Bayes Nets

Chapter 14

Probabilistic Models

* Describe how the world works
* Models are always simplifications! But even if not exact can be useful...
* What do agents do with Prob. Models?
    * reason about unknown variables given evidence (sensors, experience)

Probabilistic Models

* Probabilistic Model is a joint distribution over a set of variables

$$P(X_1, X_2, \ldots X_n)$$

* Posterior probabilities are what we use to reason about the world...ask queries

$$P(X_q | x_{e_1}, \ldots x_{e_k})$$

Independence

* Two variables are independent if:

$$\forall x, y : P(x, y) = P(x)P(y)$$

$$\forall x, y : P(x|y) = P(x)$$

* Notation for independent: $X \perp Y$
* This is a simplifying modeling assumption.
* EX: {Weather, Traffic, Cavity, Toothache}

Example Independence

* N fair, independent coin flips

  $P(X_1), P(X_2), \ldots, P(X_n)$

  .5 .5 5

* Always 50/50 no matter what the previous result was....
* $P(X_1, X_2 \ldots X_n) = P(X_1)P(X_2) \ldots P(X_n)$

Conditional Independence

* P(Toothache, Cavity, Catch)
* If I have a cavity, prob that probe catches doesn't depend on whether I have a toothache
    * P(catch | ache, cav) = P(catch | cav)
* The same independence holds if I don't have a cavity
    * P(catch | ache, ¬cav) = P(catch | ¬cav)
* Catch **Conditionally Indep** of Ache given Cavity
    * P(Catch | Ache, Cav) = P(Catch | Cav)
* Effects are Independent given their Common Cause

Overview

* Bayes Nets (Graphical Models)
    * Syntax, Semantics
    * How to compactly represent Joint Distributions
    * How to efficiently do inference
* Dynamic Bayes Nets
    * How to adapt BN to reason over time
    * Markov Models, Hidden MM, Particle Filters
* Project 3 will use these concepts!

Bayes Rule

* Prod. Rule: $P(a \wedge b) = P(a|b)P(b) = P(b|a)P(a)$
* Bayes Rule: $P(a|\ b) = \frac{P(b|a)P(a)}{P(b)}$
* Good for talking Causal Probability

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

* Example Meningitis, Stiff Neck

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

Conditional Independence

* Conditional Independence is our most basic and robust knowledge about uncertain environments

$$\forall x, y, z : P(x, y|z) = P(x|z)P(y|z)$$

$$\forall x, y, z : P(x|z, y) = P(x|z)$$

$$X \perp\!\!\!\perp Y | Z$$

* What about this domain:
  * Traffic
  * Need an Umbrella
  * Raining

Bayes Nets

* Two problems with full joint distribution tables for prob. models
  * gets WAY too big
  * awkward to specify joint prob for more than a few vars
* Bayes Nets are a technique for describing complex joint distributions with simple local distributions (Conditional Probabilities)

The Chain Rule

Remember....the definition of conditional probability, also called the Product Rule:

$$P(a \wedge b) = P(a \mid b)\, P(a)$$

$$P(x_1, \ldots, x_n) = P(x_n \mid x_{n-1}, \ldots, x_1)\, P(x_{n-1}, \ldots, x_1)$$

$$P(x_1, \ldots, x_n) = P(x_n \mid x_{n-1}, \ldots, x_1)\, P(x_{n-1} \mid x_{n-2}, \ldots, x_1)\, P(x_{n-2}, \ldots, x_1)$$

$$P(x_1, \ldots, x_n) = P(x_n \mid x_{n-1}, \ldots, x_1)\, P(x_{n-1} \mid x_{n-2}, \ldots, x_1) \ldots P(x_2 \mid x_1)\, P(x_1)$$

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid x_{i-1}, \ldots, x_1)$$

Chain rule is the product rule applied multiple times, turning a joint probability into conditional probabilities

Traffic,Rain,Umbrella

* Trivial decomposition

$$P(\text{Traffic}, \text{Rain}, \text{Umbrella}) =$$

$$P(\text{Rain})P(\text{Traffic}|\text{Rain})P(\text{Umbrella}|\text{Rain}, \text{Traffic})$$

* Conditional Independence assumptions

$$P(\text{Traffic}, \text{Rain}, \text{Umbrella}) =$$

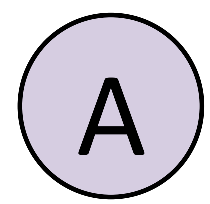
$$P(\text{Rain})P(\text{Traffic}|\text{Rain})P(\text{Umbrella}|\text{Rain})$$

* Bayes Nets let us compactly express conditional independence

Example: Factorization and Graph

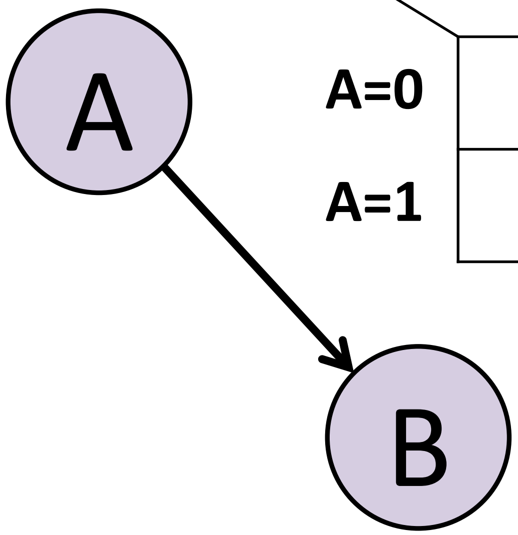
(chain rule of probability)

