Combining Events: The **union** $A \cup B$ is the event consisting of all outcomes in $A$ or in $B$ or in both.

The **intersection** $A \cap B$ is the event consisting of all outcomes in both $A$ and $B$.

If $A \cap B$ contains no outcomes (i.e., if $A$ and $B$ have no elements in common), then $A$, $B$ are said to be **mutually exclusive**.

The **Complement** $\overline{A}$ of the event $A$ consists of all outcomes in the sample space $S$ which are not in $A$. 
• \( P(A \cup B) = P(A \text{ or } B \text{ or both occur}) \)

• \( P(A \cap B) = P(A \text{ and } B \text{ both occur}) \)

• \( P(\overline{A}) = P(A \text{ does not occur}) \).

**Eg 1:** In a study of soft drink preferences, a subject is given Coke, Pepsi and Sprite, and asked to state their preference, if any.

The sample space of possible outcomes is

\[ S = \{ \text{Coke, Pepsi, Sprite, No Preference} \}. \]

Suppose \( A = \{ \text{Coke} \} \), \( B = \{ \text{Cola} \} = \{ \text{Coke, Pepsi} \} \), \( C = \{ \text{Pepsi} \} \).

What are the events \( A \cup C \), \( A \cap B \), \( A \cap C \), and \( \overline{B} \)?
• Complement Rule: \( P(\overline{A}) = 1 - P(A) \)

• Addition Rule: \( P( A \cup B ) = P(A) + P(B) - P(A \cap B) \).

• If \( A, B \) are mutually exclusive, then \( P(A \cap B) = 0 \), since \( A \cap B \) contains no outcomes. This gives:

  Addition Rule for mutually exclusive events:

  \[
P(A \cup B ) = P(A) + P(B),
  \]

  if \( A, B \) are mutually exclusive.
Eg 2: In the soft drink preference study, suppose that:

<table>
<thead>
<tr>
<th>Preference</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coke</td>
<td>0.3</td>
</tr>
<tr>
<td>Pepsi</td>
<td>0.4</td>
</tr>
<tr>
<td>Sprite</td>
<td>0.2</td>
</tr>
<tr>
<td>None</td>
<td>0.1</td>
</tr>
</tbody>
</table>

If \( A = \{\text{Coke}\} \), \( B = \{\text{Cola}\} \), \( C = \{\text{Pepsi}\} \), compute \( P(A \cup C) \), \( P(A \cap B) \), \( P(A \cap C) \), \( P(A \cup B) \), \( P(\overline{B}) \).
Eg 3: Performance of mutual funds (Goetzmann and Ibbotson)

<table>
<thead>
<tr>
<th></th>
<th>Current Winner</th>
<th>Current Loser</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Past Winner</td>
<td>482</td>
<td>296</td>
<td>778</td>
</tr>
<tr>
<td>Past Loser</td>
<td>285</td>
<td>493</td>
<td>778</td>
</tr>
<tr>
<td>Total</td>
<td>767</td>
<td>789</td>
<td>1556</td>
</tr>
</tbody>
</table>

Compute the probability that a randomly selected mutual fund is a current winner.

If we are told that the fund was a past winner, does this partial knowledge change the probability that the fund is a current winner?
**Conditional Probability**: If A, B are events with \( P(B) > 0 \), then the conditional probability that A occurs, given that B has occurred is

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

**Independence**: A and B are independent if \( P(A|B) = P(A) \). Otherwise they are dependent.

- A and B are independent if the knowledge that B has occurred doesn't change the probability that A will occur.
Multiplication Rules:

• If A and B are independent, then \( P(A \cap B) = P(A)P(B) \).

• In General, \( P(A \cap B) = P(A) P(B|A) \)
Examples of conditional probability:

• Click-through rate (given search/visit history).
• Default probability on a loan (given credit score).
• The current value of a prediction market contract.

• ”The probability of a Fed rate increase at its December meeting moved up to 67 percent, according to futures data compiled by Bloomberg.”

Models that yield conditional probabilities:

• Linear Regression
• Logistic Regression
Eg 4: Classification of Managers (Past 3 Years)

<table>
<thead>
<tr>
<th></th>
<th>Promoted</th>
<th>Not Promoted</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>46</td>
<td>184</td>
<td>230</td>
</tr>
<tr>
<td>Female</td>
<td>8</td>
<td>32</td>
<td>40</td>
</tr>
<tr>
<td>Total</td>
<td>54</td>
<td>216</td>
<td>270</td>
</tr>
</tbody>
</table>

Note that only 8 out of 54 promotions went to women. Are the women being unfairly passed over for promotions?

Sol: No, since Prob\{Promoted|Female\}=8/40=.2=54/270 =Prob\{Promoted\}.

Alternatively:
Prob\{Male|Promoted\}=46/54=.852=230/270=Prob\{Male\}.

Promotion status and gender are independent.
Eg: A coin is tossed twice. Suppose all outcomes (HH, HT, TH, TT) are equally likely.

