Uncertainty

Inference

- All kings who are greedy are Evil
- **$\forall x$ king(x) $\land$ greedy (x) $\rightarrow$ Evil (x)**

- **King(John) $\land$ Greedy (John) $\rightarrow$ Evil (John)**
- **King(Richard) $\land$ Greedy (Richard) $\rightarrow$ Evil (Richard)**
- **King(Father(John)) $\land$ Greedy (Father(John)) $\rightarrow$ Evil (Father(John))**

- **Example:** Let's say that,
- "Priyanka got good marks in English."
- "Therefore, someone got good marks in English."

Uncertainty

- Let action *A(t)* denote leaving for the airport *t* minutes before the flight
- For a given value of *t*, will *A(t)* get me there on time?

- **Problems:**
  - Partial observability (roads, other drivers' plans, etc.)
  - Noisy sensors (traffic reports)
  - Uncertainty in action outcomes (flat tire, etc.)
  - Immense complexity of modeling and predicting traffic

How To Deal With Uncertainty

– **Implicit methods:**
– Ignore uncertainty as much as possible
– Build procedures that are robust to uncertainty
– This is the approach in the planning methods studied so far (e.g. monitoring and replanning)
– **Explicit methods**
– Build a model of the world that describes the uncertainty (about the system's state, dynamics, sensors, model)
– Reason about the effect of actions given the model

Methods for Handling Uncertainty

- *Default (non-monotonic) logic*: make assumptions unless contradicted by evidence.

  –E.g. “Assume my car doesnt have a flat tire.

  What assumptions are reasonable? What about contradictions?
- *Rules with fudge factor:*

  –E.g. “Sprinkler $\rightarrow_{0.99}$ WetGrass”, “WetGrass $\rightarrow_{0.7}$ Rain”

  But: Problems with combination (e.g. Sprinkler causes rain?)
- *Probability*:

  –E.g. Given what I know, *A(25)* succeed with probability 0.2
- *Fuzzy logic*:

  –E.g. WetGrass is true to degree 0.2

  But: Handles degree of truth, NOT uncertainty.

Why Not Use First-Order Logic?

- A purely logical approach has two main problems:
  - Risks falsehood
    - A(25) will get me there on time.
  - Leads to conclusions that are too weak:
    - *A(25)* will get me there on time if there is no accident on the bridge and it does not rain and my tires remain intact, etc. etc.
    - *A(1440)* might reasonably be said to get me there on time (but I would have to stay overnight at the airport!)

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

Probability

- A well-known and well-understood framework for dealing with uncertainty
- Has a clear semantics
- Provides principled answers for:
  - Combining evidence
  - Predictive and diagnostic reasoning
  - Incorporation of new evidence
- Can be learned from data
- Intuitive to human experts (arguably?)

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

Need for Reasoning w/ Uncertainty

- The world is full of uncertainty
  - chance nodes/sensor noise/actuator error/partial info..
  - Logic is brittle
    - can’t encode exceptions to rules
    - can‘t encode statistical properties in a domain
  - Computers need to be able to handle uncertainty
- Probability: new foundation for AI (& CS!)

- Massive amounts of data around today
  - Statistics and CS are both about data
  - Statistics lets summarize and understand it
  - Statistics is the basis for most learning
- Statistics lets data do our work for us

Beliefs (Bayesian Probabilities)

- We use probability to describe uncertainty due to:
  - Laziness: failure to enumerate exceptions, qualifications etc.
  - Ignorance: lack of relevant facts, initial conditions etc.
  - True randomness? Quantum effects? ...
- *Beliefs (Bayesian or subjective probabilities)* relate propositions to one's current state of knowledge
  - E.g. $P(A(25) \mid \text{no reported accident}) = 0.1$

  These are *not assertions about the world / absolute truth*
- Beliefs change with new evidence:
  - E.g. $P(A(25) \mid \text{no reported accident, 5am}) = 0.2$
- This is analogous to logical entailment: *KB* *given* the *KB*, but may not be true in general.

Making Decisions Under Uncertainty

- Suppose I believe the following:

$P(A(25)\ \textit{gets me there on time} \mid \ldots) = 0.04$

$P(A(90)\ \textit{gets me there on time} \mid \ldots) = 0.70$

$P(A(120)\ \textit{gets me there on time} \mid \ldots) = 0.95$

$P(A(1440)\ \textit{gets me there on time} \mid \ldots) = 0.9999$

- Which action should I choose?