$$P(A,B,C,D) = P(D \mid C,B,A)P(C \mid B,A)P(B \mid A)\boxed{P(A)}$$



Example: Factorization and Graph

$$P(A,B,C,D) = P(D \mid C,B,A)P(C \mid B,A)\boxed{P(B \mid A)}P(A)$$

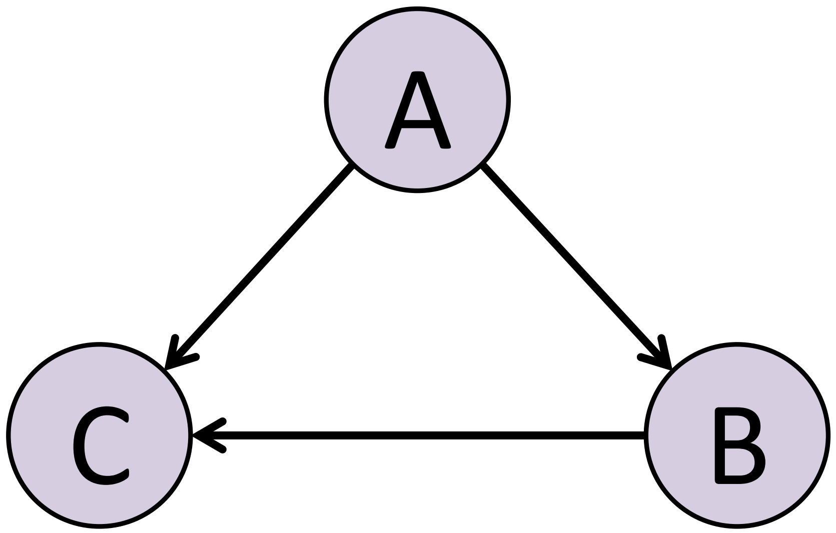


	B=0	B=1
A=0	0.4	0.6
A=1	0.7	0.3

Conditional Probability Table (CPT)

