CSPB3202 Artificial Intelligence

Reasoning with Uncertainty

University of Colorado Boulder

Note: Slides were adapted from the materials from Prof. Pieter Abbeel and Prof. Dan Klein at UC Berkeley, and Russell & Norvig's AI book.

Bayes Net: Representation

Review: Probability Recap

- Conditional probability $$P(x|y) = \frac{P(x,y)}{P(y)}$$
- Product rule $$P(x,y) = P(x|y)P(y)$$
- Bayes rule $$P(x|y) = \frac{P(y|x)}{P(y)}P(x)$$
- X, Y independent if and only if:

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

Review: Conditional Independence

- X is conditionally independent of Y given Z $\quad X \perp\!\!\!\perp Y | Z$

  if and only if:

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

  or, equivalently, if and only if

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

Review: The Chain Rule

- More generally, can always write any joint distribution as an incremental product of conditional distributions

$$P(x_1, x_2, x_3) = P(x_1)P(x_{\cancel{2}3}|x_1)P(x_{\cancel{3}2}|x_1, x_{\cancel{2}3})$$

$$P(x_1, x_2, \ldots x_n) = \prod_i P(x_i|x_1 \ldots x_{i-1})$$

Review: Inference by Enumeration

- P(W)?

Q → weather < sun, rain

E → None

H → S, T

0.3 + 0.1 + 0.1 + 0.15 = 0.65

= 1 − 0.65 = 0.35

- P(W | winter)?

s 0.5

r 0.5

$\frac{0.25}{0.5} = 0.5$

- P(W | winter, hot)?

s | w, h = $\frac{0.1}{0.15} = 2/3$

r | w, h = 1 − 2/3 = $\frac{0.05}{0.15}$ = 1/3

S	T	W	P
summer	hot	sun	0.30
summer	hot	rain	0.05
summer	cold	sun	0.10
summer	cold	rain	0.05
winter	hot	sun	0.10
winter	hot	rain	0.05
winter	cold	sun	0.15
winter	cold	rain	0.20

Z

Review: Inference using Bayes' Rule

P(A|B) ↔ P(B|A)

- Given:

$P(W)$

R	P
sun	0.8
rain	0.2

$P(D|W)$

D	W	P
wet	sun	0.1
dry	sun	0.9
wet	rain	0.7
dry	rain	0.3

×0.8

×0.2

- What is P(W | dry) ?

$$P(W=s \mid dry) = \frac{\overset{0.9}{P(dry|sun)} \cdot \overset{0.8}{P(sun)}}{P(dry) = 0.72 + 0.06}$$

$$P(W=r \mid dry) = \frac{\overset{0.3}{P(dry|rain)} \cdot \overset{0.2}{P(rain)}}{P(dry) = 0.78}$$

