CSCI 561 - Foundation for Artificial Intelligence

Discussion Section (Week 13) Reason about uncertainties

PROF WEI-MIN SHEN SHEN@ISI.EDU

Uncertainty:
Coping with What You Don't Grasp

Outside scope of awareness
Too complex to reason about in complete detail
Too expensive or risky to ensure certainty
Problem/information has inherent randomness

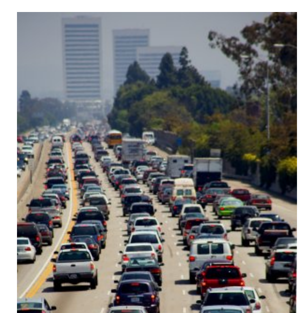
Example
When to Leave for Airport?

Requires understanding how long it is *likely* to take to get there (*mean*), and how *likely* there is to be a significant deviation from this expectation (*variance*)

Decision making under significant uncertainty

A range of possible causes for this uncertainty

- Partial observability
  - State of roads, other drivers' plans, etc.
- Noisy sensors (introducing errors, not just lack of info)
  - Traffic reports, etc.
- Uncertainty in action outcomes
  - Flat tire, accident, etc.
- Uncertainty about how traffic will evolve over time
  - Complexity of modeling and predicting traffic



Logic and Uncertainty

Logic generally talks in terms of *certainty*

- A fact is either *true* or *false*
  - *A*lthough may be *unknown* as a very generic form of uncertainty
- A rule applies to all cases or not
- But often need more shades of meaning than this
  - Risk either stating things as true that are false or leaving as unknown something about which key information is known

Need at times to be able to talk instead about how (un)certain you are about something

Probability

Where is probability?

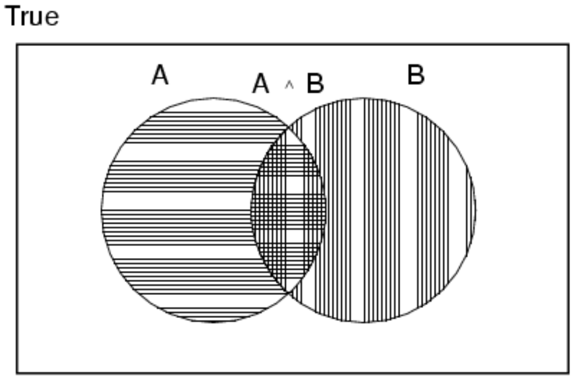
“*Probability* is the chance that something is likely to happen or be the case.” (Wikipedia)

- This is *objective probability:* P(*heads*)=P(*tails*)=1/2
- Our focus will be on *subjective probability:*
  - An agent’s estimated likelihood of the truth of a sentence
- Modal/meta-level knowledge about a sentence, rather than knowledge directly about the world
- E.g., “It will take 25 minutes to get to the airport” vs. ‘I believe that the chance is .35 that the sentence “It will take 25 minutes to get to the airport” is true’

Basic Concepts of Probability

P(x) is a real number in [0, 1]
that represents the likelihood that x is true

- P(x)=1 $\Rightarrow$ Certainty
  - P(*true*)=1
- P(x)=0 $\Rightarrow$ Impossibility
  - P(*false*)=0
- 0<P(x)<1 $\Rightarrow$ Uncertainty (varying levels)
  - For a fair coin, P(heads)=P(tails)=.5
  - For a fair die, P(1)=…=P(6)=.166…
- $P(A \wedge B) = P(A)P(B|A) = P(B)P(A|B)$ // remember this!
- $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$

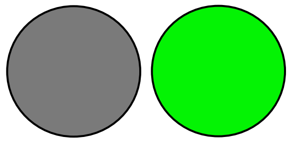


More on Probability

- **Product Rule: $P(xy)=P(x)P(y|x)=P(y)P(x|y)$**
  - **Remember this product rule, if nothing else!**

- **$P(x \wedge y)=P(x)P(y)$ if *x* and *y* are *independent***
  - $P(heads^1 \wedge heads^2)=P(heads^1)P(heads^2)=.5*.5=.25$
  - $P(AinCourse \wedge Raining)=P(AinCourse)P(Raining)$
  - *No general way to compute when not independent, except…*

- **$P(x \wedge y)=0$ if *x* and *y* are *mutually exclusive***
  - $P(AinCourse \wedge BinCourse)=0$
  - $P(x \wedge \neg x)=0$
  - Makes $P(x \vee y)=P(x)+P(y)$



