Uncertainty

Chapter 13

Uncertain Knowledge

- Agents don't have complete knowledge about the world
- Agents need to make decisions based on their uncertainty
- It isn't enough to assume what the world is like

  Example: wearing a seat belt
- An agent needs to reason about its uncertainty
- When an agent makes an action under uncertainty, it is gambling ➔probability

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 …)

Probability

- Probability is an agent's measure of belief in some proposition → subjective probability

- **Example:** Your probability of a bird flying is your measure of belief in the flying ability of an individual based only on the knowledge that the individual is a bird
  - Other agents may have different probabilities, as they may have had different experiences with birds or different knowledge about this particular bird.
  - An agent's belief in a bird's flying ability is a affected by what the agent knows about that bird

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

  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:

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

Making decisions under uncertainty

Suppose I believe the following:

$P(A_{25}$ gets me there on time | …) = 0.04
$P(A_{90}$ gets me there on time | …) = 0.70
$P(A_{120}$ gets me there on time | …) = 0.95
$P(A_{1440}$ gets me there on time | …) = 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

Numerical Measures of Belief

- Belief in proposition, f , can be measured in terms of a number between 0 and 1 – this is the probability off
  - The probability f is 0 means that f is believed to be definitely false
  - The probability f is 1 means that f is believed to be definitely true
- Using 0 and 1 is purely a convention
- f has a probability between 0 and 1, doesn't mean f is true to some degree, but means you are ignorant of its truth value. Probability is a measure of your ignorance

Random Variables

- A random variable is a term in a language that can take one of a number of different values
- The **domain** of a variable *X*, written *dom(X)*, is the set of values X can take
- A tuple of random variables $\langle X_1,\ldots X_n \rangle$ is a complex random variable with domain $\langle dom(X_1) \times \ldots \times dom(X_n) \rangle$

Often the tuple is written as $X_1,\ldots X_n$

- Assignment **X = x** means variable *X* has value *x*
- A ***proposition*** is a Boolean formula made from assignments of values to variables

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 <*sunny,rainy,cloudy,snow*>
- Domain values must be exhaustive and mutually exclusive
- Elementary proposition constructed by assignment of a value to a random variable: e.g., *Weather = sunny*, *Cavity = false* (abbreviated as $\neg$*cavity*)
- Complex propositions formed from elementary propositions and standard logical connectives e.g., *Weather = sunny* $\vee$ *Cavity = 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:

  $Cavity = false \wedge Toothache = false$

  $Cavity = false \wedge Toothache = true$

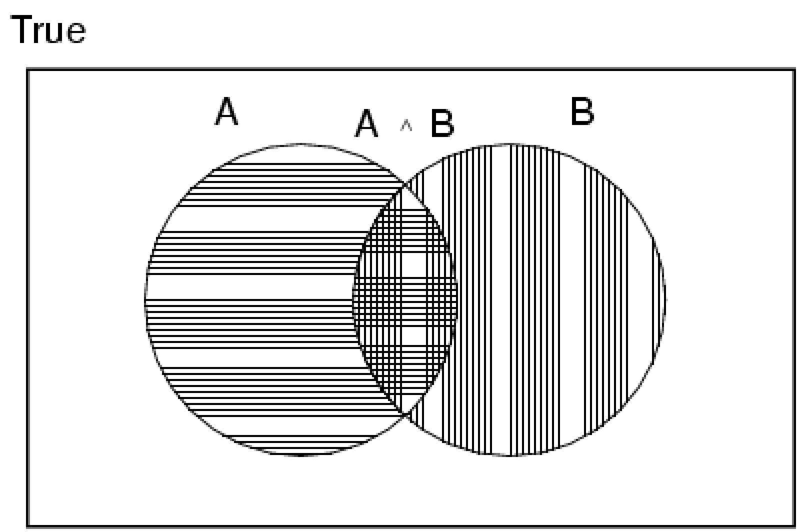
  $Cavity = true \wedge Toothache = false$

  $Cavity = true \wedge Toothache = true$

- Atomic events are mutually exclusive and exhaustive

Axioms of probability

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



Axioms of probability

Three axioms define what follows from a set of probabilities:

1. $0 \leq P(f)$ for any formula f
2. $P(\tau) = 1$ if τ is a tautology
3. $P(A \vee B) = P(A) + P(B)$ if $\neg(A \wedge B)$ is a tautology

