$.

The CMF is given as $P(X \leq x) = \sum_{k \leq x} P(k)$.

<table>
<thead>
<tr>
<th>Number of Heads</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>1.0</td>
</tr>
</tbody>
</table>

CMF of 2 Heads $P = 0.50$
A cumulative mass function (CMF) accumulates probability up to the value of $x$. 
A cumulative mass function (CMF) accumulates probability up to the value of $x$.

Take the previous example: Assume an experiment of two tosses of a fair coin and let $X$ be the number of heads.
A cumulative mass function (CMF) accumulates probability up to the value of $x$.

Take the previous example: Assume an experiment of two tosses of a fair coin and let $X$ be the number of heads.

Sample space $\Omega = \{HH, TH, HT, TT\}$
Cumulative Mass Function

- A cumulative mass function (CMF) accumulates probability up to the value of $x$.
- Take the previous example: Assume an experiment of two tosses of a fair coin and let $X$ be the number of heads.
  - Sample space $\Omega = \{HH, TH, HT, TT\}$
- The CMF is given as $P(X \leq x) = \sum_{k \leq x} P(k)$
A cumulative mass function (CMF) accumulates probability up to the value of $x$.

Take the previous example: Assume an experiment of two tosses of a fair coin and let $X$ be the number of heads.

- Sample space $\Omega = \{HH, TH, HT, TT\}$
- The CMF is given as $P(X \leq x) = \sum_{k \leq x} P(k)$
We often want to summarize the information for a given PMF by a single number. The expectation of the random variable $X$ is a weighted average of all possible values of $X$.

Let's take an example. Assume that we flip one fair coin and let $X$ be the payout on getting a head. If the coin is heads you get $40 and if tails you get $1. What is the expected value of heads?

$$E(X) = 0.5 \times 40 + 0.5 \times 1 = 20.50$$

Now you try assume coin is not fair $P(\{H\}) = 0.25$ and $P(\{T\}) = 0.75$. What is the expected value of heads?

$$E(X) = 0.25 \times 40 + 0.75 \times 1 = 10.75$$
We often want to summarize the information for a given PMF by a single number.

$$E(X) = 0.5 \times 40 + 0.5 \times 1 = 20.50$$

Now you try assume coin is not fair $P(\{H\}) = 0.25$ and $P(\{T\}) = 0.75$. What is the expected value of heads?

$$E(X) = 0.25 \times 40 + 0.75 \times 1 = 10.75$$
Expectation and Mean

- We often want to summarize the information for a given PMF by a single number.
- The expectation of the random variable $X$.
Expectation and Mean

- We often want to summarize the information for a given PMF by a single number.
- The expectation of the random variable $X$
  - The expectation of $X$ is a weighted average of all possible values of $X$.

Let's take an example. Assume that we flip one fair coin and let $X$ be the payout on getting a head. If the coin is heads you get $40 and if tails you get $1. What is the expected value of heads?

$$E(X) = 0.5 \times 40 + 0.5 \times 1 = 20.50$$

Now you try assume coin is not fair $P(\{H\}) = 0.25$ and $P(\{T\}) = 0.75$. What is the expected value of heads?

$$E(X) = 0.25 \times 40 + 0.75 \times 1 = 10.75$$
We often want to summarize the information for a given PMF by a single number.

The expectation of the random variable $X$ is a weighted average of all possible values of $X$.

Let’s take an example. Assume that we flip one fair coin and let $X$ be the payout on getting a head.
Expectation and Mean

- We often want to summarize the information for a given PMF by a single number.
- The expectation of the random variable $X$ is a weighted average of all possible values of $X$.
- Let’s take an example. Assume that we flip one fair coin and let $X$ be the payout on getting a head.
- If the coin is heads you get $40 and if tails you get $1.
Expectation and Mean

- We often want to summarize the information for a given PMF by a single number.
- The expectation of the random variable $X$
  - The expectation of $X$ is a weighted average of all possible values of $X$.
- Let’s take an example. Assume that we flip one fair coin and let $X$ be the payout on getting a head.
- If the coin is heads you get $40 and if tails you get $1.
- What is the expected value of heads?

$$E(X) = 0.5 \times 40 + 0.5 \times 1 = 20.5$$
We often want to summarize the information for a given PMF by a single number.

The expectation of the random variable $X$ is a weighted average of all possible values of $X$.

Let's take an example. Assume that we flip one fair coin and let $X$ be the payout on getting a head.

If the coin is heads you get $40 and if tails you get $1.

What is the expected value of heads?

$$E(X) = .5 \times 40 + .5 \times 1 = 20.50$$
We often want to summarize the information for a given PMF by a single number.

The expectation of the random variable $X$

- The expectation of $X$ is a weighted average of all possible values of $X$.

Let’s take an example. Assume that we flip one fair coin and let $X$ be the payout on getting a head.

If the coin is heads you get $40 and if tails you get $1.

What is the expected value of heads?

- $\mathbb{E}(X) = .5 \times 40 + .5 \times 1 = 20.50$

Now you try assume coin is not fair $P(\{H\}) = .25$ and $P(\{T\}) = .75$. What is the expected value of heads?
We often want to summarize the information for a given PMF by a single number.

The expectation of the random variable $X$

- The expectation of $X$ is a weighted average of all possible values of $X$.

Let’s take an example. Assume that we flip one fair coin and let $X$ be the payout on getting a head.

If the coin is heads you get $40 and if tails you get $1.

What is the expected value of heads?

