Uncertainty

Chapter 13

Mausam

(Based on slides by UW-AI faculty, Stuart Russell and Subbarao Kambhampati)

Knowledge Representation

KR Language	Ontological Commitment	Epistemological Commitment
Propositional Logic	facts	true, false, unknown
First Order Logic	facts, objects, relations	true, false, unknown
Temporal Logic	facts, objects, relations, times	true, false, unknown
Probability Theory	facts	degree of belief
Fuzzy Logic	facts, degree of truth	known interval values

Probabilistic Relational Models

- combine probability and first order logic

Propositional Logic Problem Solving

- Need to write what you know as propositional formulas
- Theorem proving will then tell you whether a given new sentence will hold given what you know

- Three kinds of queries
  - Is my knowledgebase consistent? (i.e. is there at least one world where everything I know is true?) *Satisfiability*
  - Is the sentence S *entailed* by my knowledge base? (i.e., is it true in every world where my knowledge base is true?)
  - Is the sentence S *consistent/possibly true* with my knowledge base? (i.e., is S true in at least one of the worlds where my knowledge base holds?)
    - S is consistent if ~S is not entailed

- But cannot differentiate between degrees of likelihood among possible sentences

Example

- Pearl lives in Los Angeles. It is a high-crime area. Pearl installed a burglar alarm. He asked his neighbors John & Mary to call him if they hear the alarm. This way he can come home if there is a burglary. Los Angeles is also earth-quake prone. Alarm goes off when there is an earth-quake.

Burglary => Alarm

Earth-Quake => Alarm

Alarm => John-calls

Alarm => Mary-calls

If there is a burglary, will Mary call?

Check KB & E |= M

If Mary didn't call, is it possible that Burglary occurred?

Check KB & ~M *doesn't entail* ~B

Example (Real)

- Pearl lives in Los Angeles. It is a high-crime area. Pearl installed a burglar alarm. He asked his neighbors John & Mary to call him if they hear the alarm. This way he can come home if there is a burglary. Los Angeles is also earth-quake prone. Alarm goes off when there is an earth-quake.
- Pearl lives in real world where (1) burglars can sometimes disable alarms (2) some earthquakes may be too slight to cause alarm (3) Even in Los Angeles, Burglaries are more likely than Earth Quakes (4) John and Mary both have their own lives and may not always call when the alarm goes off (5) Between John and Mary, John is more of a slacker than Mary.(6) John and Mary may call even without alarm going off

Burglary => Alarm
Earth-Quake => Alarm
Alarm => John-calls
Alarm => Mary-calls

If there is a burglary, will Mary call?

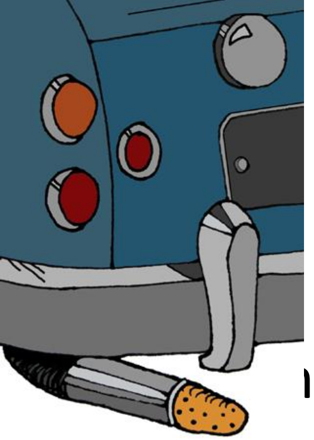
Check KB & E |= M

If Mary didn't call, is it possible that Burglary occurred?

Check KB & ~M *doesn't entail* ~B

John already called. If Mary also calls, is it more likely that Burglary occurred?

You now also hear on the TV that there was an earthquake. Is Burglary more or less likely now?



How do we handle Real Pearl?

•*Potato in the tail-pipe*

- iscient & Eager way:
  - Model everything!
  - E.g. Model exactly the conditions under which John will call
    - He shouldn’t be listening to loud music, he hasn’t gone on an errand, he didn’t recently have a tiff with Pearl etc etc.

A & c1 & c2 & c3 &..cn => J

(also the exceptions may have interactions

c1&c5 => ~c9)

- Ignorant (non-omniscient) and Lazy (non-omnipotent) way:
  - Model the likelihood
  - In 85% of the worlds where there was an alarm, John will actually call
  - How do we do this?
    - Non-monotonic logics
    - “certainty factors”
    - “fuzzy logic”
    - “probability” theory?