Making Decisions Under Uncertainty

- Suppose I believe the following:

$P(A(25) \text{ gets me there on time} \mid \ldots) = 0.04$

$P(A(90) \text{ gets me there on time} \mid \ldots) = 0.70$

$P(A(120) \text{ gets me there on time} \mid \ldots) = 0.95$

$P(A(1440) \text{ gets me there on time} \mid \ldots) = 0.9999$

- Which action should I choose?
- Which action should I 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

Random Variables

- A *random variable* X describes an outcome that cannot be determined in advance
  - –E..g. The roll of a die
  - –E.g. Number of e-mails received in a day
- The *sample space (domain)* S of a random variable X is the set of all possible values of the variable
  - – E.g. For a die, $S = \{1, 2, 3, 4, 5, 6\}$
  - –E.g. For number of emails received in a day, S is the natural numbers
- An *event* is a subset of S.
  - –E.g. $e = \{1\}$ corresponds to a die roll of 1
  - –E.g. number of e-mails in a day more than 100

Probability for Discrete Random Variables

- Usually, random variables are governed by some "law of nature", described as a *probability function* P defined on S.
- $P(x)$ defines the chance that variable X takes value $x \in S$.
  – E.g. for a die roll with a fair die, $P(1) = P(2) = \cdots = P(6) = 1/6$
- Note that we still cannot determine the value of X, just the chance of encountering a given value
- If X is a discrete variable, then a probability space $P(x)$ has the following properties:

$$0 \leq P(x) \leq 1, \forall x \in S \text{ and } P(x) = 1$$

$x \in S$

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

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. $0 \le P(x) \le 1$ 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

Axioms of Probability

- Beliefs satisfy the axioms of probability.
- For any propositions A, B:

1. $0 \leq P(A) \leq 1$
2. $P(\text{True}) = 1$ (hence $P(\text{False}) = 0$)
3. $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$
4. Alternatively, if A and B are mutually exclusive ($A \wedge B = F$) then:

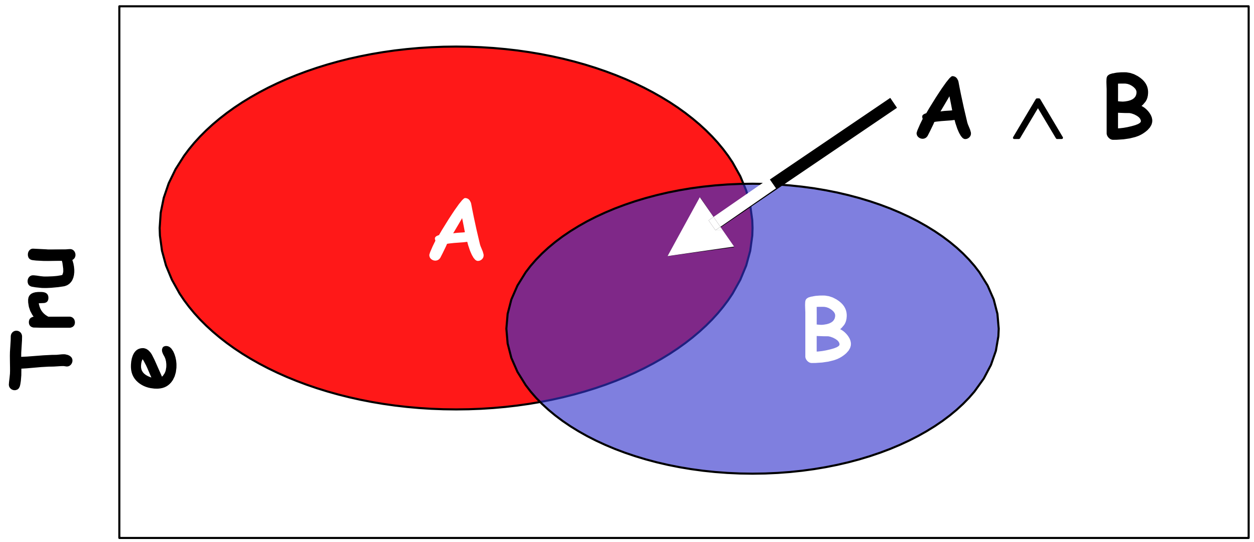
$$P(A \vee B) = P(A) + P(B)$$

- The axioms of probability limit the class of functions that can be considered probability functions.

