# Quantifying Uncertainty

## Acting Under Uncertainty

### Why are agents uncertain?

1. Nondeterminism in the world, and in their behavior
2. Partial observability of the world. If the agent can't see what is going on, the agent can't think about its actions.

### How do agents deal with their uncertainty?

Most agents keep around a *belief state*, which is a list of all possible worlds the agent might currently be in. This works for small problem spaces but has some issues overall.

1. An agent would need to keep track of *every* possible world, no matter how unlikely that world is.
2. Contingency plans the agent might create become exceedingly large as a result of not measuring likelihood.
3. Sometimes, there is no promise that any plan will actually acheive the agent's goal, but the agent must act regardless.

The agent's goal is always to maximize its performance measure. To do that, it must weigh the relative importance of its goals with considerations on the likelihood it will acheive those goals. To summarize its uncertainty and make a rational decision, we use probability theory.

## Basic Probability Notation

Probabilities are about asserting if something will happen, relative to other events. Here are some basic definitions.

1. **sample space**: the set of all possible worlds that are under our consideration.
   (a) Denoted as $\Omega$.
   (b) Each $\omega \in \Omega$ gets a probability assigned to it.
   (c) $0 \leq P(\omega) \leq 1, \forall \omega \in \Omega$.
   (d) $\sum_{\omega \in \Omega} P(\omega) = 1$.
2. **events** are sets of possible worlds that satisfy a given property.
3. The probability of an event is the sum of the probabilities of the events satisfying the property. As an example, given two dice, $P(roll = 11) = P(5,6) + P(6,5) = \frac{1}{18}$.
4. Probabilities of events with no other information are called **priors** or **unconditional probabilities**.
5. **Posterior (conditional) probabilities** are those probabilities given some other piece of information.
6. $P(a \wedge b) = P(a|b)P(b)$. This rearrainges to the more familiar fractional equation of conditional probabilities.
7. A **probability distribution** of a discrete random variable is a listing of the values the random variable can take.
8. $P(\neg a) = 1 - \sum P(a)$
9. $P(a \vee b) = P(a) + P(b) - P(a \wedge b)$
10. **Joint probability distribution**: the probability of random variables $X$ and $Y$ occurring.
11. **Full joint probability distribution**: the joint distribution of all random variables.

## Inference Using Full Joint Distributions

A **marginal probability** is one that is gathered from a full joint probability distribution, wherein all other random variables have been ignored. Symbolically,

$$P(Y) = \sum_{z \in Z} P(Y \wedge z).$$

Following Russell and Norvig's example,

$$P(cavity) = P(cavity \wedge toothache) + P(cavity \wedge catch).$$

This process is called *marginalization*. Analogously, *conditioning* is similar but uses conditional probabilities instead.

$$P(Y) = \sum_{z \in Z} P(Y|z)P(z).$$

This follows directly from point 6 above.

## Independence

Sometimes things are not related at all; for example, the weather does not influence whether not you have a cavity, toothache, and catch at the dentist's office. Random variables that are not related are called **independent**. Mathematically, $P(a|b) = P(a)$. But since $P(a \wedge b) = P(a|b)P(b)$, the probability of two independent events happening is $P(a \wedge b) = P(a)P(b)$.

A slight modification to this is when two events are *conditionally independent*. That means that given some condition, the condition applies to both events. $P(a \wedge b|c) = P(a|c)P(b|c)$.

## Bayes' Rule and Its Use

Note the equivalence of the following two equations. $P(a \wedge b) = P(a|b)P(b)$ and $P(a \wedge b) = P(b|a)P(a)$. Equating the two and dividing gives Bayes' rule (theorem, law)

$$P(b|a) = \frac{P(a|b)P(b)}{P(a)}.$$