Uncertainty

Dr. Ganesh Bhutkar
VIT, Pune INDIA
ganesh.bhutkar@vit.edu

TY BTech Comp - 2020-21

Uncertainty

- **Logical agents** - Make the epistemological commitment that propositions are true, false or unknown.
- Unfortunately, agents almost never have access to the whole truth about their environment.
- **Qualification problem**
  "A_{90} will get me there on-time if there is no accident on the bridge, it doesn't rain and my tires remain intact."

Real World Uncertainty

Let action A_t = leave for airport t minutes before flight

Will A_t get me there on time?

"A_{90} will get me there on-time if there is no accident on the bridge and it doesn't rain and my tires remain intact."

"A_{120} might reasonably be said to get me there on-time but I'd have to wait …"

Uncertainty Knowledge

- **Medical Diagnosis** always involves uncertainty –
  - Symptom cannot conclude disease.
  - Disease cannot conclude symptom.

- **First-Order Logic (FOL)** fails for three reasons:
  - Laziness
    - Too much work to list
  - Theoretical ignorance
    - No complete theory for diagnosis
  - Practical ignorance
    - Not all the necessary tests have been run.
- How to solve?
  - Degree of Belief, Probability Theory & Fuzzy Logic

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

Rational Decisions

- Preferences of outcomes
- Utility theory
- Maximum Expected Utility

Decision Theory = Probability Theory + Utility Theory

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

  - **Utility theory** is used to represent and infer preferences.
  - **Decision theory** = **Probability Theory + Utility Theory**

Probability 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 you have a dental 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*

- 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:

  $Cavity = false \wedge Toothache = false$

  $Cavity = false \wedge Toothache = true$

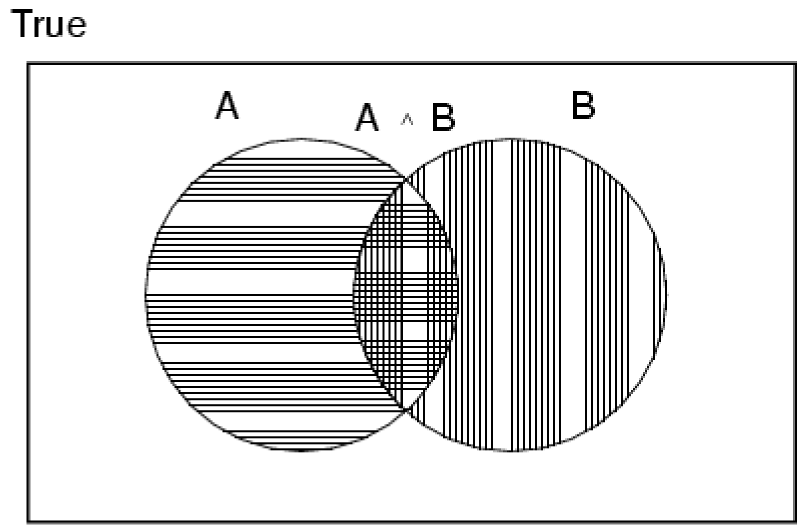
  $Cavity = true \wedge Toothache = false$

  $Cavity = true \wedge Toothache = true$

- Atomic events are mutually exclusive and exhaustive.

Axioms of Probability

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



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.**

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:

$$\boxed{P(a \mid b) = \frac{P(a \wedge b)}{P(b)} \qquad \textit{where} \qquad P(b) > 0}$$

- **Product rule** gives an alternative formulation:

$$\boxed{P(a \wedge b) = P(a \mid b)\, P(b) = P(b \mid a)\, P(a)}$$

Axioms of Probability

- Kolmogorov's three axioms

1. All probabilities are between 0 and 1

$$0 \leqq P(a) \leqq 1$$

2. Necessarily true is 1, necessarily false is 0

$$P(true)=1, \qquad P(false)=0$$

3. Disjunction

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

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

- P(*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

- P(cavity ∨*toothache*) = 0.108 + 0.012 + 0.072 + 0.008 + 0.016 + 0.064 = 0.28

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

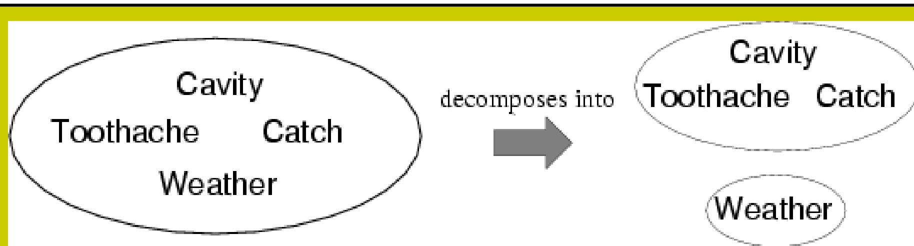
- Denominator can be viewed as a **normalization constant** α

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**

Independence

- *A* and *B* are independent iff
  P(*A*|*B*) = **P**(*A*) or **P**(*B*|*A*) = **P**(*B*) or **P**(A, B) = **P**(*A*) **P**(*B*)



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

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

- The same independence holds if I haven't got a cavity:
- (2) **P**(*catch* | *toothache*,¬*cavity*) = **P**(*catch* | ¬*cavity*)

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

- Equivalent statements:
  P(*Toothache* | *Catch, Cavity*) = **P**(*Toothache* | *Cavity*)
  P(*Toothache, Catch* | *Cavity*) = **P**(*Toothache* | *Cavity*) **P**(*Catch* | *Cavity*)

Conditional Independence

- 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

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

Bayes' Rule

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

  $\Rightarrow$ **Bayes' rule:** $$P(a \mid b) = \frac{P(b \mid a)\, P(a)}{P(b)}$$

- 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:
  - $P(m|s) = P(s|m)\, P(m) / P(s) = 0.5 \times 0.00002 / 0.05 = 0.0002$

Combining Evidence

Assume independence of catch and toothache

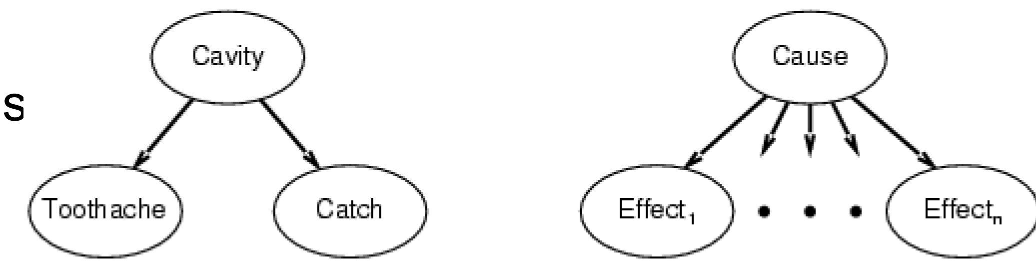
$\mathbf{P}(Cavity \mid toothache \wedge catch)$

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

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

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

P(Caus



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