Probability in Computing

• **LECTURE 5**
  - Conditional Probability
  - Product Rule
  - Law of total probability
  - Independent events
  - Bayes' Rule

**Reminders**
- Submit HW1 regrade requests by Thursday
- HW2 is out.

Prof. Tiago Januario
Slides by Prof. Sofya Raskhodnikova
The annual death rate among people who know this statistic is one in six.

https://xkcd.com/795/
Incorporating New Information

• A patient has some unknown disease
• Based on the symptoms, the doctor estimates that the patient has:


• Then a test reveals that the disease is neither b nor d
• Based on this information, what is the chance of having each disease?
Incorporating New Information

• A test eliminated diseases b and d

<table>
<thead>
<tr>
<th>Disease</th>
<th>Before the test</th>
<th>After the test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease a</td>
<td>0.05</td>
<td>?</td>
</tr>
<tr>
<td>Disease b</td>
<td>0.1</td>
<td>?</td>
</tr>
<tr>
<td>Disease c</td>
<td>0.3</td>
<td>?</td>
</tr>
<tr>
<td>Disease d</td>
<td>0.55</td>
<td>?</td>
</tr>
</tbody>
</table>

Tiago Januario: Probability in Computing
Incorporating New Information

- A test eliminated diseases b and d

<table>
<thead>
<tr>
<th>Disease</th>
<th>Probability Before the test</th>
<th>Probability After the test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease a</td>
<td>0.05</td>
<td>?</td>
</tr>
<tr>
<td>Disease b</td>
<td>0.1</td>
<td>0</td>
</tr>
<tr>
<td>Disease c</td>
<td>0.3</td>
<td>?</td>
</tr>
<tr>
<td>Disease d</td>
<td>0.55</td>
<td>0</td>
</tr>
</tbody>
</table>
Incorporating New Information

- A test eliminated diseases b and d

<table>
<thead>
<tr>
<th>Disease</th>
<th>Probability Before the Test</th>
<th>Probability After the Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>0.05</td>
<td>0.05x</td>
</tr>
<tr>
<td>b</td>
<td>0.1</td>
<td>0</td>
</tr>
<tr>
<td>c</td>
<td>0.3</td>
<td>0.3x</td>
</tr>
<tr>
<td>d</td>
<td>0.55</td>
<td>0</td>
</tr>
</tbody>
</table>

- In the absence of any other information, the new probabilities of a and b should be proportional to their original probabilities.
Incorporating New Information

- A test eliminated diseases b and d

<table>
<thead>
<tr>
<th>Disease</th>
<th>Probability Before the Test</th>
<th>Probability After the Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease a</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>Disease b</td>
<td>0.1</td>
<td>0</td>
</tr>
<tr>
<td>Disease c</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>Disease d</td>
<td>0.55</td>
<td>0</td>
</tr>
</tbody>
</table>

- Probabilities should add up to 1: \(0.05x + 0.3x = 1\)

\[
x = \frac{1}{0.35}
\]
Incorporating New Information

- A test eliminated diseases b and d

<table>
<thead>
<tr>
<th>Disease</th>
<th>Probability Before the Test</th>
<th>Probability After the Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease a</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>Disease b</td>
<td>0.1</td>
<td>0</td>
</tr>
<tr>
<td>Disease c</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>Disease d</td>
<td>0.55</td>
<td>0</td>
</tr>
</tbody>
</table>

- Note: \( \Pr(\text{"the disease is neither b nor d"}) = \Pr(\{a, c\}) = 0.35 \)
Conditional Probability

• The information provided by the test is that the following event has happened:

\[ B = "the\ disease\ is\ neither\ b\ nor\ d" = \{a, c\} \]

• We define a new probability function that assigns a probability \( \Pr(A \mid B) \) to every event \( A \subseteq \Omega \)

\[
\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}
\]
Conditional Probability

• The information provided by the test is that the following event has happened:
  
  \[ B = "the\ disease\ is\ neither\ b\ nor\ d" = \{a, c\} \]

• We define a new probability function that assigns a probability \( \Pr(A \mid B) \) to every event \( A \subseteq \Omega \)

\[
\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}
\]
Conditional Probability

- The information provided by the test is that the following event has happened:

\[ B = \text{"the disease is neither } b \text{ nor } d\" = \{a, c\} \]

- We define a new probability function that assigns a probability \( \Pr(A | B) \) to every event \( A \subseteq \Omega \)

\[
\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}
\]

Read: "Probability of \( A \) given \( B \)"

If \( A = \{\omega\} \) for an outcome \( \omega \), we also write: \( \Pr(\omega \mid B) \)