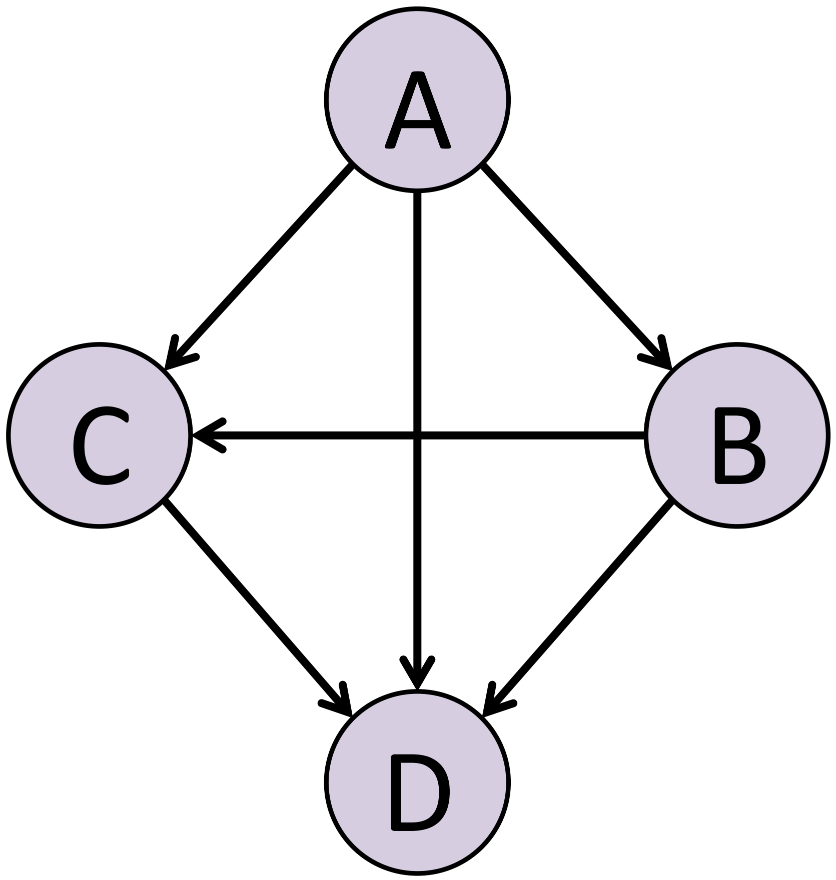
Example: Factorization and Graph

$$P(A,B,C,D) = P(D \mid C,B,A)\boxed{P(C \mid B,A)}P(B \mid A)P(A)$$



Example: Factorization and Graph

$$P(A,B,C,D) = \boxed{P(D \mid C,B,A)}P(C \mid B,A)P(B \mid A)P(A)$$



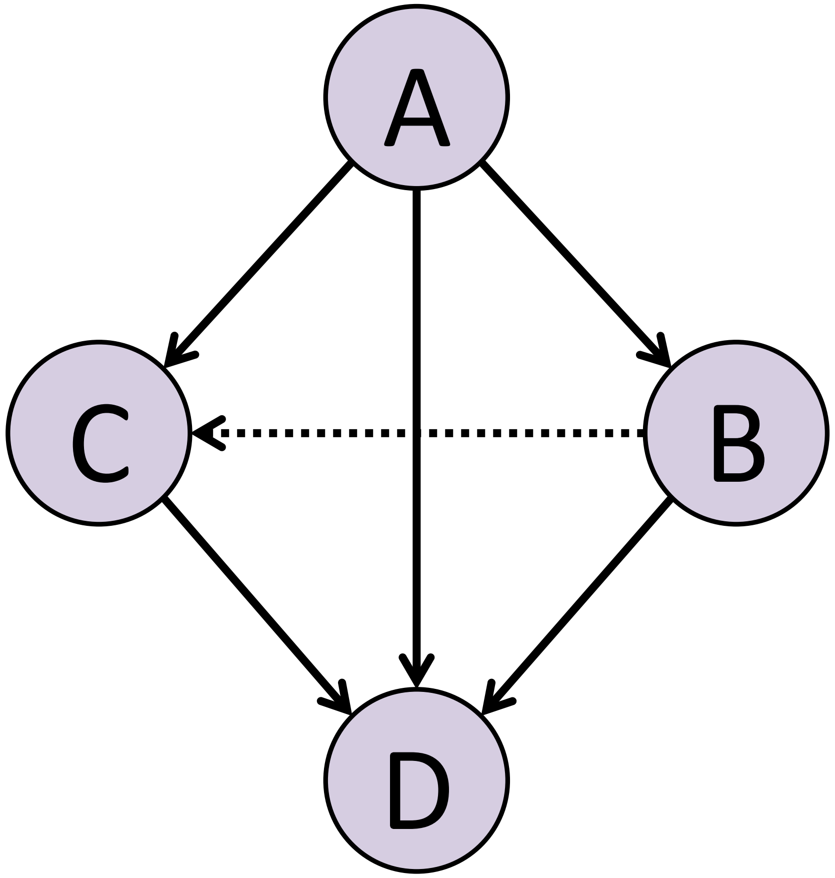
Complete graph

Example: Conditional Independencies

$$P(A,B,C,D) = P(D \mid C,B,A)\cancel{P(C \mid B,A)}P(B \mid A)P(A)$$

$$P(C \mid A)$$

$$\frac{I_\ell(P)}{C \perp B \mid A}$$

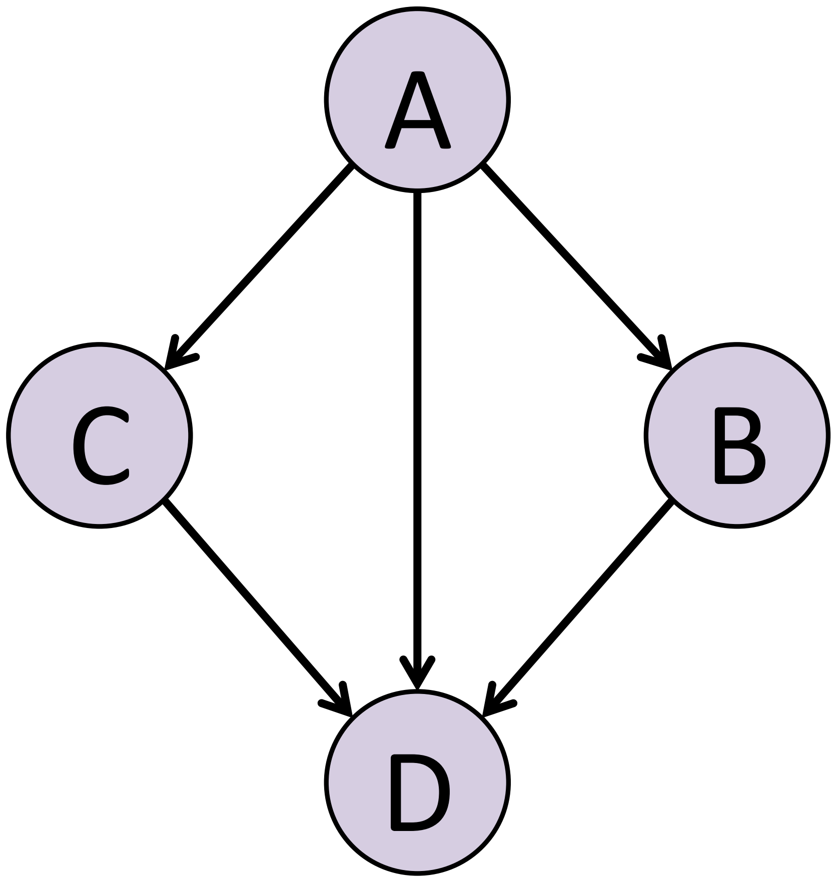


Example: Conditional Independencies

$$P(A,B,C,D) = P(D \mid C,B,A)\cancel{P(C \mid B,A)}P(B \mid A)P(A)$$

$$P(C \mid A)$$

$$\frac{I_{\ell}(P)}{C \perp B \mid A}$$



Example: Conditional Independencies

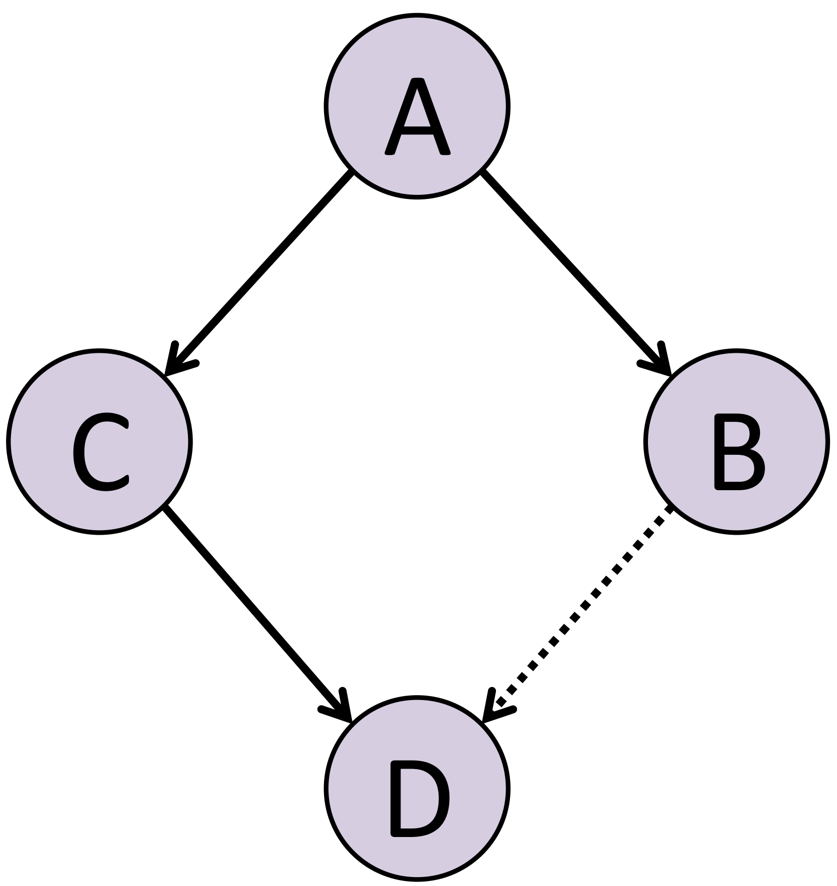
$$P(A,B,C,D) = \cancel{P(D \mid C,B,A)}\,\cancel{P(C \mid B,A)}\,P(B \mid A)P(A)$$

