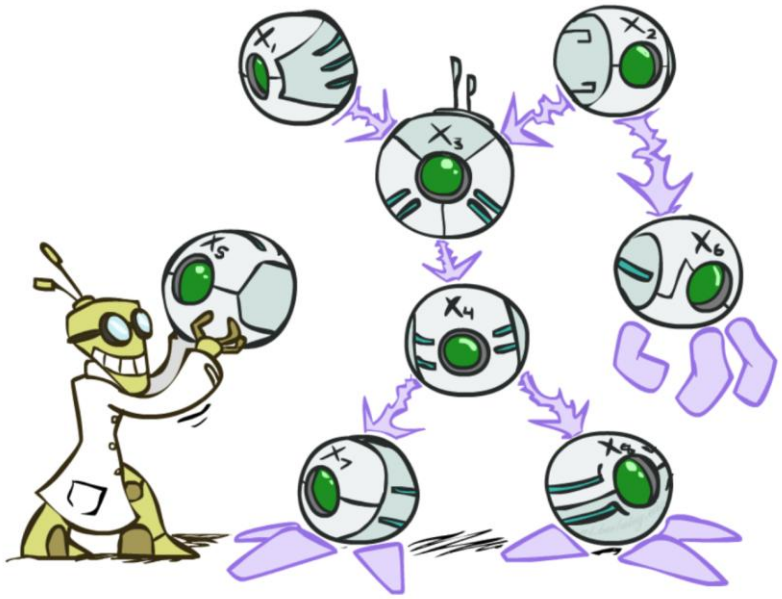
Bayes Nets

CSE 473: Introduction to Artificial Intelligence



Megan Hazen(mh75@cs)
cs.uw.edu/473
acknowledgements

Welcome!
Do the drawings enhance the slides?

Administrivia

ANNOUNCEMENTS

- **Test 1** scores released today

ASSIGNMENTS

- **Project 2** due **tonight**
- **Homework 3** posted, due Mon 11/16
- **Project 3** posted Wednesday



Probability Recap

CONTEXT

- Laws of Probability
  - 'Worlds' ω in universe Ω
  - $0 \leq P(\omega) \leq 1$
  - $\sum_{\omega \in \Omega} P(\omega) = 1$
  - $P(A) = \sum_{\omega \in A} P(\omega)$
- Random Variable X
  - $X(\omega)$ maps outcomes ω to range $x \epsilon X$
  - $P(X = x)$ is a deterministic function from $x \in X$ to $[0,1]$
- Distributions
  - **Joint Distribution** $P(X, Y)$ gives total probability for each x, y combination
  - Marginal Distribution $P(X) = \sum_{y \in Y} P(X, Y = y)$
  - Conditional Distribution $P(X|Y) = \frac{P(X,Y)}{P(Y)}$
- Product Rule
  - $P(X, Y) = P(X|Y) \cdot P(Y)$
  - Chain Rule: $P(X_1, \ldots, X_n) = \prod_i P(X_i | X_1, \ldots, X_{i-1})$
- Bayes Rule
  - $P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$

Inferring the Weather 1

EXAMPLE

- $P(W)$?
  - $P(W) = \sum_{s \in S, t \in T} P(S = s, T = t, W)$
  - $P(W = \text{sun}) = \sum_{s \in S, t \in T} P(S = s, T = t, W = \text{sun})$
  - $P(W = \text{sun}) = 0.30 + 0.08 + 0.10 + 0.15 = .63$

$W = w$	$P(w)$
sun	0.63
rain	0.17
fog	0.20
meteor	0.0

S	T	W	$P(S,T,W)$
summer	hot	sun	0.30
summer	hot	rain	0.01
summer	hot	fog	0.01
summer	hot	meteor	0.00
summer	cold	sun	0.08
summer	cold	rain	0.05
summer	cold	fog	0.05
summer	cold	meteor	0.00
winter	hot	sun	0.10
winter	hot	rain	0.01
winter	hot	fog	0.02
winter	hot	meteor	0.00
winter	cold	sun	0.15
winter	cold	rain	0.10
winter	cold	fog	0.12
winter	cold	meteor	0.00

Inferring the Weather 2

EXAMPLE

- $P(W|\text{summer})$?
  - $P(S = \text{summer}, W) = \sum_{t \in T} P(S = \text{summer}, T = t, W)$
  - $\alpha = \frac{1}{\sum_{w \in W} P(S=\text{summer}, W=w)}$
  - $P(W|\text{summer}) = \alpha \cdot P(S = \text{summer}, W)$

- $P(W = \text{sun}|\text{summer}) = \frac{.30+.08}{.30+.01+.01+.08+.05+.05}$
  - $= \frac{.38}{.5} = .76$

$W = w$	$P(w)$
sun	0.76
rain	0.12
fog	0.12
meteor	0.0

S	T	W	$P(S,T,W)$
summer	hot	sun	0.30
summer	hot	rain	0.01
summer	hot	fog	0.01
summer	hot	meteor	0.00
summer	cold	sun	0.08
summer	cold	rain	0.05
summer	cold	fog	0.05
summer	cold	meteor	0.00
winter	hot	sun	0.10
winter	hot	rain	0.01
winter	hot	fog	0.02
winter	hot	meteor	0.00
winter	cold	sun	0.15
winter	cold	rain	0.10
winter	cold	fog	0.12
winter	cold	meteor	0.00

Inferring the Weather 3

EXAMPLE

- $P(W|\text{summer, hot})$?
  - $P(S = \text{summer}, T = \text{hot}, W)$
  - $\alpha = \frac{1}{\sum_{w \in W} P(S=\text{summer}, T=\text{hot}, W=w)}$
  - $P(W|\text{summer, hot}) = \alpha \cdot P(S = \text{summer}, T = \text{hot}, W)$
  - $\text{P(W = sun|summer, hot)} = \frac{.30}{.30+.01+.01} = \frac{.30}{.32} = 0.94$

$W = w$	$P(w)$
sun	0.94
rain	0.03
fog	0.03
meteor	0.0

S	T	W	$P(S,T,W)$
summer	hot	sun	0.30
summer	hot	rain	0.01
summer	hot	fog	0.01
summer	hot	meteor	0.00
summer	cold	sun	0.08
summer	cold	rain	0.05
summer	cold	fog	0.05
summer	cold	meteor	0.00
winter	hot	sun	0.10
winter	hot	rain	0.01
winter	hot	fog	0.02
winter	hot	meteor	0.00
winter	cold	sun	0.15
winter	cold	rain	0.10
winter	cold	fog	0.12
winter	cold	meteor	0.00

Problems with Inference by Enumeration

CONTEXT

Inference by Enumeration calculates probabilities from a full-join-probability distribution

