General information

- 7 credit units
- Can be included in post-graduate studies
- Lectures (4h per week), 7 + 7 weeks
- Student presentations instead of lectures towards the end
- Course Assignment: Essay, presentation, and (preferably) a software demonstration
- Lectures based on the AIMA course book
Organization & timetable

- Lectures: prof. Tapio Elomaa
- Tue 12–14 TB219 & Thu 12–14 TB223
- Jan. 12 – Apr. 28
  - Period break: Feb. 29 – Mar. 6
  - Easter break: Mar. 23 – 29
- Course assignments guided by M.Sc. Juho Lauri
- Exam: Wed May 4, 2016

Examples of AI Systems

- Boston Dynamics robots
- IBM Watson for Jeopardy!
- Driver assisting systems in VW Golf
13 QUANTIFYING UNCERTAINTY

• In practice agents almost never have full access to the whole truth of their environment and, therefore, must act under uncertainty
• A logical agent may fail to acquire certain knowledge that it would require
• If the agent cannot conclude that any particular course of action achieves its goal, then it will be unable to act
• Conditional planning can overcome uncertainty to some extent, but it does not resolve it
• An agent based solely on logics cannot choose rational actions in an uncertain environment

• Logical knowledge representation requires rules without exceptions
• In practice, we can typically at best provide some degree of belief for a proposition
• In dealing with degrees of belief we will use probability theory
  - Probability 0 corresponds to an unequivocal belief that the sentence is false and, respectively, 1 to an unequivocal belief that the sentence is true
  - Probabilities in between correspond to intermediate degrees of belief in the truth of the sentence, not on its relative truth
• Utilities that have been weighted with probabilities give the agent a chance of acting rationally by preferring the action that yields the highest expected utility
• Principle of Maximum Expected Utility (MEU)
13.2 Basic Probability Notation

- Probabilistic assertions are about possible worlds just like logical assertions
- However, they talk about how probable the various worlds are
- The set of all possible worlds $\Omega$ is called the **sample space**
- The possible worlds are **mutually exclusive** and **exhaustive**
- If we roll two (indistinguishable) dice, there are 36 worlds to consider: \((1,1), (1,2), \ldots, (6,6)\)
- A fully specified probability model associates a numerical probability $P(\omega)$ with each possible world

- The basic axioms of probability theory say that
  - every possible world has a probability between 0 and 1
  
  \[0 \leq P(\omega) \leq 1 \text{ for every } \omega\]

  - the total probability of the set of possible worlds is 1:
  
  \[\sum_{\omega \in \Omega} P(\omega) = 1\]

- Probabilistic assertions and queries are not usually about particular possible worlds, but about sets of them
For example, we might be interested in the cases where the two dice add up to 11

In probability theory these sets are called **events**

In AI the sets are always described by **propositions** in a formal language

The probability associated with a proposition is defined to be the sum of the probabilities of the worlds in which it holds:

- For any proposition \( \varphi \)

\[
P(\varphi) = \sum_{\omega \in \varphi} P(\omega)
\]

### Prior and posterior probability

- Rolling fair dice, we have

\[
P(\text{Total} = 11) = P((5,6)) + P((6,5)) = 1/36 + 1/36 = 1/18
\]

- Probability such as \( P(\text{Total} = 11) \) is called unconditional or **prior probability**

- \( P(a) \) is the degree of belief accorded to proposition \( a \) in the absence of any other information

- Once the agent has obtained some **evidence**, we have to switch to using **conditional (posterior) probabilities**

\[
P(\text{doubles} \mid \text{Die 1} = 5)
\]

- \( P(\text{cavity}) = 0.2 \) is interesting when visiting a dentist for regular checkup, but \( P(\text{cavity} \mid \text{toothache}) = 0.6 \) matters when visiting the dentist because of a toothache
• \( P(\text{cavity}) = P(\text{cavity} \mid \) \\
• We can express conditional probabilities in terms of unconditional probabilities:

\[
P(a \mid b) = \frac{P(a \land b)}{P(b)}
\]

whenever \( P(b) > 0 \)

• E.g.,

\[
P(\text{doubles} \mid \text{Die}_1 = 5) = \frac{P(\text{doubles} \land \text{Die}_1 = 5)}{P(\text{Die}_1 = 5)}
\]

• Rewriting the definition of conditional probability yields the product rule

\[
P(a \land b) = P(a \mid b) \cdot P(b)
\]

• We can, of course, have the rule the other way around

\[
P(a \land b) = P(b \mid a) \cdot P(a)
\]

• A random variable refers to a part of the world, whose status is initially unknown

• Random variables play a role similar to proposition symbols in propositional logic

• E.g., \( \text{Cavity} \) might refer whether the lower left wisdom tooth has a cavity

• The domain of a random variable may be of type
  
  – **Boolean:**
    
    we write \( \text{Cavity} = \text{true} \Leftrightarrow \text{cavity} \) and
    
    \( \text{Cavity} = \text{false} \Leftrightarrow \neg \text{cavity} \);

  – **discrete:** e.g., \( \text{Weather} \) might have the domain \{sunny, rain, cloudy, snow\};

  – **continuous:** then one usually examines the cumulative distribution function; e.g., \( X \leq 4.02 \)
Elementary propositions can be combined to form complex propositions using all the standard connectives: $cavity \land \neg toothache$

An atomic event is a complete specification of the world, i.e., an assignment of values to all the variables.