$$P(D \mid C) \qquad P(C \mid A)$$

$$\frac{I_\ell(P)}{}$$

$C \perp B \mid A$

$D \perp \{A,B\} \mid C$



Example: Conditional Independencies

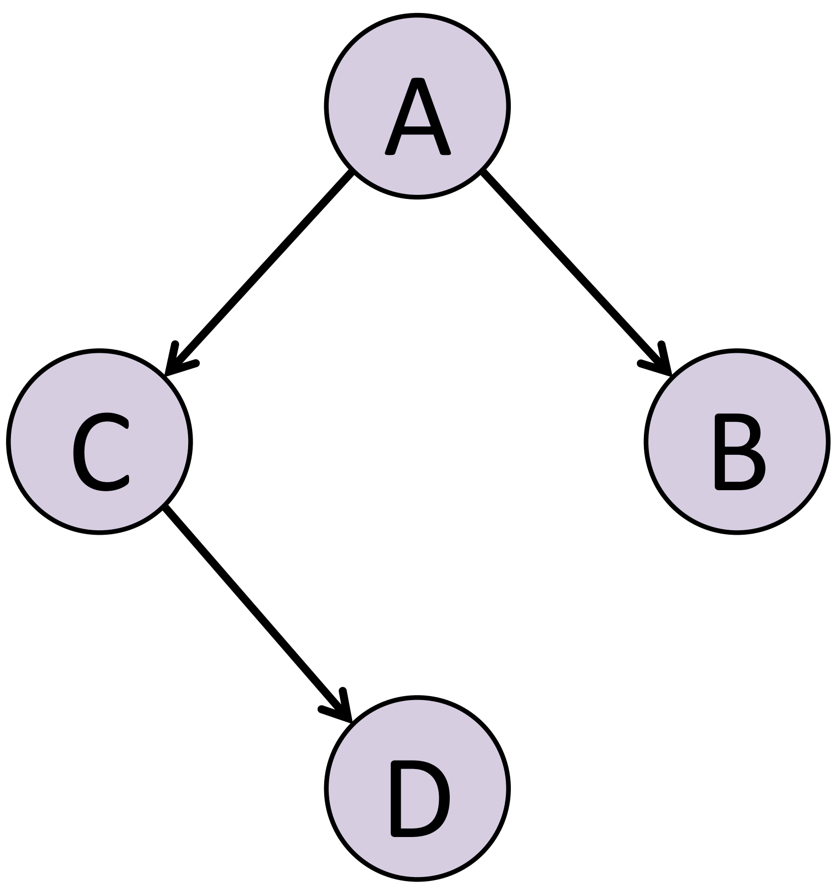
$$P(A,B,C,D) = \cancel{P(D \mid C,B,A)}\,\cancel{P(C \mid B,A)}\,P(B \mid A)\,P(A)$$