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:

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



Beliefs

- We use probability to describe the world and existing uncertainty
- Agents will have *beliefs* based on their current state of knowledge
  - E.g. *P(Some day AI agents will rule the world)=0.2* reflects a personal belief, based on one's state of knowledge about current AI, technology trends etc.
- Different agents may hold different beliefs, as these are *subjective*
- Beliefs may change over time as agents get new evidence
- *Prior (unconditional) beliefs* denote belief prior to the arrival of any new evidence.

Defining Probabilistic Models

- We define the world as a set of random variables $\Omega = \{X_1 \ldots X_n\}$.
- A *probabilistic model* is an encoding of probabilistic information that allows us to compute the probability of any event in the world
- The world is divided into a set of elementary, mutually exclusive events, called *states*
  - E.g. If the world is described by two Boolean variables A, B, a state will be a complete assignment of truth values for A and B.
- A *joint probability distribution function* assigns non-negative weights to each event, such that these weights sum to 1.

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

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:

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

Joint distribution can answer any question

Inference using Joint Distributions

E.g. Suppose *Happy* and *Rested* are the random variables:

	$Happy = true$	$Happy = false$
$Rested = true$	0.05	0.1
$Rested = false$	0.6	0.25

The *unconditional probability* of any proposition is computable as the sum of entries from the full joint distribution

- E.g. *P(Happy)* = *P(Happy, Rested)* +*P(Happy,* ¬ *Rested)* = *0.65*

Conditional Probability

- The basic statements in the Bayesian framework talk about *conditional probabilities*.
  - $P(A|B)$ is the belief in event A given that event B is known with certainty
  - Conditional probabilities are written like P(A|B), which can be read to mean, "the **probability** that A happens **Given** B has happened."
- The *product rule* gives an alternative formulation:

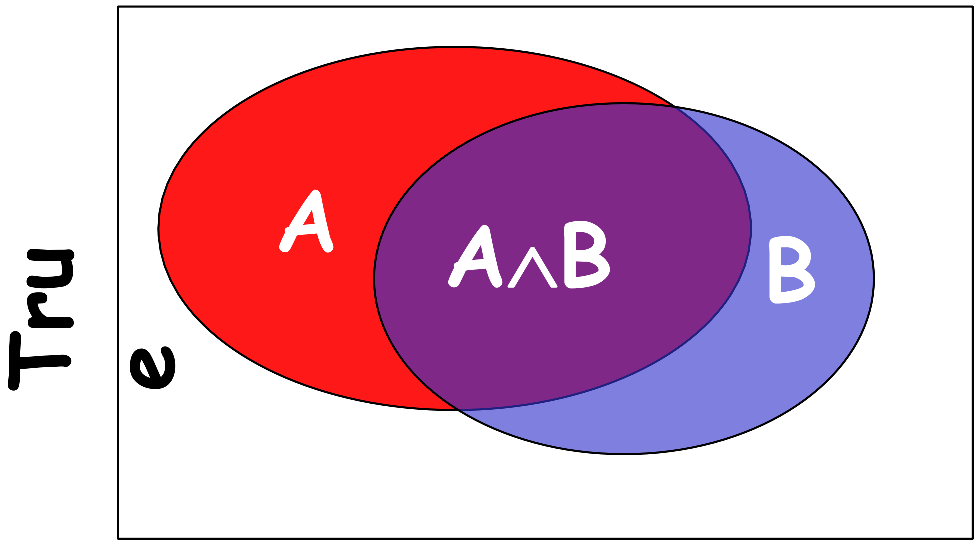
$$P(A \wedge B) = P(A|B)P(B) = P(B|A)P(A)$$