- $E(X) = .5 \times 40 + .5 \times 1 = 20.50$

Now you try assume coin is not fair $P(\{H\}) = .25$ and $P(\{T\}) = .75$. What is the expected value of heads?
Expectation and Mean

- We often want to summarize the information for a given PMF by a single number.
- The expectation of the random variable $X$
  - The expectation of $X$ is a weighted average of all possible values of $X$.
- Let’s take an example. Assume that we flip one fair coin and let $X$ be the payout on getting a head.
- If the coin is heads you get $40 and if tails you get $1.
- What is the expected value of heads?
  - $E(X) = .5 \times 40 + .5 \times 1 = 20.50$
- Now you try assume coin is not fair $P(\{H\}) = .25$ and $P(\{T\}) = .75$. What is the expected value of heads?
  - $E(X) = .25 \times 40 + .75 \times 1 = 10.75$
The expected value or the mean of a discrete random variable $X$ is formally defined as

$$E(X) = \sum x P(X = x)$$

Another way to visualize the expected value is by viewing it as the center of gravity of the PMF. The expectation of $X$ is a weighted average of all possible values of $X$. 

\footnote{Image from Bertsekas and Tsitsiklis (2008, p. 83)}
The expected value or the mean of a discrete random variable $X$ is formally defined as

$$\mathbb{E}(X) = \sum_{x} x P(X = x)$$
The expected value or the mean of a discrete random variable $X$ is formally defined as

$$\mathbb{E}(X) = \sum_{x} x P(X = x)$$

Another way to visualize the expected value is by viewing it as the center of gravity of the PMF.
The expected value or the mean of a discrete random variable $X$ is formally defined as

$$\mathbb{E}(X) = \sum_x xP(X = x)$$

Another way to visualize the expected value is by viewing it as the center of gravity of the PMF.

The expectation of $X$ is a weighted average of all possible values of $X$.

---

\(^1\)Image from Bertsekas and Tsitsiklis (2008, p. 83)
The expected value or the mean of a discrete random variable $X$ is formally defined as

$$\mathbb{E}(X) = \sum_x xP(X = x)$$

Another way to visualize the expected value is by viewing it as the center of gravity of the PMF.

The expectation of $X$ is a weighted average of all possible values of $X$.

\[\text{Center of gravity}
\]
\[c = \text{mean} = \mathbb{E}[X]\]

\[\text{Image from Bertsekas and Tsitsiklis (2008, p. 83)}\]
Let's take another example for expectation. Assume that we are playing a dice game with a 4 sided fair dice. The payouts are as follows:

- roll of a 1 or 4 = $10
- roll of a 2 or 3 = $5

Let $X$ be the expected value of the payout. What is $E(X)$?

**Hint:**

$$E(X) = \sum x P(X = x)$$

Since $P(X = x) = 1/4$ for $x = 1, 2, 3, 4$,

$$E(X) = \frac{1}{4} \times 10 + \frac{1}{4} \times 5 + \frac{1}{4} \times 5 + \frac{1}{4} \times 10 = \$7.50$$
Let’s take another example for expectation
Let’s take another example for expectation.

Assume that we are playing a dice game with a 4 sided fair dice.

The payouts are as follows:
- roll of a 1 or 4 = $10
- roll of a 2 or 3 = $5

Let $X$ be the expected value of the payout. What is $E(X)$?

Hint: $E(X) = \sum x P(X=x)$ and $P(X=x) = 1/4$, for $x = 1, 2, 3, 4$.

$E(X) = 0.25 \times 10 + 0.25 \times 5 + 0.25 \times 5 + 0.25 \times 10 = 7.5$. 
Expectation and Mean Ex.

- Let’s take another example for expectation
- Assume that we are playing a dice game with a 4 sided fair dice.
- The payouts are as follows:

\[
E(X) = \sum x \cdot P(X = x) = \frac{1}{4} \cdot 10 + \frac{1}{4} \cdot 5 + \frac{1}{4} \cdot 5 + \frac{1}{4} \cdot 10 = \frac{25}{4} = 6.25
\]
Let’s take another example for expectation

Assume that we are playing a dice game with a 4 sided fair dice.

The payouts are as follows:

- roll of a 1 or 4 = $10
Let’s take another example for expectation

Assume that we are playing a dice game with a 4 sided fair dice.

The payouts are as follows:

- roll of a 1 or 4 = $10
- roll of a 2 or 3 = $5
Let’s take another example for expectation

Assume that we are playing a dice game with a 4 sided fair dice.

The payouts are as follows:

- roll of a 1 or 4 = $10
- roll of a 2 or 3 = $5

Let $X$ be the expected value of the payout. What is $\mathbb{E}(X)$?
Let’s take another example for expectation

Assume that we are playing a dice game with a 4 sided fair dice.

The payouts are as follows:
- roll of a 1 or 4 = $10
- roll of a 2 or 3 = $5

Let $X$ be the expected value of the payout. What is $\mathbb{E}(X)$?

Hint: $\mathbb{E}(X) = \sum_x xP(X = x)$ and $P(X = x) = 1/4, x = 1, 2, 3, 4$
Let’s take another example for expectation

Assume that we are playing a dice game with a 4 sided fair dice.

The payouts are as follows:
- roll of a 1 or 4 = $10
- roll of a 2 or 3 = $5

Let $X$ be the expected value of the payout. What is $\mathbb{E}(X)$?