But Inference by Enumeration requires *a lot* of enumeration…

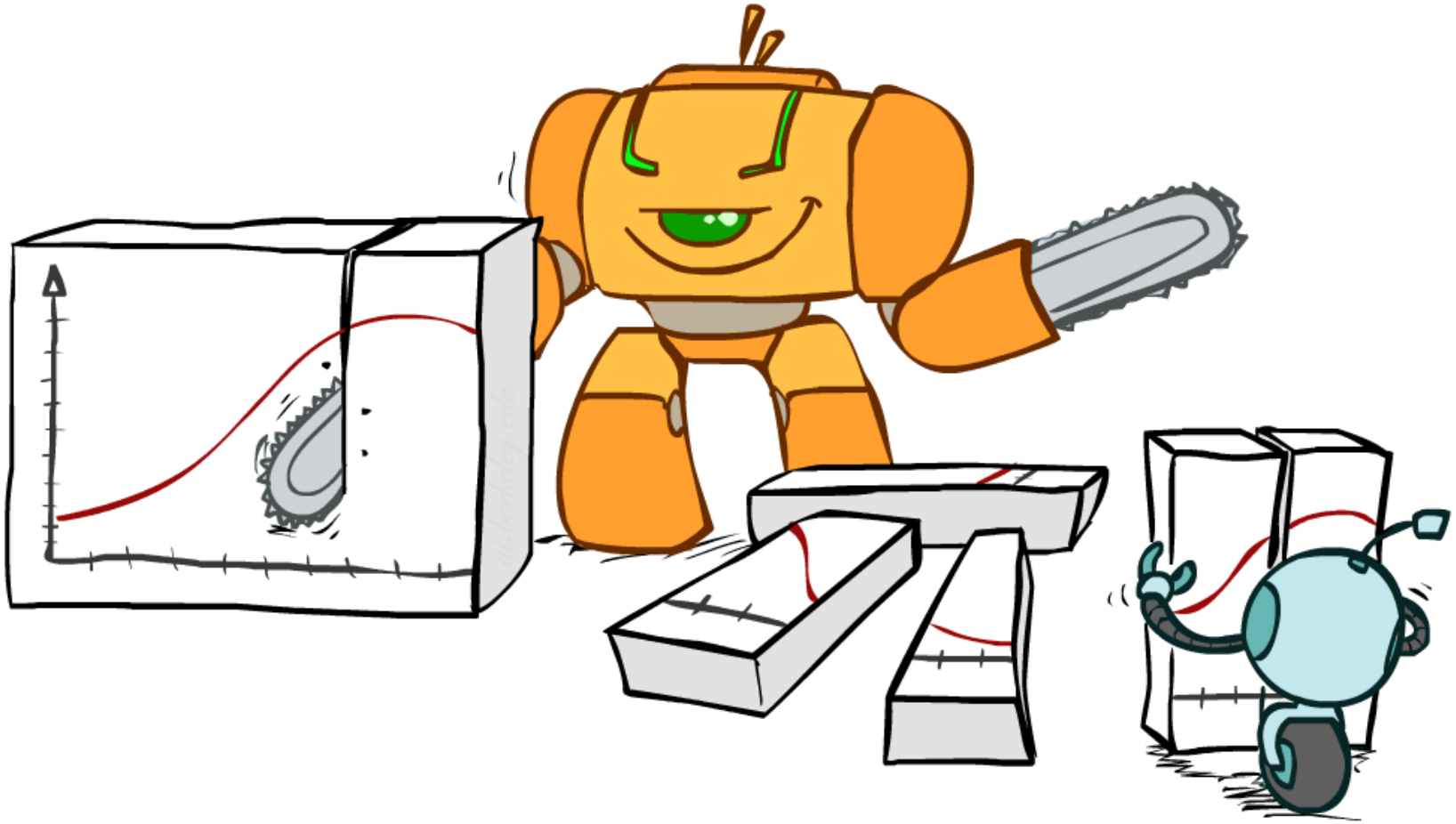
- Worst-Case Time Complexity: $O(d^n)$
- Space Complexity: $O(d^n)$ to store joint distribution
- Need $O(d^n)$ data points to estimate the entries in the joint distribution

Size of distribution for n variables with range d:
$O(d^n)$

Questions?

A Better Approach?

CONTEXT



Bayes' Rule

DEFINITION

Formulate conditional probability another way…

- From the product rule:
  - $P(A, B) = P(A|B) \cdot P(B)$
  - $P(A, B) = P(B|A) \cdot P(A)$
- Combine and rearrange:

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

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

Implications of Bayes' Rule

DEFINITION

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

- How does this help?
  - Calculate a condition from its reverse, which may be easy to calculate
  - Describes update to **posterior** $P(A|B)$ given **prior** $P(A)$ and **likelihood** $P(B|A)$

One of the most important equations in all of artificial intelligence!

Inference with Bayes' Rule

EXAMPLE

- Diagnostic probability from causal probability

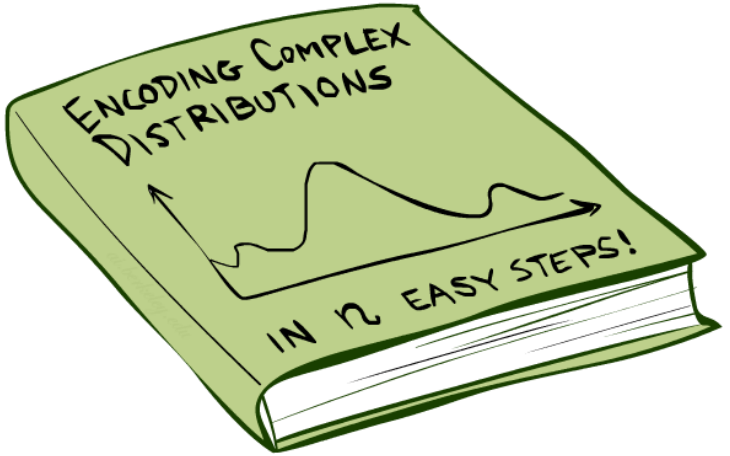
$$P(\text{cause}|\text{effect}) = \frac{P(\text{effect}|\text{cause}) \cdot P(\text{cause})}{P(\text{effect})}$$

- For example:
  - M is meningitis, S is stiff neck
  - Given $P(m) = 0.0001, P(s) = 0.01, P(s|m) = 0.8$

$$P(m|s) = \frac{P(s|m) \cdot P(m)}{P(s)} = \frac{0.8 \cdot 0.0001}{0.01} = 0.008$$

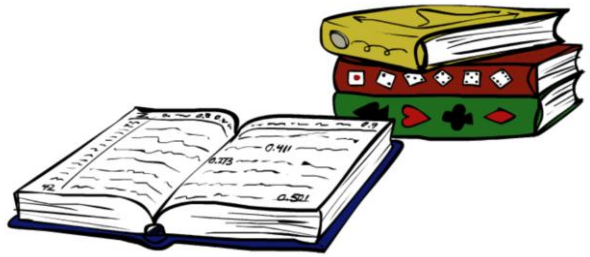
  - Posterior probability is still small, but 80x higher given evidence of stiff neck

How to manage complex probability distributions?



Bayesian Networks

FOUNDATIONS

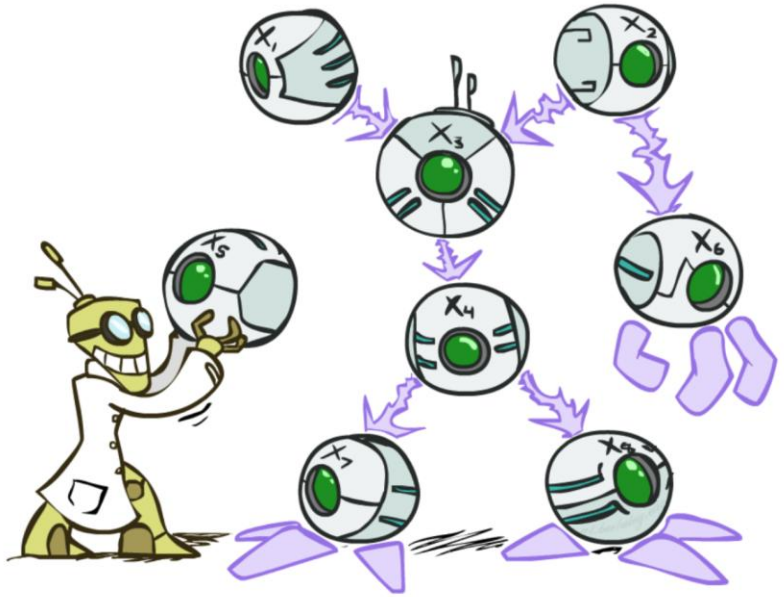


Bayes Nets provide a technique describing complex joint distributions using simple conditional distributions

- Subset of a general class of graphical models
- Also called “belief networks”

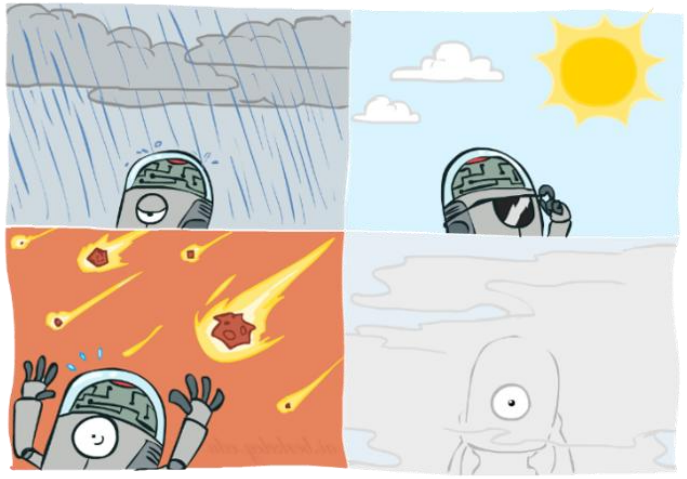
- Use local causality / conditional independence
  - The world is composed of many variables
  - Each interacting locally with a few others

- Expressivity vs. Efficiency
  - While conditional probabilities are less expressive (comprehensive) than joint probabilities, they are faster and more space-efficient to work with