$$P(D \mid C) \qquad P(C \mid A)$$

$$I_\ell(P)$$

$C \perp B \mid A$

$D \perp \{A,B\} \mid C$



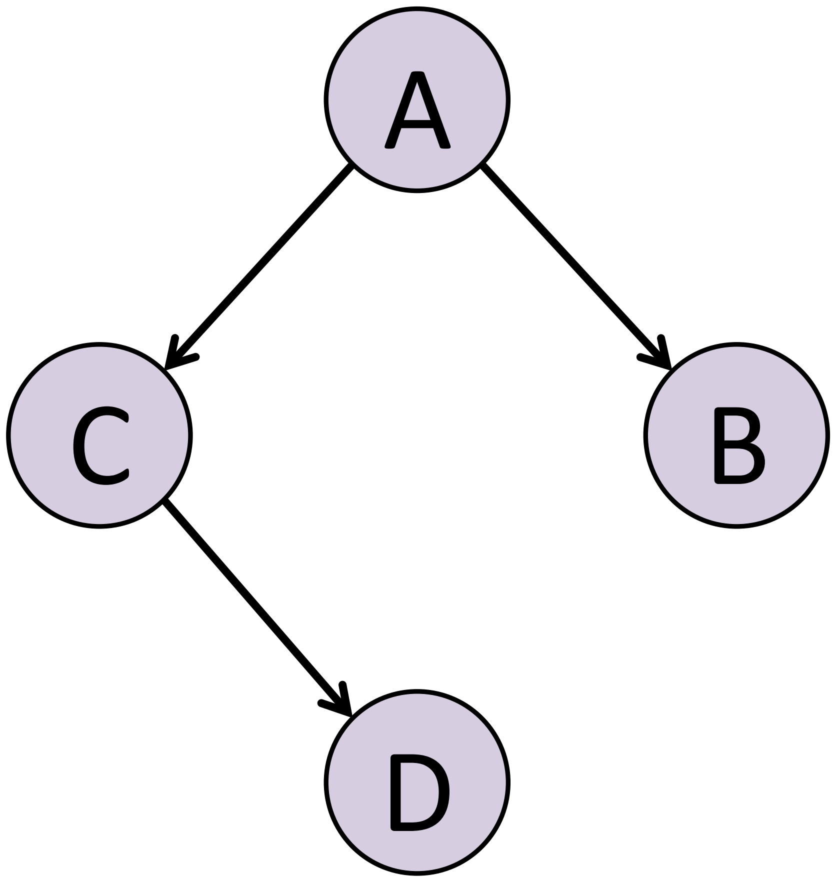
Chain Rule of Bayesian Networks

$$P(A,B,C,D) = P(D \mid C)P(C \mid A)P(B \mid A)P(A)$$

$$I_\ell(P)$$

$$C \perp B \mid A$$

$$D \perp \{A,B\} \mid C$$

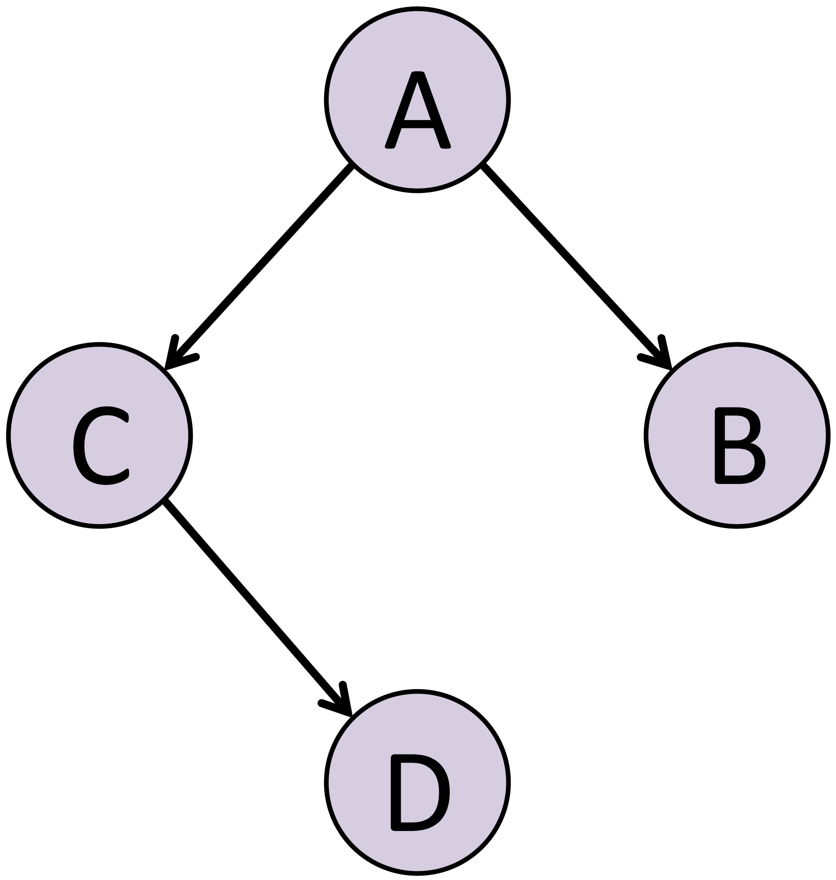


Chain Rule of Bayesian Networks

$$P(A,B,C,D) = P(D \mid C)P(C \mid A)P(B \mid A)P(A)$$

In general: $$P(X) = \prod_{i=1}^{N} P(x_i \mid \mathrm{pa}(x_i))$$ *Chain rule of Bayes nets*

$$\frac{I_\ell(P)}{\begin{array}{l} C \perp B \mid A \\ D \perp \{A,B\} \mid C \end{array}}$$