Review: Conditional Independence

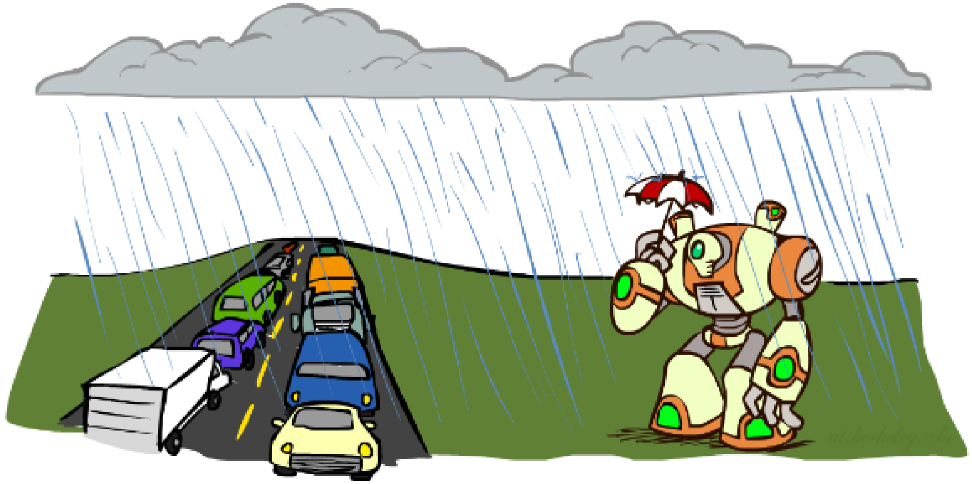
- Traffic
- Umbrella
- Raining

$R \to T$ ✓

$R \to U$ ✓

$\boxed{T \perp\!\!\!\perp U \mid R}$

$R \to T$, $R \to U$

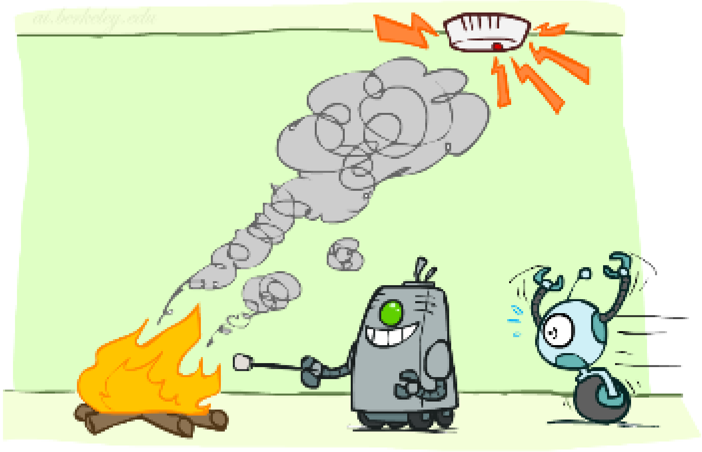


$T \leftarrow R \rightarrow U$

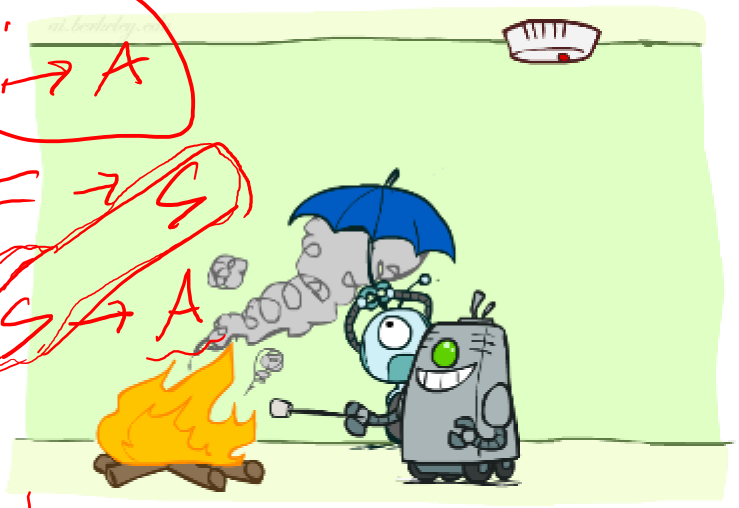
$F \to S \to A$

2)

- Fire
- Smoke
- Alarm



$F \to S \to A$



$F \to S \to A$

$F \perp\!\!\!\perp A \mid S$

Conditional Independence and the Chain Rule

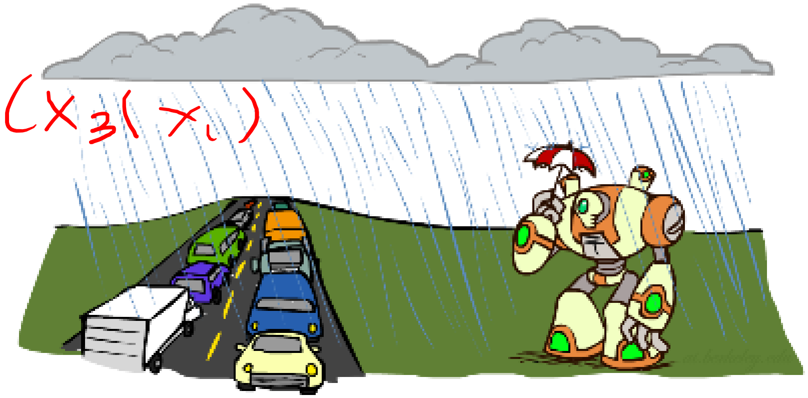
- Chain rule: $P(X_1, X_2, \ldots X_n) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)\ldots$

