Bayesian Networks

CSE 573

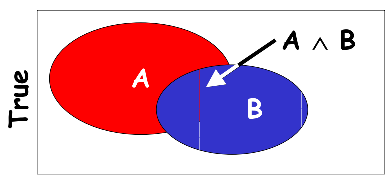
Last Time

- Basic notions
  - Atomic events
  - Probabilities
  - Joint distribution
    - Inference by enumeration
  - Independence & conditional independence
  - Bayes' rule
- Bayesian networks
- Statistical learning
- Dynamic Bayesian networks (DBNs)
- Markov decision processes (MDPs)

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:

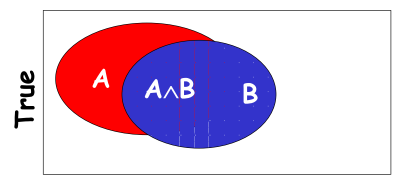
$$P(A \vee B) = P(A) + P(B) - P(A \wedge B)$$



Conditional Probability

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

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



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

Can also compute conditional probabilities:

$$P(\neg cavity \mid toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)} = \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

Problems ??

- Worst case time: $O(n^d)$
  Where d = max arity
  And n = number of random variables
- Space complexity also $O(n^d)$
  Size of joint distribution
- How get $O(n^d)$ entries for table??

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

Value of cavity & catch irrelevant - When computing P(toothache)

Independence

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

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

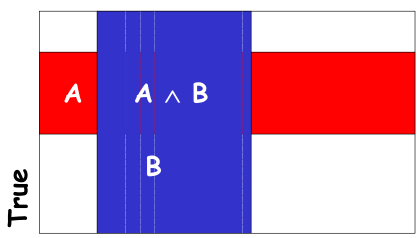
These two constraints are logically equivalent

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

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

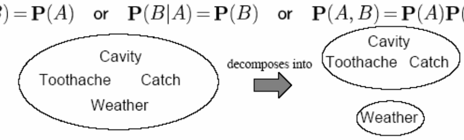
$$P(A \wedge B) = P(A)P(B)$$

Independence



Independence

A and B are independent iff
$\mathbf{P}(A|B) = \mathbf{P}(A)$ or $\mathbf{P}(B|A) = \mathbf{P}(B)$ or $\mathbf{P}(A,B) = \mathbf{P}(A)\mathbf{P}(B)$



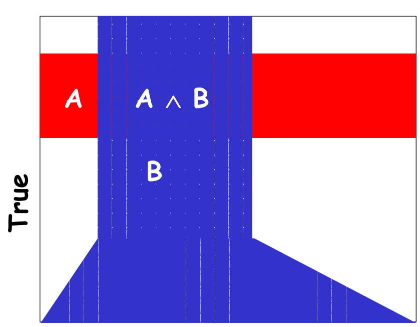
$\mathbf{P}(Toothache, Catch, Cavity, Weather)$
$= \mathbf{P}(Toothache, Catch, Cavity)\mathbf{P}(Weather)$

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?

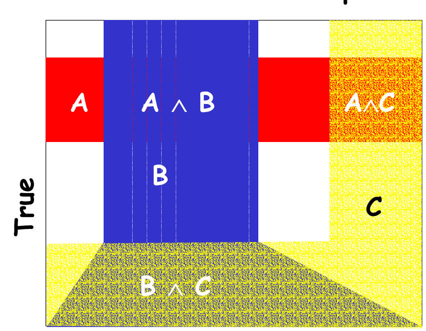
Conditional Independence

A&B *not* independent, since P(A|B) < P(A)



Conditional Independence

But: A&B are *made* independent by ¬C



P(A|¬C) = P(A|B,¬C)

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

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

Simple proof from def of conditional probability:

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

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

$$P(H \wedge E) = P(E \mid H)P(H) \quad \text{(Mult by P(H) in line 1)}$$

QED: $$P(H \mid E) = \frac{P(E \mid H)P(H)}{P(E)} \quad \text{(Substitute \#3 in \#2)}$$

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) $\equiv \dfrac{P(s|m)P(m)}{P(s)} = \dfrac{0.8 \times 0.0001}{0.1} = 0.0008$

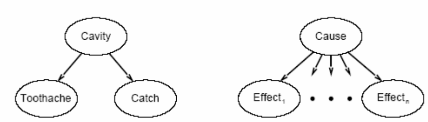
Note: posterior probability of meningitis still very small!

