Foundations of Artificial Intelligence

46. Uncertainty: Introduction and Quantification

Malte Helmert and Gabriele Röger

University of Basel

May 24, 2017

Uncertainty: Overview

chapter overview:

- 46. Introduction and Quantification
- 47. Representation of Uncertainty

Introduction

Motivation

Uncertainty in our knowledge of the world caused by

- partial observability,
- unreliable information (e.g. from sensors),
- nondeterminism,
- laziness to collect more information,
- ...

Yet we have to act!

Option 1: Try to find solution that works in all possible worlds
$\rightsquigarrow$ often there is no such solution

Option 2: Quantify uncertainty (degree of belief) and maximize expected utility

Example

Have to get from Aarau to Basel to attend a lecture at 9:00.

Different options:

- 7:36–8:12 IR2256 to Basel
- 7:40–7:53 S 23 to Olten, 8:05–9:29 IC 1058 to Basel
- 7:40–7:57 IR 2160 to Olten, 8:05–9:29 IC 1058 to Basel
- 8:13–8:24 RE 4760 to Olten, 8:30–8:55 IR 2310 to Basel
- leave by car at 8:00 and drive approx. 45 minutes
- ...

Different utilities (travel time, cost, slack time, convenience, ...) and different probabilities of actually achieving the goal (traffic jams, accidents, broken trains, missed connections, ...).

Uncertainty and Logical Rules

Example: diagnosing a dental patient's toothache

- $\mathit{toothache} \rightarrow \mathit{cavity}$
  Wrong: not all patients with toothache have a cavity.
- $\mathit{toothache} \rightarrow \mathit{cavity} \vee \mathit{gumproblem} \vee \mathit{abscess} \vee \ldots$
  Almost unlimited list of possible problems.
- $\mathit{cavity} \rightarrow \mathit{pain}$
  Wrong: not all cavities cause pain.

$\rightsquigarrow$ Logic approach not suitable for domain like medical diagnosis.

Instead: Use probabilities to express degree of belief, e.g. there is a 80% chance that the patient with toothache has a cavity.

Probability Theory

Probability Model

Sample space Ω is countable set of possible worlds

Definition

A probability model associates a numerical probability $P(\omega)$ with each possible world such that

$$0 \leq P(\omega) \leq 1 \text{ for every } \omega \in \Omega \text{ and}$$

$$\sum_{\omega \in \Omega} P(\omega) = 1.$$

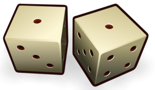
For $\Omega' \subseteq \Omega$ the probability of Ω' is defined as

$$P(\Omega') = \sum_{\omega \in \Omega'} P(\omega).$$

Factored Representation of Possible Worlds

- Possible worlds defined in terms of random variables.
  - variables Die_1 and Die_2 with domain $\{1, \dots, 6\}$ for the values of two dice.
- Describe sets of possible worlds by logical formulas (called propositions) over random variables.
  - $\mathrm{Die}_1 = 1$
  - $(\mathrm{Die}_1 = 2 \lor \mathrm{Die}_1 = 4 \lor \mathrm{Die}_1 = 6) \land (\mathrm{Die}_2 = 2 \lor \mathrm{Die}_2 = 4 \lor \mathrm{Die}_2 = 6)$
  - also use informal descriptions if meaning is clear, e.g. "both values even"

Probability Model: Example



Two dice

- $\Omega = \{\langle 1,1\rangle, \dots, \langle 1,6\rangle, \dots, \langle 6,1\rangle, \dots, \langle 6,6\rangle\}$
- $P(\langle x,y\rangle) = 1/36$ for all $x, y \in \{1, \dots, 6\}$ (fair dice)
- $P(\{\langle 1,1\rangle, \langle 1,2\rangle, \langle 1,3\rangle, \langle 1,4\rangle, \langle 1,5\rangle, \langle 1,6\rangle\}) = 6/36 = 1/6$
- Propositions to describe sets of possible worlds
  - $P(\text{Die}_1 = 1) = 1/6$
  - $P(\text{both values even}) = P(\{\langle 2,2\rangle, \langle 2,4\rangle, \langle 2,6\rangle, \langle 4,2\rangle, \langle 4,4\rangle, \langle 4,6\rangle, \langle 4,2\rangle, \langle 4,4\rangle, \langle 4,6\rangle\}) = 9/36 = 1/4$
  - $P(\text{Total} \geq 11) = P(\{\langle 6,5\rangle, \langle 5,6\rangle, \langle 6,6\rangle\}) = 3/36 = 1/12$

Relationships

The following rules can be derived from the definition of a probability model:

- $P(a \vee b) = P(a) + P(b) - P(a \wedge b)$
- $P(\neg a) = 1 - P(a)$

Probability Distribution

Convention: names of random variables begin with uppercase letters and names of values with lowercase letters.

Random variable *Weather* with

$$P(\mathit{Weather} = \text{sunny}) = 0.6$$
$$P(\mathit{Weather} = \text{rain}) = 0.1$$
$$P(\mathit{Weather} = \text{cloudy}) = 0.29$$
$$P(\mathit{Weather} = \text{snow}) = 0.01$$

Abbreviated: $\mathbf{P}(\mathit{Weather}) = \langle 0.6, 0.1, 0.29, 0.01 \rangle$

A probability distribution **P** is the vector of probabilities for the (ordered) domain of a random variable.

Joint Probability Distribution

For multiple random variables, the joint probability distribution defines values for all possible combinations of the values.