- Trivial decomposition:

X_1, X_2

$P(X_3 | X_1)$

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



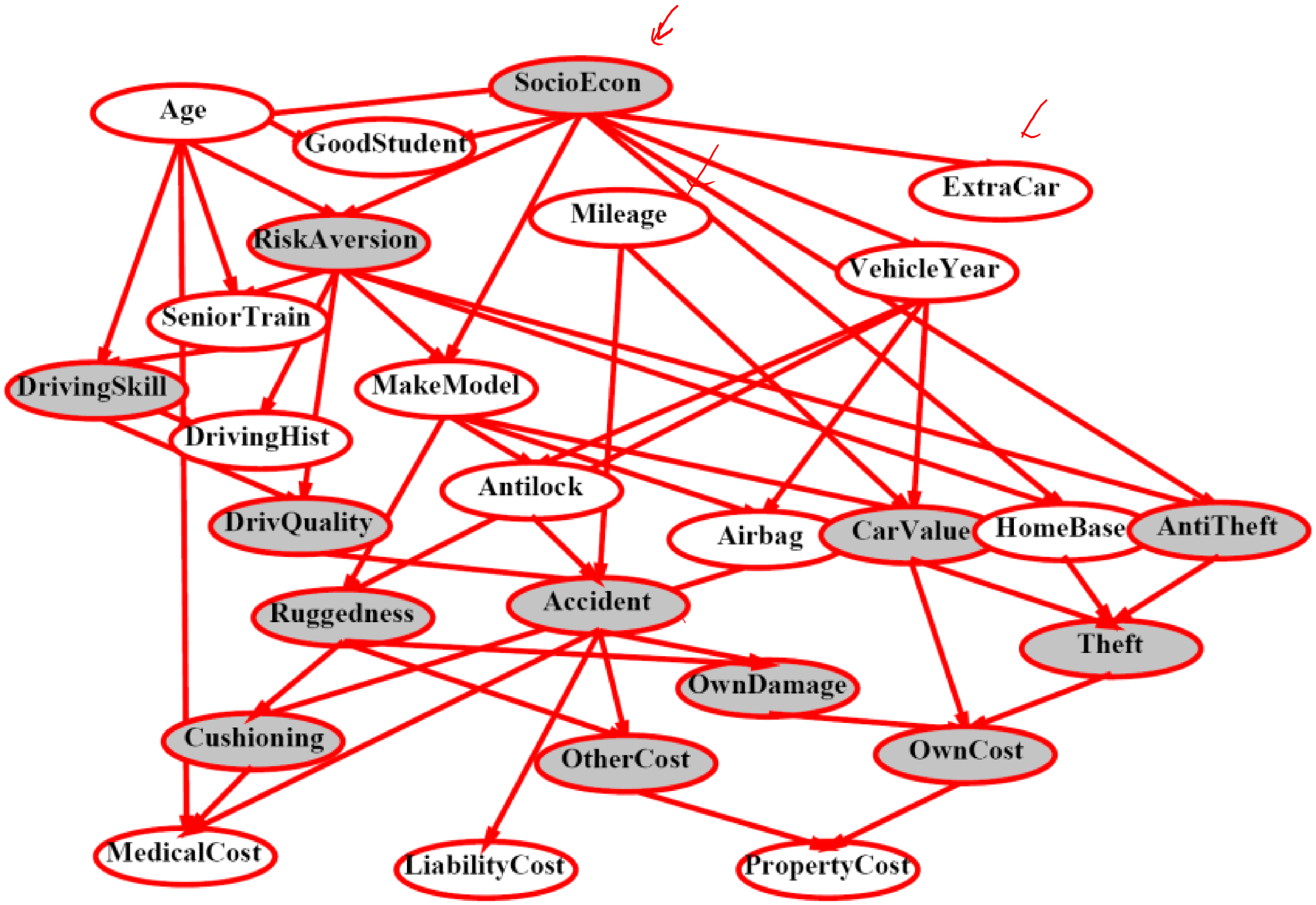
- With assumption of conditional independence:

