Law of Total Probability

- **Theorem:** For every partition \( \{B_1, B_2, B_3, \ldots\} \) of \( S \), it holds that

\[
P[A] = \sum_i P[A \cap B_i], \ \forall A \subseteq S.
\]

- Using \( P[A \cap B_i] = P[A|B_i]P[B_i] \), we have

\[
P[A] = \sum_i P[A|B_i]P[B_i]
\]
Law of Total Probability

• Why is this expression, i.e., \( P[A] = \sum_i P[A|B_i]P[B_i] \), useful?

• Makes sense in practice when you do not have first hand knowledge of \( A \) but some ‘derived knowledge’.

• **Example** I wish to calculate the probability that the weight of a student is over 200 pounds in the class, denoted by \( A \). I know the probabilities of \( B_1 = \) “the height of a student in the class is less than or equal to 6 feet” and \( B_2 = \) “the height of a student in the class is larger than 6 feet”. I also know that if a student is more than 6 feet, the probability that the student has more than 200 pounds is 0.8, and 0.2 otherwise.

• Hence, I know \( P[A|B_1], P[A|B_2], P[B_1], \) and \( P[B_2] \), which means I can easily compute \( P[A] \)?
Law of Total Probability

• Wait...is $B_1, B_2$ a partition of the sample space?

• What is the sample space?
**Law of Total Probability**

- Wait...is $B_1, B_2$ a partition of the sample space?
- What is the sample space?

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(h&gt;6 feet, w\leq200 pounds)</td>
<td>(h&gt;6 feet, w&gt;200 pounds)</td>
</tr>
<tr>
<td>(h\leq6 feet, w\leq200 pounds)</td>
<td>(h\leq6 feet, w&gt;200 pounds)</td>
</tr>
</tbody>
</table>
Law of Total Probability

- Wait...is $B_1, B_2$ a partition of the sample space?

- What is the sample space?

\[
\begin{array}{c|c}
B_2 & (h>6 \text{ feet, } w\leq 200 \text{ pounds}) & (h>6 \text{ feet, } w>200 \text{ pounds}) \\
B_1 & (h\leq 6 \text{ feet, } w\leq 200 \text{ pounds}) & (h\leq 6 \text{ feet, } w>200 \text{ pounds}) \\
\end{array}
\]
Example

- **Problem**: A radar system detects a target w.p. 0.9 when the target is present, and falsely detects that the target is present w.p. 0.15 when in fact no target is present. If the target is present w.p. 0.05, what is the probability of alarm?

- To solve the problem, the first thing is to find out what is the sample space \( S \) of this particular problem.
Example

- **Problem**: A radar system detects a target w.p. 0.9 when the target is present, and falsely detects that the target is present w.p. 0.15 when in fact no target is present. If the target is present w.p. 0.05, what is the probability of alarm?

- To solve the problem, the first thing is to find out what is the **sample space** $S$ of this particular problem.

<table>
<thead>
<tr>
<th></th>
<th>(present, alarm)</th>
<th>(present, no alarm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(present, alarm)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(present, no alarm)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>(absent, alarm)</th>
<th>(absent, no alarm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(absent, alarm)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(absent, no alarm)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Example

• **Problem**: A radar system detects a target w.p. 0.9 when the target is present, and falsely detects that the target is present w.p. 0.15 when in fact no target is present. If the target is present w.p. 0.05, what is the probability of alarm?

• To solve the problem, the first thing is to find out what is the **sample space** $S$ of this particular problem.

![Sample Space Diagram]

- (present, alarm)
- (present, no alarm)
- (absent, alarm)
- (absent, no alarm)
• What do we know? \( P[\text{alarm}|\text{present}] = 0.9, \ P[\text{alarm}|\text{absent}] = 0.15, \ P[\text{present}] = 0.05, \ P[\text{absent}] = 0.95. \)

• Using the law of total probability, we have

\[
P[\text{alarm}] = P[\text{alarm}, \text{present}] + P[\text{alarm}, \text{absent}]
= P[\text{alarm}|\text{present}]P[\text{present}] + P[\text{alarm}|\text{absent}]P[\text{absent}] = 0.1875
\]
Bayes' Theorem

- **Bayes' Theorem**: \( P[B|A] = \frac{P[A|B]P[B]}{P[A]} \).

- **Proof**:

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

\[
\implies P[B|A] = \frac{P[A|B]P[B]}{P[A]}
\]

- **Importance**: allows us to express \( P[A|B] \) using \( P[B|A] \).