$\mathbf{P}(Weather, Headache)$

	$headache$	$\neg headache$
sunny	$P(sunny \wedge headache)$	$P(sunny \wedge \neg headache)$
rain		
cloudy	...	
snow		

Conditional Probability: Intuition

$P(x)$ denotes the unconditional or prior probability that x will appear in the absence of any other information, e.g.

$$P(cavity) = 0.6.$$

The probability of a cavity increases if we know that a patient has toothache.

$$P(cavity \mid toothache) = 0.8$$

$\rightsquigarrow$ conditional probability (or posterior probability)

Conditional Probability

Definition

The conditional probability for proposition a given proposition b with $P(b) > 0$ is defined as

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

Example: $P(both\ values\ even | Die_2 = 4) = ?$

Product Rule: $P(a \wedge b) = P(a \mid b)P(b)$

Independence

- X and Y are independent if $P(X \wedge Y) = P(X)P(Y)$.
- For independent variables X and Y with $P(Y) > 0$ it holds that $P(X \mid Y) = P(X)$.

Inference from Full Joint Distributions

Full Joint Distribution

full joint distribution: joint distribution for all random variables.

	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

Sum of entries is always 1. (Why?)
Sufficient for calculating the probability of any proposition.

Marginalization

For any sets of variables $\mathbf{Y}$ and $\mathbf{Z}$:

$$\mathbf{P}(\mathbf{Y}) = \sum_{\mathbf{z} \in \mathbf{Z}} \mathbf{P}(\mathbf{Y}, \mathbf{z}),$$

where $\sum_{\mathbf{z} \in \mathbf{Z}}$ means to sum over all possible combinations of values of the variables in $\mathbf{Z}$.

$\mathbf{P}(Cavity) =$ $\rightsquigarrow$ blackboard

Conditioning

To determine conditional probabilities express them as unconditional probabilities and evaluate the subexpressions from the full joint probability distribution.

$$\begin{aligned} P(cavity \mid toothache) &= \frac{P(cavity \wedge toothache)}{P(toothache)} \\ &= \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6 \end{aligned}$$

Normalization: Idea

$$P(cavity \mid toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)} = \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6$$

$$P(\neg cavity \mid toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)} = \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

Term $1/P(toothache)$ remains constant.

Probabilities from complete case analysis always sum up to 1.

Idea: Use normalization constant α instead of constant term.

Normalization: Example

$$\begin{aligned}
\mathbf{P}(\mathit{Cavity} \mid \mathit{toothache}) &= \alpha \mathbf{P}(\mathit{Cavity}, \mathit{toothache}) \\
&= \alpha \big(\mathbf{P}(\mathit{Cavity}, \mathit{toothache}, \mathit{catch}) + \mathbf{P}(\mathit{Cavity}, \mathit{toothache}, \neg \mathit{catch}) \big) \\
&= \alpha \big(\langle 0.108, 0.016 \rangle + \langle 0.012, 0.064 \rangle \big) = \alpha \langle 0.12, 0.08 \rangle = \langle 0.6, 0.4 \rangle
\end{aligned}$$

With normalization, we can compute the probabilities without knowing $P(\mathit{toothache})$.

Full Joint Probability Distribution: Discussion

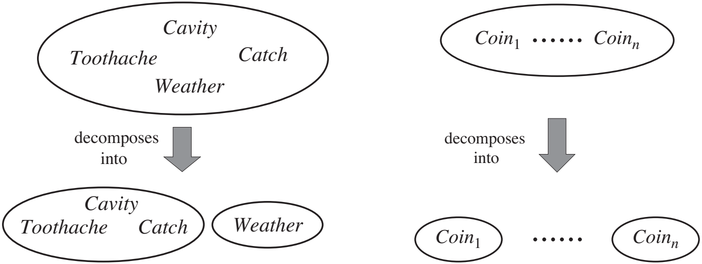
Advantage: Contains all necessary information

Disadvantage: Prohibitively large in practice:
Table for n Boolean variables has size $O(2^n)$.

Good for theoretical foundations, but what to do in practice?

Possible Solution: Factorization

Idea: Exploit independence $\mathbf{P}(X, Y) = \mathbf{P}(X)\mathbf{P}(Y)$ to factorize large joint distribution into smaller distributions.



Problem: Independence is quite rare.
(We will come back to this idea later.)

Bayes' Rule

Bayes' Rule

Product rule: $P(a \wedge b) = P(a \mid b)P(b)$, $P(a \wedge b) = P(b \mid a)P(a)$

Combination gives Bayes' rule

$$P(b \mid a) = \frac{P(a \mid b)P(b)}{P(a)}$$

General version with multivalued variables and conditioned on some background evidence **e**:

$$\mathbf{P}(Y \mid X, \mathbf{e}) = \frac{\mathbf{P}(X \mid Y, \mathbf{e})\mathbf{P}(Y \mid \mathbf{e})}{\mathbf{P}(X \mid \mathbf{e})}$$

Bayes' Rule: Example

Meningitis causes a stiff neck 70% of the time.
The prior probability that a patient has meningitis is 1/50,000, and the prior probability that a patient has a stiff neck is 1%.
What is the probability that a patient with a stiff neck has meningitis?

$$P(s \mid m) = 0.7$$
$$P(m) = 1/50000$$
$$P(s) = 0.01$$
$$P(m \mid s) = \frac{P(s \mid m)P(m)}{P(s)} = \frac{0.7 \cdot 1/50000}{0.01} = 0.0014$$

Summary

Summary

- Uncertainty is inescapable in complex, nondeterministic or partially observable environments.
- Probabilities summarize the agent's beliefs relative to the evidence.
- The full joint probability distribution specifies probabilities for each possible world.
- The full joint probability distribution contains sufficient information to calculate the probability of any proposition.
- An explicit representation of the full joint probability distribution is usually too large, but in the presence of independence it can be factored into smaller distributions.
- With Bayes' rule we can compute unknown probabilities from known conditional probabilities.