Hint: $\mathbb{E}(X) = \sum_x xP(X = x)$ and

$P(X = x) = 1/4, x = 1, 2, 3, 4$

$\mathbb{E}(X) = .25 \times 10 + .25 \times 5 + .25 \times 5 + .25 \times 10 = 7.5$
Another important quantity associated with a random variable is the variance. The variance is the average dispersion of the data from the mean/expected value. The variance of random variable $X$, denoted $\text{var}(X)$, is defined as:

$$\text{var}(X) = E[(X - E(X))^2]$$

or

$$\text{var}(X) = \sum x (x - E(X))^2 P(X = x)$$

The variance is non-negative and provides a measure of dispersion of $X$ around the mean/expected value.
Another important quantity associated with a random variable is the variance
Another important quantity associated with a random variable is the variance.

The variance is the average dispersion of the data from the mean/expected value.
Another important quantity associated with a random variable is the variance.

The variance is the average dispersion of the data from the mean/expected value.

The variance of random variable $X$ denoted $\text{var}(X)$ is

$$\text{var}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2]$$

$$\text{var}(X) = \sum_{x}(x - \mathbb{E}(X))^2 P(X = x)$$
Variance

- Another important quantity associated with a random variable is the variance.
- The variance is the average dispersion of the data from the mean/expected value.
- The variance of random variable $X$ denoted $\text{var}(X)$ is

$$\text{var}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2]$$

$$\text{var}(X) = \sum_{x}(x - \mathbb{E}(X))^2 P(X = x)$$

- The variance is non-negative.
Another important quantity associated with a random variable is the variance.

The variance is the average dispersion of the data from the mean/expected value.

The variance of random variable $X$ denoted $\text{var}(X)$ is

$$\text{var}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2]$$

$$\text{var}(X) = \sum_x (x - \mathbb{E}(X))^2 P(X = x)$$

The variance is non-negative.

The variance provides a measure of dispersion of $X$ around the mean/expected value.
A software engineering company tested a new product of theirs and found that the number of errors per 100 CDs of the new software had the following probability distribution:

<table>
<thead>
<tr>
<th>$x$</th>
<th>$P(X=x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.01</td>
</tr>
<tr>
<td>3</td>
<td>0.25</td>
</tr>
<tr>
<td>4</td>
<td>0.40</td>
</tr>
<tr>
<td>5</td>
<td>0.30</td>
</tr>
<tr>
<td>6</td>
<td>0.04</td>
</tr>
</tbody>
</table>

What is the variance of $X$?

$$\text{Var}(X) = \sum x (x - \mu_x)^2 P(X=x)$$

---

A software engineering company tested a new product of theirs and found that the number of errors per 100 CDs of the new software had the following probability distribution:

<table>
<thead>
<tr>
<th>x</th>
<th>P(X=x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.01</td>
</tr>
<tr>
<td>3</td>
<td>0.25</td>
</tr>
<tr>
<td>4</td>
<td>0.40</td>
</tr>
<tr>
<td>5</td>
<td>0.30</td>
</tr>
<tr>
<td>6</td>
<td>0.04</td>
</tr>
</tbody>
</table>

\[ \text{var}(X) = \sum (x - \mu_X)^2 P(X=x) \]

\(^1\)Example from http://www.wyzant.com/resources/lessons/math/statistics_and_probability/expected_value/variance
Variance Ex.

A software engineering company tested a new product of theirs and found that the number of errors per 100 CDs of the new software had the following probability distribution:

<table>
<thead>
<tr>
<th>x</th>
<th>P(X=x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.01</td>
</tr>
<tr>
<td>3</td>
<td>0.25</td>
</tr>
<tr>
<td>4</td>
<td>0.40</td>
</tr>
<tr>
<td>5</td>
<td>0.30</td>
</tr>
<tr>
<td>6</td>
<td>0.04</td>
</tr>
</tbody>
</table>

What is the variance of X?

\[ \text{var}(X) = \sum_{x}(x - \mu_x)^2 P(X = x) \]

1Example from http://www.wyzant.com/resources/lessons/math/statistics_and_probability/expected_value/variance
Variance Ex. Cont...

From previous slide:

\[ P(X=x) \]

<table>
<thead>
<tr>
<th>x</th>
<th>P(X=x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.01</td>
</tr>
<tr>
<td>3</td>
<td>0.25</td>
</tr>
<tr>
<td>4</td>
<td>0.40</td>
</tr>
<tr>
<td>5</td>
<td>0.30</td>
</tr>
<tr>
<td>6</td>
<td>0.04</td>
</tr>
</tbody>
</table>

Step 1 find

\[ \mu = E(X) = \sum x \cdot P(X=x) \]

\[ E(X) = 2 \cdot 0.01 + 3 \cdot 0.25 + 4 \cdot 0.40 + 5 \cdot 0.30 + 6 \cdot 0.04 = 4.11 \]

Step 2 find

\[ \text{var}(X) = \sum (x - E(X))^2 \cdot P(X=x) \]

\[ \text{var}(X) = (2 - 4.11)^2 \cdot 0.01 + (3 - 4.11)^2 \cdot 0.25 + (4 - 4.11)^2 \cdot 0.40 + (5 - 4.11)^2 \cdot 0.30 + (6 - 4.11)^2 \cdot 0.04 = 0.74 \]
Variance Ex. Cont...

From previous slide:

<table>
<thead>
<tr>
<th>x</th>
<th>P(X=x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.01</td>
</tr>
<tr>
<td>3</td>
<td>0.25</td>
</tr>
<tr>
<td>4</td>
<td>0.40</td>
</tr>
<tr>
<td>5</td>
<td>0.30</td>
</tr>
<tr>
<td>6</td>
<td>0.04</td>
</tr>
</tbody>
</table>
Variance Ex. Cont...