- e.g., \( P[\text{absent}|\text{alarm}] = \frac{P[\text{alarm}|\text{absent}]P[\text{absent}]}{P[\text{alarm}]} = \frac{0.15 \times 0.95}{0.1875} = 0.76 \).
Independence of Events

• **Definition:** Events $A, B$ are said to be independent if and only if $P[A \cap B] = P[A]P[B]$.

• wired ... why do we define it this way?
Independence of Events

- The insight of this definition goes back to the definition of conditional probability.

- Recall that conditional probability reveals the ‘side information of $A$ given that $B$ happens’.

- Independence, intuitively, means that $B$ contains no side information about $A$. Therefore, we should have $P[A|B] = P[A]$.

- Using the definition of independence, we have

  $$P[A|B] = \frac{P[A \cap B]}{P[B]} = \frac{P[A]P[B]}{P[B]} = P[A],$$

  which is consistent with our intuition.

- If one of $A$ and $B$ is empty, they are always independent ($P[A \cap B] = P[A]P[B] = 0$).
Independent and Disjoint

- Disjoint and independent are **very different**.

- Assume that $A \cap B = \emptyset$.

\[
P[A \cap B] = P[\emptyset] = 0 \neq P[A]P[B]
\]

- To be independent, $A$ and $B$ must have nonempty intersection.
• Note: independence hinges on the particular values of the probabilities.

\[
\begin{array}{cccc}
  & S & & \\
R & 0.1 & 0.4 & \\
  & 0.4 & 0.1 & S \\
\end{array}
\]

- Is the weather of Portland independent with that of Corvallis?

- Let us denote \( A \) as “it’s sunny in Portland” and \( B \) “it’s sunny in Corvallis”.

\[
P[A \cap B] = 0.4, \quad P[A]P[B] = (0.4 + 0.1) \times (0.4 + 0.1) = 0.25.
\]
More Remarks on Independence

- How to cook up an example such that we have an independent case?

- Check it up by yourselves.
More Remarks on Independence

• **Definition:** $A, B, C$ are independent if and only if
  
  – i) any two of them are pairwise independent; and

• How to define independence of four events?
Sequential Experiments and Tree Diagrams

- Many experiments consist of a sequence of *subexperiments*.

- The procedure followed by each subexperiment may depend on the results of the previous experiments.

- It is useful to represent such experiments using tree diagrams.

**Example:** Suppose you have two coins, one biased, one fair, but you don’t know which coin is which. Coin 1 is biased. It comes up heads with probability 3/4. Suppose you pick a coin at random and flip it. Let $C_i$ denote the event that coin $i$ is picked. Let $H$ and $T$ denote the possible outcomes of the flip. Given that the outcome of the flip is a head, what is $P[C_1|H]$? What is $P[C_1|T]$?
Sequential Experiments and Tree Diagrams

- Tree Diagram:

\[
P[C_1|H] = \frac{P[C_1 \cap H]}{P[H]} = \frac{P[C_1 \cap H]}{P[C_1 \cap H] + P[C_2 \cap H]} = \frac{3/8}{3/8 + 1/4} = 1/3
\]
Sequential Experiments

- Consider a case where you keep paging a person for several times. Every time the probability that you successfully find the person is 0.2.

- This can be represented by the following picture:
What is $P[\text{success}]$?

- **Sol 1**: $P[\text{success}] = 1 - P[\text{failure}] = 1 - 0.8 \times 0.8 \times 0.8 = 0.488$.
- **Sol 2**: Elementary outcomes of the experiment are

  $$\{s, fs, ffs, fff\}$$

- Both solutions use independence between outcomes of the subexperiments.
Counting Methods

- Suppose that you have \( n \) entities to choose from, and you pick \( k \) out of them. How many different sets of \( k \) entities can you draw, if the order in which they are drawn is important?

- Note: the order is important means that “Alice, Bob, John” and “John, Alice, Bob” are considered two sets.

- First draw: \( n \) possibilities. Second draw: \( n - 1 \) possibilities. ... \( k \)th draw: \( n - k + 1 \) possibilities.

- Hence, we have in total

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

possibilities.
Counting Methods

- What if order does not matter? This means that “Alice, Bob, John” and “Alice, John, Bob” are considered the same set.

- To count the number of such sets, we should take away all the possible permutations of a \( k \)-element set.

- How many permutations are there? There are \( k! \) permutations.

- Hence, without regard to order, the number of sets that I can pick is

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

- What if sampling with replacement?

\[
n \cdot n \ldots n = n^k.
\]

Does order matter in this case?