Qualification and Ramification problems make this an infeasible enterprise

Logic vs. Probability

Symbol: Q, R …	Random variable: Q …
Boolean values: T, F	Domain: you specify e.g. {heads, tails} [1, 6]
State of the world: Assignment to Q, R … Z	Atomic event: complete specification of world: Q… Z • Mutually exclusive • Exhaustive
	Prior probability (aka Unconditional prob: P(Q)
	Joint distribution: Prob. of every atomic event

Probability Basics

- Begin with a set S: the sample space
  - e.g., 6 possible rolls of a die.
- $x \in S$ is a sample point/possible world/atomic event
- A probability space or probability model is a sample space with an assignment P(x) for every x s.t. $0 \leq P(x) \leq 1$ and $\sum P(x) = 1$
- An event A is any subset of S
  - e.g. A= ‘die roll < 4’
- A random variable is a function from sample points to some range, e.g., the reals or Booleans

Types of Probability Spaces

Propositional or Boolean random variables

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

Discrete random variables (*finite* or *infinite*)

e.g., $Weather$ is one of $\langle sunny, rain, cloudy, snow \rangle$

$Weather = rain$ is a proposition

Values must be exhaustive and mutually exclusive

Continuous random variables (*bounded* or *unbounded*)

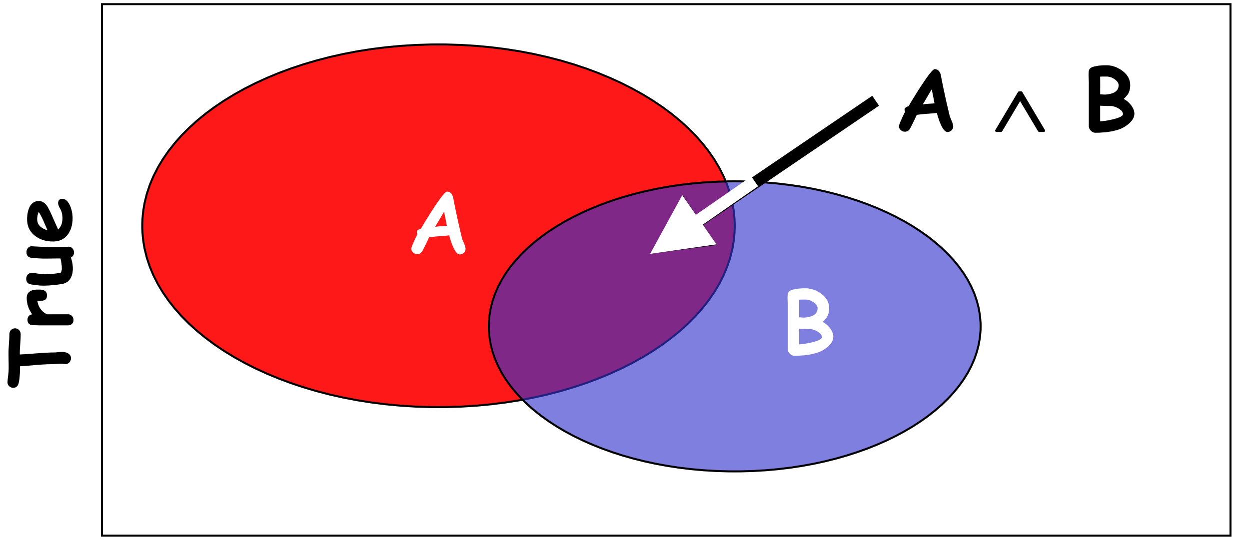
e.g., $Temp = 21.6$; also allow, e.g., $Temp < 22.0$.

