Artificial Intelligence

CSE 440/EEE 333/ETE333

Chapter 13
Fall 2017

Mirza Mohammad Lutfe Elahi
Department of Electrical and Computer Engineering
North South University

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.

Uncertainty

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

“A_{90} will get us to the airport in time, as long as the car doesn’t break down or run out of gas, and I don’t get into an accident, and there are no accidents on the bridge, and the plane doesn’t leave early, and no meteorite hits the car, and . . .”

“A1440 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_{90} works unless contradicted by evidence

Issues: What assumptions are reasonable? How to handle contradiction?

- Probability
  - Given the available evidence,
    - A_{90} will get me there on time with probability 0.04
- (Fuzzy logic handles degree of truth NOT uncertainty
  - e.g., WetGrass is true to degree 0.2)

Probability

- Probabilistic assertions summarize effects 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., $P(A_{90} | \text{no reported accidents}) = 0.06$
- 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., $P(A_{90} | \text{no reported accidents, 5 a.m.}) = 0.15$
- (Analogous to logical entailment status $KB \models \alpha$, not truth.)

Making decisions under uncertainty

Suppose I believe the following:

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

$P(A_{90}$ 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

Probability basics

- Let A be a proposition, P(A) denotes the unconditional probability that A is true.

- Example: if Male denotes the proposition that a particular person is male, then P(Male) = 0.5 mean that without any other information, the probability of that person being male is 0.5 (a 50% chance).

- Alternatively, if a population is sampled, then 50% of the people will be male.

- Of course, with additional information (e.g. that the person is a CSE440 student), the “posterior probability” will likely be different.

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.

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

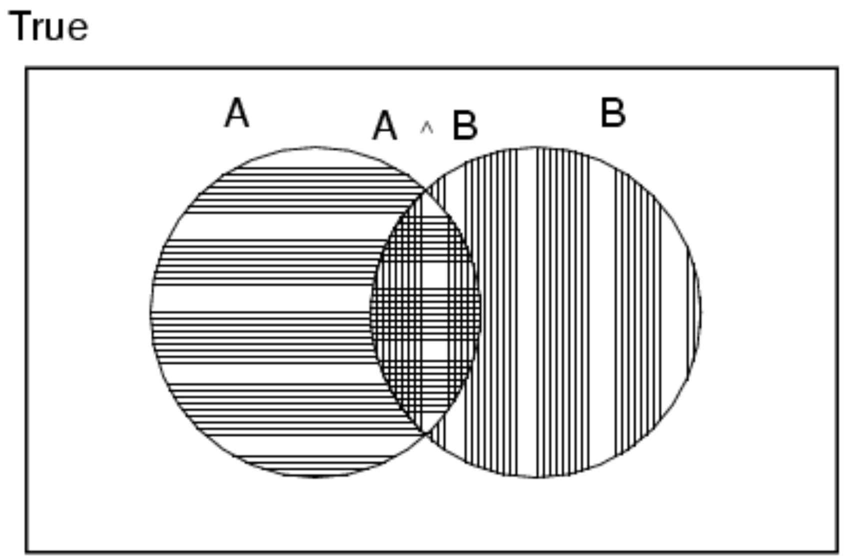
  $\sum_{\omega} P(\omega) = 1$

  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

Axiom of probability

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



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 $A(\omega)$ = true

    event $\neg a$ = set of sample points where $A(\omega)$ = false

    event $a \wedge b$ = points where $A(\omega)$ = true and $B(\omega)$ = true
- Often in AI applications, the sample points are defined by the values 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 $a \wedge \neg b$.
- Proposition = disjunction of atomic events in which it is true

    e.g., $(a \vee b) \equiv (\neg a \wedge b) \vee (a \wedge \neg b) \vee (a \wedge b)$

    $\Rightarrow P(a \vee b) = P(\neg a \wedge b) + P(a \wedge \neg b) + P(a \wedge b)$

Syntax

- Basic element: random variable (function from sample points to some range, e.g., the reals or Booleans)

- 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

Syntax

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

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)

Prior probability

- 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	0.15	0.03	0.01	0.02
Cavity = false	0.57	0.08	0.06	0.08

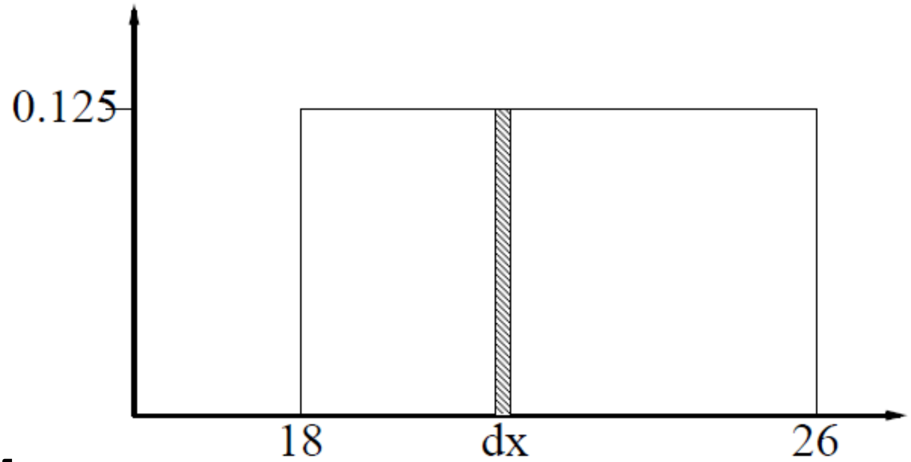
- Every question about a domain can be answered by the joint distribution because every event is a sum of sample points

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 ≤ X ≤ 20.5 + dx)/dx = 0.125

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:
  $P(a \mid b) = P(a \wedge b) / P(b)$ 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.,
  $\mathbf{P}(\textit{Weather}, \textit{Cavity}) = \mathbf{P}(\textit{Weather} \mid \textit{Cavity})\, \mathbf{P}(\textit{Cavity})$
- (View as a set of 4×2 equations, not matrix mult.)