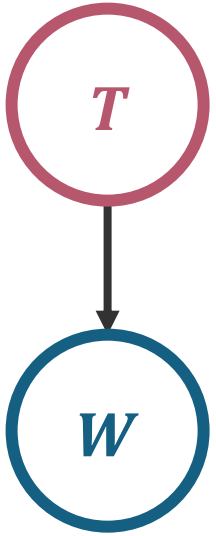


Graphical Models

DEFINITION



- Nodes: Variables
  - Assigned (observed)
  - Unassigned (unobserved)
- Arcs: Interactions
  - Indicate “direct influence” between variables
  - Formally: encode conditional (in)dependence



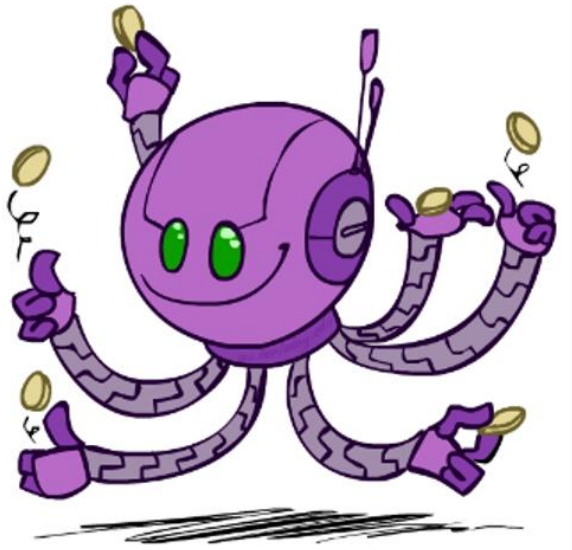
Coin Flips

EXAMPLE

N independent coin flips

- No interactions between variables $\Longrightarrow$ absolute independence





Traffic

EXAMPLE

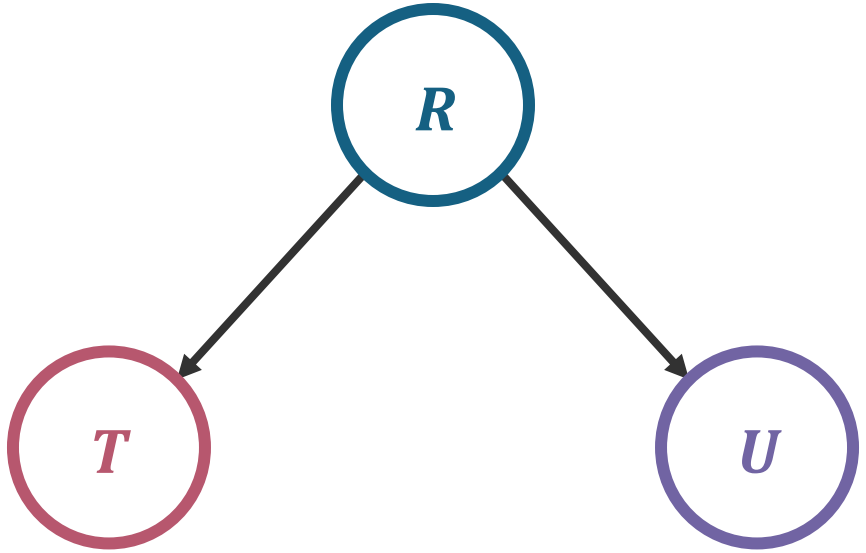


Variables

- R it rains
- T there's traffic
- U I'm holding my umbrella







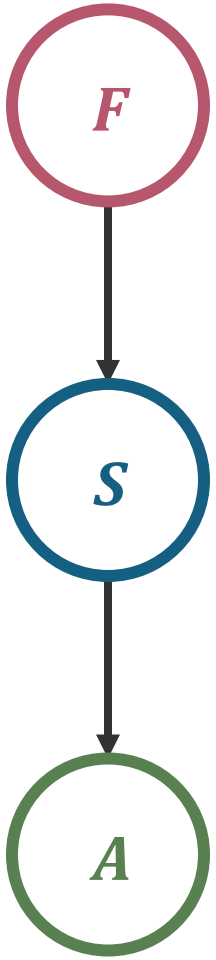


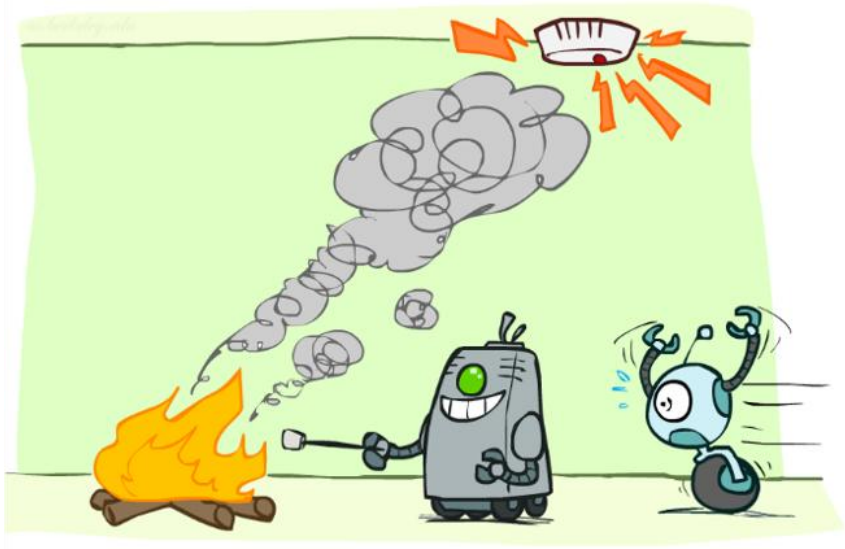
Smoke Alarm

EXAMPLE

Variables

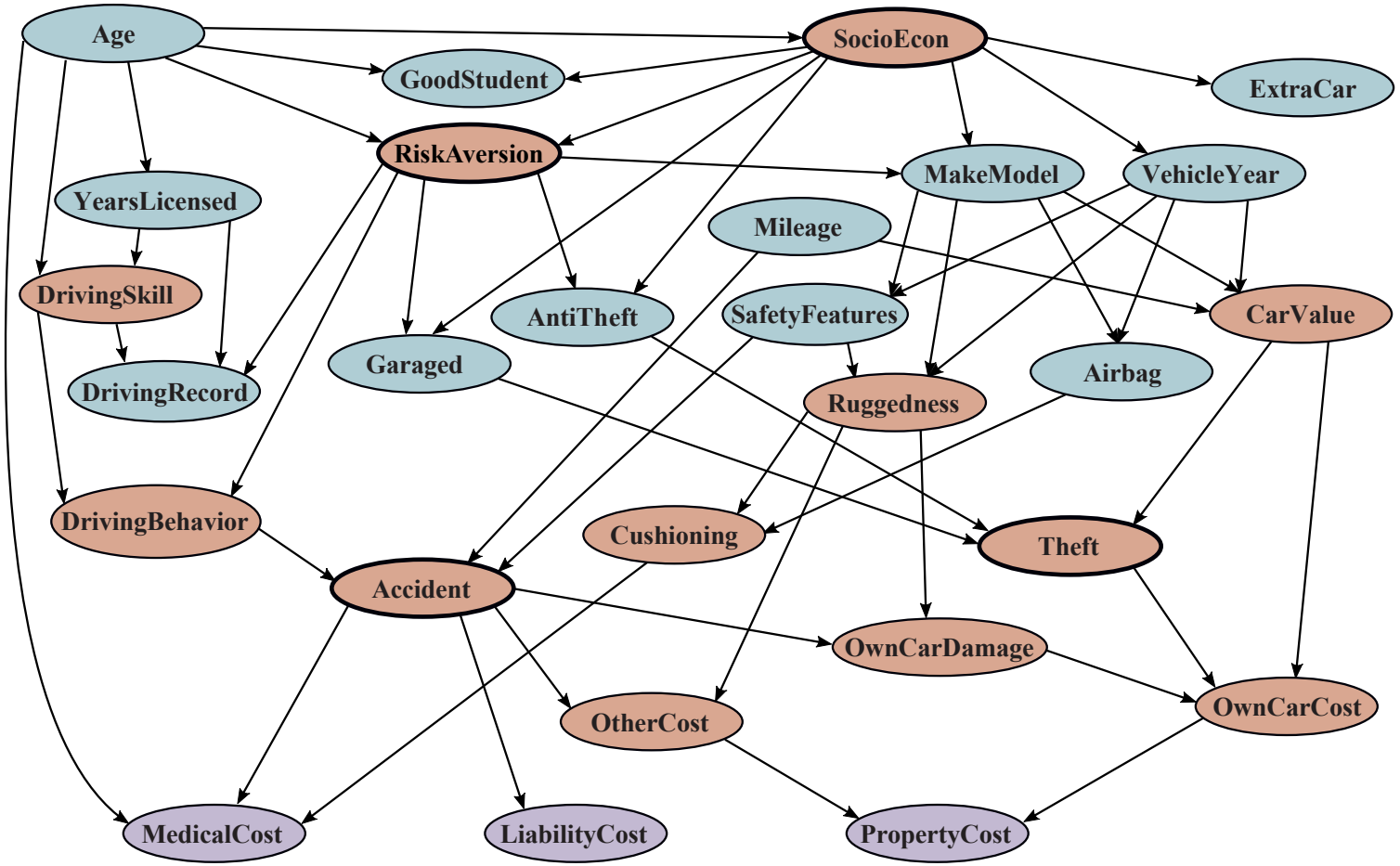
- F there is a fire
- S there is smoke
- A alarm sounds