If we specify \( \Pr(\omega \mid B) \) for every outcome \( \omega \in \Omega \), then \( \Pr(A \mid B) \) is determined for all events \( A \) (as the sum of probabilities of all outcomes in \( A \) )
Conditional Probability

- The information provided by the test is that the following event has happened:

\[ B = "\text{the disease is neither } b \text{ nor } d" = \{a, c\} \]

- We define a new probability function that assigns a probability \( \Pr(A \mid B) \) to every event \( A \subseteq \Omega \)

\[
\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}
\]

\[
\Pr(a \mid B) = \frac{\Pr(a)}{\Pr(B)} = \frac{0.05}{0.35}
\]
Conditional Probability

- The information provided by the test is that the following event has happened:
  \[ B = "the\ disease\ is\ neither\ b\ nor\ d" = \{a, c\} \]
- We define a new probability function that assigns a probability \( \Pr(A \mid B) \) to every event \( A \subseteq \Omega \)
  \[
  \Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}
  \]
  \[
  \Pr(c \mid B) = \frac{\Pr(c)}{\Pr(B)} = \frac{0.3}{0.35}
  \]

\[ \Omega = \{a, b, c, d\} \]
Conditional Probability

• The information provided by the test is that the following event has happened:

\[ B = \text{"the disease is neither } b \text{ nor } d\text{"} = \{a, c\} \]

• We define a new probability function that assigns a probability \( \Pr(A \mid B) \) to every event \( A \subseteq \Omega \)

\[
\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}
\]

\[
\Pr\{a, b\} | B = \frac{\Pr\{a\}}{\Pr(B)} = \frac{0.05}{0.35} = 0.1429
\]

\( \Omega = \{a, b, c, d\} \)

\( \Omega \)

\[ \Omega \]

\( A \)

\( B \)

\( a \)

\( b \)

\( c \)

\( d \)
Conditional Probability

- The information provided by the test is that the following event has happened:

  \[ B = "the \text{ disease is neither } b \text{ nor } d" = \{a, c\} \]

- We define a new probability function that assigns a probability \( \Pr(A \mid B) \) to every event \( A \subseteq \Omega \)

  \[
  \Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}
  \]

  \[
  \Pr(\{a, c\} \mid B) = \frac{\Pr(\{a, c\})}{\Pr(B)} = 1
  \]
Definition: Conditional Probability

Let $B$ be an event such that $\Pr(B) \neq 0$. For every event $A$, we define the conditional probability of event $A$ given event $B$:

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}.$$

$\Pr(A \mid B)$ is undefined if $\Pr(B) = 0$. 

Tiago Januario: Probability in Computing
Experiment: roll two fair 6-sided dice

A = “first die is 6”

B = “sum of the two rolls is 9”

What is \( \Pr(A \mid B) \)?

A. \( \frac{1}{36} \)

B. \( \frac{1}{6} \)

C. \( \frac{1}{4} \)

D. None of the above
Experiment: toss a fair coin three times
A = “more H than T”
B = “first toss is H”
What is \( \Pr(A \mid B) \)?

A. \( \frac{3}{8} \)
B. \( \frac{1}{2} \)
C. \( \frac{3}{4} \)
D. None of the above
• Experiment: spin the dial of the spinner
• A = “dial stopped in the upper half of the circle”
• B = “outcome is in [1/3, 2/3]”

What is $\Pr(A \mid B)$?

A. $\frac{1}{3}$
B. $\frac{1}{2}$
C. $\frac{3}{4}$
D. None of the above
Top Hat question (Join Code: 258641)

- Experiment: spin the dial of the spinner
- A = “dial stopped in the upper half of the circle”
- C = “outcome is in \([1/6, 1/3]\)”

What is \(\Pr(A \mid C)\)?

A. \(\frac{1}{12}\)
B. \(\frac{1}{6}\)
C. 1
D. None of the above
Mr. Smith has two children
(boys and girls are equally likely)

• A = “both children are boys”
• B = “at least one child is a boy”
• C = “older child is a boy”

True or false: $\Pr(A \mid B) = \Pr(A \mid C)$?

