

Vorlesung

Grundlagen der

Künstlichen Intelligenz

Reinhard Lafrenz / Prof. A. Knoll

Robotics and Embedded Systems
Department of Informatics – I6
Technische Universität München

www6.in.tum.de

lafrenz@in.tum.de

089-289-18136

Room 03.07.055



Wintersemester 2012/13

11.1.2013



Chapter 13

Reasoning under Uncertainty – Probability Theory

Outline

- Uncertainty
- Probability
- Syntax and Semantics
- Inference
- Independence and Bayes' Rule



Uncertainty

Let action A_t = leave for airport t minutes before flight
Will A_t get me there on time?

Problems:

1. partial observability (road state, other drivers' plans, etc.)
2. noisy sensors (traffic reports)
3. uncertainty in action outcomes (flat tire, etc.)
4. immense complexity of modeling and predicting traffic

Hence a purely logical approach either

1. risks falsehood: “ A_{25} will get me there on time”, or
2. leads to conclusions that are too weak for decision making:

“ A_{25} will get me there on time if there's no accident on the bridge and it doesn't rain and my tires remain intact etc etc.”

(A_{1440} might reasonably be said to get me there on time but I'd have to stay overnight in the airport ...)



Methods for handling uncertainty

- **Default or nonmonotonic logic:**
 - Assume my car does not have a flat tire
 - Assume A_{25} works unless contradicted by evidence
- Issues: What assumptions are reasonable? How to handle contradiction?
- **Rules with fudge factors:**
 - $A_{25} \dashv\rightarrow_{0.3}$ get there on time
 - $Sprinkler \dashv\rightarrow_{0.99} WetGrass$
 - $WetGrass \dashv\rightarrow_{0.7} Rain$
- Issues: Problems with combination, e.g., *Sprinkler causes Rain??*
- **Probability**
 - Model agent's degree of belief
 - Given the available evidence,
 - A_{25} will get me there on time with probability 0.04



Probability

Probabilistic assertions **summarize** effects of

- **laziness**: failure to enumerate exceptions, qualifications, etc.
- **ignorance**: lack of relevant facts, initial conditions, etc.

Subjective probability:

- Probabilities relate propositions to agent's own state of knowledge

$$\text{e.g., } P(A_{25} \mid \text{no reported accidents}) = 0.06$$

These are **not** assertions about the world

Probabilities of propositions change with new evidence:

$$\text{e.g., } P(A_{25} \mid \text{no reported accidents, 5 a.m.}) = 0.15$$



Making decisions under uncertainty

Suppose I believe the following:

$$P(A_{25} \text{ gets me there on time} \mid \dots) = 0.04$$

$$P(A_{90} \text{ gets me there on time} \mid \dots) = 0.70$$

$$P(A_{120} \text{ gets me there on time} \mid \dots) = 0.95$$

$$P(A_{1440} \text{ gets me there on time} \mid \dots) = 0.9999$$

- Which action to choose?

Depends on my **preferences** for missing flight vs. time spent waiting, etc.

- **Utility theory** is used to represent and infer preferences
- **Decision theory** = probability theory + utility theory



Syntax

- Basic element: **random variable**
- Similar to propositional logic: possible worlds defined by assignment of values to random variables.
- **Boolean** random variables
 - e.g., *Cavity* (do I have a cavity?)
- **Discrete** random variables
 - e.g., *Weather* is one of $\langle \text{sunny}, \text{rainy}, \text{cloudy}, \text{snow} \rangle$
 - Domain values must be exhaustive and mutually exclusive
- Elementary proposition constructed by assignment of a value to a random variable: e.g., $\text{Weather} = \text{sunny}, \text{Cavity} = \text{false}$
 - (abbreviated as $\neg \text{cavity}$)
- Complex propositions formed from elementary propositions and standard logical connectives e.g., $\text{Weather} = \text{sunny} \vee \text{Cavity} = \text{false}$



Syntax

- **Atomic event:** A **complete** specification of the state of the world about which the agent is uncertain
- E.g., if the world consists of only two Boolean variables *Cavity* and *Toothache*, then there are 4 distinct atomic events:

Cavity = false \wedge *Toothache = false*

Cavity = false \wedge *Toothache = true*

Cavity = true \wedge *Toothache = false*

Cavity = true \wedge *Toothache = true*

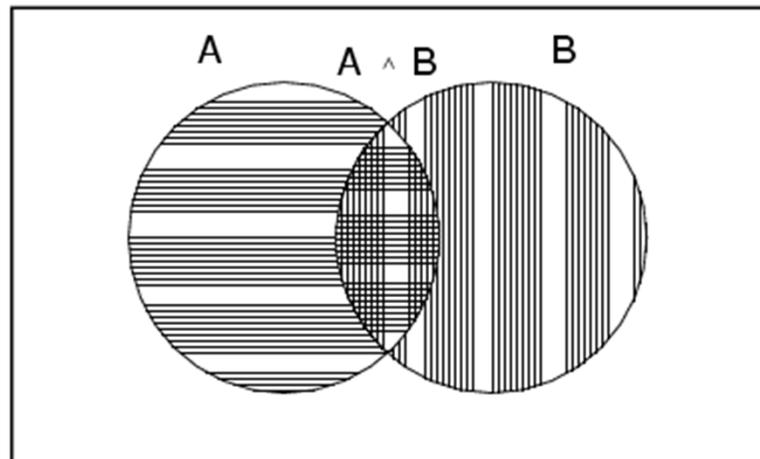
- Atomic events are mutually exclusive and exhaustive



Axioms of probability

- For any propositions A, B
 - $0 \leq P(A) \leq 1$
 - $P(\text{true}) = 1$ and $P(\text{false}) = 0$
 - $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$

True



Prior probability

- Prior or unconditional probabilities of propositions
- e.g., $P(\text{Cavity} = \text{true}) = 0.1$ and $P(\text{Weather} = \text{sunny}) = 0.72$ correspond to belief prior to arrival of any (new) evidence

- Probability distribution gives values for all possible assignments:

$$P(\text{Weather}) = \langle 0.72, 0.1, 0.08, 0.1 \rangle \text{ (normalized, i.e., sums to 1)}$$

- Joint probability distribution for a set of random variables gives the probability of every atomic event on those random variables

$$P(\text{Weather}, \text{Cavity}) = \text{a } 4 \times 2 \text{ matrix of values:}$$

<i>Weather</i> =	sunny	rainy	cloudy	snow
<i>Cavity</i> = true	0.144	0.02	0.016	0.02
<i>Cavity</i> = false	0.576	0.08	0.064	0.08

- Every question about a domain can be answered by the joint distribution



Conditional probability

- Conditional or posterior probabilities

e.g., $P(\text{cavity} \mid \text{toothache}) = 0.8$

i.e., given that *toothache* is all I know

- (Notation for conditional distributions:

$\mathbf{P}(\text{Cavity} \mid \text{Toothache}) = 2\text{-element vector of } 2\text{-element vectors}$)

- If we know more, e.g., *cavity* is also given, then we have

$P(\text{cavity} \mid \text{toothache}, \text{cavity}) = 1$

- New evidence may be irrelevant, allowing simplification, e.g.,

$P(\text{cavity} \mid \text{toothache}, \text{sunny}) = P(\text{cavity} \mid \text{toothache}) = 0.8$

- This kind of inference, sanctioned by domain knowledge, is crucial



Conditional probability

- Definition of conditional probability:
 $P(a | b) = P(a \wedge b) / P(b)$ if $P(b) > 0$
- **Product rule** gives an alternative formulation:
 $P(a \wedge b) = P(a | b) P(b) = P(b | a) P(a)$
- A general version holds for whole distributions, e.g.,
 $\mathbf{P}(\textit{Weather}, \textit{Cavity}) = \mathbf{P}(\textit{Weather} | \textit{Cavity}) \mathbf{P}(\textit{Cavity})$

(View as a set of 4×2 equations, **not** matrix mult.)

- **Chain rule** is derived by successive application of product rule:

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



Inference by enumeration

- Start with the joint probability distribution:

	<i>toothache</i>		\neg <i>toothache</i>	
	<i>catch</i>	\neg <i>catch</i>	<i>catch</i>	\neg <i>catch</i>
<i>cavity</i>	.108	.012	.072	.008
\neg <i>cavity</i>	.016	.064	.144	.576

- For any proposition φ , sum the atomic events where it is true: $P(\varphi) = \sum_{\omega: \omega \models \varphi} P(\omega)$



Inference by enumeration

- Start with the joint probability distribution:

	<i>toothache</i>		\neg <i>toothache</i>	
	<i>catch</i>	\neg <i>catch</i>	<i>catch</i>	\neg <i>catch</i>
<i>cavity</i>	.108	.012	.072	.008
\neg <i>cavity</i>	.016	.064	.144	.576