Car Insurance

EXAMPLE



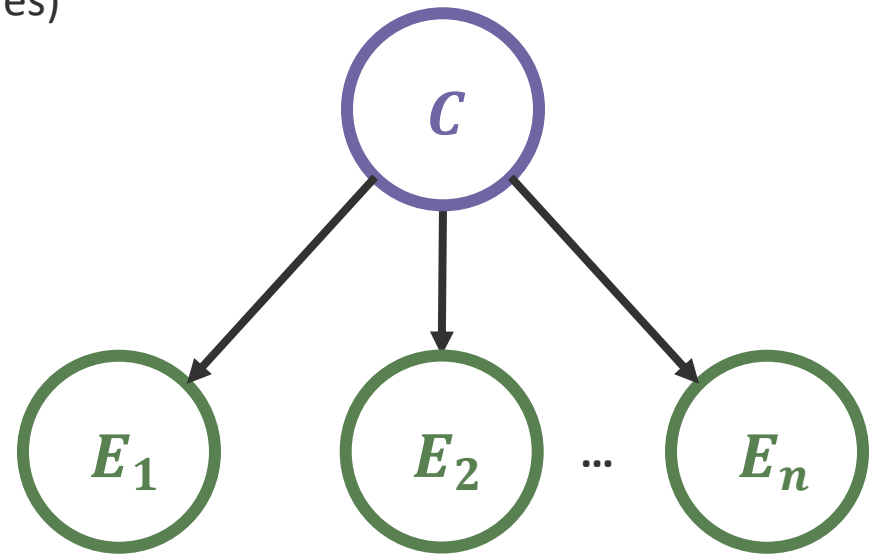
Naïve Bayes

EXAMPLE

Variables

- C class we want to predict
- $E_1, \ldots, E_n$ evidence (or features)

"**Naïve**" because of the simplifying assumption that evidence variables are **conditionally independent** given C



Questions?

Syntax and Semantics

FOUNDATIONS

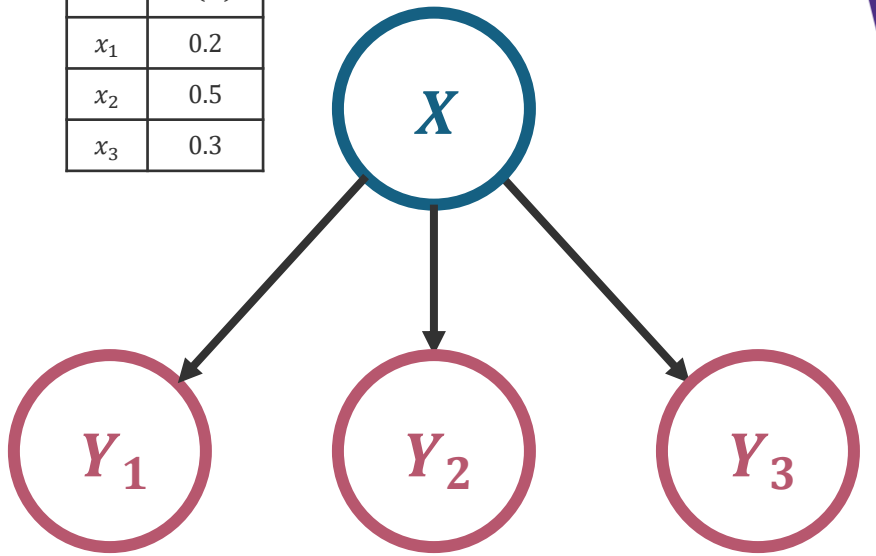


Bayes Net Syntax

DEFINITION

- Set of nodes, one per variable X_i
- Directed, acyclic graph
- Conditional distribution for each node given parent
  - Conditional Probability Table (CPT)

X	$P(X)$
x_1	0.2
x_2	0.5
x_3	0.3



	$P(Y_1 \mid X)$		
X	$Y_1 = a$	$Y_1 = b$	$Y_1 = c$
x_1	0.2	0.6	0.2
x_2	0.1	0.0	0.9
x_3	0.4	0.3	0.3

	$P(Y_3 \mid X)$		
X	$Y_3 = d$	$Y_3 = e$	$Y_3 = f$
x_1	0.0	0.1	0.9
x_2	0.25	0.25	0.5
x_3	0.8	0.1	0.1

Bayes Net = Graph + Local Probability Tables

Bayes Net Semantics

DEFINITION

Bayes nets encode joint distributions as products of conditional distributions on each variable

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | \text{parents}(X_i))$$

- Compare with chain rule: $P(X_1, \ldots, X_n) = \prod_i P(X_i | X_1, \ldots, X_{i-1})$
  - Assume $X_1, \ldots, X_n$ sorted in topological order, so $\text{parents}(X_i) \subseteq X_1, \ldots, X_{i-1}$
  - Therefore, bayes net asserts the conditional independence:

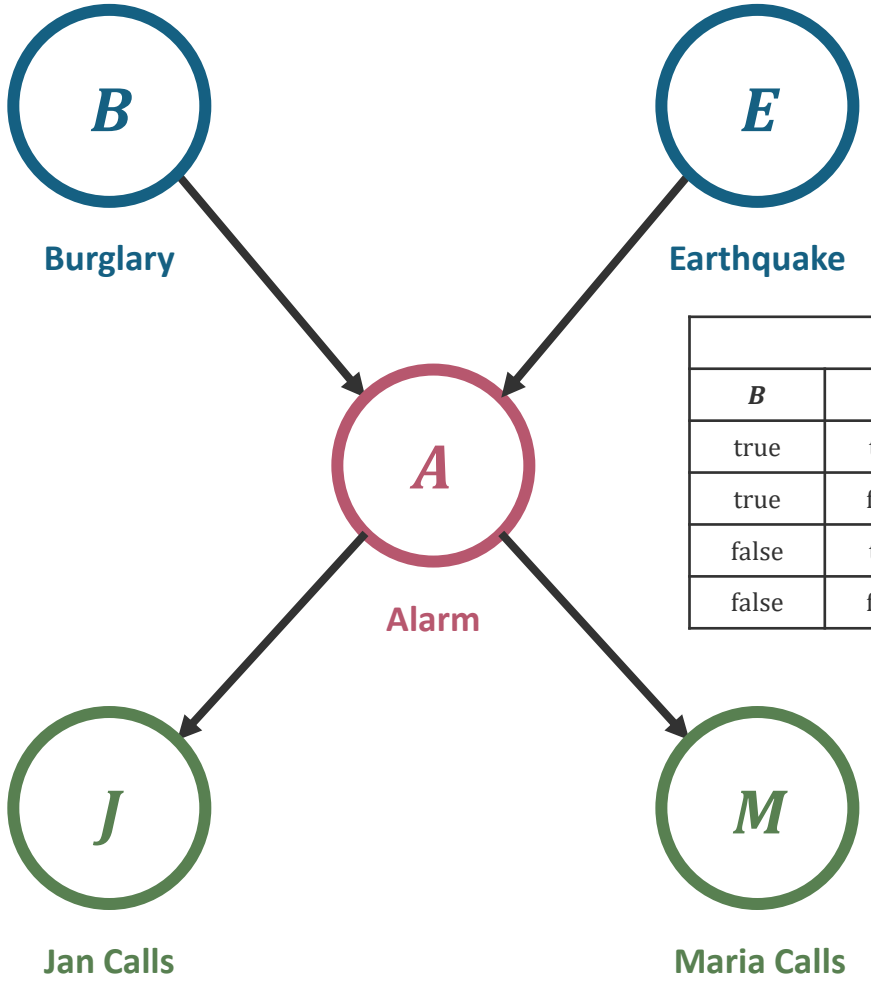
$$P(X_i | X_1, \ldots, X_{i-1}) = P(X_i | \text{parents}(X_i))$$

