Artificial Intelligence

Quantifying Uncertainty

CS 444 – Spring 2019
Dr. Kevin Molloy
Department of Computer Science
James Madison University

JAMES MADISON UNIVERSITY

Uncertainty

Let action A_t = leave for airport t minutes before your flight. What A_t will get me there on time?

Problems:

1) Partial observability (road state, other drivers' plans, etc.).
2) Noisy sensors (WTOP traffic reports)
3) Uncertainty in action outcomes (flat tire, etc.)
4) Immense complexity of modelling and predicting traffic

Hence a purely logic 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} wll get me there on time if there's no accident on the bridge and it doesn't rai and my tires remain intact, etc).

(A_{1440} might reasonable be said to get me there on time, but I'd have to stay overnight at 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:

$A_{25} \mapsto_{0.3} AtAirportOnTime$

Sprinkler $\mapsto_{0.99}$ $Wetgrass$

Sprinkler $\mapsto_{0.7}$ Rain

Issues: Problems with combinations, e.g., Sprinkler causes Rain?

Probability

Given the available evidence, A_{25} will get me there on time with probability 0.04.

Mahaviracara (9th C.), Cardamo (1565) theory of gambling

(**Fuzzy logic** handles **degrees of truth** NOT uncertainty, e.g

WetGrass is true to degree 0.2)

Probability

Probabilistic assertions summarize efforts of:

Laziness: failure to enumerate exceptions, qualifications, etc.

Ignorance: lack of relevant facts, initial conditions, etc.

Subjective or Bayesian probability:

Probabilities relate propositions to one's own state of knowledge.

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

These are **not** claims of a "probabilistic tendency" in the current situation (but might be learned from past experience of similar situations)

Probabilities of propositions change with new evidence:

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

Analogous to logical entailment status ($KB \vDash \alpha$, not truth).

Making Decisions under Uncertainty

Suppose I believe the following:

$P(A_{25}$ gets me there on time | …) = 0.04

$P(A_{90}$ get me there on time | ...) = 0.70

$P(A_{120}$ get 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 airport cuisine, etc.

Utility theory is used to represent and infer preferences

Decision theory = utility theory + probability theory

Probability Basics

Begin with a set Ω – the sample space. E.g. 6 possible rolls of a die.

$\omega \in \Omega$ is a sample point/possible world/atomic event. A probability space or probability model is a sample space with an assignment $P(\omega)$ for every $\omega \in \Omega$ s.t.

$0 \leq P(\omega) \leq 1$

$$\sum_{\omega} P(\omega) = 1$$

e.g., P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = 1/6.

An event A is any subset of Ω

$$P(A) = \sum_{\omega \in A} P(\omega)$$

e.g. P(die roll < 4) = P(1) + P(2) + P(3) = 1/6 + 1/6 + 1/6 = 1/2

Random Variables

A random variable is a function from sample points to some range, e.g. the reals or Booleans

e.g. Odd(1) = true

P induces a probability distribution for any r.v. X: $P(X = x_i) = \sum_{\{\omega : X(\omega) = x_i\}} P(\omega)$

e.g. P(Odd = true) = P(1) + P(3) + P(5) = 1/6 + 1/6 + 1/6 = 1/2

Propositions

Think of a proposition as the event (set of sample points), where the proposition is true

Given Boolean random variables A and B:

event *a* = set of sample points where A(ω) = true

event ¬*a* = points where A(ω) = true and B(ω) = true

Often in AI applications, the sample points are **defined** by the value of a set of random variables, i.e., the sample space is the Cartesian product of the ranges of the variables