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

- Bayes'nets / graphical models help us express conditional independence assumptions

Bayes’ Nets

- A Bayes’ net is an efficient encoding of a probabilistic model of a domain

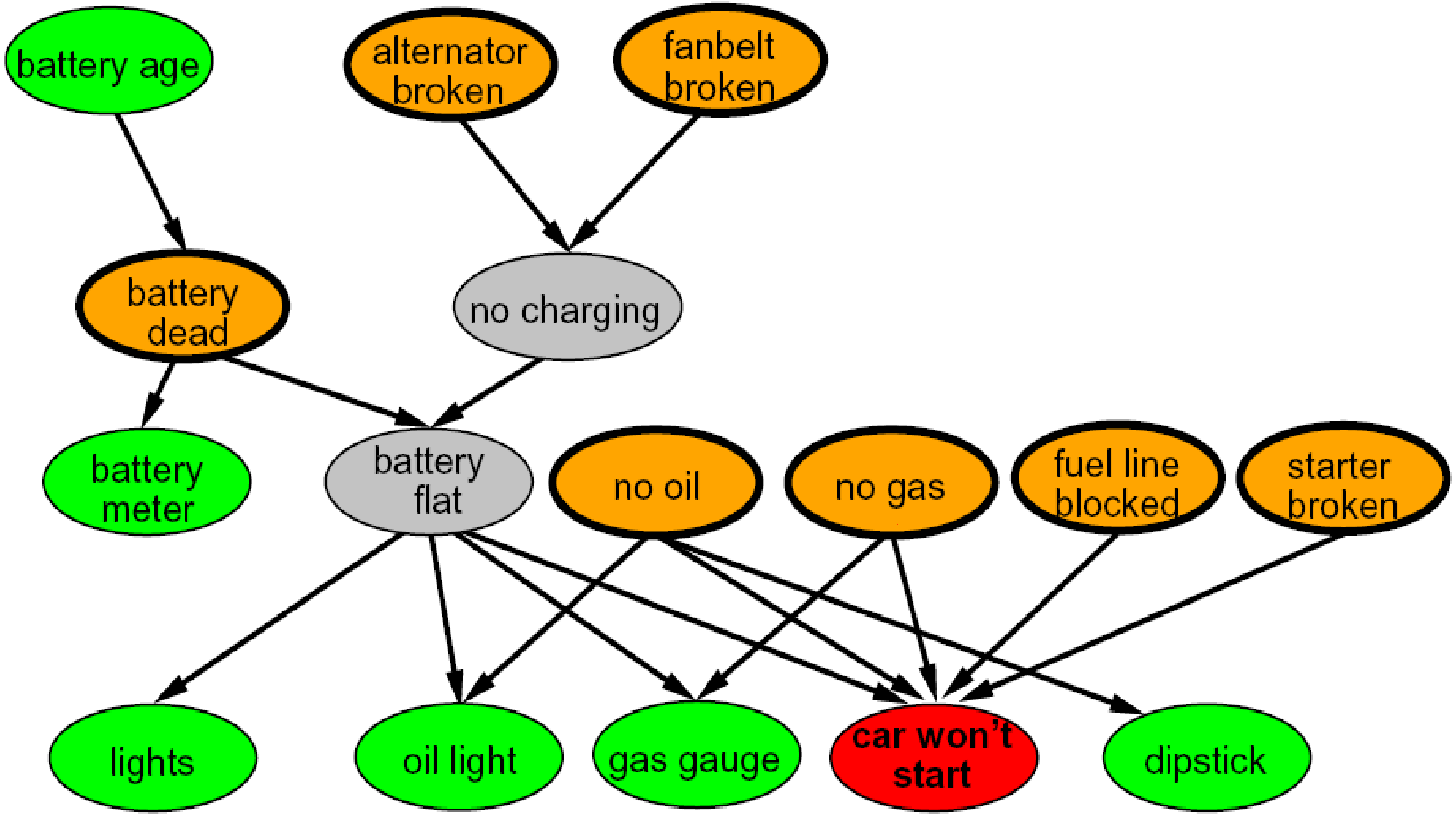


Bayes' Nets: Big Picture

- Two problems with using full joint distribution tables as our probabilistic models:
  - Unless there are only a few variables, the joint is WAY too big to represent explicitly
  - Hard to learn (estimate) anything empirically about more than a few variables at a time

- Bayes' nets: a technique for describing complex joint distributions (models) using simple, local distributions (conditional probabilities)
  - More properly called graphical models