A. True
B. False
Conditional probabilities satisfy the axioms of probability

- **Non-negativity**: \( \Pr(A \mid E) \geq 0 \) for all events \( A \subseteq \Omega \)

- **Additivity**: if \( A \) and \( B \) are disjoint events then
  \[
  \Pr(A \cup B \mid E) = \Pr(A \mid E) + \Pr(B \mid E)
  \]

- **Normalization**: \( \Pr(\Omega \mid E) = 1 \)

This applies only to conditional probabilities with the same conditioning event: \( \Pr(\cdot \mid E) \)
Conditional probabilities satisfy the axioms of probability

- **Non-negativity**: $\Pr(A \mid E) \geq 0$ for all events $A \subseteq \Omega$

  **Proof**: 
  
  $$\Pr(A \mid E) = \frac{\Pr(A \cap E)}{\Pr(E)}$$

  *by definition of conditional probability*

- **Normalization**: $\Pr(\Omega \mid E) = 1$

  **Proof**: 
  
  $$\Pr(\Omega \mid E) = \frac{\Pr(\Omega \cap E)}{\Pr(E)}$$

  $$= \frac{\Pr(E')}{\Pr(E)} = 1$$

  *by definition of conditional probability*

  *since $E \subseteq \Omega$*
Conditional probabilities satisfy the axioms of probability

- **Additivity:** if $A$ and $B$ are disjoint events then
  \[
  \Pr(A \cup B \mid E) = \Pr(A \mid E) + \Pr(B \mid E)
  \]

**Proof:**

\[
\Pr(A \cup B \mid E) = \frac{\Pr((A \cup B) \cap E)}{\Pr(E)}
\]

\[
= \frac{\Pr((A \cap E) \cup (B \cap E))}{\Pr(E)}
\]

\[
= \frac{\Pr(A \cap E) + \Pr(B \cap E)}{\Pr(E)}
\]

\[
= \Pr(A \mid E) + \Pr(B \mid E)
\]
Conditional Probability Rules

Since the conditional probability function is a valid probability function, all the probability rules remain valid

- Example: Inclusion-Exclusion Principle

  Standard: \( \text{Pr}(A \cup B) = \text{Pr}(A) + \text{Pr}(B) - \text{Pr}(A \cap B) \)

  Conditional: \( \text{Pr}(A \cup B \mid E) = \text{Pr}(A \mid E) + \text{Pr}(B \mid E) - \text{Pr}(A \cap B \mid E) \)
Hoping to avoid the difficulties of using conditional probability, Thomas Jefferson writes the Declaration of Independence.
Tree Diagrams: Example

- Best-of-three tournament between BU Terriers and Harvard Huskies
- First game: Terriers win with probability $\frac{1}{2}$
- Subsequent games:

  \[
  \text{Terriers win a game with probability } \begin{cases} 
  \frac{2}{3} & \text{if they won the previous game} \\
  \frac{1}{3} & \text{if they lost the previous game} 
  \end{cases}
  \]

What is the probability that the Terriers win the tournament \text{given} that they win the first game?
Event $T$ = "Terriers win the tournament"
Event $E_1$ = "Terriers win the first game"
**Event** $T$ = ``Terriers win the tournament’’

**Event** $E_1$ = ``Terriers win the first game’’

$$\Pr(T|E_1) = \frac{\Pr((T \cap E_1))}{\Pr(E_1)} = \frac{\Pr(\text{WW, WLW, WLL})}{\Pr(\text{WW})} = \frac{\frac{1}{2} + \frac{1}{3} + \frac{1}{18}}{\frac{1}{2}} = \frac{\frac{1}{2} + \frac{1}{3} + \frac{1}{18}}{\frac{1}{2}} = \frac{\frac{1}{2} + \frac{1}{3} + \frac{1}{18}}{\frac{1}{2}} = \frac{\frac{9}{18} + \frac{6}{18} + \frac{1}{18}}{\frac{1}{2}} = \frac{\frac{16}{18}}{\frac{1}{2}} = \frac{8}{9}$$
• Why do we multiply probabilities along each branch in a tree diagram?

• Example:
  – Event $E_1 = \text{``Terriers win the first game''}$
  – Event $E_2 = \text{``Terriers win the second game''}$

• We calculated $\Pr(WW) = \Pr(E_1 \cap E_2)$
  $= \Pr(E_1) \cdot \Pr(E_2 | E_1)$

• Product rule follows from the definition of conditional probability:

$$\Pr(E_2 | E_1) = \frac{\Pr(E_1 \cap E_2)}{\Pr(E_1)}$$
• Why do we multiply probabilities along each branch in a tree diagram?
• Example:
  – Event $E_1$ = “Terriers win the first game”
  – Event $E_2$ = “Terriers win the second game”
• We calculated $\Pr(WW) = \Pr(E_1 \cap E_2)$
  $$\Pr(E_1) \cdot \Pr(E_2 \mid E_1)$$
• Product rule follows from the definition of conditional probability:
  $$\Pr(E_2 \mid E_1) = \frac{\Pr(E_1 \cap E_2)}{\Pr(E_1)}$$
Product Rule