Conditional Independence:
$\boldsymbol{P(x|y,z) = P(x|z)}$

Alarm Bayes Net

EXAMPLE

B	$P(B)$
true	0.001
false	0.999

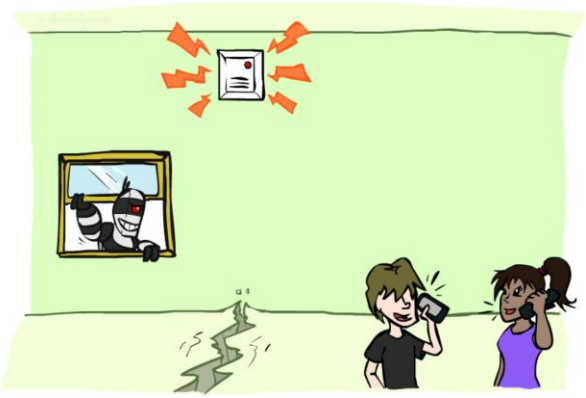


E	$P(E)$
true	0.002
false	0.998

		$P(A\|B,E)$	
B	E	true	false
true	true	0.95	0.05
true	false	0.94	0.06
false	true	0.29	0.71
false	false	0.001	0.999

	$P(J\|A)$	
A	true	false
true	0.9	0.1
false	0.05	0.95

	$P(M\|A)$	
A	true	false
true	0.7	0.3
false	0.01	0.99



Factor Size per CPT:

$$= d_c \prod_{i \in \text{parents}(X_c)} d_i$$

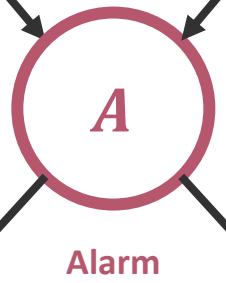
Alarm Bayes Net Practice

EXAMPLE

B	$P(B)$
true	0.001
false	0.999

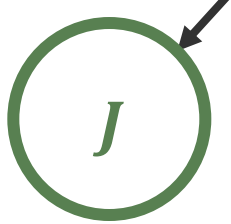


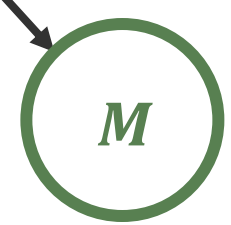
E	$P(E)$
true	0.002
false	0.998



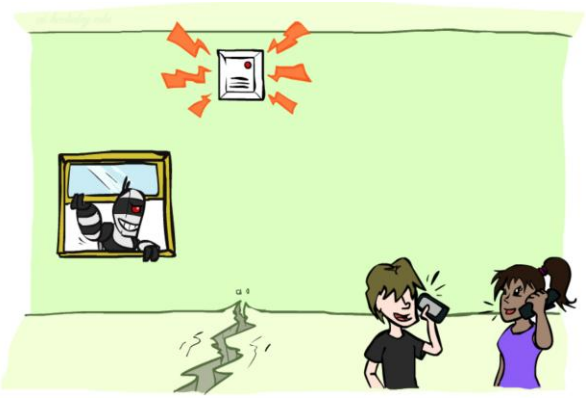
		$P(A\|B,E)$	
B	E	true	false
true	true	0.95	0.05
true	false	0.94	0.06
false	true	0.29	0.71
false	false	0.001	0.999

	$P(J\|A)$	
A	true	false
true	0.9	0.1
false	0.05	0.95





	$P(M\|A)$	
A	true	false
true	0.7	0.3
false	0.01	0.99



Calculate $P(b, \neg e, a, \neg j, \neg m)$

Product Rule:

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i|\text{parents}(X_i))$$

$$P(b, \neg e, a, \neg j, \neg m)$$
$$= P(b)P(\neg e)P(a|b, \neg e)P(\neg j|a)P(\neg m|a)$$

Alarm Bayes Net Practice (1)

EXAMPLE

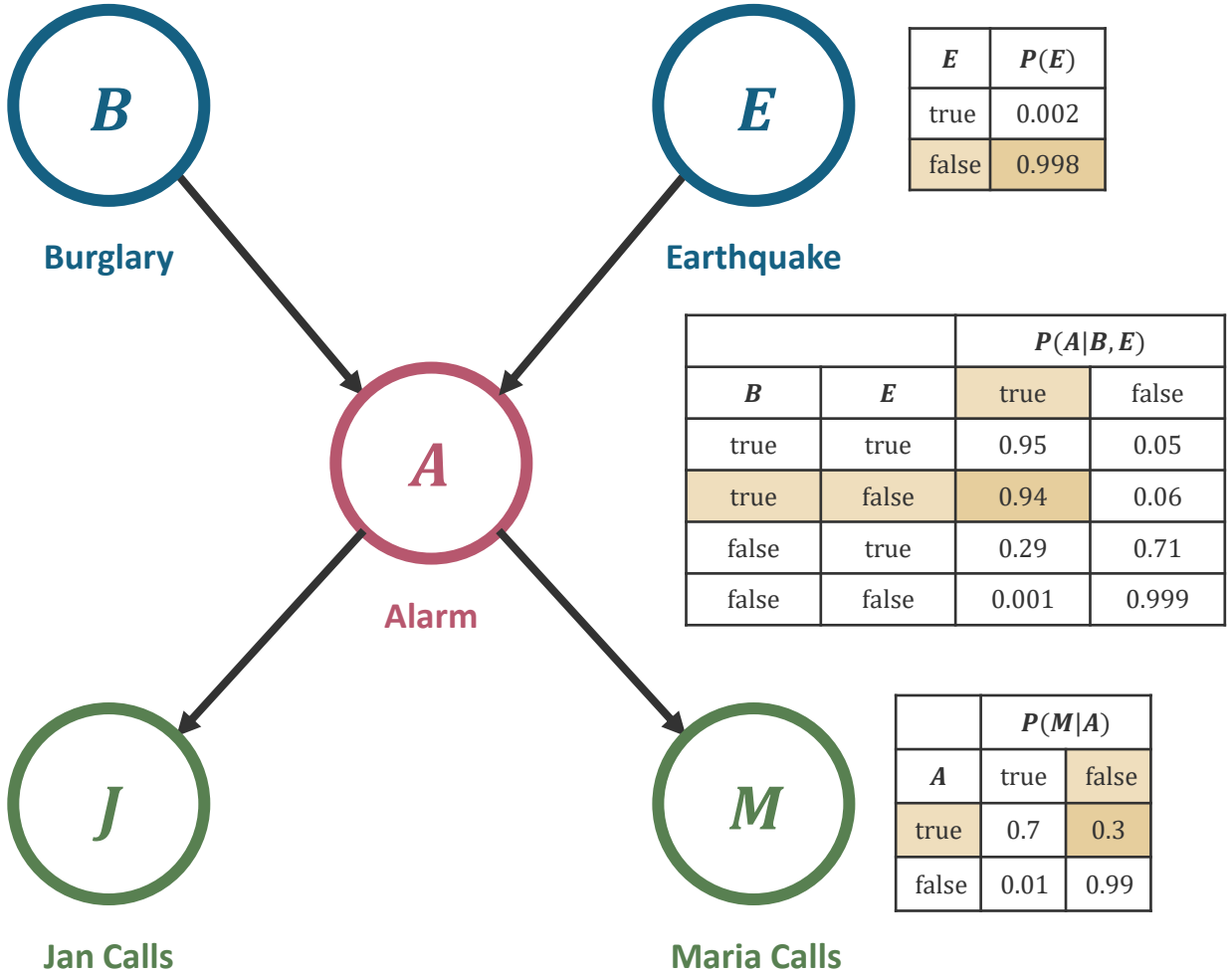
B	$P(B)$
true	0.001
false	0.999

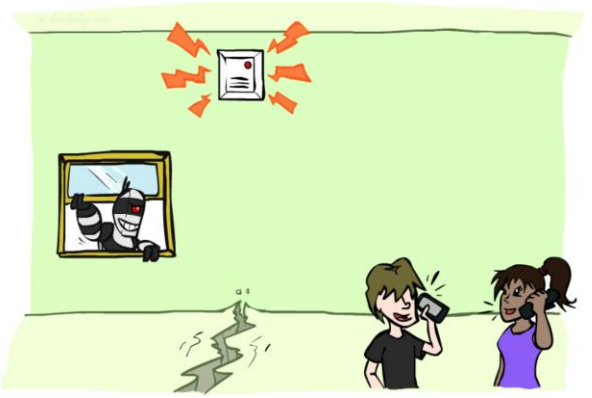
E	$P(E)$
true	0.002
false	0.998

		$P(A\|B,E)$	
B	E	true	false
true	true	0.95	0.05
true	false	0.94	0.06
false	true	0.29	0.71
false	false	0.001	0.999