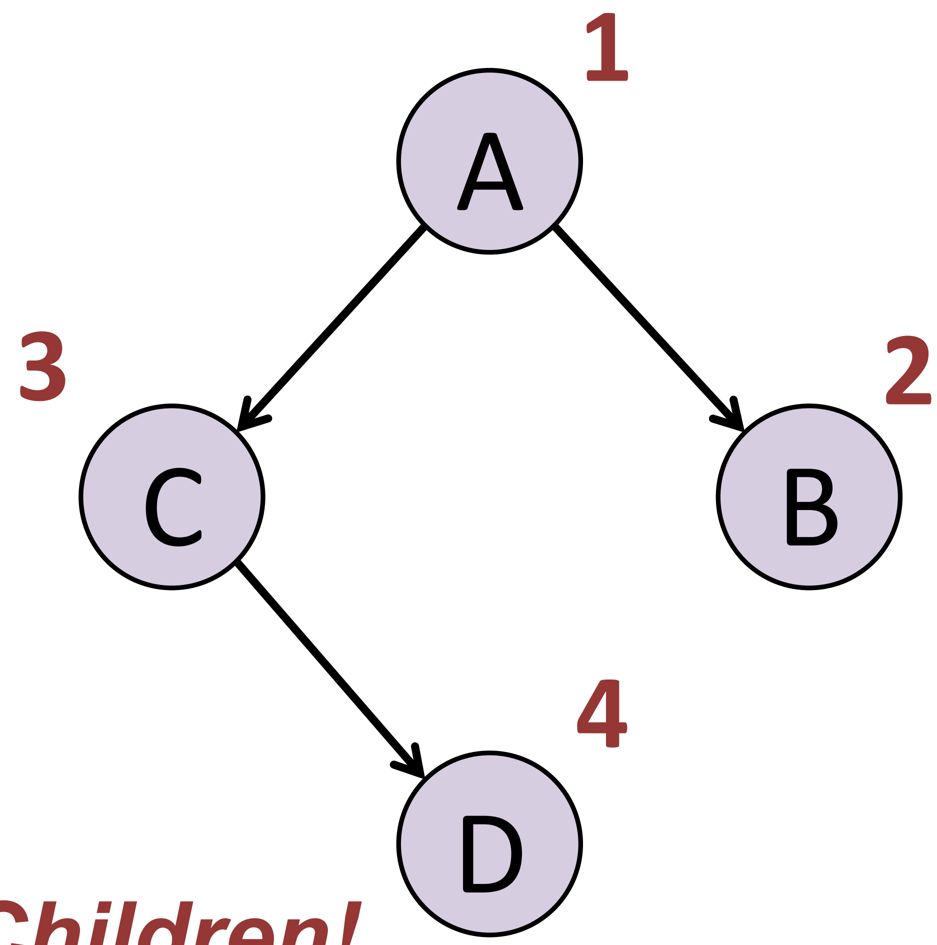


Chain Rule of Bayesian Networks

$$P(A,B,C,D) = P(D \mid C)P(C \mid A)P(B \mid A)P(A)$$

In general: $$P(X) = \prod_{i=1}^{N} P(x_i \mid \mathrm{pa}(x_i))$$ *Chain rule of Bayes nets*

$$\frac{I_\ell(P)}{\begin{array}{l} C \perp B \mid A \\ D \perp \{A,B\} \mid C \end{array}}$$

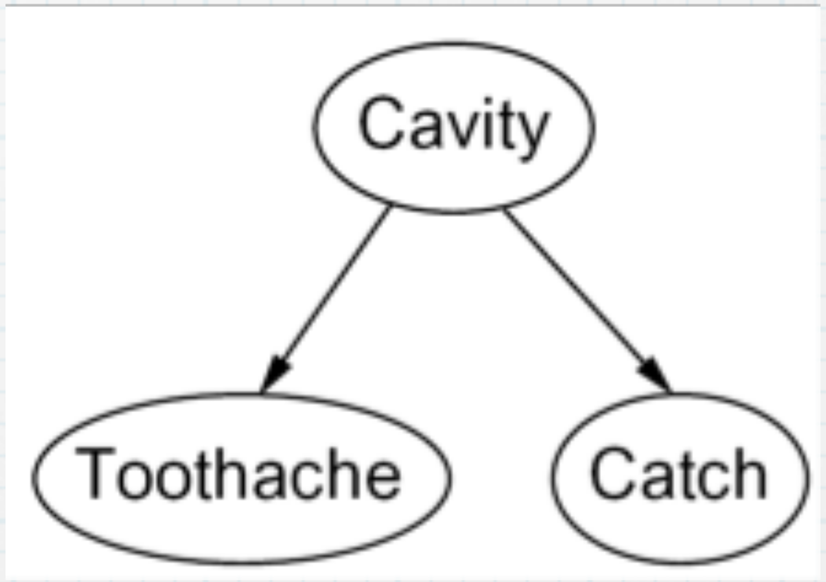


Topological Order:
A, B, C, D
Parents come before Children!

Variable Elimination Example

Bayes Net Notation

- Nodes: variables (with domains)
- Arcs: interactions
  - Directional
  - "Direct Influence" between vars
  - Formally: conditional indep.



Example: Coin Flips

- N independent flips

- No interactions between variables

Example: Traffic

- Variables:
  - R: It rains
  - T: There's traffic
- Model 1: independence
- Model 2: rain causes traffic
- Which is better for an agent to use?

Example: Traffic 2

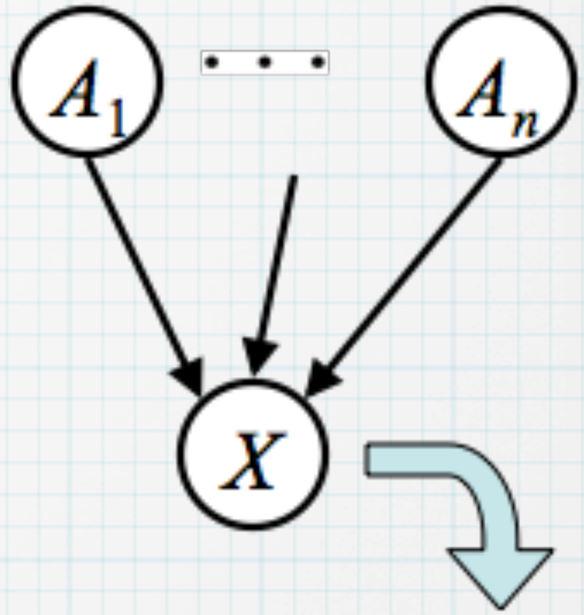
* Let's build a causal graphical model
* Variables:
  * **T**: Traffic
  * **R**: It rains
  * **L**: Low air pressure
  * **D**: Roof drips
  * **B**: Ballgame
  * **C**: Cavity

Example: Burglar Alarm

* Variables:
  * B: Burglary
  * A: Alarm goes off
  * M: Mary calls
  * J: John calls
  * E: Earthquake

Bayes' Net Semantics

- Formalizing the semantics of a BN
- A set of nodes, one per variable X
- A directed, acyclic graph
- A conditional distribution for each node
    - Local conditional prob tables (CPT)
    - P(X | parent nodes)