• The product rule for two events $A, B$:
  \[ \Pr(A \cap B) = \Pr(A) \cdot \Pr(B \mid A) \]

• It can be generalized to three or more events:
  \[
  \Pr(A \cap B \cap C) = \Pr(A \cap B) \cdot \Pr(C \mid A \cap B)
  = \Pr(A) \cdot \Pr(B \mid A) \cdot \Pr(C \mid A \cap B)
  \]
Product Rule

• The product rule for two events $A, B$:

$$Pr(A \cap B) = Pr(A) \cdot Pr(B | A)$$

• It can be generalized to three or more events:

$$Pr(A \cap B \cap C) = Pr(A \cap B) \cdot Pr(C | A \cap B)$$

$$= Pr(A) \cdot Pr(B | A) \cdot Pr(C | A \cap B)$$
Product Rule

• The **product rule** for two events $A, B$:
  \[
  \Pr(A \cap B) = \Pr(A) \cdot \Pr(B \mid A)
  \]
  
  • It can be generalized to three or more events:
  \[
  \Pr(A \cap B \cap C) = \Pr(A \cap B) \cdot \Pr(C \mid A \cap B)
  = \Pr(A) \cdot \Pr(B \mid A) \cdot \Pr(C \mid A \cap B)
  \]
Top Hat question (Join Code: 033357)

• We have two urns with colored balls:
  – Urn 1 has 2 red balls and 3 green balls
  – Urn 2 has 1 red ball and 1 green ball

• We pick an urn uniformly at random, and then pick a ball uniformly at random from the chosen urn

Given that the ball we picked is red, what is the probability that it came from Urn 1?

A. 4/9
B. 1/2
C. 2/3
D. None of the above
### Tree Diagram Method

<table>
<thead>
<tr>
<th>urn</th>
<th>ball</th>
<th>outcome</th>
<th>probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urn 1</td>
<td>1/2</td>
<td>(U₁, R)</td>
<td>1/5</td>
</tr>
<tr>
<td>Urn 1</td>
<td>3/5</td>
<td>(U₁, G)</td>
<td></td>
</tr>
<tr>
<td>Urn 2</td>
<td>1/2</td>
<td>(U₂, R)</td>
<td>1/4</td>
</tr>
<tr>
<td>Urn 2</td>
<td>1/2</td>
<td>(U₂, G)</td>
<td></td>
</tr>
</tbody>
</table>

\[
\begin{align*}
(U₁, R) &= \frac{1}{2} \cdot \frac{2}{5} = \frac{1}{5} \\
(U₁, G) &= \frac{1}{2} \cdot \frac{3}{5} = \frac{3}{10} \\
(U₂, R) &= \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} \\
(U₂, G) &= \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}
\end{align*}
\]
Tree Diagram Method

Urn 1

\[
\begin{align*}
\text{Urn 1} & : & \frac{1}{2} & \rightarrow \frac{3}{5} \\
\text{Urn 2} & : & \frac{1}{2} & \rightarrow \frac{1}{2}
\end{align*}
\]

Ball

\[
\begin{align*}
\text{Urn 1} & : & \frac{2}{5} \\
\text{Urn 2} & : & \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}
\end{align*}
\]

Outcome

\[
\begin{align*}
(U_1, R) & : & 1 \cdot \frac{2}{5} = \frac{1}{5} \\
(U_1, G) & : & 1 \cdot \frac{3}{5} = \frac{3}{5} \\
(U_2, R) & : & 1 \cdot \frac{1}{2} = \frac{1}{2} \\
(U_2, G) & : & 1 \cdot \frac{1}{2} = \frac{1}{2}
\end{align*}
\]

Probability

\[
\begin{align*}
\Pr(U_1 | R) & = \frac{\Pr(U_1 \cap R)}{\Pr(R)} \\
& = \frac{\Pr(\{U_1, R\})}{\Pr(\{(U_1, R), (U_2, R)\})} \\
& = \frac{1/5}{1/5 + 1/4} = \frac{4}{9}
\end{align*}
\]

Tiago Januario, Sofya Raskhodnikova; Probability in Computing
Top Hat question (Join Code: 033357)

• We have two urns with colored balls:
  – Urn 1 has 2 red balls and 3 green balls
  – Urn 2 has 1 red ball and 1 green ball

• We pick an urn uniformly at random, and then pick a ball uniformly at random from the chosen urn

Given that the ball we picked is red, what is the probability that it came from Urn 1?

