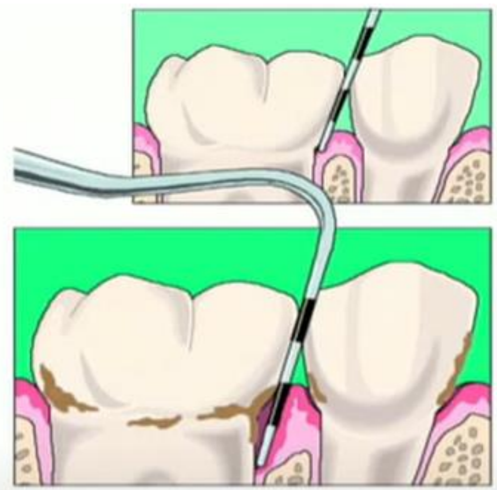
Reasoning under Uncertainty

Instructors: Dr. Durgesh Singh

CSE Discipline, PDPM IIITDM, Jabalpur -482005

Example

A domain consisting of just the three Boolean variables $Toothache$, $Cavity$, and $Catch$ (the dentist's nasty steel probe catches in my tooth).



Inference Using Full Joint Distributions

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 Using Full Joint Distributions

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) = .20 + .072 + .008

.28

Inference Using Full Joint Distributions

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:

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

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

Problems with joint distribution ??

- 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

Independence

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

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

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

Therefore, if A and B are independent:

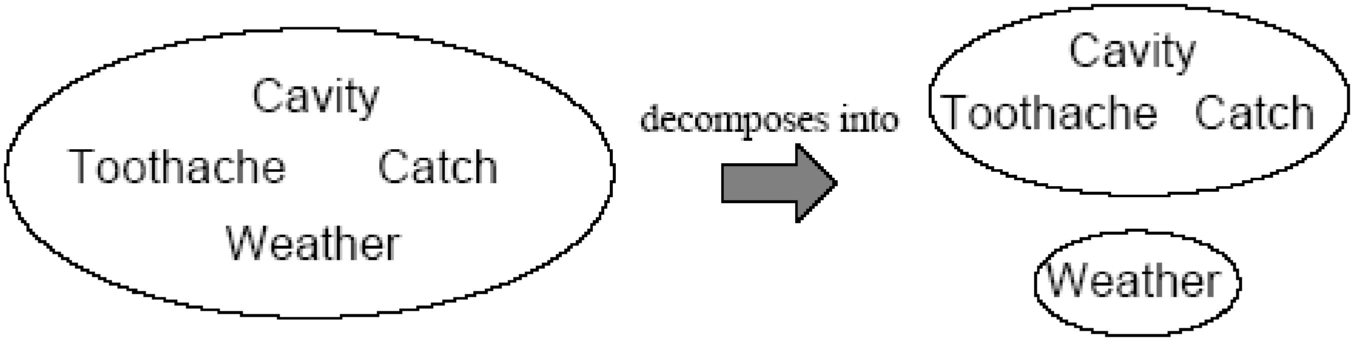
$$P(A \mid B) = \frac{P(A \wedge B)}{P(B)} = P(A)$$

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

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;

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 $Cavity$:

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

Conditional Independence

- The general definition of **conditional independence** of two variables X and Y, given a third variable Z is

(I) $$\mathbf{P}(X, Y \mid Z) = \mathbf{P}(X \mid Z)\mathbf{P}(Y \mid Z).$$

(II) $$\mathbf{P}(X \mid Y, Z) = \mathbf{P}(X \mid Z) \quad \text{and} \quad \mathbf{P}(Y \mid X, Z) = \mathbf{P}(Y \mid Z)$$

Conditional Independence II

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

Equivalent statements:

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

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

$$P(H \mid E) = \frac{P(E \mid H)P(H)}{P(E)}$$

Simple proof from def of conditional probability:

$$P(H \mid E) = \frac{P(H \wedge E)}{P(E)}$$ (Def. cond. prob.)

$$P(E \mid H) = \frac{P(H \wedge E)}{P(H)}$$ (Def. cond. prob.)

$$P(H \wedge E) = P(E \mid H)P(H)$$ (Mult by P(H) in line 2)

$$P(H \mid E) = \frac{P(E \mid H)P(H)}{P(E)}$$ (Substitute #3 in #1)

Use to Compute *Diagnostic* Probability from *Causal* Probability

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

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!

Bayes Rule

- Does patient have cancer or not?