Arbitrary Boolean combinations of basic propositions

Axioms of Probability Theory

- All probabilities between 0 and 1
  - $0 \leq P(A) \leq 1$
  - P(true) = 1
  - P(false) = 0.
- The probability of disjunction is:

$$P(A \vee B) = P(A) + P(B) - P(A \wedge B)$$



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:

$\mathbf{P}(Weather) = \langle 0.72, 0.1, 0.08, 0.1 \rangle$ (*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

$\mathbf{P}(Weather, Cavity) =$ a 4×2 matrix of values:

Joint distribution can answer any question

Conditional probability

- Conditional or posterior probabilities
  e.g., P(*cavity* | *toothache*) = 0.8
  i.e., given that *toothache* is all I know there is 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

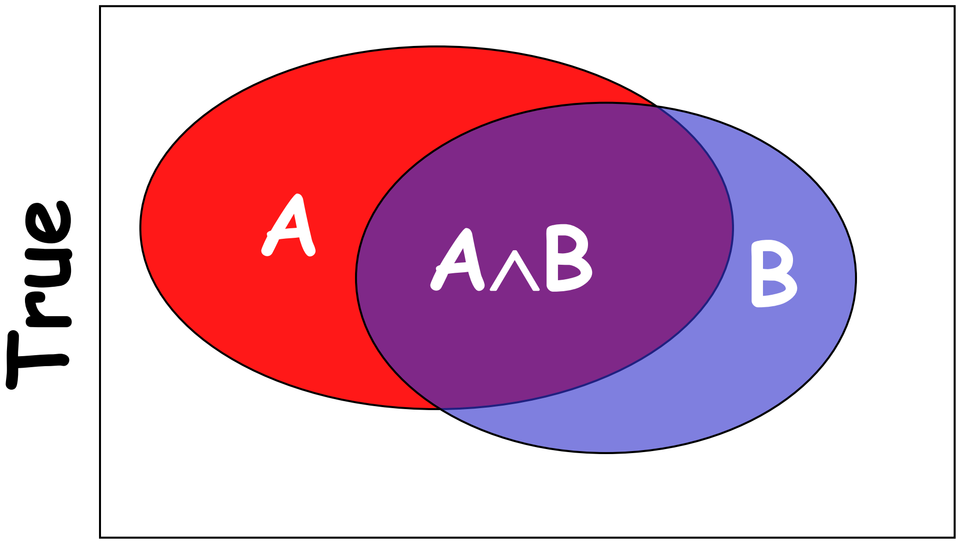
- New evidence may be irrelevant, allowing simplification:
  P(*cavity* | *toothache, sunny*) = P(*cavity* | *toothache*) = 0.8

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

Conditional Probability

- P(*A* | *B*) is the probability of *A* given *B*
- Assumes that *B* is the only info known.
- Defined by:

$$P(A|B) = \frac{P(A \wedge B)}{P(B)}$$

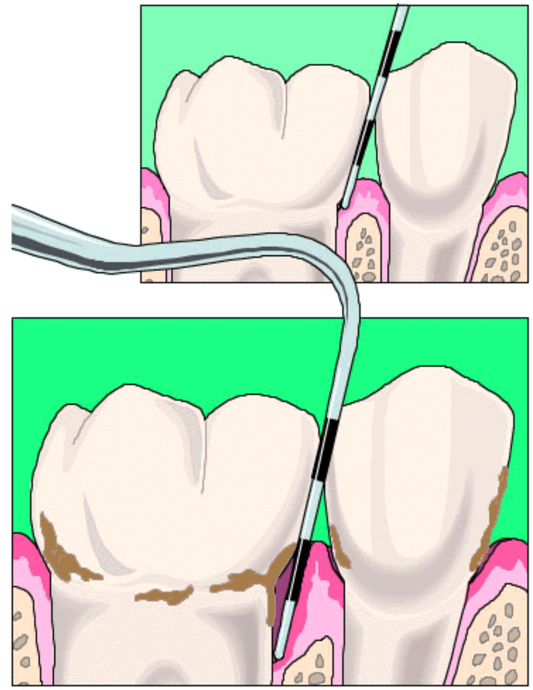


Chain Rule/Product Rule

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

  $= \Pi P(X_i | X_1,..X_{i-1})$

Dilemma at the Dentist's



What is the probability of a cavity given a toothache?
What is the probability of a cavity given the probe catches?