- From previous slide:

<table>
<thead>
<tr>
<th>$x$</th>
<th>$P(X=x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.01</td>
</tr>
<tr>
<td>3</td>
<td>0.25</td>
</tr>
<tr>
<td>4</td>
<td>0.40</td>
</tr>
<tr>
<td>5</td>
<td>0.30</td>
</tr>
<tr>
<td>6</td>
<td>0.04</td>
</tr>
</tbody>
</table>

- Step 1 find $\mu_x = E(X) = \sum_x xP(X = x)$
Variance Ex. Cont...

- From previous slide:

<table>
<thead>
<tr>
<th>x</th>
<th>P(X=x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.01</td>
</tr>
<tr>
<td>3</td>
<td>0.25</td>
</tr>
<tr>
<td>4</td>
<td>0.40</td>
</tr>
<tr>
<td>5</td>
<td>0.30</td>
</tr>
<tr>
<td>6</td>
<td>0.04</td>
</tr>
</tbody>
</table>

- Step 1 find $\mu_x = \mathbb{E}(X) = \sum_x xP(X = x)$
  
  $\mathbb{E}(X) = 2 \times 0.01 + 3 \times 0.25 + 4 \times 0.40 + 5 \times 0.30 + 6 \times 0.04 = 4.11$
Variance Ex. Cont...

From previous slide:

<table>
<thead>
<tr>
<th>x</th>
<th>P(X=x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.01</td>
</tr>
<tr>
<td>3</td>
<td>0.25</td>
</tr>
<tr>
<td>4</td>
<td>0.40</td>
</tr>
<tr>
<td>5</td>
<td>0.30</td>
</tr>
<tr>
<td>6</td>
<td>0.04</td>
</tr>
</tbody>
</table>

Step 1 find $\mu_x = \mathbb{E}(X) = \sum_x xP(X = x)$

$\mathbb{E}(X) = 2 \cdot 0.01 + 3 \cdot 0.25 + 4 \cdot 0.40 + 5 \cdot 0.30 + 6 \cdot 0.04 = 4.11$

Step 2 find $var(X) = \sum_x (x - \mathbb{E}(X))^2 P(X = x)$
Variance Ex. Cont...

- From previous slide:

<table>
<thead>
<tr>
<th>x</th>
<th>P(X=x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.01</td>
</tr>
<tr>
<td>3</td>
<td>0.25</td>
</tr>
<tr>
<td>4</td>
<td>0.40</td>
</tr>
<tr>
<td>5</td>
<td>0.30</td>
</tr>
<tr>
<td>6</td>
<td>0.04</td>
</tr>
</tbody>
</table>

- Step 1 find $\mu_x = \mathbb{E}(X) = \sum_x xP(X=x)$
  - $\mathbb{E}(X) = 2 \times .01 + 3 \times .25 + 4 \times .40 + 5 \times .30 + 6 \times .04 = 4.11$

- Step 2 find $\text{var}(X) = \sum_x (x - \mathbb{E}(X))^2 P(X = x)$
  - $\text{var}(X) = (2 - 4.11)^2(.01) + (3 - 4.11)^2(.25) + (4 - 4.11)^2(.40) + (5 - 4.11)^2(.30) + (6 - 4.11)^2(.04) = 0.74$
Another measure of dispersion is the standard deviation of $X$, which is the square-root of the variance.

$$\sigma_X = \sqrt{\text{var}(X)}$$

The standard deviation is easier to interpret because it has the same units as $X$.

Example: If $X$ measures height in inches the units of variance are in square inches, but the units of the standard deviation are in inches.
Standard Deviation

Another measure of dispersion is the standard deviation of $X$, which is the square-root of the variance.

$$\sigma_X = \sqrt{\text{var}(X)}$$
Another measure of dispersion is the standard deviation of $X$, which is the square-root of the variance.

$$\sigma_X = \sqrt{\text{var}(X)}$$

The standard deviation is easier to interpret because it has the same units as $X$. 
Standard Deviation

- Another measure of dispersion is the standard deviation of $X$, which is the square-root of the variance.

$$\sigma_X = \sqrt{\text{var}(X)}$$

- The standard deviation is easier to interpret because it has the same units as $X$.

- Example: If $X$ measures height in inches the units of variance are in square inches, but the units of the standard deviation are in inches.
Learning Check - Expectation and Variance

Compute the expectation and variance using the following table

<table>
<thead>
<tr>
<th>x</th>
<th>P(X=x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.35</td>
</tr>
<tr>
<td>14</td>
<td>0.25</td>
</tr>
<tr>
<td>20</td>
<td>0.40</td>
</tr>
</tbody>
</table>
Joint Probability Mass Function - Overview

1

Up to now we have looked at one random variable. Real problems will require more than one random variable. Here we consider the case of two to give you insight.

Let's take an example. We toss a pair of fair 4-sided dice. One die is red and the other is black.

$$X \text{(red die)} = \{1, 2, 3, 4\}$$

$$Y \text{(black die)} = \{1, 2, 3, 4\}$$

We want to know the probability that $$X = x$$ and $$Y = y$$ or $$P(X = x, Y = y)$$.

To solve this problem, we need to use the fundamental counting rule. $$X$$ has 4 values and $$Y$$ has 4 values, so the total number of possible pairs of $$(x, y)$$ outcomes is $$4 \times 4 = 16$$.