Properties of atomic events:
- They are mutually exclusive
- The set of all atomic events is exhaustive — at least one must be the case
- Any particular atomic event entails the truth or falsehood of every proposition
- Any proposition is logically equivalent to the disjunction of all atomic events that entail the truth of the proposition

$cavity \equiv (cavity \land toothache) \lor (cavity \land \neg toothache)$

The probability distribution (PD) $P(X)$ of a random variable $X$ is a vector of values for probabilities of the elements in its (ordered) domain.

E.g., when

\[
\begin{align*}
P(sunny) &= 0.02, \\
P(rain) &= 0.2, \\
P(cloudy) &= 0.7 \\
P(snow) &= 0.08, \text{then} \\
P(Weather) &= (0.02, 0.2, 0.7, 0.08)
\end{align*}
\]

Conditional distributions:

$P(X \mid Y) \equiv P(X = x_i \mid Y = y_j) \forall \text{ pair } i, j$

By the product rule

$P(X, Y) = P(X \mid Y) P(Y)$

(entry-by-entry, not a matrix multiplication)
The joint probability distribution of two random variables is the product of their domains. 

E.g., \( P(\text{Weather, Cavity}) \) is a \( 4 \times 2 \) table of probabilities. 

Full joint PD covers the complete set of random variables used to describe the world.

For continuous variables it is not possible to write out the entire distribution as a table, one has to examine probability density functions instead.

Rather than examine point probabilities (that have value 0), we examine probabilities of value ranges.

We will concentrate mostly on discrete-valued random variables.

\[
\begin{align*}
P(W = \text{sunny}, C = \text{true}) &= P(W = \text{sunny}|C = \text{true})P(C = \text{true}) \\
P(W = \text{rain}, C = \text{true}) &= P(W = \text{rain}|C = \text{true})P(C = \text{true}) \\
P(W = \text{cloudy}, C = \text{true}) &= P(W = \text{cloudy}|C = \text{true})P(C = \text{true}) \\
P(W = \text{snow}, C = \text{true}) &= P(W = \text{snow}|C = \text{true})P(C = \text{true}) \\
P(W = \text{sunny}, C = \text{false}) &= P(W = \text{sunny}|C = \text{false})P(C = \text{false}) \\
P(W = \text{rain}, C = \text{false}) &= P(W = \text{rain}|C = \text{false})P(C = \text{false}) \\
P(W = \text{cloudy}, C = \text{false}) &= P(W = \text{cloudy}|C = \text{false})P(C = \text{false}) \\
P(W = \text{snow}, C = \text{false}) &= P(W = \text{snow}|C = \text{false})P(C = \text{false})
\end{align*}
\]
13.2.3 Probability Axioms

- The axiomatization of probability theory by Kolmogorov (1933) based on three simple axioms:

1. For any proposition \( a \) the probability is in between 0 and 1:
   \[ 0 \leq P(a) \leq 1 \]

2. Necessarily true (i.e., valid) propositions have probability 1 and necessarily false (i.e., unsatisfiable) propositions have probability 0:
   \[ P(\text{true}) = 1 \quad P(\text{false}) = 0 \]

3. The probability of a disjunction is given by the inclusion-exclusion principle:
   \[ P(a \lor b) = P(a) + P(b) - P(a \land b) \]

- We can derive a variety of useful facts from the basic axioms; e.g.:
  \[
P(a \lor \neg a) = P(a) + P(\neg a) - P(a \land \neg a) \\
P(\text{true}) = P(a) + P(\neg a) - P(\text{false}) \\
1 = P(a) + P(\neg a) \\
P(\neg a) = 1 - P(a)
\]

- The fact of the third line can be extended for a discrete variable \( D \) with the domain \( d_1, ..., d_n \):
  \[
  \sum_{i=1}^{n} P(D = d_i) = 1
  \]

- For a continuous variable \( X \) the summation is replaced by an integral:
  \[
  \int_{-\infty}^{\infty} P(X = x) \, dx = 1
  \]
The PD on a single variable must sum to 1.

It is also true that any joint PD on any set of variables must sum to 1.

Recall that any proposition \( a \) is equivalent to the disjunction of all the atomic events in which \( a \) holds.

Call this set of events \( e(a) \).

Atomic events are mutually exclusive, so the probability of any conjunction of atomic events is zero, by axiom 2.

Hence, from axiom 3

\[
P(a) = \sum_{e \in e(a)} P(e)
\]

Given a full joint PD that specifies the probabilities of all atomic events, this equation provides a method for computing the probability of any proposition.

### 13.3 Inference Using Full Joint Distribution

<table>
<thead>
<tr>
<th></th>
<th>toothache</th>
<th>¬toothache</th>
</tr>
</thead>
<tbody>
<tr>
<td>catch</td>
<td>0.108</td>
<td>0.012</td>
</tr>
<tr>
<td>¬catch</td>
<td>0.016</td>
<td>0.064</td>
</tr>
</tbody>
</table>

E.g., there are six atomic events for \( cavity \lor\) \( toothache \):

\[
0.108 + 0.012 + 0.072 + 0.008 + 0.016 + 0.064 = 0.28
\]

Extracting the distribution over a variable (or some subset of variables), marginal probability, is attained by adding the entries in the corresponding rows or columns.

E.g., \( P(cavity) = 0.108 + 0.012 + 0.072 + 0.008 = 0.2 \)