- For any proposition φ , sum the atomic events where it is true: $P(\varphi) = \sum_{\omega: \omega \models \varphi} P(\omega)$
- $P(\textit{toothache}) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2$



Inference by enumeration

- Start with the joint probability distribution:

	<i>toothache</i>		\neg <i>toothache</i>	
	<i>catch</i>	\neg <i>catch</i>	<i>catch</i>	\neg <i>catch</i>
<i>cavity</i>	.108	.012	.072	.008
\neg <i>cavity</i>	.016	.064	.144	.576

- For any proposition φ , sum the atomic events where it is true: $P(\varphi) = \sum_{\omega:\omega \models \varphi} P(\omega)$
- $P(\textit{toothache}) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2$



Inference by enumeration

- Start with the joint probability distribution:

	<i>toothache</i>		\neg <i>toothache</i>	
	<i>catch</i>	\neg <i>catch</i>	<i>catch</i>	\neg <i>catch</i>
<i>cavity</i>	.108	.012	.072	.008
\neg <i>cavity</i>	.016	.064	.144	.576

- Can also compute conditional probabilities:

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



Normalization

	<i>toothache</i>		\neg <i>toothache</i>	
	<i>catch</i>	\neg <i>catch</i>	<i>catch</i>	\neg <i>catch</i>
<i>cavity</i>	.108	.012	.072	.008
\neg <i>cavity</i>	.016	.064	.144	.576

- Denominator can be viewed as a **normalization constant** α

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

General idea: compute distribution on query variable by fixing **evidence variables** and summing over **hidden variables**



Inference by enumeration, cont'd.

Typically, we are interested in

- the posterior joint distribution of the **query variables** \mathbf{Y}
- given specific values \mathbf{e} for the **evidence variables** \mathbf{E}

Let the **hidden variables** be $\mathbf{H} = \mathbf{X} - \mathbf{Y} - \mathbf{E}$

Then the required summation of joint entries is done by summing out the hidden variables:

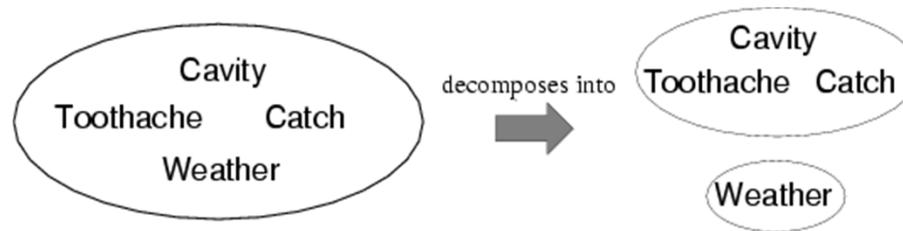
$$\mathbf{P}(\mathbf{Y} \mid \mathbf{E} = \mathbf{e}) = \alpha \mathbf{P}(\mathbf{Y}, \mathbf{E} = \mathbf{e}) = \alpha \sum_{\mathbf{h}} \mathbf{P}(\mathbf{Y}, \mathbf{E} = \mathbf{e}, \mathbf{H} = \mathbf{h})$$

- The terms in the summation are joint entries because \mathbf{Y} , \mathbf{E} and \mathbf{H} together exhaust the set of random variables
- Obvious problems:
 1. Worst-case time complexity $O(d^n)$ where d is the largest arity
 2. Space complexity $O(d^n)$ to store the joint distribution
 3. How to find the numbers for $O(d^n)$ entries?



Independence

- A and B are independent iff
 $\mathbf{P}(A/B) = \mathbf{P}(A)$ or $\mathbf{P}(B/A) = \mathbf{P}(B)$ or $\mathbf{P}(A, B) = \mathbf{P}(A) \mathbf{P}(B)$



$$\begin{aligned} & \mathbf{P}(\textit{Toothache}, \textit{Catch}, \textit{Cavity}, \textit{Weather}) \\ &= \mathbf{P}(\textit{Toothache}, \textit{Catch}, \textit{Cavity}) \mathbf{P}(\textit{Weather}) \end{aligned}$$

- 32 entries reduced to 12; for n independent biased coins,

$$O(2^n) \rightarrow O(n)$$

- Absolute independence powerful but rare
- Dentistry is a large field with hundreds of variables, none of which are independent. What to do?