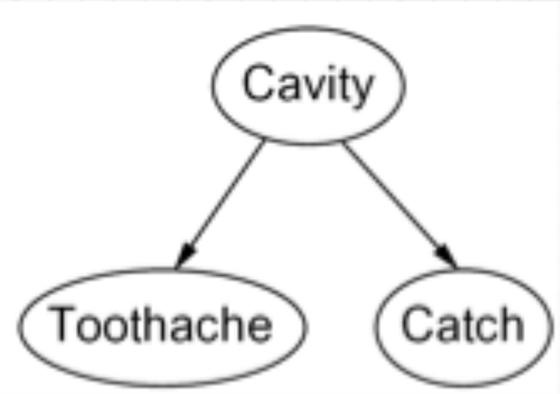


$P(X | A_1 \ldots A_n)$

Bayes' Net = Graph Topology + CPTs

Probabilities in BNs

* Bayes' nets implicitly encode joint distributions
    * As a product of local cond. distrib.
    * Can multiply all relevant conditionals to get any full joint

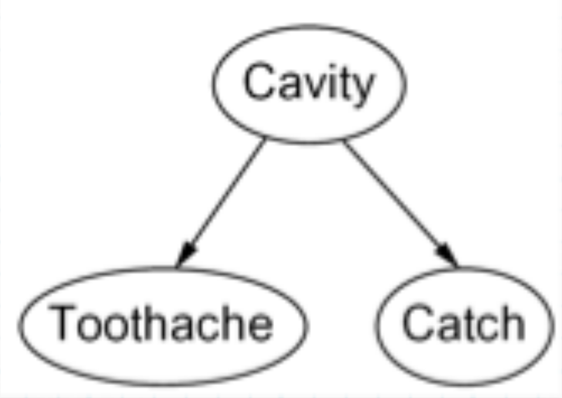


$$P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i | parents(X_i))$$

* Example: $P(+cavity, +catch, \neg toothache)$
* This let's us construct any entry in the full joint distribution table!

Probabilities in BNs

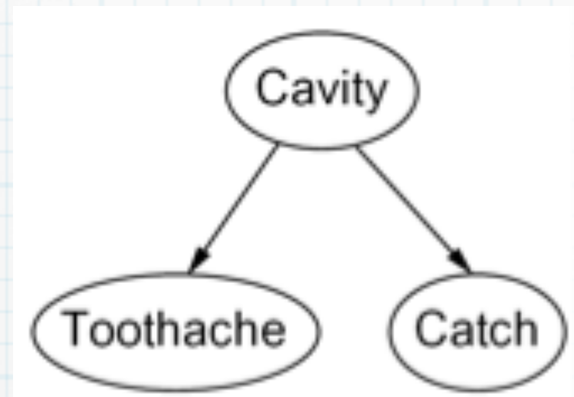
Bayes Net:
Structure + CPTs



P(Cavity)
P(Toothache | Cavity)
P(Catch | Cavity)

Probabilities in BNs

Bayes Net:
Structure + CPTs



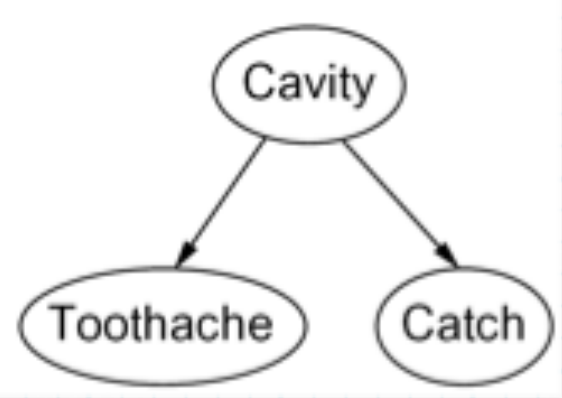
P(Cavity)
P(Toothache | Cavity)
P(Catch | Cavity)

$$P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i | parents(X_i))$$

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

Probabilities in BNs

Bayes Net:
Structure + CPTs



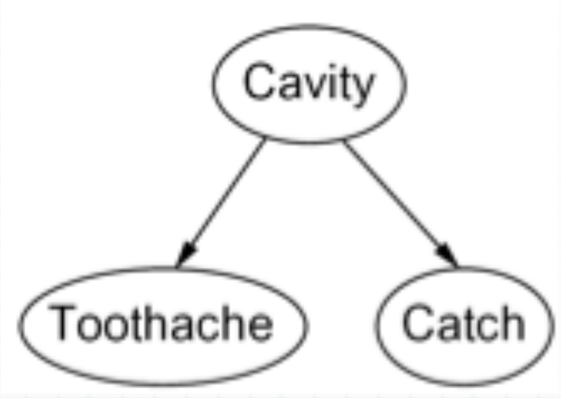
P(Cavity)
P(Toothache | Cavity)
P(Catch | Cavity)

$$P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i | parents(X_i))$$

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

Probabilities in BNs

Bayes Net:
Structure + CPTs



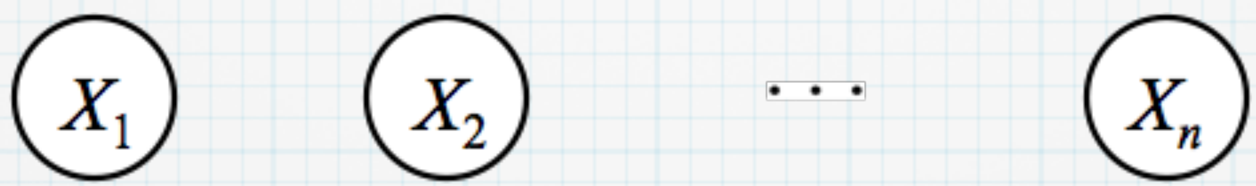
P(Cavity)
P(Toothache | Cavity)
P(Catch | Cavity)