Example from https://onlinecourses.science.psu.edu/stat414/node/104
Joint Probability Mass Function - Overview

- Up to know we have looked at one random variable

1 Example from https://onlinecourses.science.psu.edu/stat414/node/104
Joint Probability Mass Function - Overview

- Up to know we have looked at one random variable
- Real problems will require more than one random variable

---

Example from https://onlinecourses.science.psu.edu/stat414/node/104
Joint Probability Mass Function - Overview

- Up to know we have looked at one random variable
- Real problems will require more than one random variable
- Here we consider the case of two to give you insight

Example from https://onlinecourses.science.psu.edu/stat414/node/104
Joint Probability Mass Function - Overview

- Up to now we have looked at one random variable
- Real problems will require more than one random variable
- Here we consider the case of two to give you insight
- Let’s take an example¹:

¹Example from https://onlinecourses.science.psu.edu/stat414/node/104
Joint Probability Mass Function - Overview

- Up to now we have looked at one random variable
- Real problems will require more than one random variable
- Here we consider the case of two to give you insight
- Let’s take an example:\n  - We toss a pair of fair 4-sided dice. One die is red and the other is black

---

\(^1\)Example from https://onlinecourses.science.psu.edu/stat414/node/104
Joint Probability Mass Function - Overview

- Up to know we have looked at one random variable
- Real problems will require more than one random variable
- Here we consider the case of two to give you insight
- Let’s take an example:\n  - We toss a pair of fair 4-sided dice. One die is red and the other is black
  - \( X \) (red die) \( X = \{1, 2, 3, 4\} \)

---

\(^{1}\) Example from https://onlinecourses.science.psu.edu/stat414/node/104
Joint Probability Mass Function - Overview

- Up to know we have looked at one random variable
- Real problems will require more than one random variable
- Here we consider the case of two to give you insight
- Let’s take an example\(^1\):
  - We toss a pair of fair 4-sided dice. One die is red and the other is black
  - \(X\) (red die) \(X = \{1, 2, 3, 4\}\)
  - \(Y\) (black die) \(Y = \{1, 2, 3, 4\}\)

\(^1\)Example from https://onlinecourses.science.psu.edu/stat414/node/104
Joint Probability Mass Function - Overview

- Up to know we have looked at one random variable
- Real problems will require more than one random variable
- Here we consider the case of two to give you insight
- Let’s take an example\(^1\):
  - We toss a pair of fair 4-sided dice. One die is red and the other is black
  - \(X\) (red die) \(X = \{1, 2, 3, 4\}\)
  - \(Y\) (black die) \(Y = \{1, 2, 3, 4\}\)
- We want to know the probability that \(X = x\) and \(Y = y\) or \(P(X = x, Y = y)\)

\(^1\text{Example from https://onlinecourses.science.psu.edu/stat414/node/104}\)
Up to know we have looked at one random variable
Real problems will require more than one random variable
Here we consider the case of two to give you insight
Let’s take an example:

- We toss a pair of fair 4-sided dice. One die is red and the other is black
  - \( X \) (red die) \( X = \{1, 2, 3, 4\} \)
  - \( Y \) (black die) \( Y = \{1, 2, 3, 4\} \)

We want to know the probability that \( X = x \) and \( Y = y \) or \( P(X = x, Y = y) \)
To solve this problem we need to use the fundamental counting rule

---

\(^1\)Example from https://onlinecourses.science.psu.edu/stat414/node/104
Up to know we have looked at one random variable
Real problems will require more than one random variable
Here we consider the case of two to give you insight
Let’s take an example:\(^1\):
We toss a pair of fair 4-sided dice. One die is red and the other is black
\(X\) (red die) \(X = \{1, 2, 3, 4\}\)
\(Y\) (black die) \(Y = \{1, 2, 3, 4\}\)
We want to know the probability that \(X = x\) and \(Y = y\) or \(P(X = x, Y = y)\)
To solve this problem we need to use the fundamental counting rule
\(X = 4\) values and \(Y = 4\) values so the total number of possible pairs of \((x, y)\) outcomes is \(4 \times 4 = 16\)

\(^1\)Example from https://onlinecourses.science.psu.edu/stat414/node/104
We toss a pair of fair 4-sided dice. One die is red and the other is black. 

\[ X \text{(red die)} = \{1, 2, 3, 4\} \]

\[ Y \text{(black die)} = \{1, 2, 3, 4\} \]

Now we need to enumerate the pairs...in this example this is easy to do using a contingency table.

<table>
<thead>
<tr>
<th>Black (Y)</th>
<th>P(X=x, Y=y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1/16</td>
</tr>
<tr>
<td>2</td>
<td>1/16</td>
</tr>
<tr>
<td>3</td>
<td>1/16</td>
</tr>
<tr>
<td>4</td>
<td>1/16</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Red (X)</th>
<th>P(x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1/4</td>
</tr>
<tr>
<td>2</td>
<td>1/4</td>
</tr>
<tr>
<td>3</td>
<td>1/4</td>
</tr>
<tr>
<td>4</td>
<td>1/4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>P(y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/4</td>
</tr>
<tr>
<td>1/4</td>
</tr>
<tr>
<td>1/4</td>
</tr>
<tr>
<td>1/4</td>
</tr>
</tbody>
</table>
Joint Probability Mass Function - Overview

- We toss a pair of fair 4-sided dice. One die is red and the other is black
We toss a pair of fair 4-sided dice. One die is red and the other is black.

\[ X \ (\text{red die}) \ X = \{1, 2, 3, 4\} \]
Joint Probability Mass Function - Overview

- We toss a pair of fair 4-sided dice. One die is red and the other is black
- \( X \) (red die) \( X = \{1, 2, 3, 4\} \)
- \( Y \) (black die) \( Y = \{1, 2, 3, 4\} \)
We toss a pair of fair 4-sided dice. One die is red and the other is black.

\[ X \text{ (red die)} \quad X = \{1, 2, 3, 4\} \]

\[ Y \text{ (black die)} \quad Y = \{1, 2, 3, 4\} \]

Now we need to enumerate the pairs...in this example this is easy to do using a contingency table.