These axioms are sound and complete with respect to the semantics

Prior probability

- Prior or unconditional probabilities of propositions

e.g., P(*Cavity* = true) = 0.1 and P(*Weather* = sunny) = 0.72 correspond to belief prior to arrival of any (new) evidence

Probability distribution gives values for all possible assignments:

- **P**(*Weather*) = <0.72,0.1,0.08,0.1> (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(*Weather,Cavity*) = a 4 × 2 matrix of values:

Weather =	sunny	rainy	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 domain can be answered by the joint distribution

Conditioning

- Probabilistic conditioning specifies how to revise beliefs based on new information
- You build a probabilistic model taking all background information into account. This gives the prior probability
- All other information must be conditioned on
- If evidence e is the all of the information obtained subsequently, the conditional probability $P(h|e)$ of h given e is the posterior probability of h

Conditional probability

- Conditional or posterior probabilities

e.g., P(*cavity* | *toothache*) = 0.8 i.e., given that *toothache* is all I know

- (Notation for conditional distributions:

P(*Cavity* | *Toothache*) = 2-element vector of 2-element vectors)

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

P(*cavity* | *toothache,cavity*) = 1

- New evidence may be irrelevant, allowing simplification,

e.g., P(*cavity* | *toothache, sunny*) = P(*cavity* | *toothache*) = 0.8

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

Conditional probability

- Definition of conditional probability:
  $P(a \mid b) = P(a \wedge b) / P(b)$ if $P(b) > 0$

- Product rule gives an alternative formulation:
  $P(a \wedge b) = P(a \mid b) P(b) = P(b \mid a) P(a)$

- A general version holds for whole distributions, e.g.,
  $\mathbf{P}(Weather, Cavity) = \mathbf{P}(Weather \mid Cavity)\, \mathbf{P}(Cavity)$

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

Semantics of conditional probability

Evidence e rules out possible worlds incompatible with e.
Evidence e induces a new measure, μ_e, over possible worlds

$$\mu_e(S) = \begin{cases} c \times \mu(S) & \text{if } \omega \models e \text{ for all } \omega \in S \\ 0 & \text{if } \omega \not\models e \text{ for all } \omega \in S \end{cases}$$

We can show $c = \frac{1}{P(e)}$.
The conditional probability of formula h given evidence e is

$$\begin{aligned} P(h|e) &= \mu_e(\{\omega : \omega \models h\}) \\ &= \frac{P(h \wedge e)}{P(e)} \end{aligned}$$

Bayes Theorem

The chain rule and commutativity of conjunction ($h \wedge e$ is equivalent to $e \wedge h$) gives us:

$$\begin{aligned} P(h \wedge e) &= P(h|e) \times P(e) \\ &= P(e|h) \times P(h). \end{aligned}$$

If $P(e) \neq 0$, you can divide the right hand sides by $P(e)$:

$$P(h|e) = \frac{P(e|h) \times P(h)}{P(e)}.$$

This is Bayes' theorem.

Why Bayes Theorem?

- Often you have causal knowledge
  - P(symptom | disease)
  - P(light is off | state of switch)
  - P(alarm | fire)
  - P(image looks like  | tree in front of a car)

- But want to do evidential reasoning
  - P(disease | symptom)
  - P(state of switch | light is off)
  - P(fire | alarm)
  - P(tree in front of a car | image looks like )

Chain Rule

$$
\begin{aligned}
&P(f_1 \wedge f_2 \wedge \ldots \wedge f_n) \\
&= P(f_n | f_1 \wedge \cdots \wedge f_{n-1}) \times \\
&\quad P(f_1 \wedge \cdots \wedge f_{n-1}) \\
&= P(f_n | f_1 \wedge \cdots \wedge f_{n-1}) \times \\
&\quad P(f_{n-1} | f_1 \wedge \cdots \wedge f_{n-2}) \times \\
&\quad P(f_1 \wedge \cdots \wedge f_{n-2}) \\
&= P(f_n | f_1 \wedge \cdots \wedge f_{n-1}) \times \\
&\quad P(f_{n-1} | f_1 \wedge \cdots \wedge f_{n-2}) \\
&\quad \times \cdots \times P(f_3 | f_1 \wedge f_2) \times P(f_2 | f_1) \times P(f_1) \\
&= \prod_{i=1}^{n} P(f_i | f_1 \wedge \cdots \wedge f_{i-1})
\end{aligned}
$$

Inference by enumeration

- Start with the joint probability distribution:

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	**.108**	**.012**	**.072**	**.008**
¬ cavity	**.016**	**.064**	**.144**	**.576**

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

Inference by enumeration

- Start with the joint probability distribution:

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

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

Inference by enumeration

- Start with the joint probability distribution:

	toothache		$\neg$ *toothache*	
	catch	$\neg$ *catch*	*catch*	$\neg$ *catch*
cavity	.108	.012	.072	.008
$\neg$ *cavity*	.016	.064	.144	.576

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

Inference by enumeration

- Start with the joint probability distribution:

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

- Can also compute conditional probabilities:

$$P(\neg cavity \mid toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)}$$

$$= \frac{0.016+0.064}{0.108 + 0.012 + 0.016 + 0.064}$$

$$= 0.4$$

Normalization

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

- Denominator can be viewed as a normalization constant α

P(*Cavity* | *toothache*) = α, **P**(*Cavity,toothache*)

= α, [**P**(*Cavity,toothache,catch*) + **P**(*Cavity,toothache,*¬ *catch*)]

= α, [<0.108,0.016> + <0.012,0.064>]

= α, <0.12,0.08> = <0.6,0.4>

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

Inference by enumeration, contd.

Typically, we are interested in
the posterior joint distribution of the query variables **Y**
given specific values **e** for the evidence variables **E**

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

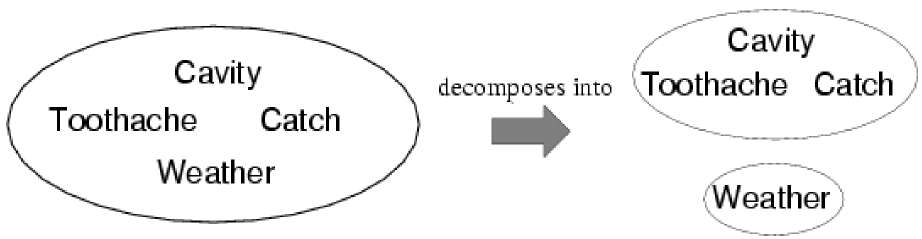
Then the required summation of joint entries is done by summing out the hidden variables:

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

- The terms in the summation are joint entries because **Y**, **E** and **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 $\mathbf{P}(A, B) = \mathbf{P}(A)\,\mathbf{P}(B)$



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

- 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

Random variable X is independent of random variable Y given random variable Z if, for all $x_i \in dom(X)$, $y_j \in dom(Y)$, $y_k \in dom(Y)$ and $z_m \in dom(Z)$,

$$\begin{aligned} & P(X = x_i | Y = y_j \wedge Z = z_m) \\ = \quad & P(X = x_i | Y = y_k \wedge Z = z_m) \\ = \quad & P(X = x_i | Z = z_m). \end{aligned}$$

That is, knowledge of Y's value doesn't affect your belief in the value of X, given a value of Z.

Conditional independence

- $\mathbf{P}(Toothache, Cavity, Catch)$ 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) $\mathbf{P}(catch \mid toothache, cavity) = \mathbf{P}(catch \mid cavity)$

- The same independence holds if I haven't got a cavity:
  (2) $\mathbf{P}(catch \mid toothache, \neg cavity) = \mathbf{P}(catch \mid \neg cavity)$

- *Catch* is conditionally independent of *Toothache* given *Cavity*:
  $\mathbf{P}(Catch \mid Toothache, Cavity) = \mathbf{P}(Catch \mid Cavity)$

- Equivalent statements:
  $\mathbf{P}(Toothache \mid Catch, Cavity) = \mathbf{P}(Toothache \mid Cavity)$
  $\mathbf{P}(Toothache, Catch \mid Cavity) = \mathbf{P}(Toothache \mid Cavity)\, \mathbf{P}(Catch \mid Cavity)$

Conditional independence contd.