  - We describe how variables locally interact
  - Local interactions chain together to give global, indirect interactions
  - For about 10 min, we'll be vague about how these interactions are specified

Example Bayes' Net: Car



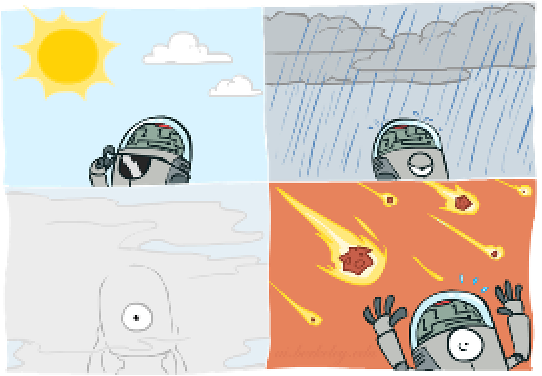
Graphical Model Notation

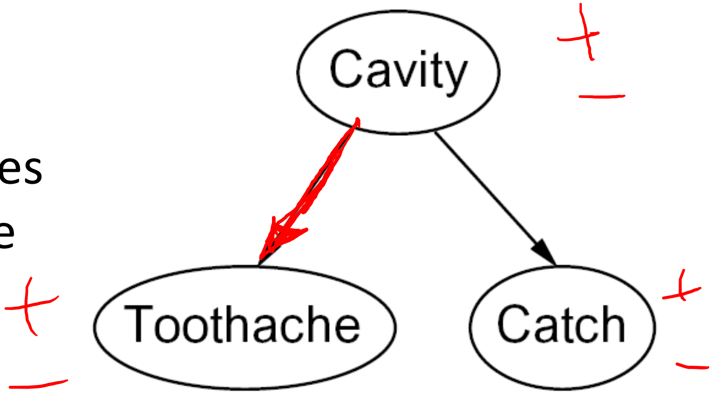
- Nodes: variables (with domains)
  - Can be assigned (observed) or unassigned (unobserved)

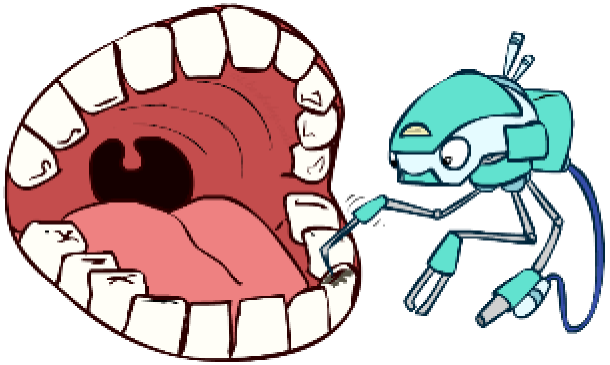
Edge

- Arcs: interactions
  - Similar to CSP constraints
  - Indicate "direct influence" between variables
  - Formally: encode conditional independence (more later)

- For now: imagine that arrows mean direct causation (in general, they don't!)