Inference by Enumeration

Start with the joint 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)=.108+.012+.016+.064

= .20 or 20%

Inference by Enumeration

Start with the joint 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∨cavity) = .20 + .072 + .008

.28

Inference by Enumeration

Start with the joint distribution:

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

Can also compute conditional probabilities:

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

$$= \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

Complexity of Enumeration

- Worst case time: $O(d^n)$
  - Where d = max arity
  - And n = number of random variables
- Space complexity also $O(d^n)$
  - Size of joint distribution
- Prohibitive!

Independence

- *A* and *B* are *independent* iff:

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

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

These two constraints are logically equivalent

- Therefore, if *A* and *B* are independent:

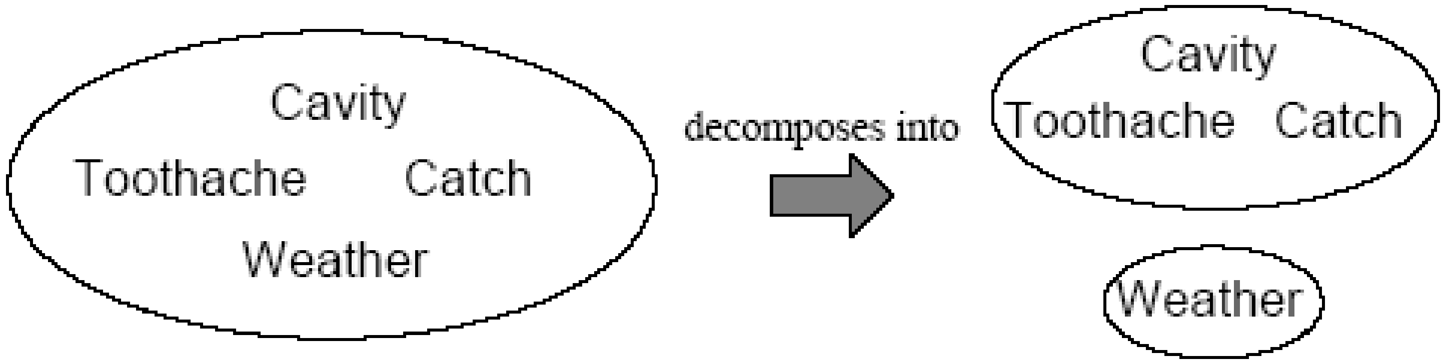
$$P(A \mid B) = \frac{P(A \wedge B)}{P(B)} = P(A)$$

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

Independence

A and B are independent iff

$$\mathbf{P}(A|B) = \mathbf{P}(A) \quad \text{or} \quad \mathbf{P}(B|A) = \mathbf{P}(B) \quad \text{or} \quad \mathbf{P}(A, B) = \mathbf{P}(A)\mathbf{P}(B)$$



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

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

Complete independence is powerful but rare

What to do if it doesn't hold?

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) $P(catch|toothache, cavity) = P(catch|cavity)$

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 $Cavity$:

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

Instead of 7 entries, only need 5

Conditional Independence II

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

P(catch | toothache,¬cavity) = P(catch |¬cavity)

Equivalent statements:

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

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

Why only 5 entries in table?