<table>
<thead>
<tr>
<th>Red (X)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>P_x(x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(\frac{1}{16})</td>
<td>(\frac{1}{16})</td>
<td>(\frac{1}{16})</td>
<td>(\frac{1}{16})</td>
<td>(\frac{4}{16})</td>
</tr>
<tr>
<td>2</td>
<td>(\frac{1}{16})</td>
<td>(\frac{1}{16})</td>
<td>(\frac{1}{16})</td>
<td>(\frac{1}{16})</td>
<td>(\frac{4}{16})</td>
</tr>
<tr>
<td>3</td>
<td>(\frac{1}{16})</td>
<td>(\frac{1}{16})</td>
<td>(\frac{1}{16})</td>
<td>(\frac{1}{16})</td>
<td>(\frac{4}{16})</td>
</tr>
<tr>
<td>4</td>
<td>(\frac{1}{16})</td>
<td>(\frac{1}{16})</td>
<td>(\frac{1}{16})</td>
<td>(\frac{1}{16})</td>
<td>(\frac{4}{16})</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Black (Y)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>P_y(y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(\frac{1}{16})</td>
<td>(\frac{1}{16})</td>
<td>(\frac{1}{16})</td>
<td>(\frac{1}{16})</td>
<td>(\frac{4}{16})</td>
</tr>
<tr>
<td>2</td>
<td>(\frac{4}{16})</td>
<td>(\frac{4}{16})</td>
<td>(\frac{4}{16})</td>
<td>(\frac{4}{16})</td>
<td>(1)</td>
</tr>
<tr>
<td>3</td>
<td>(\frac{4}{16})</td>
<td>(\frac{4}{16})</td>
<td>(\frac{4}{16})</td>
<td>(\frac{4}{16})</td>
<td>(1)</td>
</tr>
<tr>
<td>4</td>
<td>(\frac{4}{16})</td>
<td>(\frac{4}{16})</td>
<td>(\frac{4}{16})</td>
<td>(\frac{4}{16})</td>
<td>(1)</td>
</tr>
</tbody>
</table>
# Joint Probability Mass Functions - Overview

<table>
<thead>
<tr>
<th>$P(X = x, Y = y)$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>$P_x(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red (X)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{4}{16}$</td>
</tr>
<tr>
<td>2</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{4}{16}$</td>
</tr>
<tr>
<td>3</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{4}{16}$</td>
</tr>
<tr>
<td>4</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{4}{16}$</td>
</tr>
</tbody>
</table>

| $P_y(y)$          | 4/16 | 4/16 | 4/16 | 4/16 | 1        |

Now let's note some properties of the joint PMF:

- A joint PMF must satisfy the axioms of probability.

Marginal distributions:

$P_x(x) = \sum_y P(x, y)$

$P_y(y) = \sum_x P(x, y)$
### Joint Probability Mass Functions - Overview

<table>
<thead>
<tr>
<th>$P(X = x, Y = y)$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>$P_x(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1/16</td>
<td>1/16</td>
<td>1/16</td>
<td>1/16</td>
<td>1/4</td>
</tr>
<tr>
<td>2</td>
<td>1/16</td>
<td>1/16</td>
<td>1/16</td>
<td>1/16</td>
<td>1/4</td>
</tr>
<tr>
<td>3</td>
<td>1/16</td>
<td>1/16</td>
<td>1/16</td>
<td>1/16</td>
<td>1/4</td>
</tr>
<tr>
<td>4</td>
<td>1/16</td>
<td>1/16</td>
<td>1/16</td>
<td>1/16</td>
<td>1/4</td>
</tr>
</tbody>
</table>

| $P_y(y)$ | 4/16 | 4/16 | 4/16 | 4/16 | 1       |

- Now let’s note some properties of the joint PMF

Now let’s note some properties of the joint PMF.
Joint Probability Mass Functions - Overview

<table>
<thead>
<tr>
<th>Red (X)</th>
<th>P(X = x, Y = y)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>P_x(x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>P_Y(y)</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td></td>
</tr>
</tbody>
</table>

- Now let’s note some properties of the joint PMF
  - A joint PMF must satisfy the axioms of probability
Joint Probability Mass Functions - Overview

<table>
<thead>
<tr>
<th>Red (X)</th>
<th>Black (Y)</th>
<th></th>
<th></th>
<th></th>
<th>P_x(x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>1/16</td>
<td>1/16</td>
<td>1/16</td>
<td>1/16</td>
</tr>
<tr>
<td>2</td>
<td>1/16</td>
<td>1/16</td>
<td>1/16</td>
<td>1/16</td>
<td>1/16</td>
</tr>
<tr>
<td>3</td>
<td>1/16</td>
<td>1/16</td>
<td>1/16</td>
<td>1/16</td>
<td>1/16</td>
</tr>
<tr>
<td>4</td>
<td>1/16</td>
<td>1/16</td>
<td>1/16</td>
<td>1/16</td>
<td>1/16</td>
</tr>
</tbody>
</table>

Now let’s note some properties of the joint PMF

- A joint PMF must satisfy the axioms of probability
- Marginal distributions
### Joint Probability Mass Functions - Overview