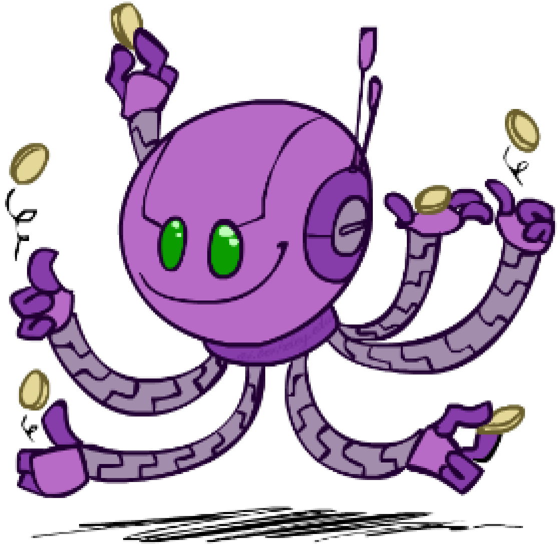






Example: Coin Flips

- N independent coin flips



X_1 X_2 $\cdots$ X_n

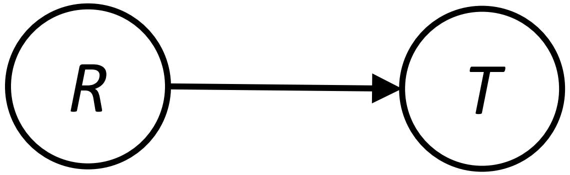
- No interactions between variables: absolute independence