	$P(J\|A)$	
A	true	false
true	0.9	0.1
false	0.05	0.95

	$P(M\|A)$	
A	true	false
true	0.7	0.3
false	0.01	0.99





Calculate $P(b, \neg e, a, \neg j, \neg m)$

$P(b, \neg e, a, \neg j, \neg m)$
$= P(b)P(\neg e)P(a|b, \neg e)P(\neg j|a)P(\neg m|a)$

Alarm Bayes Net Practice (2)

EXAMPLE

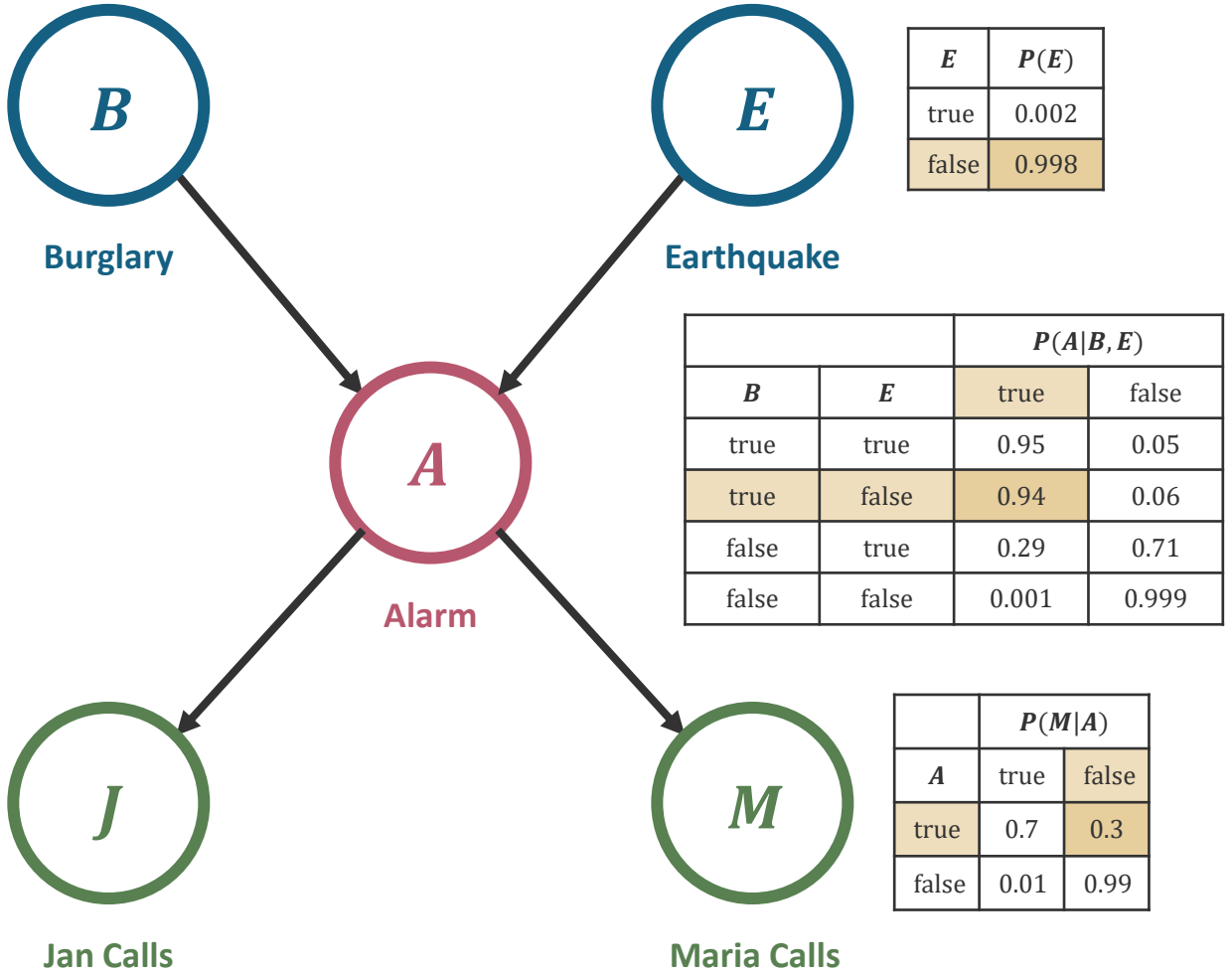
B	$P(B)$
true	0.001
false	0.999

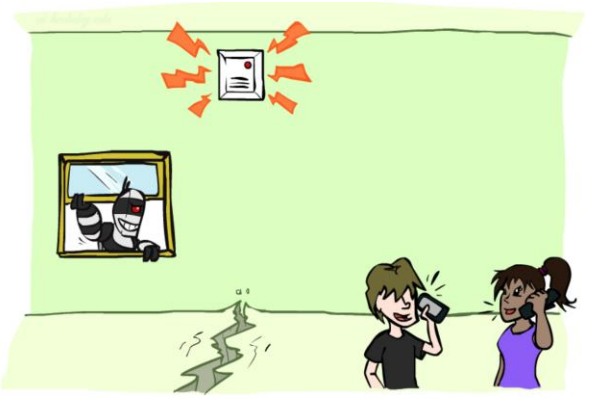
E	$P(E)$
true	0.002
false	0.998

		$P(A\|B,E)$	
B	E	true	false
true	true	0.95	0.05
true	false	0.94	0.06
false	true	0.29	0.71
false	false	0.001	0.999

	$P(J\|A)$	
A	true	false
true	0.9	0.1
false	0.05	0.95

	$P(M\|A)$	
A	true	false
true	0.7	0.3
false	0.01	0.99





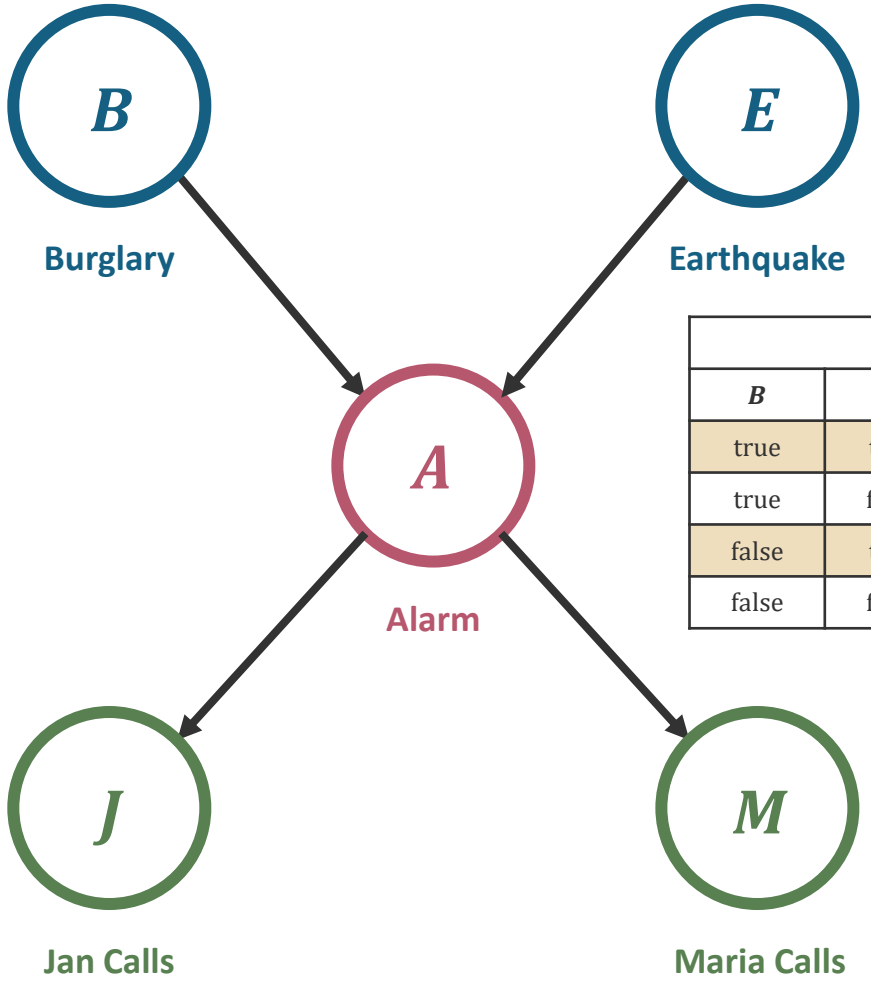
Calculate $P(b, \neg e, a, \neg j, \neg m)$

$$= P(b) \cdot P(\neg e) \cdots$$
$$\cdot P(a|b, \neg e) \cdots$$
$$\cdot P(\neg j|a) \cdot P(\neg m|a) \cdots$$
$$= 0.001 \cdot 0.998 \cdot 0.94 \cdot 0.1 \cdot 0.3 = 0.000028$$