Given: A patient takes a lab test, and the result comes back positive. The test returns a correct positive result in only 98% of the cases in which the disease is present, and a correct negative result in only 97% of the cases in which the disease is not present. Furthermore, 0.008 of the entire population have this cancer.

$$P(cancer) = \qquad P(\neg cancer) =$$

$$P(+ \mid cancer) = \qquad P(- \mid cancer) =$$

$$P(+ \mid \neg cancer) = \qquad P(- \mid \neg cancer) =$$

$$P(cancer) = 0.008 \qquad P(\neg cancer) = 0.992$$

$$P(+ \mid cancer) = 0.98 \qquad P(- \mid cancer) = 0.02$$

$$P(+ \mid \neg cancer) = 0.03 \qquad P(- \mid \neg cancer) = 0.97$$

$$P(cancer \mid +) = \frac{P(+ \mid cancer) P(cancer)}{P(+)};$$

$$P(\neg cancer \mid +) = \frac{P(+ \mid \neg cancer) P(\neg cancer)}{P(+)}$$

$$P(cancer \mid +) P(+) = 0.98 \times 0.008 = 0.0078;$$

$$P(\neg cancer \mid +) P(+) = 0.03 \times 0.992 = 0.0298$$

$$P(+) = 0.0078 + 0.0298$$

$$P(cancer \mid +) = 0.21; \qquad P(\neg cancer \mid +) = 0.79$$

The patient, more likely than not, does not have cancer

Bayesian Networks

- In general, joint distribution over set of variables ($X_1, X_1, \ldots, X_n$) requires exponential space for representation & inference.
- We also saw that independence and conditional independence relationships among variables can greatly reduce the number of probabilities that need to be specified in order to define the full joint distribution.
- BNs(a graphical representation) is a data structure
  - represents the dependencies among variables and
  - give a concise specification of any full joint probability distribution

Chain rule in Bayesian Networks

$$\begin{aligned} P(x_1,\ldots,x_n) &= P(x_n|x_{n-1},\ldots,x_1)P(x_{n-1}|x_{n-2},\ldots,x_1)\cdots P(x_2|x_1)P(x_1) \\ &= \prod_{i=1}^{n} P(x_i|x_{i-1},\ldots,x_1). \end{aligned}$$

The general assertion that, for every variable Xi in the Bayesian network,

$$\mathbf{P}(X_i|X_{i-1},\ldots,X_1) = \mathbf{P}(X_i|Parents(X_i))$$

$$P(x_1,\ldots,x_n) = \prod_{i=1}^{n} P(x_i|parents(X_i)).$$

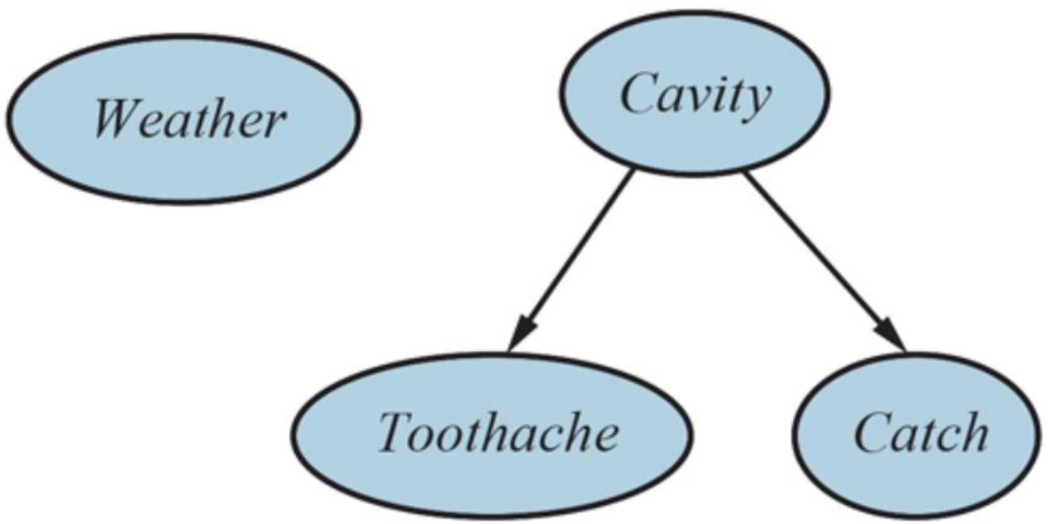
Bayes Networks

- A Bayesian network is a directed graph in which each node is annotated with quantitative probability information.
- The full specification is as follows:

1. Each node corresponds to a random variable, which may be discrete or continuous.
2. Directed links or arrows connect pairs of nodes. If there is an arrow from node X to node Y , X is said to be a parent of Y.
3. Each node X_i, has a conditional probability distribution P (Xi | Parents (Xi)) that quantifies the effect of the parents on the node.
4. The graph has no directed cycles (and hence is a directed, acyclic graph, or DAG).

Example

Topology of network encodes conditional independence assertions:



A simple Bayesian network in which *Weather* is independent of the other three variables and *Toothache* and *Catch* are conditionally independent, given *Cavity*.

Example: Burglar Alarm

- You have a new burglar alarm installed at home.
- It is reliable at detecting a burglary, but also responds on occasion to minor earthquakes.
- You also have two neighbors, John and Mary, who have promised to call you at work when they hear the alarm.
- John always calls when he hears the alarm, but sometimes confuses the telephone ringing with the alarm and calls then, too.
- Mary, on the other hand, likes loud music and sometimes misses the alarm altogether.

Given the evidence of who has or has not called, we would like to estimate the probability of a burglary.

Example: Burglar Alarm