Example: Traffic

- Variables:
  - R: It rains
  - T: There is traffic

- Model 1: independence

- Model 2: rain causes traffic



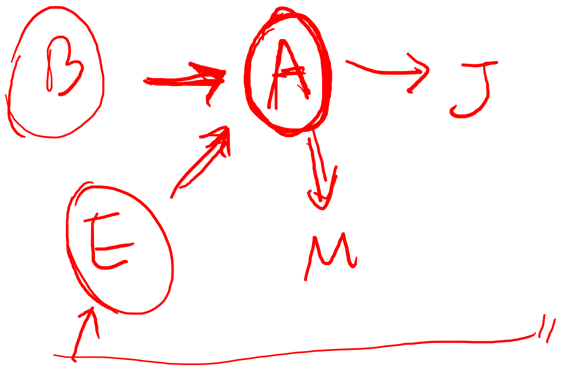
Example: Traffic II

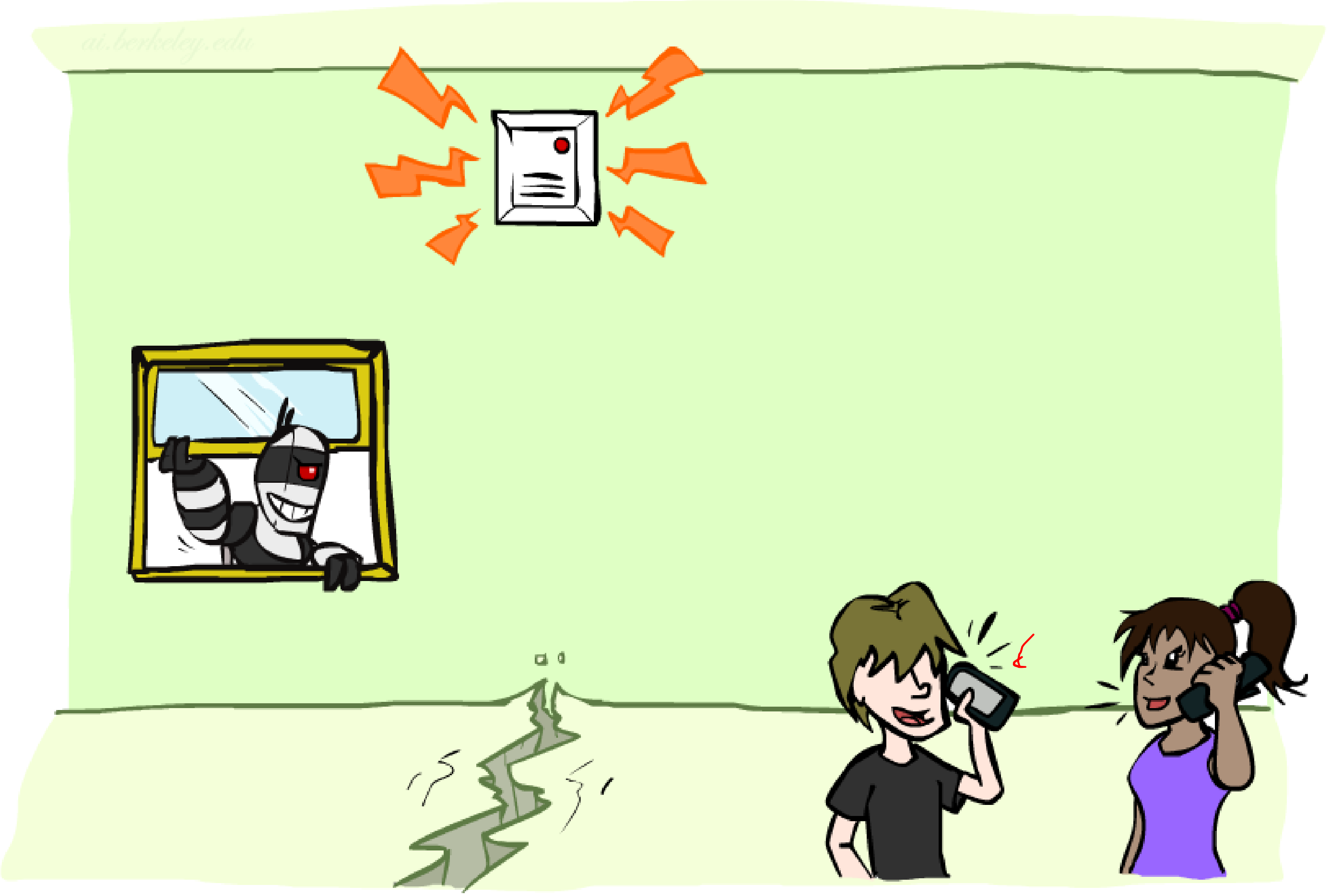
- Let's build a causal graphical model!
- Variables
  - T: Traffic
  - R: It rains
  - L: Low pressure
  - D: Roof drips
  - B: Ballgame
  - C: Cavity

$$\begin{tikzcd} & & & B \arrow[dl] \\ L \arrow[r] & R \arrow[r] \arrow[d] & T & \\ & D & & C \end{tikzcd}$$

Example: Alarm Network

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





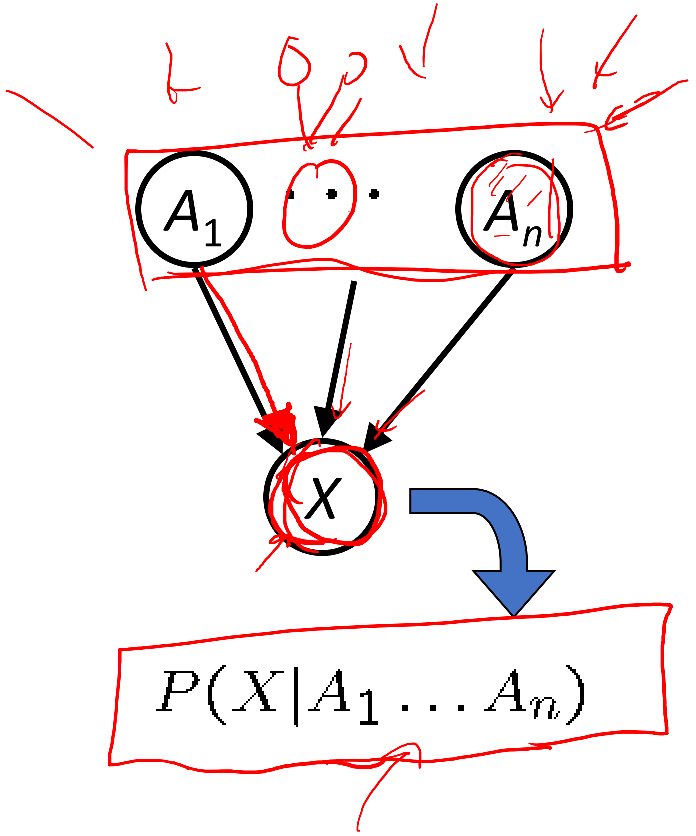
Bayes' Net Semantics

Bayes' Net Semantics

- A set of nodes, one per variable X
- A directed, acyclic graph
- A conditional distribution for each node
  - A collection of distributions over X, one for each combination of parents' values