Conditional independence

- $\mathbf{P}(\textit{Toothache}, \textit{Cavity}, \textit{Catch})$ has $2^3 - 1 = 7$ independent entries
- If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:

$$(1) \mathbf{P}(\textit{catch} \mid \textit{toothache}, \textit{cavity}) = \mathbf{P}(\textit{catch} \mid \textit{cavity})$$

- The same independence holds if I haven't got a cavity:

$$(2) \mathbf{P}(\textit{catch} \mid \textit{toothache}, \neg \textit{cavity}) = \mathbf{P}(\textit{catch} \mid \neg \textit{cavity})$$

- *Catch* is **conditionally independent** of *Toothache* given *Cavity*:

$$\mathbf{P}(\textit{Catch} \mid \textit{Toothache}, \textit{Cavity}) = \mathbf{P}(\textit{Catch} \mid \textit{Cavity})$$

- Equivalent statements:

$$\mathbf{P}(\textit{Toothache} \mid \textit{Catch}, \textit{Cavity}) = \mathbf{P}(\textit{Toothache} \mid \textit{Cavity})$$

$$\mathbf{P}(\textit{Toothache}, \textit{Catch} \mid \textit{Cavity}) = \mathbf{P}(\textit{Toothache} \mid \textit{Cavity}) \mathbf{P}(\textit{Catch} \mid \textit{Cavity})$$



Conditional independence cont'd.

- Write out full joint distribution using chain rule:

$$\mathbf{P}(\textit{Toothache}, \textit{Catch}, \textit{Cavity})$$

$$= \mathbf{P}(\textit{Toothache} \mid \textit{Catch}, \textit{Cavity}) \mathbf{P}(\textit{Catch}, \textit{Cavity})$$

$$= \mathbf{P}(\textit{Toothache} \mid \textit{Catch}, \textit{Cavity}) \mathbf{P}(\textit{Catch} \mid \textit{Cavity}) \mathbf{P}(\textit{Cavity})$$

$$= \mathbf{P}(\textit{Toothache} \mid \textit{Cavity}) \mathbf{P}(\textit{Catch} \mid \textit{Cavity}) \mathbf{P}(\textit{Cavity})$$

I.e., $2 + 2 + 1 = 5$ independent numbers

- In most cases, the use of conditional independence reduces the size of the representation of the joint distribution from exponential in n to linear in n .
- Conditional independence is our most basic and robust form of knowledge about uncertain environments.



Bayes' Rule

- Product rule $P(a \wedge b) = P(a | b) P(b) = P(b | a) P(a)$

⇒ Bayes' rule: $P(a | b) = P(b | a) P(a) / P(b)$

- or in distribution form

$$P(Y|X) = P(X|Y) P(Y) / P(X) = \alpha P(X|Y) P(Y)$$

- Useful for assessing **diagnostic** probability from **causal** probability:

- $P(\text{Cause}|\text{Effect}) = P(\text{Effect}|\text{Cause}) P(\text{Cause}) / P(\text{Effect})$

- E.g., let M be meningitis, S be stiff neck:

$$P(m|s) = P(s|m) P(m) / P(s) = 0.8 \times 0.0001 / 0.1 = 0.0008$$

- Note: posterior probability of meningitis still very small!



Bayes' Rule and conditional independence

$$\begin{aligned} & \mathbf{P}(\text{Cavity} \mid \text{toothache} \wedge \text{catch}) \\ &= \alpha \mathbf{P}(\text{toothache} \wedge \text{catch} \mid \text{Cavity}) \mathbf{P}(\text{Cavity}) \\ &= \alpha \mathbf{P}(\text{toothache} \mid \text{Cavity}) \mathbf{P}(\text{catch} \mid \text{Cavity}) \mathbf{P}(\text{Cavity}) \end{aligned}$$

- This is an example of a **naïve Bayes** model:

$$\mathbf{P}(\text{Cause}, \text{Effect}_1, \dots, \text{Effect}_n) = \mathbf{P}(\text{Cause}) \prod_i \mathbf{P}(\text{Effect}_i \mid \text{Cause})$$



- Total number of parameters is **linear** in n



Summary

- Probability is a rigorous formalism for uncertain knowledge
- Joint probability distribution specifies probability of every atomic event
- Queries can be answered by summing over atomic events
- For nontrivial domains, we must find a way to reduce the joint size
- Independence and conditional independence provide the tools