<table>
<thead>
<tr>
<th></th>
<th>Black (Y)</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(X = x, Y = y)$</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>$P_x(x)$</td>
</tr>
<tr>
<td>1</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{8}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{4}{16}$</td>
</tr>
<tr>
<td>2</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
</tr>
<tr>
<td>3</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
</tr>
<tr>
<td>4</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
</tr>
<tr>
<td>$P_y(y)$</td>
<td>$\frac{4}{16}$</td>
<td>$\frac{4}{16}$</td>
<td>$\frac{4}{16}$</td>
<td>$\frac{4}{16}$</td>
<td>$1$</td>
</tr>
</tbody>
</table>

- Now let’s note some properties of the joint PMF
  - A joint PMF must satisfy the axioms of probability
  - Marginal distributions
    - $P_x(X = x) = \sum_y P(X = x, Y = y)$
Joint Probability Mass Functions - Overview

<table>
<thead>
<tr>
<th>Red (X)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>P_x(x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1/16</td>
<td>1/16</td>
<td>1/16</td>
<td>1/16</td>
<td>4/16</td>
</tr>
<tr>
<td>2</td>
<td>1/16</td>
<td>1/16</td>
<td>1/16</td>
<td>1/16</td>
<td>4/16</td>
</tr>
<tr>
<td>3</td>
<td>1/16</td>
<td>1/16</td>
<td>1/16</td>
<td>1/16</td>
<td>4/16</td>
</tr>
<tr>
<td>4</td>
<td>1/16</td>
<td>1/16</td>
<td>1/16</td>
<td>1/16</td>
<td>4/16</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Black (Y)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>P_y(y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(X = x, Y = y)</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>1/16</td>
</tr>
<tr>
<td>P_x(x)</td>
<td>4/16</td>
<td>4/16</td>
<td>4/16</td>
<td>4/16</td>
<td>1</td>
</tr>
<tr>
<td>P_y(y)</td>
<td>1/16</td>
<td>1/16</td>
<td>1/16</td>
<td>1/16</td>
<td>1</td>
</tr>
</tbody>
</table>

Now let’s note some properties of the joint PMF

- A joint PMF must satisfy the axioms of probability
- Marginal distributions
  - \( P_x(X = x) = \sum_y P(X = x, Y = y) \)
  - \( P_y(Y = y) = \sum_x P(X = x, Y = y) \)
Now let’s assume that you have two dice $X$ (red) and $Y$ (black)

Assume $X = \{1, 2, 3\}$ and $Y = \{1, 2, 3\}$

Write out the contingency table of the joint PMF $P(X = x, Y = y)$
Joint Probability Mass Functions (PMFs) also have conditioning. Here are two things we won’t cover, but you should have in the background of your mind:

- Conditional expectation
- Conditional variance

**Example of conditional distribution**

\[
P(X = x | Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}
\]

Let’s take an example...
Conditional Probability - Joint PMFs

- Joint PMFs also have conditioning
Conditional Probability - Joint PMFs

- Joint PMFs also have conditioning
- Here are two things we won’t cover but you should have in the back of your brain
Conditional Probability - Joint PMFs

- Joint PMFs also have conditioning
- Here are two things we won’t cover but you should have in the back of your brain
  - Conditional expectation
Conditional Probability - Joint PMFs

- Joint PMFs also have conditioning
- Here are two things we won't cover but you should have in the back of your brain
  - Conditional expectation
  - Conditional variance
Conditional Probability - Joint PMFs

- Joint PMFs also have conditioning
- Here are two things we won’t cover but you should have in the back of your brain
  - Conditional expectation
  - Conditional variance
- Example of conditional distribution
Conditional Probability - Joint PMFs

- Joint PMFs also have conditioning
- Here are two things we won't cover but you should have in the back of your brain
  - Conditional expectation
  - Conditional variance
- Example of conditional distribution
  \[ P(X = x|Y = y) = \frac{P(X=x, Y=y)}{P(Y=y)} = \frac{\text{joint pmf}}{\text{marginal pmf}} \]
Conditional Probability - Joint PMFs

- Joint PMFs also have conditioning
- Here are two things we won’t cover but you should have in the back of your brain
  - Conditional expectation
  - Conditional variance
- Example of conditional distribution
  - $P(X = x|Y = y) = \frac{P(X=x,Y=y)}{P(Y=y)} = \frac{\text{joint pmf}}{\text{marginal pmf}}$
- Let’s take an example
### Conditional Probability - Joint PMF Ex.

<table>
<thead>
<tr>
<th>$P(X = x, Y = y)$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>$P_x(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{4}{16}$</td>
</tr>
<tr>
<td>2</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{4}{16}$</td>
</tr>
<tr>
<td>3</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{4}{16}$</td>
</tr>
<tr>
<td>4</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{4}{16}$</td>
</tr>
</tbody>
</table>

Red (X)

<table>
<thead>
<tr>
<th>$P_y(y)$</th>
<th>4</th>
<th>4</th>
<th>4</th>
<th>4</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\frac{4}{16}$</td>
<td>$\frac{4}{16}$</td>
<td>$\frac{4}{16}$</td>
<td>$\frac{4}{16}$</td>
<td>$\frac{1}{1}$</td>
</tr>
</tbody>
</table>

Black (Y)
### Conditional Probability - Joint PMF Ex.