$$P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i | parents(X_i))$$

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

Example: Coin Flips

* N independent flips



$P(X_1)$

h	0.5
t	0.5

$P(X_2)$

h	0.5
t	0.5

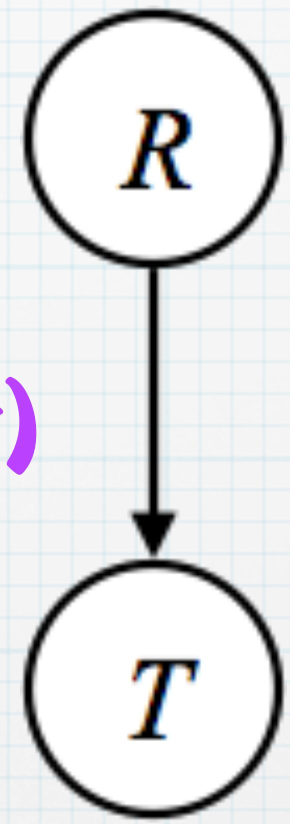
$P(X_n)$

h	0.5
t	0.5

* P(h,h,t,h) = .5 * .5 * .5 * .5

Example: Traffic

* $P(r, \neg t) = P(r) * P(\neg t \mid r)$
  $1/4 * \quad 1/4$



P(R)

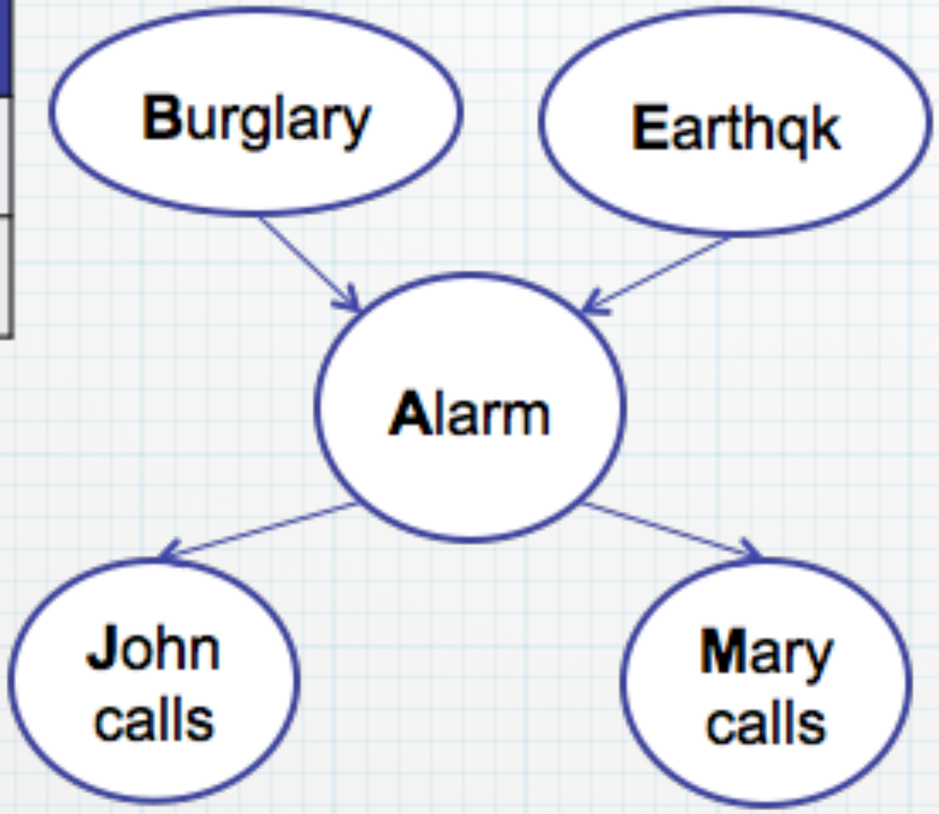
r	1/4
¬r	3/4

P(T | R)

r	t	3/4
	¬t	1/4
¬r	t	1/2
	¬t	1/2

Example CTPs: Alarm

B	P(B)
+b	0.001
¬b	0.999



E	P(E)
+e	0.002
¬e	0.998

B	E	A	P(A|B,E)
+b	+e	+a	0.95
+b	+e	¬a	0.05
+b	¬e	+a	0.94
+b	¬e	¬a	0.06
¬b	+e	+a	0.29
¬b	+e	¬a	0.71
¬b	¬e	+a	0.001
¬b	¬e	¬a	0.999

A	J	P(J|A)
+a	+j	0.9
+a	¬j	0.1
¬a	+j	0.05
¬a	¬j	0.95

A	M	P(M|A)
+a	+m	0.7
+a	¬m	0.3
¬a	+m	0.01
¬a	¬m	0.99

Building (Entire) Joint

* We can use the Bayes' Net to build any entry from full distribution it encodes

$$P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i | parents(X_i))$$

* Typically no reason to build everything, just calc what we need on the fly

* Every BN over a domain of variables **Implicitly Defines a Joint Distribution**

Size of a Bayes' Net

* Size of joint dist table of **N** boolean vars
  * 2^N
* Size of **N**-node net where each node has up to **k** parents
  * $O(N*2^{k+1})$
* BN can be huge savings if **k << N**
* Easier to find local CPTs vs global joints

Bayes' Nets So Far...

* What we know:
  * Syntax and Semantics of BNs
* Next: properties of the joint distribution
  * Formalizing the notion of conditional independence and causality
  * Goal: answer queries about conditional independence and influence
  * Need to calc posterior probabilities quickly!