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

Bayes Nets

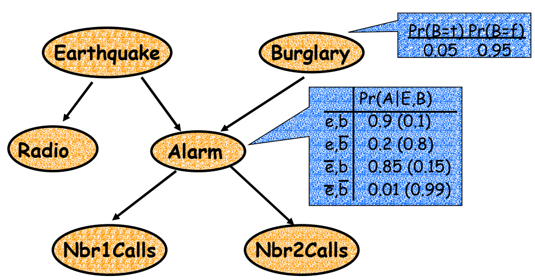
- In general, joint distribution *P* over set of variables ($X_1 \times ... \times X_n$) requires exponential space for representation & inference
- BNs provide a graphical representation of *conditional independence* relations in *P*
  - usually quite compact
  - requires assessment of fewer parameters, those being quite natural (e.g., causal)
  - efficient (usually) inference: query answering and belief update

Independence (in the extreme)

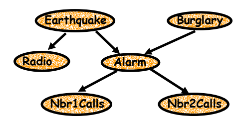
- If $X_1, X_2, ... X_n$ are mutually independent, then

  $P(X_1, X_2, ... X_n) = P(X_1)P(X_2) ... P(X_n)$
- Joint can be specified with *n* parameters
  - cf. the usual 2^n-1 parameters required
- While extreme independence is unusual,
  - Conditional independence is common
- BNs exploit this conditional independence

An Example Bayes Net



Earthquake Example (con't)

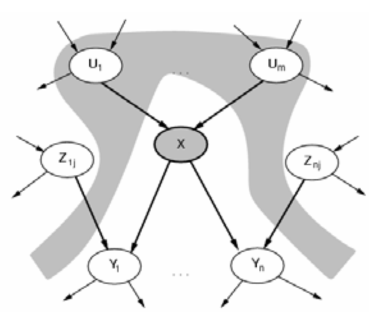


- **If I know if *Alarm*, no other evidence influences my degree of belief in *Nbr1Calls***
  - $P(N1|N2,A,E,B) = P(N1|A)$
  - also: $P(N2|N2,A,E,B) = P(N2|A)$ and $P(E|B) = P(E)$
- **By the chain rule we have**
  - $P(N1,N2,A,E,B) = P(N1|N2,A,E,B) \cdot P(N2|A,E,B) \cdot P(A|E,B) \cdot P(E|B) \cdot P(B)$
  - $= P(N1|A) \cdot P(N2|A) \cdot P(A|B,E) \cdot P(E) \cdot P(B)$
- **Full joint requires only 10 parameters (cf. 32)**

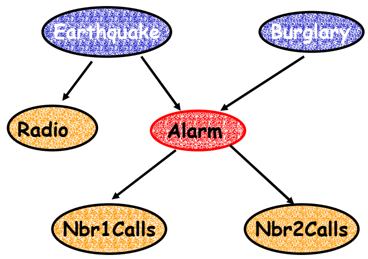
BNs: Qualitative Structure

- Graphical structure of BN reflects conditional independence among variables
- Each variable *X* is a node in the DAG
- Edges denote *direct probabilistic influence*
  - usually interpreted *causally*
  - parents of *X* are denoted *Par(X)*
- ***X* is conditionally independent of all nondescendants given its parents**
  - Graphical test exists for more general independence
  - "Markov Blanket"

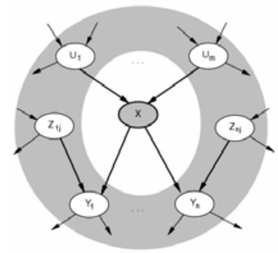
Given Parents, X is Independent of Non-Descendants



For Example

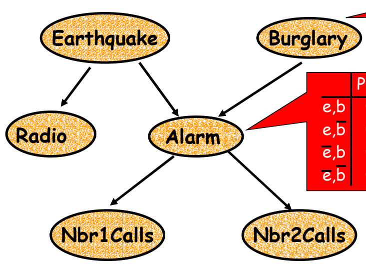


Given Markov Blanket, X is Independent of All Other Nodes



MB(X) = Par(X) ∪ Childs(X) ∪ Par(Childs(X))