With Boolean variables, sample point = propositional logic model

e.g., A = true, B = false, or $a \land \neg b$

Propositional = disjunction of atomic events in which it is true

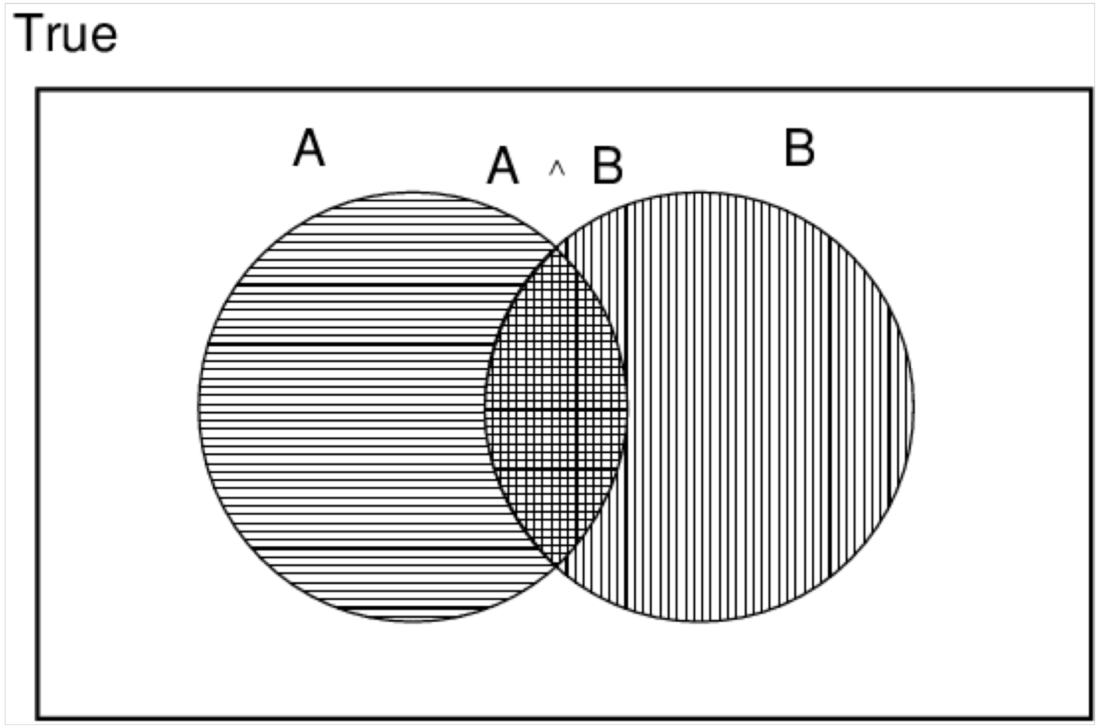
e..g $(a \lor b) \equiv (\neg a \land b) \lor (a \land \neg b) \lor (a \land b)$

$\implies P(a \lor b) = P(\neg a \land b) + P(a \land \neg b) + P(a \land b)$

Why Use Probability?

The definitions imply that certain logically related events must have related probabilities

E.g., $P(a \vee b) = P(a) + P(b) - P(a \wedge b)$



de Finetti (1931): an agent who bets according to probabilities that violate these axioms can be forced to bet so as to lose money regardless of outcome.

Syntax for Propositions

Propositional or Boolean random variables

e.g., Cavity (do I have a cavity?)

Cavity = true is a proposition, also written cavity

Discrete random variables (finite or infinite).

e.g. Weather is one of <sunny, rain, cloudy, snow>

Weather = rain is a proposition

Values must be exhaustive and mutually exclusive

Continuous random variables (bounded or unbounded)

e.g., Temp = 21.6, also allowed, Temp < 22.0

Arbitrary Boolean combinations of basic propositions

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 r.v.s gives the probability of every atomic event on those r.v.s (i.e., every sample point)

P(Weather, Cavity) = a 4 x 2 matrix of values:

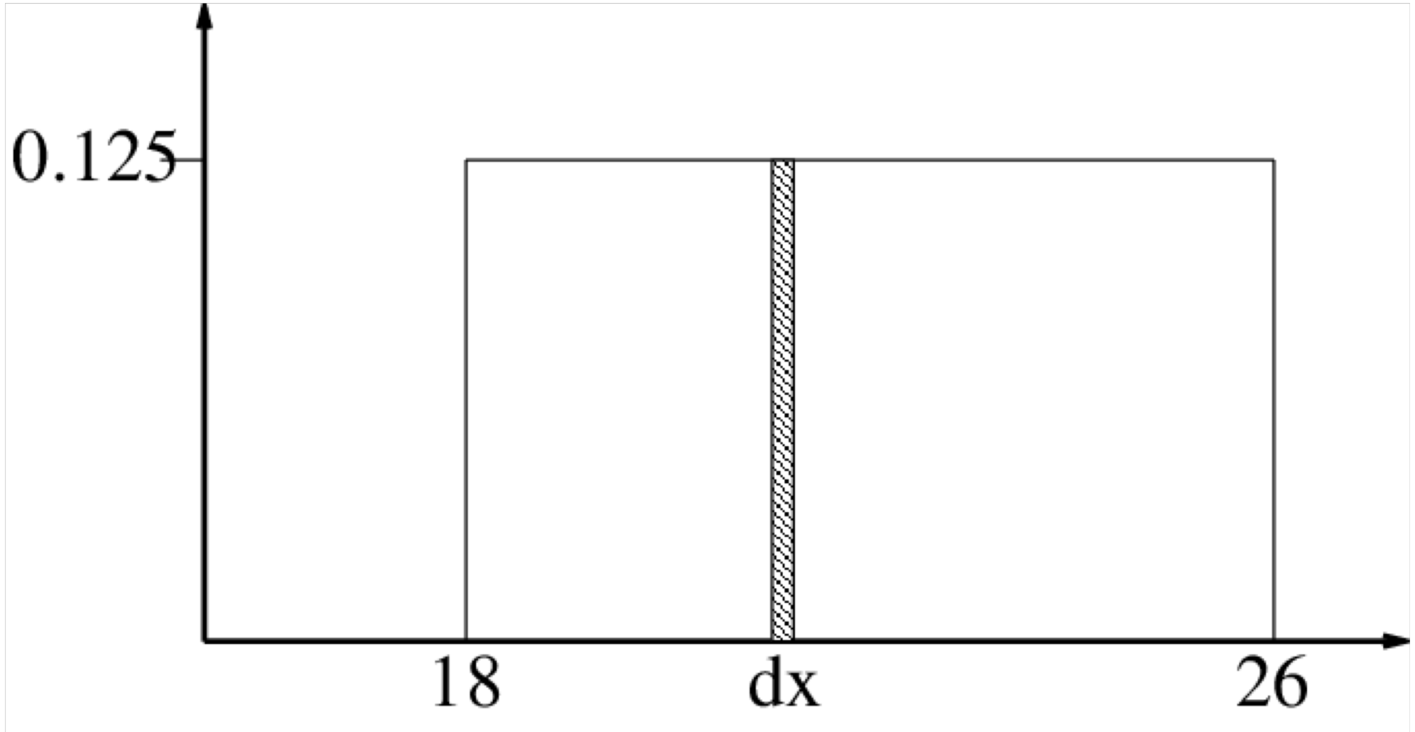
Weather =		**sunny**	**rain**	**cloudy**	**snow**
Cavity = true		0.144	0.02	0.016	0.02
Cavity = false		0.576	0.08	0.064	0.08

Every question about a domain can be answered by the joint distribution because every event is a sum of sample points

Probability for Continuous Variables

Express distribution as a parameterized function of value:

P(X = x) = U[18,26](x) = uniform density between 18 and 26

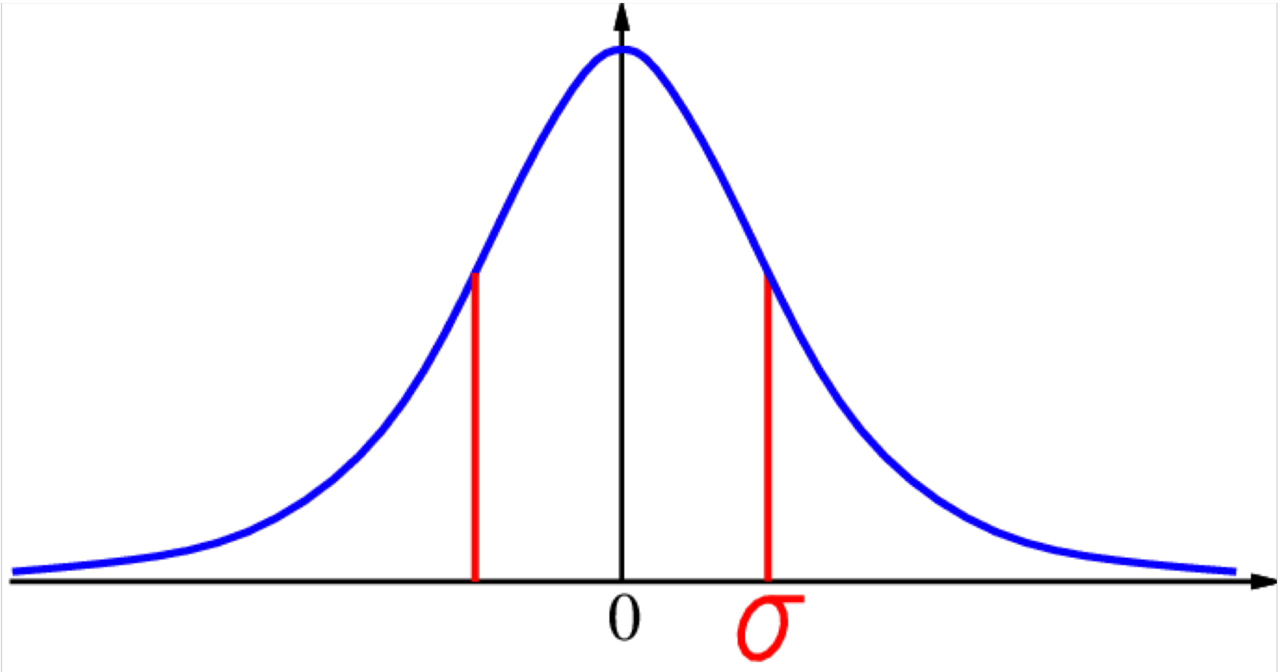


Here P is a density; integrates to 1. P(X = 20.5) = 0.125 really means

$$\lim_{dx \to 0} P(20.5 \leq X \leq 20.5 + dx)\, dx = 0.125$$

Gaussian Density

Express distribution as a parameterized function of value:



$$P(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-(x-\mu)^2 / 2\sigma^2}$$

What does P(x) represent?

Conditional Probability

Conditional or posterior probabilities

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

NOT "if *toothache* then 80% chance of *cavity*". 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

Note: the less specific belief remains valid after more evidence arrives, but is not always useful

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

P(cavity | toothache, 49ersWin) = P(cavity | toothache) = 0.8

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

Conditional Probability

Definition of conditional probability: $P(a \mid b) = \dfrac{P(a \wedge b)}{P(b)}$ $if\ P(b) \neq 0$

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.

P(Weather, Cavity) = P(Weather | Cavity)P(Cavity). View as a 4 x 2 set of equations, not matrix mult.)

Chain rule is derived by successive application of product rule:

$P(X_1, \ldots, X_n) = P(X_1, \ldots, X_{n-1})\ P(X_n \mid X_1, \ldots X_{n-1})$

$= P(X_1, \ldots., X_{n-2})P(X_{n-1} \mid X_1, \ldots.. X_{n-2})\ P(X_n, X_{1,} \ldots., X_{n-1})$

$= \prod_{i=1}^{n} P(X_i | X_1, \ldots, X_{i-1})$

Inference by Enumeration

Start with the joint distribution:

	toothache		¬ toothache	
	catch	¬catch	Catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

For any proposition $\boldsymbol{\phi}$, sum the atomic events where it is true $P(\boldsymbol{\phi}) = \sum_{\omega:w\vDash\phi} P(\omega)$

P(toothache) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2

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

P(¬cavity | toothache) = $\dfrac{P(\neg\text{cavity} \wedge \text{toothache})}{P(toothache)}$ $\dfrac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$

Normalization

	toothache		¬ toothache	
	catch	¬catch	Catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Denominator can be viewed as a normalization constant α

P(Cavity | toothache) = αP(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

Let X be all the variables. Typically, we want the posterior joint distribution of the query variables Y given specific values e for the evidence variables E.