Write out full joint distribution using chain rule:

$$\begin{aligned}
&\mathbf{P}(Toothache, Catch, Cavity) \\
&= \mathbf{P}(Toothache|Catch, Cavity)\mathbf{P}(Catch, Cavity) \\
&= \mathbf{P}(Toothache|Catch, Cavity)\mathbf{P}(Catch|Cavity)\mathbf{P}(Cavity) \\
&= \mathbf{P}(Toothache|Cavity)\mathbf{P}(Catch|Cavity)\mathbf{P}(Cavity)
\end{aligned}$$

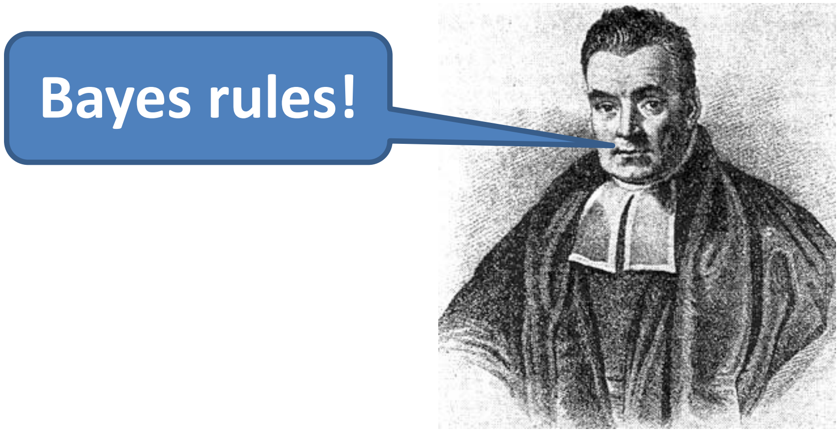
I.e., $2 + 2 + 1 = 5$ independent numbers (equations 1 and 2 remove 2)

Power of Cond. Independence

- Often, using conditional independence reduces the storage complexity of the joint distribution from exponential to linear!!

- Conditional independence is the most basic & robust form of knowledge about uncertain environments.

Bayes Rule



posterior

$$P(x,y) = P(x \mid y)P(y) = P(y \mid x)P(x)$$

$$P(x \mid y) = \frac{P(y \mid x)\ P(x)}{P(y)} = \frac{\text{likelihood} \cdot \text{prior}}{\text{evidence}}$$

Useful for assessing diagnostic probability from causal probability:

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

Computing Diagnostic Prob. from Causal Prob.

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

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

P(M) = 0.0001,
P(S) = 0.1,
P(S|M)= 0.8

P(M|S) $= \frac{P(s|m)P(m)}{P(s)} = \frac{0.8 \times 0.0001}{0.1} = 0.0008$

Note: posterior probability of meningitis still very small!

Other forms of Bayes Rule

$$P(x|y) = \frac{P(y|x)\ P(x)}{P(y)} = \frac{\text{likelihood} \cdot \text{prior}}{\text{evidence}}$$

$$P(x|y) = \frac{P(y|x)\ P(x)}{\sum_x P(y|x)\ P(x)}$$

$$P(x|y) = \alpha P(y|x) P(x)$$

$$\text{posterior} \propto \text{likelihood} \cdot \text{prior}$$

Conditional Bayes Rule

$$P(x \mid y, z) = \frac{P(y \mid x, z)\ P(x \mid z)}{P(y \mid z)}$$

$$P(x \mid y, z) = \frac{P(y \mid x, z)\ P(x, z)}{\sum_x P(y \mid x, z)\ P(x \mid z)}$$

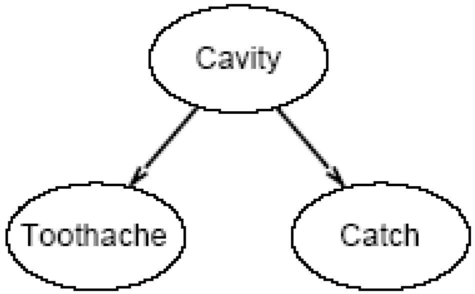
$$P(x \mid y, z) = \alpha P(y \mid x, z) P(x \mid z)$$