- There is no way to prove this mathematically without making additional, circular assumptions. It’s just an empirical fact, which can be verified by experiments with real coins.

If \( A = \{ \text{Heads on First Toss} \} \), \( B = \{ \text{Heads on Second Toss} \} \), then \( A \) and \( B \) are independent, since
\[
P(A \cap B) = P(HH) = \frac{1}{4} = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = P(A) \cdot P(B).
\]
Equivalently, \( P(B|A) = P(B) \), so the coin “doesn't remember” what the first toss was.
• In general, if a coin is tossed n times and we assume that all individual outcomes (such as HHTTHH...T) are equally likely, it follows that successive tosses are independent (i.e., the coin has no memory of previous tosses).

• If, on the other hand, we assume that successive tosses are independent, and that the probability of Heads on each toss is 1/2, it follows that all individual outcomes are equally likely. Later in the course, we will start with the assumption of independence (i.e., “lack of memory”).

• Either way, though, our initial assumption defies mathematical proof without resorting to circular reasoning. It must be checked by actually tossing coins.
So, let’s check it.

Class: Toss a coin 3 times.

What is proportion of students getting T on third toss?

What if we restrict attention to only those students who got H on first two?

Therefore, what's Prob \{T on 3^{rd} toss \mid H on first 2 tosses\}?
Eg: The Problem With Drug Testing.

Suppose that 1% of the population actually uses a certain illegal drug, and that the test has a false positive rate of 5% and a false negative rate of 5%.

If someone tests positive for the drug, does this information change the probability that the person is actually a user?

Does the probability increase? Is it 95%?

Can we be reasonably sure that someone who has tested positive is actually guilty?
Sol: If 10,000 people are tested we would “expect” to find:

<table>
<thead>
<tr>
<th></th>
<th>Drug User</th>
<th>Non User</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Positive</td>
<td>95</td>
<td>495</td>
<td>590</td>
</tr>
<tr>
<td>Test Negative</td>
<td>5</td>
<td>9405</td>
<td>9410</td>
</tr>
<tr>
<td>TOTAL</td>
<td>100</td>
<td>9900</td>
<td>10,000</td>
</tr>
</tbody>
</table>

Originally, \( \text{Prob}\{\text{User}\} = \frac{100}{10,000} = 0.01 \).
But if they tested positive, the prob becomes \( \frac{95}{590} = 0.16 \).
This is bigger than .01, but it’s still amazingly small.
Only 16% of those testing positive were actually users!
The other 84% of the positives are *false* positives.

• \( \text{Prob}\{\text{Test Positive|User}\} \) and \( \text{Prob}\{\text{User|Test Positive}\} \) are two *very different* things.
“Independent” and “Mutually Exclusive” are not the same thing.

[Eg 5, Eg 6]

**Eg:** Suppose Serena Williams and Venus Williams are tied at 40-40 (Deuce) in a tennis game. What is the probability that Venus will win the game? (The game must be won by two points).

**Sol:** We assume that outcomes on each point are independent, and that on any given point, Venus wins the point with probability $p$ and loses the point with probability $q = 1 - p$. We have

$$\text{Prob} \{\text{Venus wins the game}\} = \text{Prob} \{\text{Venus wins} \mid \text{Game ends after two more points}\}, \quad (1)$$

since knowing that the game will end after two more points tells us nothing about whether Venus will win the game. (The two events here are independent).
Now, the conditional probability (1) is

\[
\frac{\text{Prob}\{\text{Venus wins the game } \cap \text{ Game ends after two more points}\}}{\text{Prob}\{\text{Game ends after two more points}\}} \tag{2}
\]

There are four possible outcomes for the next two points:

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Venus wins both points</td>
<td>(p^2)</td>
</tr>
<tr>
<td>Serena wins both points</td>
<td>(q^2)</td>
</tr>
<tr>
<td>Venus wins first, Serena wins second</td>
<td>(pq)</td>
</tr>
<tr>
<td>Serena wins first, Venus wins second</td>
<td>(qp) (same as (pq))</td>
</tr>
</tbody>
</table>

These probabilities add to one, and (2) becomes

\[
\frac{\text{Prob}\{\text{Venus wins next two points}\}}{\text{Prob}\{\text{Game ends after two more points}\}} = \frac{p^2}{1-2pq}
\]

Thus, \(\text{Prob}\{\text{Venus wins game}\} = \frac{p^2}{1-2pq}\).
Here's a table for some specific values of $p$.

<table>
<thead>
<tr>
<th>$p$</th>
<th>Prob ${\text{Venus wins game}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.5</td>
<td>.5</td>
</tr>
<tr>
<td>.6</td>
<td>.69</td>
</tr>
<tr>
<td>.7</td>
<td>.84</td>
</tr>
<tr>
<td>.8</td>
<td>.94</td>
</tr>
<tr>
<td>.9</td>
<td>.99</td>
</tr>
</tbody>
</table>