- Write out full joint distribution using chain rule:

  P(*Toothache, Catch, Cavity*)

  = **P**(*Toothache* | *Catch, Cavity*) **P**(*Catch, Cavity*)

  = **P**(*Toothache* | *Catch, Cavity*) **P**(*Catch* | *Cavity*) **P**(*Cavity*)

  = **P**(*Toothache* | *Cavity*) **P**(*Catch* | *Cavity*) **P**(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 Revisited

- Product rule $P(a \wedge b) = P(a \mid b)\, P(b) = P(b \mid a)\, P(a)$
  $\Rightarrow$ Bayes' rule: $P(a \mid b) = P(b \mid a)\, P(a) / P(b)$

- or in distribution form
  $$\mathbf{P}(Y|X) = \mathbf{P}(X|Y)\, \mathbf{P}(Y) / \mathbf{P}(X) = \alpha \mathbf{P}(X|Y)\, \mathbf{P}(Y)$$

- Useful for assessing diagnostic probability from causal probability:
  - $P(Cause|Effect) = P(Effect|Cause)\, P(Cause) / P(Effect)$
  - 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

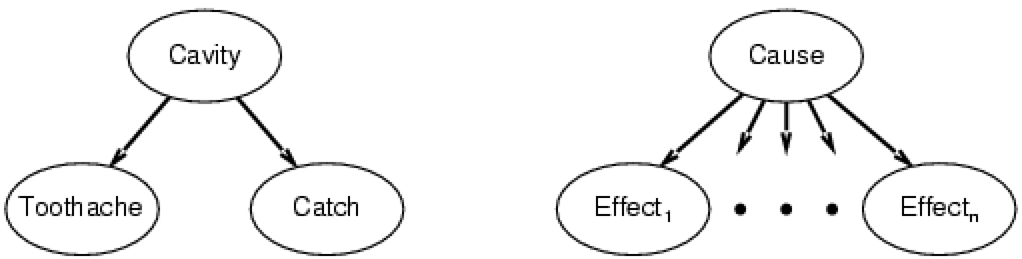
$\mathbf{P}(Cavity \mid toothache \wedge catch)$

$= \alpha\, \mathbf{P}(toothache \wedge catch \mid Cavity)\, \mathbf{P}(Cavity)$

$= \alpha\, \mathbf{P}(toothache \mid Cavity)\, \mathbf{P}(catch \mid Cavity)\, \mathbf{P}(Cavity)$

- This is an example of a naïve Bayes model:

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



- Total number of parameters is linear in n

Belief (Bayes) Nets

Totally order the variables of interest: $X_1, \ldots, X_n$

Theorem of probability theory (chain rule):
$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | X_1, \ldots, X_{i-1})$

The parents $parents(X_i)$ of X_i are those predecessors of X_i that render X_i independent of the other predecessors. That is, $parents(X_i) \subseteq X_1, \ldots, X_{i-1}$ and
$P(X_i | parents(X_i)) = P(X_i | X_1, \ldots, X_{i-1})$

So $P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | parents(X_i))$

A belief network is a graph: the nodes are random variables; there is an arc from the parents of each node into that node.

Belief Networks

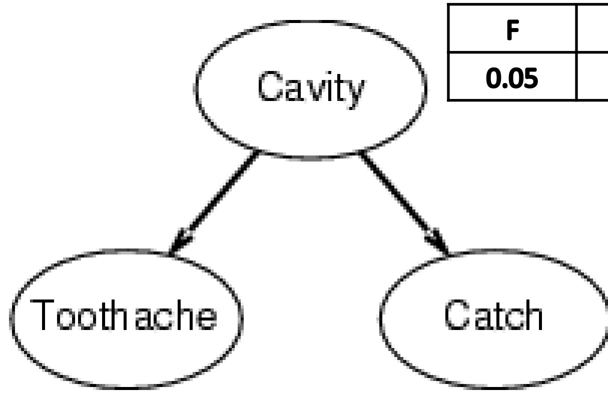
A belief network consists of:

- a directed acyclic graph with nodes labeled with random variables
- a domain for each random variable
- a set of conditional probability tables for each variable given its parents (including prior probabilities for nodes with no parents).

Belief Nets

P(Cavity)

F	T
0.05	0.95



P(Toothache|Cavity)

	F	T
F	0.2	0.8
T	0.05	0.95

P(Catch|Cavity)

	F	T
F	0.1	0.9
T	0.2	0.8

The distribution over cavity is called the ***prior*** distribution

The other two distributions are called ***conditional*** distributions

Joint distribution over the variables is the product of the conditionals (follows from chain rule)

P(Cavity,Toothache,Catch) = P(Cavity) x P(Toothache|Cavity) x P(Catch |Cavity)

Belief Network Summary

- A belief network is automatically acyclic by construction.
- A belief network is a directed acyclic graph (DAG) where nodes are random variables.
- The parents of a node n are those variables on which n directly depends.
- A belief network is a graphical representation of dependence and independence:
  - A variable is independent of its non-descendants given its parents.