

---

# 1 Uncertainty

---

Up until now we assumed that every statement is either true or false (or unknown) and that every action taken behaves exactly as expected.

In the real world, this is not always the case. Agents almost never have access to the entire state (incomplete information).

This means that in many cases, **agents must deal with uncertainty**.

## 1.1 Probabilities

---

One way to deal with uncertainty is to use **probabilities**.

We can assign a probability to every event, to summarize effects of uncertainty, due to:

- **Laziness:**
  - Agent is too lazy to consider all possible outcomes and possible events
- **Theoretical Ignorance:**
  - Some things are impossible to know (e.g. Weather cannot be definitely predicted)
- **Practical Ignorance:**
  - Some things might just not be known about an event or situation (e.g. traffic conditions)

Probabilities often are not based on objective truths but rather **subjective beliefs**: A probability  $p$  means that I believe that the statement will be true in  $p \cdot 100\%$  of cases.

- **Degree of Belief:** There is a traffic jam in 10% of the cases
- **Degree of Truth:** The street is 10% jammed (blocked off a bit)

**Probability theory** is about **degree of belief** not **degree of truth**.

### 1.1.1 Basics

---

The state or **sample space**  $\Omega$  can be seen as a set of all **samples**  $\omega$ :

$$\begin{matrix} \Omega \\ \omega \in \Omega \end{matrix}$$

The **probability space** or probability model is a sample space with an assignment of probabilities to all samples:

$$\forall \omega \in \Omega : \exists P(\omega) : \sum_{\omega \in \Omega} P(\omega) = 1$$

**An event**  $A$  is any subset of the sample space:

$$\forall A \subseteq \Omega : \exists P(A) : \sum_{\omega \in A} P(\omega) = P(A)$$

### 1.1.2 Kolmogorov's Axioms of Probability

---

1. All probabilities are between 0 and 1

$$0 \leq P(A) \leq 1$$

2. Necessarily true proposition have probability 1, false propositions have probability 0

$$P(\text{true}) = 1, P(\text{false}) = 0$$

3. The probability of a disjunction is:

$$P(A \vee B) = P(A) + P(B) - P(A \wedge B)$$

4. Axioms restrict the set of beliefs that an agent can hold:

e.g. A and  $\neg A$  cannot both be true.

Just as in logic, the violation of the axioms leads to inconsistent beliefs that may result in nonsensical or wrong conclusions and actions.

### 1.1.3 Random Variables

---

Specific events are often a bit complicated to work with. We can use **random variables** in combination with **atomic events** to map outcomes to atomic events:

**Example: Roulette**

- **Atomic Events:** Numbers 0-36
- **Random Variables:**
  - Red | Black
  - Odd | Even
  - High | Low (1-18 | 19-36)
  - Street | Square | Split
  - Dozens
  - ...

Then for every atomic event we can apply a random variable to check whether it is true (e.g. `red(36) = true`).

The probability of an event  $X$  taking a specific value  $x_i$   $P(X = x_i)$  is obtained by summing the probabilities of all atomic event  $\omega$  that result in  $X(\omega) = x_i$ :

$$P(X = x_i) = \sum_{\omega: X(\omega) = x_i} P(\omega)$$

### 1.1.4 Propositions

---

A proposition is a disjunction of atomic events in which it is true:

$$\begin{aligned} (a \vee b) &\equiv (\neg a \wedge b) \vee (a \wedge \neg b) \vee (a \wedge b) \\ \rightarrow P(a \vee b) &= P(\neg a \wedge b) + P(a \wedge \neg b) + P(a \wedge b) \end{aligned}$$

## Syntax of Propositions

- **Propositional or Boolean** random can be true or false
  - `hasUmbrella = true` is a proposition and can be written as `hasUmbrella`
- **Discrete** random variables (finite or infinite)
  - e.g. Weather is one of {sunny, cloudy, rainy, snow}
  - `Weather = rainy` is a proposition
  - Values must be mutually exclusive and exhaustive (All possible cases are covered by values)
- **Continuous** random variables (bounded or unbounded)
  - e.g. Temperature
  - `Temp = 25.5` , `Temp > 23` are propositions

## 1.2 Joint Distribution

A joint distribution is the probability of combined events e.g. The probability that  $X = x$  and  $Y = y$  is true:

$$P(x, y) \equiv P(X = x \wedge Y = y)$$

Applying this to a whole set we can obtain a **joint probability distribution truth table**:

Smoking	No	Benign	Malignant
No	0.768	0.024	0.008
Few	0.132	0.012	0.006
Many	0.035	0.010	0.005

*Joint Probability Distribution of Smoking and Cancer*

### 1.2.1 Marginalization

Often we don't want to talk about the joint distribution of two events, but get the marginal distribution of one event:

For any set of variables  $X$  and  $Y$  we can compute the probability

$$P(Y) = \sum_{i=1}^n P(x_i, Y)$$

Smoking	No	Benign	Malignant
No	0.768	0.024	0.008
Few	0.132	0.012	0.006
Many	0.035	0.010	0.005

*Joint Probability Distribution of Smoking and Cancer*

So for example:

$$P(Y = \text{few}) = P(\text{no}, \text{few}) + P(\text{benign}, \text{few}) + P(\text{malignant}, \text{few}) = 0.132 + 0.012 + 0.006 = 0.15$$

### 1.2.2 Conditional Probabilities

---

The probability of  $X = x$  under the assumption that  $Y = y$  is true:

$$P(x|y) = \frac{P(x \wedge y)}{P(y)} = \frac{P(x, y)}{P(y)}$$

The product rule yields an alternative representation for the joint probability:

$$P(x, y) = P(x|y) \cdot P(y) = P(y|x) \cdot P(x)$$

Expanding this onto the **Chain rule**:

$$\begin{aligned} P(X_1, \dots, X_n) &= P(X_1, \dots, X_{n-1})P(X_n|X_1, \dots, X_{n-1}) \\ &= P(X_1, \dots, X_{n-2})P(X_{n-1}|X_1, \dots, X_{n-2})P(X_n|X_1, \dots, X_{n-1}) \\ &= \prod_{i=1}^n P(X_i|X_1, \dots, X_{i-1}) \end{aligned}$$

### 1.2.3 Independence

---

$X$  and  $Y$  are independent from another if one of the following is true:

$$\begin{aligned} 1 : & P(X|Y) = P(X) \\ 2 : & P(Y|X) = P(Y) \\ 3 : & P(X|Y) = P(X)P(Y) \end{aligned}$$

Independent variables are not affected by the other variable. This reduces the amount of possible variables. However, absolute independent variables are rare.

### 1.2.4 Bayes Theorem

---

$$\underbrace{P(x|y)}_{\text{Posterior}} = \frac{\overbrace{P(y|x)}^{\text{Likelihood}} \overbrace{P(x)}^{\text{Prior}}}{\underbrace{P(y)}_{\text{Marginalization}}}$$

- **Posterior**: Probability of hypothesis  $X$  after evidence  $Y$
- **Likelihood**: Probability of evidence  $Y$  given hypothesis  $X$  is true
- **Prior**: Probability of the hypothesis  $X$  before considering evidence  $Y$
- **Marginalization**: Probability of evidence  $Y$

## 1.3 Uncertainty in AI

---

Joint distribution is done by enumerating everything:

- **Worst-Case Run Time**:  $O(2^n)$ 
  - $n$  = number of random variables
- **Space Complexity**:  $O(2^n)$  = size of the table

Due to this we mainly use **independencies** to compress the representation.