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:

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

TODO: proof