  - Note: we often write $P(A, B)$ as a shorthand for $P(A \wedge B)$

Conditional Probability

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

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

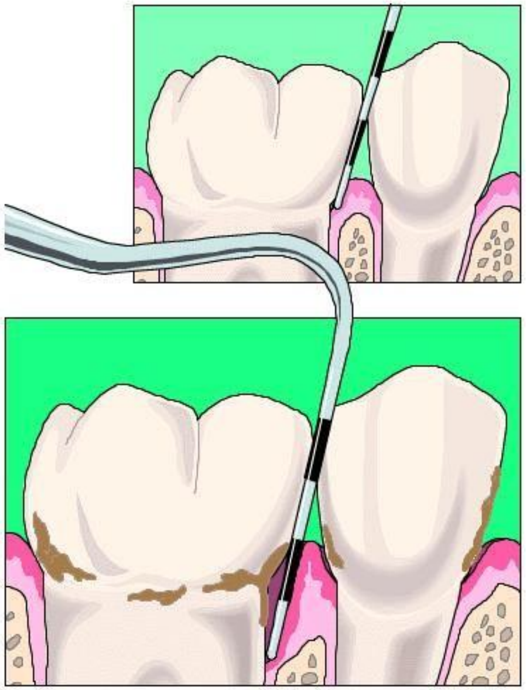


Chain Rule/Product Rule

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

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

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?

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

Inference by Enumeration

Start with the joint distribution:

	toothache		$\neg$ *toothache*	
	catch	$\neg$ *catch*	*catch*	$\neg$ *catch*
cavity	.108	.012	.072	.008
$\neg$ *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 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		$\neg$ *toothache*	
	catch	$\neg$ *catch*	*catch*	$\neg$ *catch*
cavity	.108	.012	.072	.008
$\neg$ *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$\vee$cavity) =

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:

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

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

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:

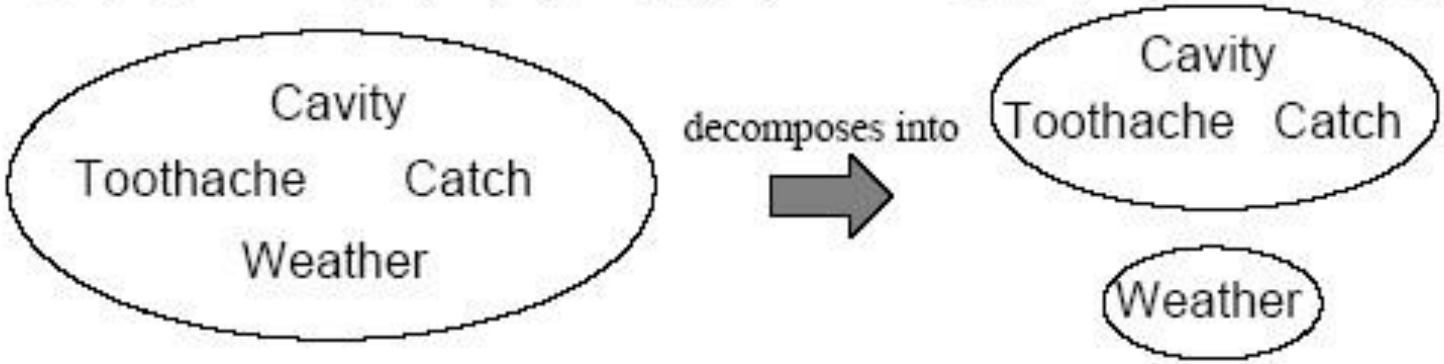
$$P(A \mid B) = \frac{P(A \wedge B)}{P(B)} = P(A)$$

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

Independence

A and B are independent iff

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



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

31 entries reduced to 10; 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 II

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

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

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

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

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

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

$$\boxed{\begin{array}{c} P(x \mid y) = \alpha P(y \mid x) P(x) \\ \text{posterior} \propto \text{likelihood} \cdot \text{prior} \end{array}}$$