Let the hidden variables be H = X – Y – E

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

$$P(Y \mid E = e) = \alpha P(Y, E = e) = \alpha \sum_h P(Y, E = e, H = h)$$

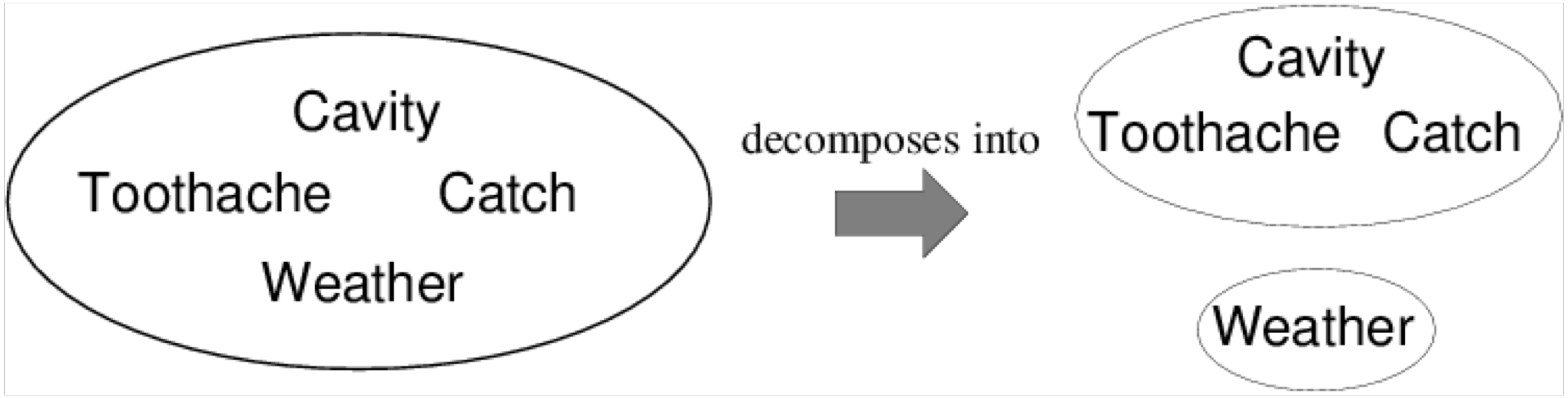
The terms in the summation are joint entries because Y, E, and H together exhaust the set of random variables.

Some 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 P(A|B) = P(A) or P(B | A) = P(B)



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

32 entries reduced to 12; for n independent biased coins, $2^n \rightarrow n$

Absolute independence is powerful, but very rare

Denistry is a large field with hundreds of variables, none of which are independent. What to do?

Conditional Independence

P(Toothache, Cavity, Catch) has 23 – 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: thus. **(1) P(catch | toothache, cavity) = P(catch | cavity)**

The same independence holds if I don't have 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)

Thus, these are 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 the 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. Big deal?

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

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 ∧ b) = P(a | b)P(b) = P(b | a)P(a)

⟹ **Bayes' rule** $$P(a\,|b) = \frac{P(b\,|a)\,P(a)}{P(b)}$$ Why is this useful?

Consider which is easier: P(Stiff Neck | Meningitis) or P(Meningitis | Stiff Neck)

Causal Direction
(usually easier to calculate)

Diagnostic
Direction

$$P(Cause\,|Effect) = \frac{P(Effect\,|Cause)\,P(Cause)}{P(Effect)}$$

Bayes' Rule

$$P(Meningitis\ |StiffNeck) = \frac{P(StiffNeck\ |Meningitis)\ P(Meningitis)}{P(StiffNeck)}$$

P(s | m) = 0.8

P(m) = 0.0001

P(s) = 0.1

$$P(Meningitis\ |StiffNeck) = \frac{0.80\ x\ 0.0001}{0.1} = 0.0008$$

If an outbreak occurs, we can update this equation rather easily.

Bayes' Rule and Naïve Bayes

$$P(Cavity \mid toothache \wedge catch) = \alpha P(toothache \wedge catch \mid Cavity)P(Cavity)$$

$$= \alpha P(toothache \mid Catch)P(catch \mid Cavity)P(Cavity)$$

This is an example of a naïve Bayes model.