# 2. Conditional Probability and Bayes' Rule

$$P(A|B) \neq P(B|A)$$

How should we update our beliefs in light of the evidence we observe? *Bayes' rule* is an extremely useful theorem that helps us perform such updates.

Together, Bayes' rule and the law of total probability can be used to solve a very wide variety of problems.

## 2.1 The importance of thinking conditionally

A useful perspective is that all probabilities are *conditional*

$A$ and $B$ events, $P(B) > 0$, then *conditional probability* of A given B:

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

$A$ is the event whose uncertainty we want to update, and B is the evidence we observe.

$P(A)$ is called *prior probability* of $A$.

$P(A|B)$ is called *posterior probability* of $A$.

$P(A|B)$ is the probability of $A$ given the evidence $B$, **not** the probability of some weird entity called $A|B$.

Note:

1. It's extremely important to be careful about which events to put on which side of the conditioning bar. Confusing these two quantities is called the *prosecutor's fallacy*.

2. Both $P(A|B)$ and $P(B|A)$ make sense. We are considering what **information** observing one event provides about another event, not whether one event **causes** another.

Frequentist interpretation:

The conditional probability of $A$ given $B$: it is the fraction of times that $A$ occurs, restricting attention to the trials where $B$ occurs.

# 2.3 Bayes' rule and the law of total probability

Just move the denominator in the definition to the other side of the equation:

$$P(A \cap B) = P(B)P(A|B) = P(A)P(B|A)$$

It often turns out to be possible to find conditional probabilities without going back to the definition.

Same for $n$ events:

$\forall A_1, \ldots A_n$

$P(A_i, A_2, \ldots, A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1, A_2)\ldots P(A_n|A_1 \ldots A_{n-1})$

Then **Bayes' rule:**

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

The **law of total probability (LOTP)** relates conditional probability to unconditional probability:

$A_1, \ldots, A_n$ a partition of the sample space $S$ ($A_i$ disjoint, their union is $S$), $P(A_i) > 0 \forall i$ then:

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$$

Proof:

Decompsition of $B$:

$$B = (B \cap A_1) \cup (B \cap A_2) \cup \ldots \cup (B \cap A_n)$$

then

$$P(B) = P(B \cap A_1) + \ldots + P(B \cap A_n) = P(B|A_1)P(A_1) \ldots P(B|A_n)P(A_n)$$

The choice of how to divide up the sample space is crucial!

# 2.4 Conditional probabilities are probabilities

When we condition on an event E, we update our beliefs to be consistent with this knowledge, effectively putting ourselves in a **universe where we know that E occurred**.

**Bayes' rule with extra conditioning**

$P(A \cap E) > 0$ and $P(B \cap E) > 0$

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

**LOTP with extra conditioning**

$A_1, \ldots, A_n$ is a partition of $S$, $P(A_i \cap E) > 0 \forall i$

$$P(B|E) = \sum_{i=1}^{n} P(B|A_i, E)P(A_i|E)$$

# 2.5 Independence of events

$A$ and $B$ are independent if

$$P(A \cap B) = P(A)P(B)$$

if $P(A) > 0$ and $P(B) > 0$,

$$P(A|B) = P(A)$$

Two events are independent if we can obtain the probability of their intersection by multiplying their individual probabilities.

If $A$ is independent of $B$, then $B$ is independent of $A$.

**Independence is not disjointness!!!**

If $A$ and $B$ are disjoint, $P(A \cap B) = 0$

Disjoint events can be independent only if $P(A) = 0$ or $P(B) = 0$.

Knowing that $A$ occurs tells us that $B$ definitely did not occur.

Independence of three or more events!

$$P(A \cap B) = P(A)P(B),$$

$$P(A \cap C) = P(A)P(C),$$

$$P(B \cap C) = P(B)P(C).$$

If these three conditions are hold: $A$, $B$, $C$ are *pairwise independent* $\neq$ *independence*.

For example, it is possible that $A$, $B$ occurred would give us knowledge about $C$ .

For full independence, 4th condition:

$$P(A \cap B \cap C) = P(A)P(B)P(C).$$

Conditional independence:

$$P(A \cap B|E) = P(A|E)P(B|E).$$

*Conditional independence* $\neq$ *independence*

(example with biased and fair coin)

*Independence* $\neq$ *conditional independence*