Alarm Bayes Net Practice (3)

EXAMPLE

B	$P(B)$
true	0.001
false	0.999

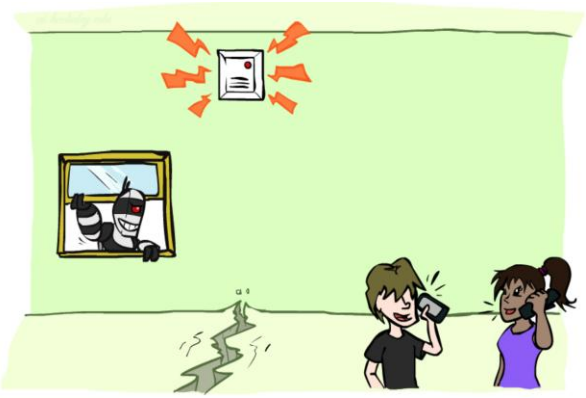


E	$P(E)$
true	0.002
false	0.998

		$P(A\|B,E)$	
B	E	true	false
true	true	0.95	0.05
true	false	0.94	0.06
false	true	0.29	0.71
false	false	0.001	0.999

	$P(J\|A)$	
A	true	false
true	0.9	0.1
false	0.05	0.95

	$P(M\|A)$	
A	true	false
true	0.7	0.3
false	0.01	0.99



What about $P(B, e, a, j, m)$?

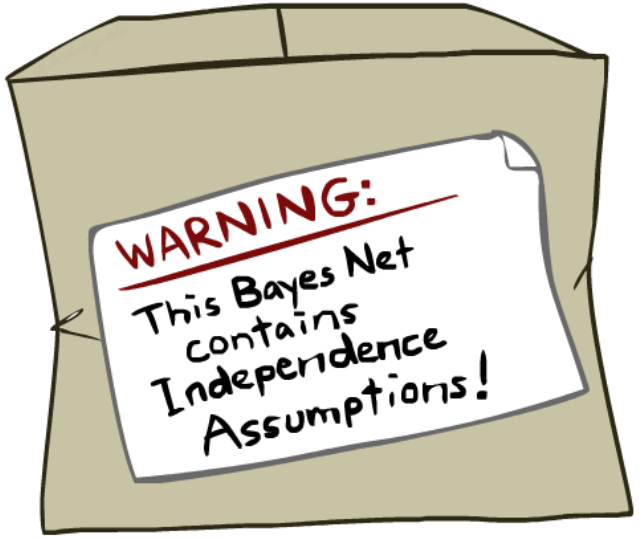
B	$P(B)$
true	$= .001 \cdot .002 \cdot 0.95 \cdot 0.9 \cdot 0.7 = 0.000001$
false	$= .999 \cdot .002 \cdot 0.29 \cdot 0.9 \cdot 0.7 = 0.00036$

Questions?

Sparse Bayes Nets

DEFINITION

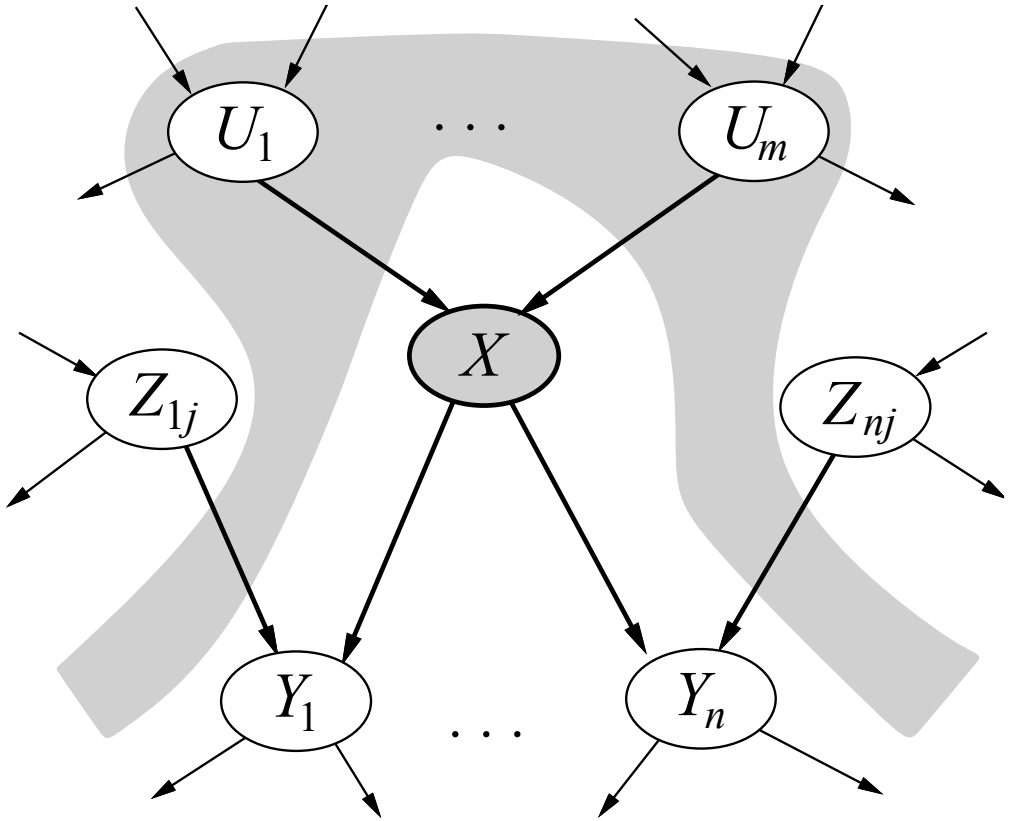
- Suppose
  - n variables
  - Maximum range size is d
  - Maximum number of parents is k
- Full Joint Distribution: $O(d^n)$
- Bayes Net: $O(n \cdot d^k)$
  - Linear scaling with n as long as causal structure is local
- Often $O(n \cdot d^k) \ll O(d^n)$

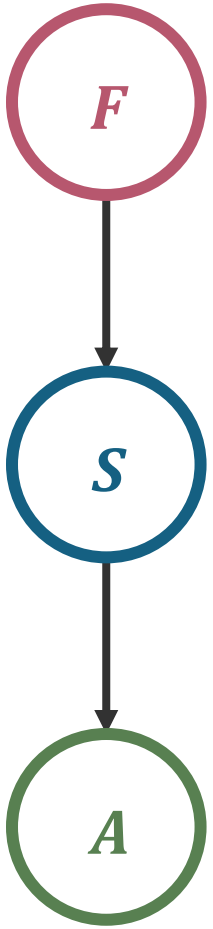


Conditional Independence

FOUNDATIONS

Every variable is conditionally independent of its non-descendants, given its parents