A. 4/9
B. 1/2
C. 2/3
D. None of the above
Law of Total Probability

- We can use conditional probability to break down probability calculations
  \[ \Pr(A) = \Pr(A \cap B) + \Pr(A \cap \overline{B}) \]
  \[ = \Pr(A|B) \cdot \Pr(B) + \Pr(A|\overline{B}) \cdot \Pr(\overline{B}) \]

- **Law of total probability:**
  \[ \Pr(A) = \Pr(A|B) \cdot \Pr(B) + \Pr(A|\overline{B}) \cdot \Pr(\overline{B}) \]

- **Example:** We have two coins: fair and double-headed
  We pick one uniformly at random and toss it.
  What is the probability we get heads?
  \( A = \text{“get heads”}, \ B = \text{“pick a fair coin”} \)
  \[ \Pr(A) = \frac{1}{2} \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{3}{4} \]
Independent Events

- Conditional probability $\Pr(A \mid B)$ captures partial information that event B provides about event A.
- An important setting is when the information that B happened does not change the probability of A:
  \[
  \Pr(A \mid B) = \Pr(A)
  \]

Then, by product rule,
\[
\Pr(A \cap B) = \Pr(A \mid B) \cdot \Pr(B) = \Pr(A) \cdot \Pr(B)
\]
Independent Events

**Definition: Independent Events**

Two events $A$ and $B$ are **independent** if

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$$

- Equivalent condition: $\Pr(A | B) = \Pr(A)$
  Alternatively, $\Pr(B | A) = \Pr(B)$
- Independence is **symmetric**: if $A$ is independent of $B$ then $B$ is independent of $A$
- This definition can be used even when $\Pr(B) = 0$
  - An event $B$ with $\Pr(B) = 0$ is independent of all events, including itself.
- **Independent ≠ disjoint!** When events $A$ and $B$ are disjoint:
  if $A$ happens then $B$ is guaranteed not to happen.

Tiago Januario, Sofya Raskhodnikova; Probability in Computing
Experiment: toss a fair coin twice
- A = “first toss is H”
- B = “second toss is H”

Are A and B independent?

A. YES

\[ \Pr(A \cap B) = \Pr(\{HH\}) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} \]

B. NO
Experiment: toss two fair coins

- A = “first toss is H”
- B = “both tosses give the same result”

Are \( A \) and \( B \) independent?

A. YES

\[
\begin{align*}
P(A) &= \frac{1}{2} \\
P(B) &= P(\{HH, TT\}) = \frac{2}{4} = \frac{1}{2}
\end{align*}
\]

B. NO

\[
P(A \cap B) = P(\{HH\}) = \frac{1}{4} = \frac{1}{2} = \frac{1}{2}
\]
Top Hat question (Join Code: 413437)

- Experiment: roll two fair 4-sided dice
- A = “first die is 1”
- B = “sum of the two rolls is 5”

Are A and B independent?

A. YES
B. NO

\[
A = \{(1,1), (1,2), (1,3), (1,4)\}
\]

\[
B = \{(1,4), (4,1), (2,3), (3,2)\}
\]

\[
A \cap B = \{(1,4)\}
\]

\[
\Pr(A) = \frac{4}{16} = \frac{1}{4}
\]

\[
\Pr(B) = \frac{4}{16} = \frac{1}{4}
\]

\[
\Pr(A \cap B) = \frac{1}{16}
\]
• Experiment: roll two fair 6-sided dice
• A = “first die is 1”
• B = “sum of the two rolls is 5”

Are $A$ and $B$ independent?

$$A = \{(1,1), (1,2), (1,3), (1,4), (1,5), (1,6)\}$$

A. YES  

$$B = \{(1,4), (4,1), (2,3), (3,2)\}$$

B. NO

$$A \cap B = \{(1,4)\}$$

$$\Pr(A) = \frac{6}{36} = \frac{1}{6}$$

$$\Pr(B) = \frac{4}{36} = \frac{1}{9}$$

$$\Pr(A \cap B) = \frac{1}{36}$$
Independent Events

**Definition: Independent Events**

Two events $A$ and $B$ are **independent** if

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$$

- Equivalent condition: $\Pr(A \mid B) = \Pr(A)$
  Alternatively, $\Pr(B \mid A) = \Pr(B)$
- Independence is **symmetric**:
  if $A$ is independent of $B$ then $B$ is independent of $A$
- This definition can be used even when $\Pr(B) = 0$
  - An event $B$ with $\Pr(B) = 0$ is independent of all events, including itself.
Conditioning the original sample space means changing the perspective: when A and B are independent, then $\Pr(A)$ does not change when reduce the sample space from $\Omega$ to $B$:

$$\Omega$$

```
A

B

A \cap B
```

$$\Omega' = B$$

```
A \cap B
```

```
A \cap B
```
Independent Events

Google Colab Simulation

A = “C1 flips heads”
B = “C2 flips heads”

\[ \Pr(A) = \frac{1}{2}, \Pr(B) = \frac{1}{2} \]

\[ \Pr(A \cap B) = \frac{1}{4} \]

\[ \Pr(A | B) = \frac{1/4}{1/2} = \frac{1}{2} = \Pr(A) \]
Independent Events

Be Careful: Independent is not the same as disjoint!

When events $A$ and $B$ are disjoint, $\Pr(A), \Pr(B) > 0$, if $A$ happens then $B$ is guaranteed not to happen and vice versa:

$$\Omega$$

$$A \cap B = \emptyset$$

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{0}{\Pr(B)} = 0$$
Independence

Let’s try some examples.....

Experiment: roll two fair 4-sided dice
   A = “first die is 1”
   B = “sum of the two rolls is 5”
Are A and B independent?

Experiment: roll two fair 4-sided dice
   A = “first die is 1”
   B = “sum of the two rolls is $\leq 5$”
Are A and B independent?

D4, Four Sided Die (Blue)
https://openclipart.org/
Experiment: roll two fair 4-sided dice

A = “first die is 1”
B = “sum of the two rolls is 5”

Are $A$ and $B$ independent?

$A = \{(1,1), (1,2), (1,3), (1,4)\} \rightarrow |A| = 4 \rightarrow \Pr(A) = \frac{1}{4}$

$B = \{(1,4), (2,3), (3,2), (4,1)\} \rightarrow |B| = 4 \rightarrow \Pr(B) = \frac{1}{4}$

$A \cap B = \{(1,4)\} \rightarrow |A \cap B| = 1 \rightarrow \Pr(A \cap B) = \frac{1}{16}$

Yes! $A$ and $B$ are independent!
Experiment: roll two fair 4-sided dice

A = “first die is 1”

B = “sum of the two rolls is ≤ 5”

Are A and B independent?

\[ A = \{(1,1), (1,2), (1,3)\} \rightarrow |A| = 3 \rightarrow \Pr(A) = \frac{3}{16} \]
\[ B = \{(1,1), (1,2), (1,3), (2,1), (2,3), (2,2), (3,1)\} \rightarrow |B| = 7 \]
\[ \rightarrow \Pr(B) = \frac{7}{16} \]

\[ A \cap B = \{(1,1), (1,2), (1,3)\} \rightarrow |A \cap B| = 3 \rightarrow \Pr(A \cap B) = \frac{3}{16} \]

No! A and B are NOT independent!
Experiment: toss two fair coins

- A = “first toss is H”
- B = “both tosses give the same result”

Are $A$ and $B$ independent?

A. YES

\[ A = \{(H, T), (H, H)\} \rightarrow |A| = 2 \rightarrow \Pr(A) = \frac{1}{2} \]

B. NO

\[ B = \{(H, H), (T, T)\} \rightarrow |B| = 2 \rightarrow \Pr(B) = \frac{1}{2} \]

\[ A \cap B = \{(H, H)\} \rightarrow |A \cap B| = 1 \rightarrow \Pr(A \cap B) = \frac{1}{4} \]

Yes! $A$ and $B$ are independent!
Experiment: toss two biased coins with \( \Pr(H) = \frac{2}{3} \)

- A = “first toss is H”
- B = “both tosses give the same result”

Are \( A \) and \( B \) independent?

A. YES
B. NO
Experiment: toss two *biased* coins with $\Pr(H) = \frac{2}{3}$

- $A =$ “first toss is H” = \{(HH,HT)\}
- $B =$ “both tosses give the same result” = \{(HH,TT)\}

Are $A$ and $B$ independent?