More on Probability

P(*x*)=P(*y*) if *x*=*y* (logically equivalent)

$P(x_1 \vee \ldots \vee x_n)=1$ if the x_i's are *exhaustive*

- $P(heads \vee tails)=1$
- $P(x \vee \neg x)=1$

$P(\neg x)=1\text{-}P(x)$

- $1=P(x \vee \neg x)=P(x)+P(\neg x)-P(x \wedge \neg x)$
- $1=P(x)+P(\neg x)-0$
- *In general, summing probabilities over any partition yields 1*

Atomic Events

An *atomic event* is a *complete* specification of state of the world about which agent is uncertain

- E.g., if the world consists of only two Boolean random variables, *Cavity* and *Toothache*, then there are 4 distinct atomic events:
  1. *Cavity=false* ∧ *Toothache=false*
  2. *Cavity=false* ∧ *Toothache=true*
  3. *Cavity=true* ∧ *Toothache=false*
  4. *Cavity=true* ∧ *Toothache=true*
- *Atomic events are analogous to models in logic*

Each atomic event entails the truth of every proposition with respect to that world

- E.g., in AE4, *cavity*⇒*toothache* is *true*

The set of possible atomic events is a *partition*

The Joint Probability Distribution

A *joint probability distribution* for a set of random variables gives the probability of every atomic event on those random variables

P(*Weather,AinClass*) = a 4 × 2 matrix of values:

***Weather* =**	***sunny***	***rainy***	***cloudy***	***snow***
AinClass* = *true	**0.144**	**0.02**	**0.016**	**0.02**
AinClass* = *false	**0.576**	**0.08**	**0.064**	**0.08**

- **A *full joint probability distribution* covers all of the random variables used to describe the world**
  - *Weather*, *Temperature*, *Cavity*, *Toothache*, *AinClass*, etc.
  - Every question about world is answerable by the full joint distribution

Probabilistic Inference by Enumeration

Can perform probabilistic inference by enumerating over (full) joint probability distribution

Start with the joint probability distribution P(*Cavity*,*Toothache*,*Catch*)

Catch refers to Dentist's instrument catching on tooth

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

For any proposition φ, sum *p* for atomic events where true

- $P(\varphi) = \Sigma_{\omega:\omega \models \varphi} P(\omega)$
- P(*toothache*)
  = .108 + .012 + .016 + .064 = .2
- P(cavity)
  = .108 + .012 + .72 +.08 = .2

Called *marginal probability* of toothache

The overall process is called *marginalization* or *summing out*: [$P(Y) = \Sigma_z P(Y,z)$]

Inference of Complex Propositions

Sum over all action events in which proposition is true

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

- P(*toothache* ∨ *cavity*)
  = .108 + .012 + .016 + .064 + .072 + .008 = .28
- P(*toothache* ∧ *cavity*)
  = .108 + .012 = .12
- P(*toothache* ⇒ *cavity*) = P(¬*toothache* ∨ *cavity)*
  = .108 +.012 +.072 + .008 + .144 + .576 = .92
- P(*cavity* ⇒ *toothache*) = P(¬*cavity* ∨ *toothache)*
  = .108 +.012 +.016 + .064 + .144 + .576 = .92

Probabilities and Evidence

Degrees of belief (subjective probabilities) are centrally affected by *evidence*

- *Prior probability* refers to degree of belief before see any evidence
  - P(*sunny*)=.5
  - P(*aincourse*)=.45
- *Conditional probability* refers to degree of belief conditioned on seeing (only) particular evidence
  - P(*sunny* | *losangeles*, *july*)=.9
  - P(*sunny* | *pittsburgh*)=.1
  - P(*aincourse* | *aonmidterm*)=.6

Inference w/Conditional Probabilities

Can compute conditional probabilities by enumerating over joint distribution via: P(*a* | *b*) = P(*a* $\wedge$ *b*) / P(*b*)

- P(cavity | toothache)
  - = P(cavity $\wedge$ toothache)/P(toothache)
  - = (.108 + .012)/(.108 + .012 + .016 + .064) = .12/.2 = .6
- P($\neg$cavity | toothache)
  - = P($\neg$cavity $\wedge$ toothache)/P(toothache)
  - = (.016 + .064)/(.108 + .012 + .016 + .064) = .08/.2 = .4

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

1/P(*toothache*) is the *normalization constant* (α) for the distribution **P**(*Cavity* | *toothache*)

- Ensures it sums to 1

Normalized Inference Procedure

General inference procedure for conditional distributions:

- $\mathbf{P}(X \mid \mathbf{e}) = \mathbf{P}(X,\mathbf{e})/\mathrm{P}(\mathbf{e}) = \alpha\mathbf{P}(X,\mathbf{e}) = \alpha\Sigma_{y}\mathbf{P}(X,\mathbf{e},\mathbf{y})$
  - X is a single query variable
  - **e** is a vector of values of the evidence variables **E**
  - **y** is a vector of values for the remaining unobserved variables **Y**
  - **P**(X,**e**,**y**) is a subset of values from the joint probability distribution

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

P(*Cavity* | toothache)

$= \mathbf{P}(Cavity,\text{toothache})/\mathrm{P}(\text{toothache})$

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

$= \alpha[\mathbf{P}(Cavity,\text{toothache},\text{catch})+\mathbf{P}(Cavity,\text{toothache},\neg\text{catch})]$

$= (1/.2)[\langle .108, 0.016\rangle + \langle .012, .064\rangle] \qquad = 5\langle .12, .08\rangle = \langle .6, .4\rangle$

Only problem, but a big one, is that this does not scale well

- If there are n Boolean variables, space and time are $O(2^n)$

Multiple Sources of Evidence

As more sources of evidence accumulate, the conditional probability of *cavity* will adjust

- E.g., if *catch* is also given, then have:

  P(*cavity* | *toothache,catch*)

  = P(*cavity* ∧ *toothache* ∧ *catch*)/P(*toothache* ∧ *catch*) = .87 (up from .6)

In general, as evidence accumulates over time, such as data from an x-ray, the conditional probability of *cavity* would continue to evolve

- *Probabilistic reasoning is inherently non-monotonic*

However, if evidence is irrelevant, may simply ignore

- P(*cavity* | *toothache,sunny*)

  = P(*cavity* | *toothache*) = 0.6

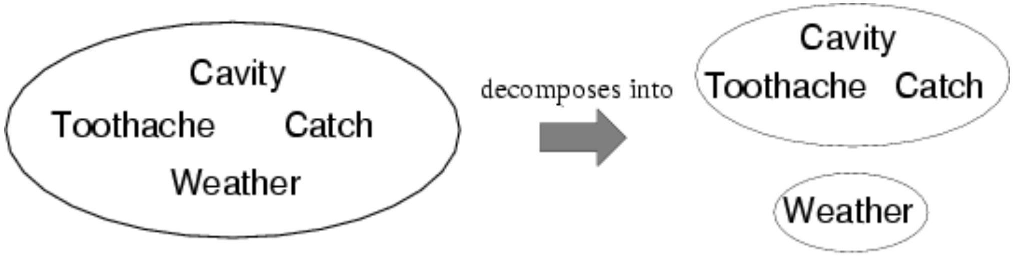
- Such reasoning about *independence* can be essential in making probabilistic reasoning tractable

Independence

***A* and *B* are independent iff**

P(A|B) = P(A) <u>or</u> P(B|A) = P(B) <u>or</u> P(AB) = P(A)P(B)

E.g., **P**(*Toothache, Catch, Cavity, Weather*)
= **P**(*Toothache, Catch, Cavity*) **P**(*Weather*)



Reduces 2x2x2x4=32 entries to 2x2x2+4=12

- For *n* independent coin tosses, $O(2^n) \rightarrow O(n)$

***Absolute* (or *marginal*) independence is powerful but rare**

- E.g., dentistry is a large field with hundreds of variables, none of which are absolutely independent from the others

Conditional Independence

Two *dependent* variables may become *conditionally independent* when *conditioned* on a third variable

$$P(YZ \mid X) = \frac{P(XYZ)}{P(X)} = \frac{P(X)P(Y \mid X)P(Z \mid X\cancel{Y})}{P(X)} = P(Y \mid X)P(Z \mid X)$$

Consider *Catch* and *Toothache*

- P(*catch*)=.34 while P(*catch* | *toothache*)=.62, so *dependent*
- The problem is that while neither causes the other, they *correlate* because both caused by single underlying hidden variable (*cavity*)
- If *condition* both *Catch* and *Toothache* on *Cavity*, conditional probabilities become independent

P(*toothache, catch* | *cavity*) = P(*toothache* | *cavity*) P(*catch* | *cavity*) = .54
P(*toothache, catch* | $\neg$*cavity*) = P(*toothache* | $\neg$*cavity*) P(*catch* | $\neg$*cavity*) = .02