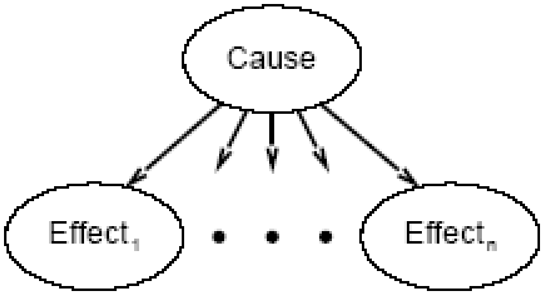
Bayes' Rule & Cond. Independence

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

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

$$\mathbf{P}(Cause, Effect_1, \ldots, Effect_n) = \mathbf{P}(Cause)\Pi_i \mathbf{P}(Effect_i|Cause)$$

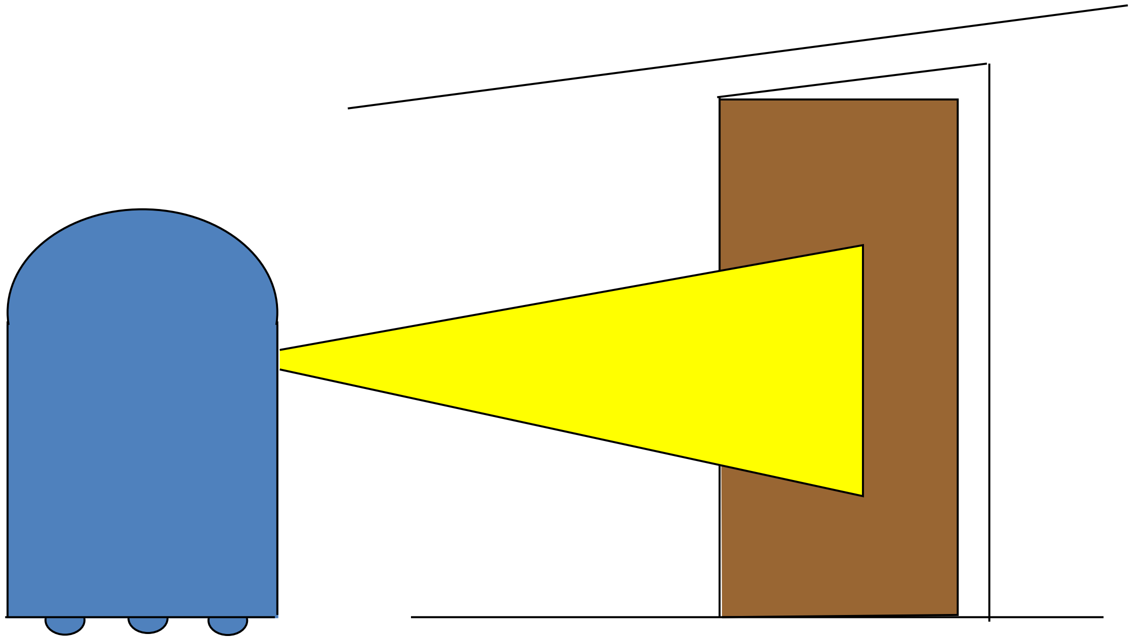




Total number of parameters is *linear* in n

Simple Example of State Estimation

- Suppose a robot obtains measurement z
- What is *P(doorOpen|z)?*



Causal vs. Diagnostic Reasoning

- *P(open|z)* is diagnostic.
- *P(z|open)* is causal.
- Often causal knowledge is easier to obtain.
- Bayes rule allows us to use causal knowledge:

count frequencies!

$$P(open \mid z) = \frac{P(z \mid open)P(open)}{P(z)}$$

Example

- $P(z|open) = 0.6$ $P(z|\neg open) = 0.3$
- $P(open) = P(\neg open) = 0.5$

$$P(open \mid z) = \frac{P(z \mid open) P(open)}{P(z \mid open) p(open) + P(z \mid \neg open) p(\neg open)}$$

$$P(open \mid z) = \frac{0.6 \cdot 0.5}{0.6 \cdot 0.5 + 0.3 \cdot 0.5} = \frac{2}{3} = 0.67$$

- *z* raises the probability that the door is open.

These calculations seem laborious to do for each problem domain –
is there a general representation scheme for probabilistic inference?

Yes – Bayesian Networks