Conditional Bayes Rule

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

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

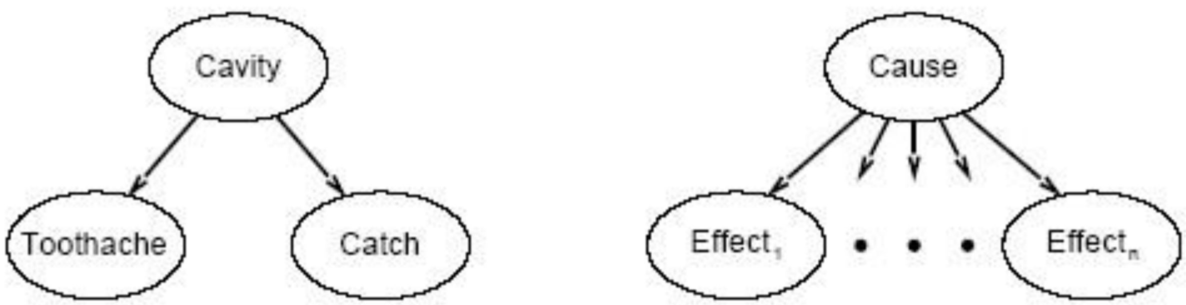
$$P(x \mid y, z) = \alpha P(y \mid x, z) P(x \mid z)$$

Bayes' Rule & Cond. Independence

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

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

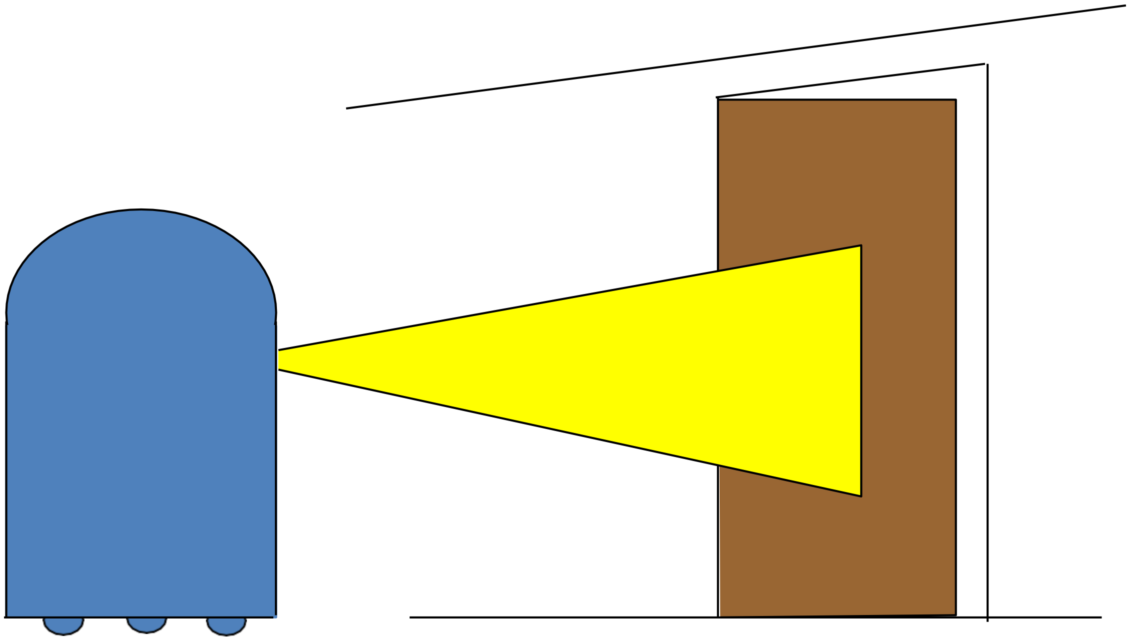
$$\mathbf{P}(Cause, Effect_1, \ldots, Effect_n) = \mathbf{P}(Cause)\Pi_i\mathbf{P}(Effect_i|Cause)$$



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

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

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

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

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

Combining Evidence

- Suppose our robot obtains another observation z_2.
- How can we integrate this new information?
- More generally, how can we estimate $P(x|\, z_1 \ldots z_n\,)$?

Example: Second Measurement

- $P(z_2|open) = 0.5$ $\quad\quad\quad P(z_2|\neg open) = 0.6$
- $P(open|z_1)=2/3$

$$P(open \mid z_2, z_1) = \frac{P(z_2 \mid open)\, P(open \mid z_1)}{P(z_2 \mid open)\, P(open \mid z_1) + P(z_2 \mid \neg open)\, P(\neg open \mid z_1)}$$

$$= \frac{\frac{1}{2} \cdot \frac{2}{3}}{\frac{1}{2} \cdot \frac{2}{3} + \frac{3}{5} \cdot \frac{1}{3}} = \frac{5}{8} = 0.625$$

- z_2 lowers 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