<table>
<thead>
<tr>
<th>$P(X = x, Y = y)$</th>
<th>Black (Y)</th>
<th>$P_X(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{4}{16}$</td>
</tr>
<tr>
<td>2</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{4}{16}$</td>
</tr>
<tr>
<td>3</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{4}{16}$</td>
</tr>
<tr>
<td>4</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{4}{16}$</td>
</tr>
</tbody>
</table>

- **Red (X)**
  - $P_Y(y)$
    - $P_Y(1) = \frac{1}{16}$
    - $P_Y(2) = \frac{1}{16}$
    - $P_Y(3) = \frac{1}{16}$
    - $P_Y(4) = \frac{1}{16}$

What is $P(X = 1|Y = 2)$?
### Conditional Probability - Joint PMF Ex.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>P(X = x, Y = y)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Red (X)</td>
<td>1/16</td>
<td>1/16</td>
<td>1/16</td>
<td>1/16</td>
<td>P_x(x)</td>
</tr>
<tr>
<td>2 Red (X)</td>
<td>1/16</td>
<td>1/16</td>
<td>1/16</td>
<td>1/16</td>
<td></td>
</tr>
<tr>
<td>3 Red (X)</td>
<td>1/16</td>
<td>1/16</td>
<td>1/16</td>
<td>1/16</td>
<td></td>
</tr>
<tr>
<td>4 Red (X)</td>
<td>1/16</td>
<td>1/16</td>
<td>1/16</td>
<td>1/16</td>
<td></td>
</tr>
<tr>
<td>P_y(y)</td>
<td>4/16</td>
<td>4/16</td>
<td>4/16</td>
<td>4/16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td></td>
</tr>
</tbody>
</table>

- What is P(X = 1|Y = 2)?
- Use this: P(X = x|Y = y) = \( \frac{P(X=x, Y=y)}{P(Y=y)} \)
## Conditional Probability - Joint PMF Ex.

<table>
<thead>
<tr>
<th>P(X = x, Y = y)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>P_x(x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black (Y)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{4}{16}$</td>
</tr>
<tr>
<td>2</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{4}{16}$</td>
</tr>
<tr>
<td>3</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{4}{16}$</td>
</tr>
<tr>
<td>4</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{4}{16}$</td>
</tr>
</tbody>
</table>

| Red (X)         |     |     |     |     |        |
| 1               | $\frac{1}{16}$ | $\frac{1}{16}$ | $\frac{1}{16}$ | $\frac{1}{16}$ | $\frac{4}{16}$ |
| 2               | $\frac{1}{16}$ | $\frac{1}{16}$ | $\frac{1}{16}$ | $\frac{1}{16}$ | $\frac{4}{16}$ |
| 3               | $\frac{1}{16}$ | $\frac{1}{16}$ | $\frac{1}{16}$ | $\frac{1}{16}$ | $\frac{4}{16}$ |
| 4               | $\frac{1}{16}$ | $\frac{1}{16}$ | $\frac{1}{16}$ | $\frac{1}{16}$ | $\frac{4}{16}$ |

| P_y(y)          | $\frac{4}{16}$ | $\frac{4}{16}$ | $\frac{4}{16}$ | $\frac{4}{16}$ | $\frac{1}{1}$ |

**What is P(X = 1|Y = 2)?**

- **Use this:** \( P(X = x|Y = y) = \frac{P(X=x, Y=y)}{P(Y=y)} \)
- \( P(X = 1|Y = 2) = \frac{P(X=1, Y=2)}{P(Y=2)} \)
Conditional Probability - Joint PMF Ex.

<table>
<thead>
<tr>
<th></th>
<th>Black (Y)</th>
<th>Red (X)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(X = x, Y = y)$</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
</tr>
<tr>
<td>2</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
</tr>
<tr>
<td>3</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
</tr>
<tr>
<td>4</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
</tr>
<tr>
<td>$P_y(y)$</td>
<td>$\frac{4}{16}$</td>
<td>$\frac{4}{16}$</td>
<td>$\frac{4}{16}$</td>
</tr>
</tbody>
</table>

- What is $P(X = 1 | Y = 2)$?
- Use this: $P(X = x | Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$.
- $P(X = 1 | Y = 2) = \frac{P(X = 1, Y = 2)}{P(Y = 2)}$.
- $P(Y = 2) = \frac{4}{16}$, $P(X = 1, Y = 2) = \frac{1}{16}$. 

Plug and Chug:

$$P(X = 1 | Y = 2) = \frac{\frac{1}{16}}{\frac{4}{16}} = \frac{1}{4}.$$
## Conditional Probability - Joint PMF Ex.

<table>
<thead>
<tr>
<th></th>
<th>Black (Y)</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(X = x, Y = y)$</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>$P_x(x)$</td>
</tr>
<tr>
<td>Red (X)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{4}{16}$</td>
</tr>
<tr>
<td>2</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{4}{16}$</td>
</tr>
<tr>
<td>3</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{4}{16}$</td>
</tr>
<tr>
<td>4</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{4}{16}$</td>
</tr>
<tr>
<td>$P_y(y)$</td>
<td>$\frac{4}{16}$</td>
<td>$\frac{4}{16}$</td>
<td>$\frac{4}{16}$</td>
<td>$\frac{4}{16}$</td>
<td>$1$</td>
</tr>
</tbody>
</table>

**What is $P(X = 1\mid Y = 2)$?**

- Use this: $P(X = x\mid Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$

- $P(X = 1\mid Y = 2) = \frac{P(X = 1, Y = 2)}{P(Y = 2)}$

- $P(Y = 2) = 4/16$, $P(X = 1, Y = 2) = 1/16$

- Plug and Chug: $P(X = 1\mid Y = 2) = (1/16)/(4/16) = 1/4$
New Section