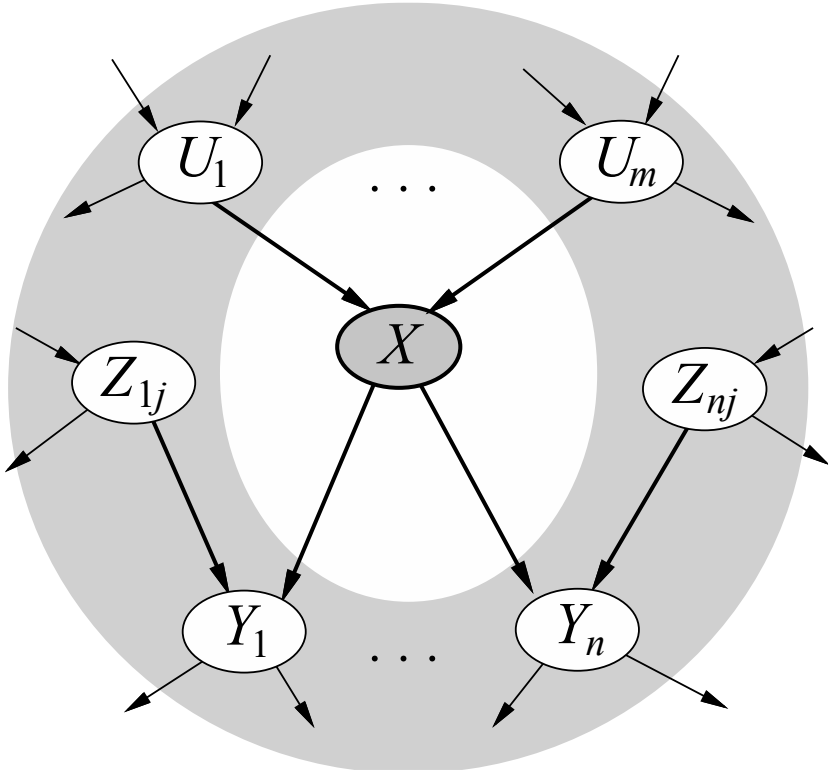




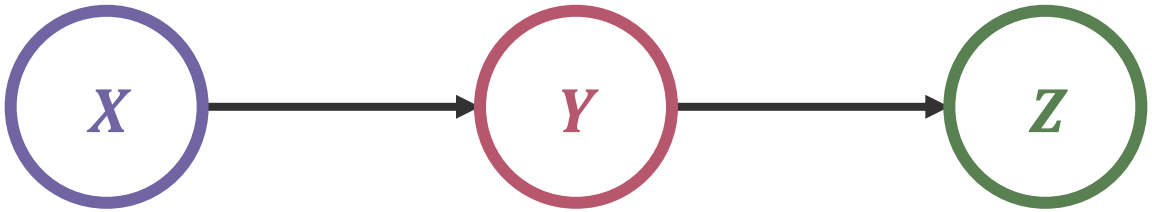
Markov Blanket

FOUNDATIONS

- Compare with Markov Assumption
  - "The future is independent of the past, given the present"
- Similar Formulation
  - **Markov blanket** consists of a node's parents, children, and children's other parents
  - Every variable is conditionally independent of all other variables, given its Markov blanket



Questions?



$X \perp Z \mid Y$

(not absolute independence)

X independent of Z?

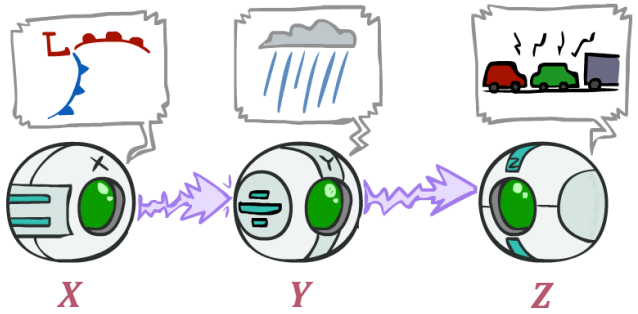
absolute independence?



Node Triples

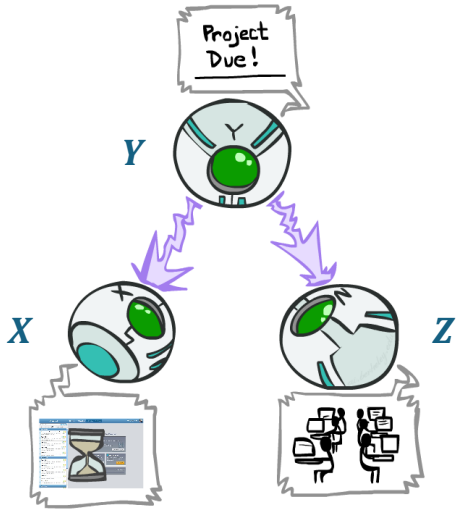
CONTEXT

How can three nodes be arranged?



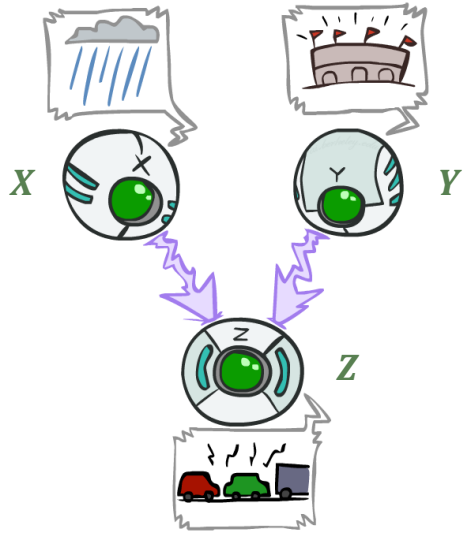
Causal Chain

$$P(X,Y,Z) = P(X) \cdot P(Y|X) \cdot P(Z|Y)$$



Common Cause

$$P(X,Y,Z) = P(Y) \cdot P(X|Y) \cdot P(Z|Y)$$



Common Effect

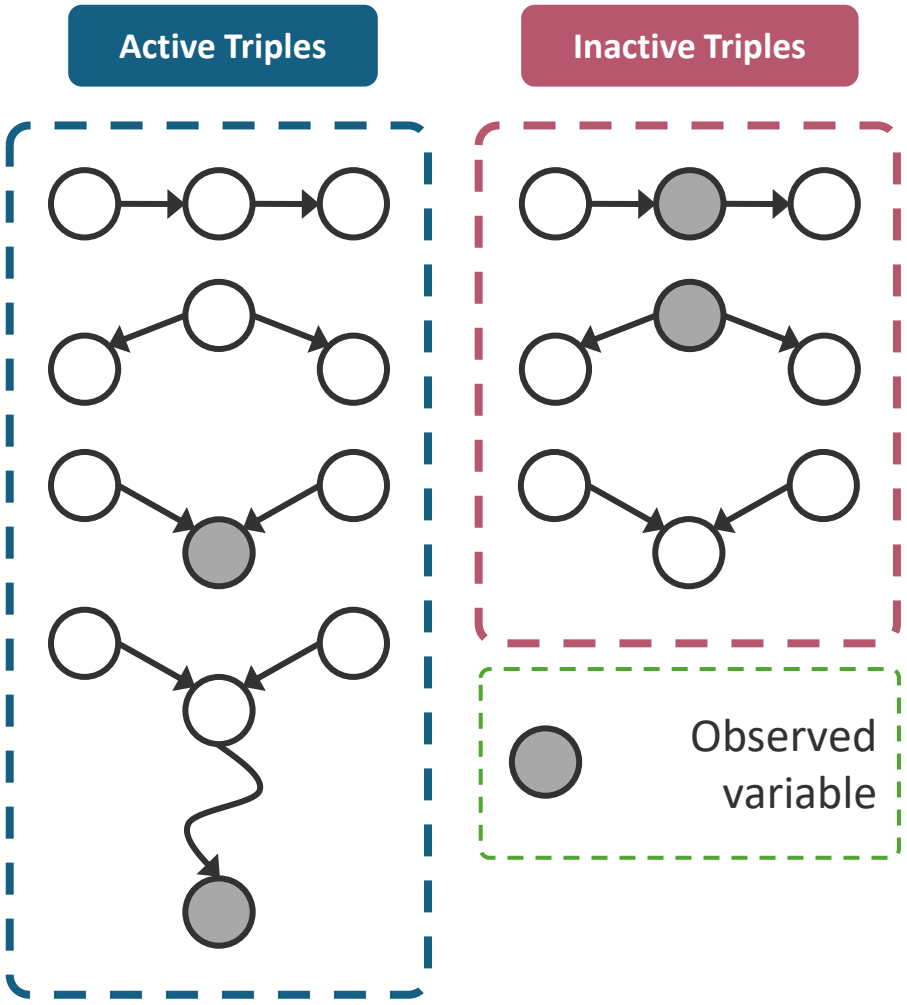
$$P(X,Y,Z) = P(X) \cdot P(Y) \cdot P(Z|X,Y)$$

D-Separation

DEFINITION

D-separation is an algorithm for checking (conditional) independence of two nodes

- **Query**: $X_i \perp X_j \mid \{X_{k_1}, \ldots, X_{k_n}\}$?
- **Test**:
  - Check all *undirected* paths between X_i and X_j:
    - If any paths are **active**: $X_i \perp\!\!\!\not\perp X_j \mid \{X_{k_1}, \ldots, X_{k_n}\}$
    - If all paths are **inactive**: $X_i \perp X_j \mid \{X_{k_1}, \ldots, X_{k_n}\}$
- **Why?**:
  - Read dependence information off a knowledge graph
  - Determine which variables must be observed to determine a value



Questions?

that's it for today!

SUMMARY

- Implications of independence and conditional independence
- Bayes nets encode joint distributions by tracking conditional probabilities

UPCOMING

- How to compute conditional probabilities of queries given evidence?

REMINDERS

- **Homework 3** posted
- **Project 2** **due tonight**