A. YES

B. NO

$\Pr(HH) = \frac{4}{9}$, $\Pr(HT) = \frac{2}{9}$, $\Pr(TH) = \frac{2}{9}$, $\Pr(TT) = \frac{1}{9}$

$\Pr(A) = \frac{4}{9} + \frac{2}{9} = \frac{6}{9}$

$\Pr(B) = \frac{4}{9} + \frac{1}{9} = \frac{5}{9}$

$\Pr(A \cap B) = \frac{4}{9} \neq \frac{30}{81} = \frac{10}{27} = \Pr(A) \cdot \Pr(B)$
Independent Repeated Trials: Example

- **Experiment**: $X =$ spin the dial of the spinner and observe the region where it stopped.
  - $\Omega = \{1, 2, 3, 4\}$
  - $\Pr(1) = \frac{1}{2}$, $\Pr(2) = \frac{1}{4}$, $\Pr(3) = \Pr(4) = \frac{1}{8}$

- Now we spin the dial twice $(X_1, X_2)$
  - Sample space is $\Omega \times \Omega = \{(i, j) : 1 \leq i, j \leq 4\}$
  - $\Pr((i, j)) = \Pr(i) \cdot \Pr(j)$

- We can spin it $n$ times
  - Each outcome is a sequence of $n$ elements, each drawn from $\Omega$
    
    $(X_1, X_2, \ldots, X_n)$
  - Sample space is $\Omega^n$
  - $\Pr((s_1, \ldots, s_n)) = \Pr(s_1) \cdot \ldots \cdot \Pr(s_n)$

Events ``1st spin is $i$'' and ``2nd spin is $j$'' are independent.

Technical note: We are overusing $\Pr$. We are talking about two different probability functions (over $\Omega^n$ and over $\Omega$).
Independent Repeated Trials

- Let $\Omega$ be a sample space for $X$ with a probability function $Pr$ and $n \in \mathbb{N}$.
- Let $\Omega^n$ denote the set of all length-$n$ sequences of elements from $\Omega$.
- Then $n$ **independent repeated trials** of the original random experiment are represented by $n$ random variables $(X_1, X_2, \ldots, X_n)$
  - sample space $\Omega^n$
  - probability function $Pr((s_1, \ldots, s_n)) = Pr(s_1) \cdot \ldots \cdot Pr(s_n)$

**Examples**

- Roll a 6-sided die twice:
  - Sample space for one roll is $\Omega = \{1, 2, 3, 4, 5, 6\}$ and $Pr(s) = \frac{1}{6}$ for all $s \in \Omega$
  - Sample space for two rolls is $\Omega^2$, and $Pr((s_1, s_2)) = \frac{1}{36}$ for all $s_1, s_2 \in \Omega$

- Toss a fair coin until you see a head:
  - Sample space for one toss is $\Omega = \{0, 1\}$ and $Pr(0) = Pr(1) = \frac{1}{2}$
  - Sample space for one trial is $\Omega^N$ in which an outcome $\Omega^k$ has probability $\frac{1}{2^k}$
Independence in Tree Diagrams

\( P(A \mid B) \) considers an event \( B \) followed by an event \( A \), and how the occurrence of \( B \) affects the occurrence of \( A \). What are the labels on a tree diagram of this random experiment?

<table>
<thead>
<tr>
<th>B occurs (or not)</th>
<th>A occurs (or not)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P(B) )</td>
<td>( P(A \mid B) )</td>
</tr>
<tr>
<td>( P(B^c) )</td>
<td>( P(A^c \mid B) )</td>
</tr>
<tr>
<td>( P(A \cap B) ) = ( P(A \mid B) \cdot P(B) )</td>
<td></td>
</tr>
<tr>
<td>( P(A \cap B^c) )</td>
<td>( P(A^c \cap B) )</td>
</tr>
<tr>
<td>( P(A \cap B^c) )</td>
<td>( P(A^c \cap B^c) )</td>
</tr>
</tbody>
</table>
How does this relate to tree diagrams?

When the events are independent, then we have the familiar tree diagram in which we simply write the probabilities of the events on each arc:

\[
\begin{array}{c|c}
\text{B occurs (or not)} & \text{A occurs (or not)} \\
P(B) & P(A) \\
P(B^c) & P(A^c) \\
& \\
\end{array}
\]

\[
P(A) 
\rightarrow 
P(A \cap B) = P(A) \cdot P(B)
\]

\[
P(A^c) = 1 - P(A)
\]

\[
P(A \cap B^c)
\]

\[
P(A^c \cap B^c)
\]
Example: Sampling with or without replacement

1. You draw 3 cards from a standard deck with replacement: what is the probability they are all Spades?

\[
\frac{13}{52} \cdot \frac{13}{52} \cdot \frac{13}{52}
\]
Example: Sampling with or without replacement

1. You draw 3 cards from a standard deck with replacement: what is the probability they are all Spades?

\[
\frac{13}{52} \cdot \frac{13}{52} \cdot \frac{13}{52}
\]

2. You draw 3 cards from a standard deck without replacement: what is the probability they are all Spades?

\[
\frac{13}{52} \cdot \frac{12}{51} \cdot \frac{11}{50}
\]
Fun fact: Dependence does not imply causality!
Chocolate consumption and Noble laureates

Aloys Leo Prinz

Highlights

- Chocolate consumption per capita is positively correlated with the stock of Nobel prizes per capita.
- A two-stage Heckman selection model is estimated.
- The correlation remains after control for scientific articles and R&D expenditures.
- It remains unclear whether the correlation is spurious or an indication for hidden variables.
Bayes’ Rule

We can rearrange the conditional probability rule in a way that makes the sequence of the events irrelevant -- which happened first, A or B? Or did they happen at the same time? Does it matter?

