Alma Mater Studiorum · Università di Bologna

Second Cycle Degree
Artificial Intelligence

Fundamentals of Artificial Intelligence and Knowledge Rapresentation

Student:
Matteo Canghiari

Academic Year 2025/2026

Capitolo 1

Acting under uncertainty

1.1 Basic probability notation

Every agent based on **decision theory** needs a formal language to use and represent probabilistic informations. Typically AI needs a more suited and consistent approach than the traditional probability theory. This section includes all the necessary definitions and examples to understand the subsequent arguments in depth.

Definition

The set of all possible worlds is called the **sample space**, denoted Ω. Any subset $A \subseteq \Omega$ is an **event**. Any element $\omega \in \Omega$ is called **sample point**.

Definition

A **probability space** is a sample space with an assignment $P(\omega)$ for every $\omega \in \Omega$ where:

- $0 \leq P(\omega) \leq 1$
- $\sum P(w) = 1$ for every $\omega \in \Omega$

Definition

A **random variable** is a function from sample points to some range, e.g., the reals or Booleans.

e.g. *Odd(1)* = *true*

Definition

P induces a **probability distribution** for any random variable ***X***:

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

A **probability distribution** gives values for all possible assignment.

Definition

Prior or **unconditional probabilities** of propositions correspond to belief prior to arrival of any new evidence.

e.g. *P(Cavity = True) = 0.1*

Definition

The **Joint Probability Distribution** for a set of random variables gives the probability of every sample point on those random variables.

e.g. ***P****(Weather, Cavity) = a* 2×4 *matrix of values:*

Weather =	*sunny*	*rain*	*cloudy*	*snow*
Cavity = True	0.144	0.02	0.016	0.02
Cavity = False	0.576	0.08	0.064	0.08

Every question about a certain domain can be answered by the joint distribution because every event is a sum of sample points.

Definition

A function $p : R \to R$ is a **probability density function** (**pdf**) for X if it is a nonnegative integrable function s.t.

$$\int_{Val(X)} p(x) dx = 1$$

Definition

Conditional or **posterior probabilities** $P(X|Evidence)$ represent a more informed distribution in the light of new **evidence**.

e.g. $P(cavity|toothache) = 0.8$

It does not mean *"if I have toothache then there is* 80% *of chance that there is also a cavity"*, instead the evidence mean *"given toothache evidence is all I know"*.

The typically definition of conditional or posterior probability is:

$$P(a|b) = \frac{P(a \wedge b)}{P(b)} \text{ if } P(b) \neq 0$$

Otherwise, numerator can be written by the **product rule**:

$$P(a \wedge b) = P(a|b)P(b) = P(b|a)P(a)$$

The product rule at the same time is applied to whole distributions, not only for single values as done previously.

$$\mathbf{P}(Weather, Cavity) = \mathbf{P}(Weather|Cavity)\mathbf{P}(Cavity)$$

1.2 Inference using full join distribution

This paragraph describes a new method to retrieve informations from data, named **probabilistic inference**. It allows the computation of conditional probabilities for query propositions by given evidence. Starting from an example is defined the **full joint distribution** as the knowledge base from which answers to all questions.

Example

e.g. (*Toothache, Cavity, Catch*) is just a domain consisting of three Boolean variables. *Catch* condition occurs when the dentist's steel probe catches in the tooth. Based on the domain, the **full joint distribution** seems like this:

	toothache		$\neg$*toothache*	
	catch	$\neg$*catch*	*catch*	$\neg$*catch*
cavity	0.108	0.012	0.072	0.008
$\neg$*cavity*	0.016	0.064	0.144	0.576

The equation

$$P(\phi) = \sum_{\omega:\omega \models \phi} P(\omega)$$

gives a direct way to calculate probabilities of any assertions, summing up all the possible worlds that satisfy the original proposition.

e.g. $P(toothache) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2$
e.g. $P(cavity \vee toothache) = 0.108+0.012+0.072+0.008+0.016+0.064 = 0.28$

It's also possible compute conditional probabilities:
e.g. $P(\neg cavity|toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)} = \frac{0.016+0.064}{0.2} = 0.4$

Notice that in this calculation the term $P(toothache)$ remains constant, no matter which value of *Cavity* is computed. In fact, it can be viewed as a **normalization constant** for the whole distribution $\mathbf{P}(Cavity|toothache)$, ensuring that the positive and negative case sum up to one, as the second probability axiom requires.

$\mathbf{P}(Cavity|toothache) = \alpha\mathbf{P}(Cavity, toothache)$
$= \alpha[\mathbf{P}(Cavity, toothache, catch) + \mathbf{P}(Cavity, toothache, \neg catch)]$
$= \alpha[\langle 0.108, 0.016\rangle + \langle 0.012, 0.064\rangle]$
$= \alpha\langle 0.12, 0.08\rangle = \langle 0.6, 0.4\rangle$

Definition

The first probability calculated $P(toothache)$ is called **marginalization**, or more simply **summing out**, because it sums up the probabilities for each possible value of the other variables.

Definition

The second one $P(\neg cavity|toothache)$ is named **conditioning**.