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

$$
\begin{aligned}
\mathbf{P}(X_1, \ldots, X_n) &= \mathbf{P}(X_1,\ldots,X_{n-1})\, \mathbf{P}(X_n \mid X_1,\ldots,X_{n-1}) \\
&= \mathbf{P}(X_1,\ldots,X_{n-2})\, \mathbf{P}(X_{n-1} \mid X_1,\ldots,X_{n-2})\, \mathbf{P}(X_n \mid X_1,\ldots,X_{n-1}) \\
&= \ldots \\
&= \prod_{i=1}^{n} \mathbf{P}(Xi \mid X_1, \ldots, Xi_{-1})
\end{aligned}
$$

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		*¬ 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: $P(\phi) = \Sigma_{\omega:\omega \models \phi} P(\omega)$
- $P(cavity \vee toothache) =$ $0.108+0.012+0.072+0.008+0.016+0.064 = 0.28$

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

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

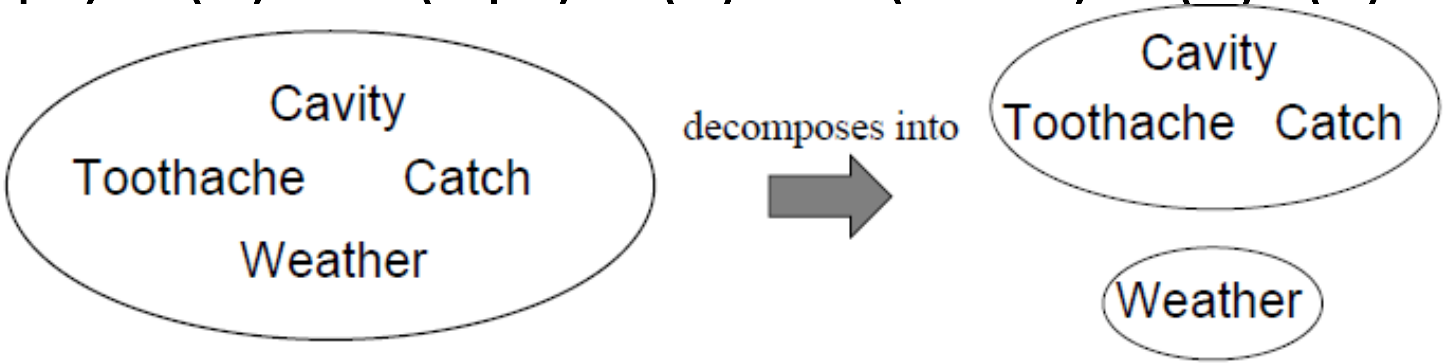
$= \alpha,\ [\mathbf{P}(\mathit{Cavity}, \mathit{toothache}, \mathit{catch}) + \mathbf{P}(\mathit{Cavity}, \mathit{toothache}, \neg\, \mathit{catch})]$
$= \alpha,\ [\langle 0.108, 0.016\rangle + \langle 0.012, 0.064\rangle]$
$= \alpha,\ \langle 0.12, 0.08\rangle = \langle 0.6, 0.4\rangle$

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 $P(A \wedge B)=P(A)P(B)$



- P(Toothache, Catch, Cavity, Weather)

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

- 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, $\neg$ cavity) = P(catch| $\neg$ 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)

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

  $\Rightarrow$ 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:

  $P(Cause|Effect) = P(Effect|Cause)P(Cause) / P(Effect)$

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

  $P(m|s) = P(s|m)P(m) / P(s)$

  $= 0.5 \times 0.00002 / 0.05$

  $= 0.0002$

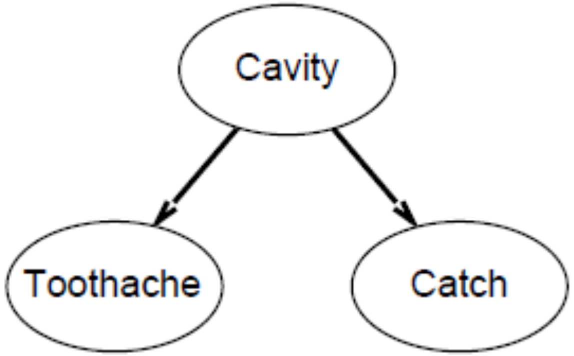
- Note: posterior probability of meningitis still very small!

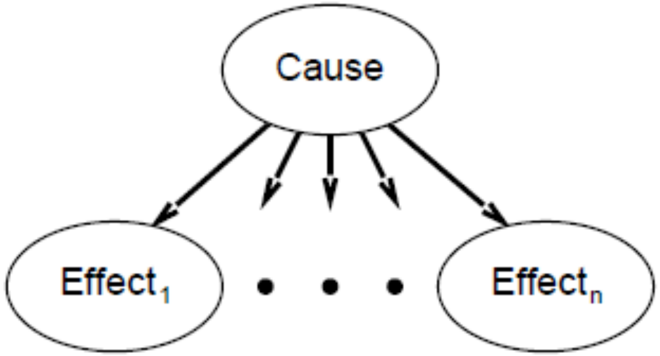
Bayes Rule

- P(Cavity|toothache $\wedge$ catch)
  = α P(toothache $\wedge$ catch|cavity)P(cavity)
  = α P(toothache|cavity)P(catch|cavity)P(cavity)

This is an example of a naive Bayes model:

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





Total number of parameters is linear in n