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

We can do a little algebra to define conditional probabilities in terms of each other:
Bayes’ Rule

We can rearrange the conditional probability rule in a way that makes the sequence of the events irrelevant -- which happened first, A or B? Or did they happen at the same time? Does it matter?

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

We can do a little algebra to define conditional probabilities in terms of each other:

\[
P(B \mid A) \cdot P(A) = P(B \cap A) = P(A \mid B) \cdot P(B)
\]

so:

\[
P(B \mid A) = \frac{P(A \mid B) \cdot P(B)}{P(A)}
\]
Bayes’ Rule

We can rearrange the conditional probability rule in a way that makes the sequence of the events irrelevant -- which happened first, A or B? Or did they happen at the same time? Does it matter?

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

We can do a little algebra to define conditional probabilities in terms of each other:

\[ P(B \mid A) \times P(A) = P(B \cap A) = P(A \mid B) \times P(B) \]

so:

\[ P(B \mid A) = \frac{P(A \mid B) \times P(B)}{P(A)} \]
Bayes’ Rule

This has an interesting flavor, because we can ask about causes of outcomes:

A Priori Reasoning -- “I randomly choose a person and observe that he is male; what the probability that it is a smoker?”

“The first toss of a pair of dice is a 5; what is the probability that the total is greater than 8?”
This has an interesting flavor, because we can ask about causes of outcomes:

A Posteriori Reasoning -- “I find a cigarette butt on the ground, what is the probability that it was left by a man?”

“The total of a pair of thrown dice is greater than 8; what is the probability that the first toss was a 5?”
Bayes’ Rule

The best way to understand this is to view it with a tree diagram!

\[ P(B|A) \] = the probability that when \ A \ happens, it was “preceeded” by \ B: 

- If \ A \ has happened, what is the probability that it did so on the path where \ B \ also occurred?

Note:

\[ \text{Pr}(A) = \text{Pr}(A \cap B) \cup \text{Pr}(A \cap B^c) \]

So what percentage of \ A \ is due to \ A \cap B \ ?

Same calculation as:

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

**Example:** Suppose that 10% of female BU students smoke cigarettes and 20% of male BU students smoke cigarettes. Suppose that 60% of BU students are female and 40% male. I see a cigarette butt on the pavement. What is the probability it was thrown there by a female student?

\[ \Pr(C|F) = 0.1 \]
\[ \Pr(C|M) = 0.2 \]
\[ \Pr(F) = 0.6 \]
\[ \Pr(M) = 0.4 \]
\[ \Pr(C) = ? \]

By law of total probability \( \Pr(C) = \Pr(C|F) \cdot \Pr(F) + \Pr(C|M) \cdot \Pr(M) = 0.1 \cdot 0.6 + 0.2 \cdot 0.4 \)
Bayes’ Rule

**Example:** Suppose that 10% of female BU students smoke cigarettes and 20% of male BU students smoke cigarettes. Suppose that 60% of BU students are female and 40% male.

I see a cigarette butt on the pavement. What is the probability it was thrown there by a female student?

\[
\begin{align*}
\text{Pr}(C|F) &= 0.1 \\
\text{Pr}(C|M) &= 0.2 \\
\text{Pr}(F) &= 0.6 \\
\text{Pr}(M) &= 0.4 \\
\text{Pr}(C) &= 0.14
\end{align*}
\]

By law of total probability \(\text{Pr}(C) = \text{Pr}(C|F) \cdot \text{Pr}(F) + \text{Pr}(C|M) \cdot \text{Pr}(M) = 0.14\)

By Bayes’ theorem \(\text{Pr}(F|C) = \frac{\text{P}(A|B) \cdot \text{P}(B)}{\text{P}(A)}\)
Bayes’ Rule

Example: We have two urns, A and B. A contains 2 red balls and 3 black balls. Urn B contains 2 red balls and 1 black ball. We select an urn uniformly at random and pick a ball and find it is red. What is the probability it came from Urn A?