- *Catch* is *conditionally independent* of *Toothache* given *Cavity*:
  P(*Toothache, Catch* | *Cavity*) = **P**(*Toothache* | *Cavity*) **P**(*Catch* | *Cavity*)
- Additional equivalent statements:
  P(*Catch* | *Toothache,Cavity*) = **P**(*Catch* | *Cavity*)
  P(*Toothache* | *Catch, Cavity*) = **P**(*Toothache* | *Cavity*)

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

Combining Evidence (for Diagnosis)

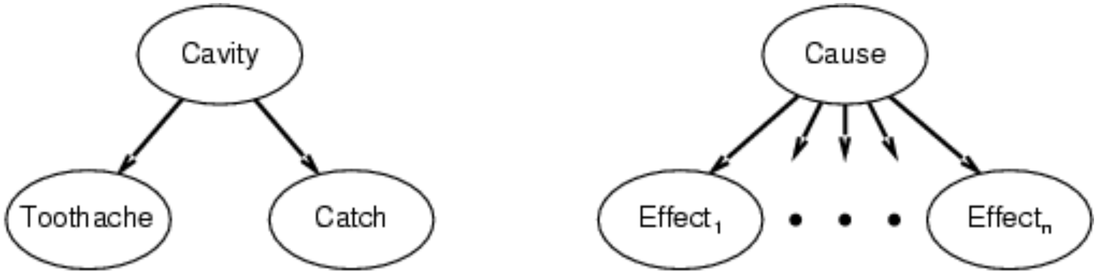
P(*Cavity* | *toothache,catch*)

= α**P**(*toothache* $\wedge$ *catch* | *Cavity*) **P**(*Cavity*) [Bayes' Rule]

= α**P**(*toothache* | *Cavity*) **P**(*catch* | *Cavity*) **P**(*Cavity*) [Cond. Ind.]

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

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

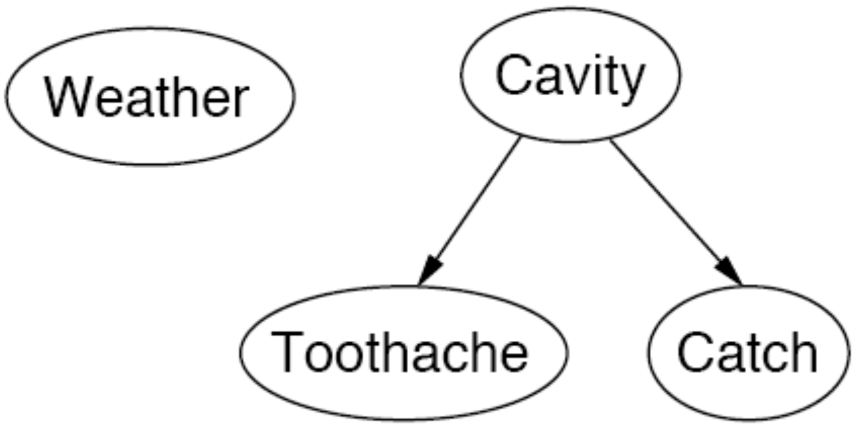


- *Cost of diagnostic reasoning now grows linearly rather than exponentially in number of conditionally independent effects*

Called naïve, because often used when the effects are not completely conditionally independent given the cause

Independence in Bayesian Networks

Topology of network encodes (conditional) independence assertions:



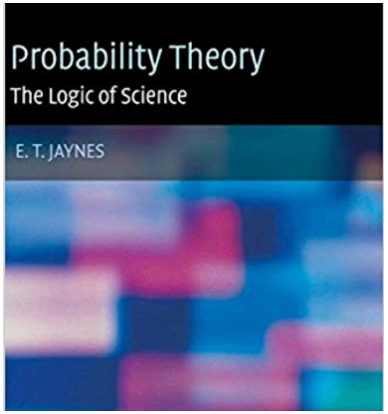
Weather **is independent of the other variables**

Toothache **and** ***Catch*** **are conditionally independent given** ***Cavity***

What you should know

- What are the sources for uncertainty?
- How can we use probability to reason about uncertainty?
- What is a random variable? Atomic events? How are they used for inference?
- What is independence? What is conditional independence? Why are they needed for reasoning about uncertainty?
- What is Bayes rule? How is this addressing combining evidence for diagnosis?

- An excellent book on probability:
  - Probability Theory: The Logic of Science, by E.T. Jaynes



Want more?

Try exercise 13.4,7,8,13,15, 14.2,8 in AIMA