Conditional Probability Tables



Pr(B=t)	Pr(B=f)
0.05	0.95

	Pr(A|E,B)
e,b	0.9 (0.1)
$e,\bar{b}$	0.2 (0.8)
$\bar{e},b$	0.85 (0.15)
$\bar{e},\bar{b}$	0.01 (0.99)

Conditional Probability Tables

- For complete spec. of joint dist., *quantify* BN
- For each variable X, specify **CPT**: $P(X \mid Par(X))$
  - number of params *locally* exponential in $|Par(X)|$
- If $X_1, X_2, \ldots X_n$ is any topological sort of the network, then we are assured:

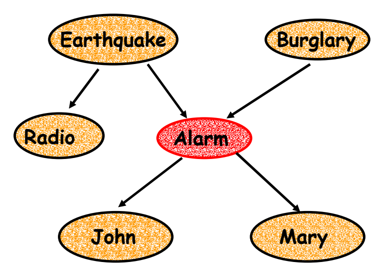
$$P(X_n, X_{n-1}, \ldots X_1) = P(X_n \mid X_{n-1}, \ldots X_1) \cdot P(X_{n-1} \mid X_{n-2}, \ldots X_1) \ldots P(X_2 \mid X_1) \cdot P(X_1)$$

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

Inference in BNs

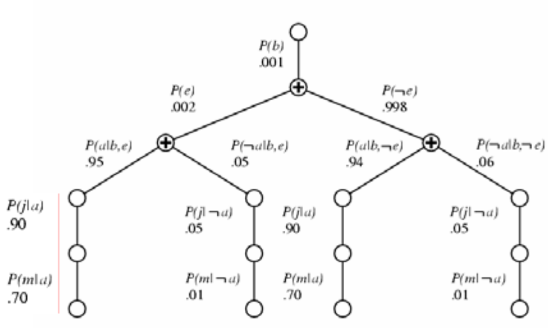
- The graphical independence representation
  - yields efficient inference schemes
- We generally want to compute
  - *Pr(X)*, or
  - *Pr(X|**E**)* where ***E*** is (conjunctive) evidence
- Computations organized by network topology
- One simple algorithm:
  - *variable elimination (VE)*

P(B | J=true, M=true)



$$P(b \mid j,m) = \alpha P(b) \sum_e P(e) \sum_a P(a \mid b,e) P(j \mid a) P(m,a)$$

Structure of Computation



Dynamic Programming

Variable Elimination

- A *factor* is a function from some set of variables into a specific value: e.g., *f(E,A,N1)*
  - CPTs are factors, e.g., *P(A|E,B)* function of *A,E,B*
- VE works by *eliminating* all variables in turn until there is a factor with only query variable
- To eliminate a variable:
  - *join* all factors containing that variable (like DB)
  - *sum out* the influence of the variable on new factor
  - exploits product form of joint distribution

Example of VE: P(N1)

P(N1)

$= \Sigma_{N2,A,B,E}$ P(N1,N2,A,B,E)

$= \Sigma_{N2,A,B,E}$ P(N1|A)P(N2|A) P(B)P(A|B,E)P(E)

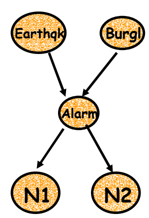
$= \Sigma_{A}$P(N1|A) Σ_{N2}P(N2|A) Σ_{B}P(B) Σ_{E}P(A|B,E)P(E)

$= \Sigma_{A}$P(N1|A) Σ_{N2}P(N2|A) Σ_{B}P(B) f1(A,B)

$= \Sigma_{A}$P(N1|A) Σ_{N2}P(N2|A) f2(A)

$= \Sigma_{A}$P(N1|A) f3(A)

= f4(N1)



Notes on VE

- Each operation is a simply multiplication of factors and summing out a variable
- Complexity determined by size of largest factor
  - e.g., in example, 3 vars (not 5)
  - linear in number of vars,
  - exponential in largest factorelimination ordering greatly impacts factor size
  - optimal elimination orderings: NP-hard
  - heuristics, special structure (e.g., polytrees)
- Practically, inference is much more
- tractable using structure of this sort