    $$P(X|a_1 \ldots a_n)$$

  - CPT: conditional probability table
  - Description of a noisy "causal" process



A Bayes net = Topology (graph) + Local Conditional Probabilities

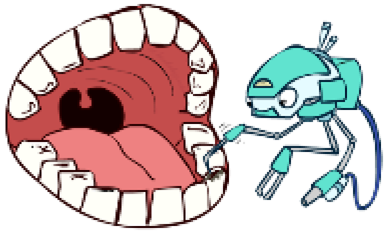
Probabilities in BNs

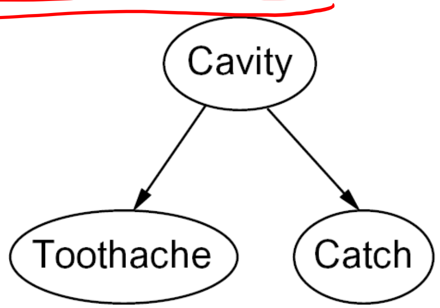
$P(X_1 \cdots X_n) =$

- Bayes' nets implicitly encode joint distributions
  - As a product of local conditional distributions
  - To see what probability a BN gives to a full assignment, multiply all the relevant conditionals together:

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

  - Example:





$P(+cavity, +catch, -toothache)$

Probabilities in BNs

- Why are we guaranteed that setting results in a proper joint distribution? $P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} \boxed{P(x_i | parents(X_i))}$

- Chain rule (valid for all distributions): $P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i | x_1 \ldots x_{i-1})$

- <u>Assume</u> conditional independences: $P(x_i | x_1, \ldots x_{i-1}) = P(x_i | parents(X_i))$

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

- Not every BN can represent every joint distribution
  - The topology enforces certain conditional independencies

Example: Coin Flips

X_1 X_2 . . . X_n

$P(X_1)$

h	0.5
t	0.5

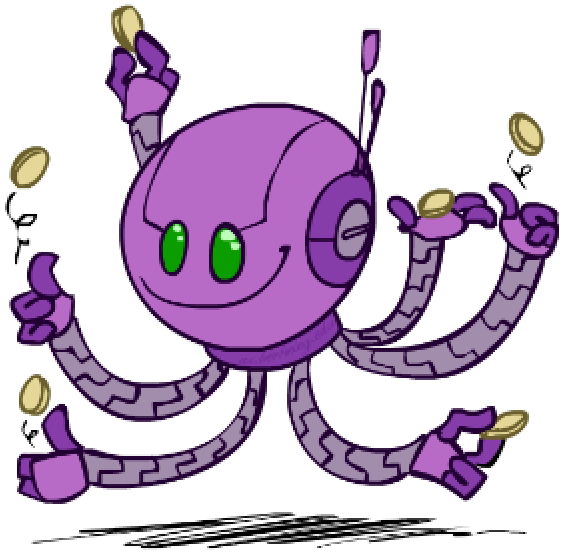
$P(X_2)$

h	0.5
t	0.5

. . .

$P(X_n)$

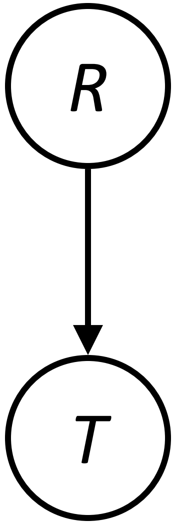
h	0.5
t	0.5



$P(h, h, t, h) =$ p(h) p(h) p(t) p(h)

Only distributions whose variables are absolutely independent can be represented by a Bayes' net with no arcs.

Example: Traffic



$P(R)$

+r	1/4
-r	3/4

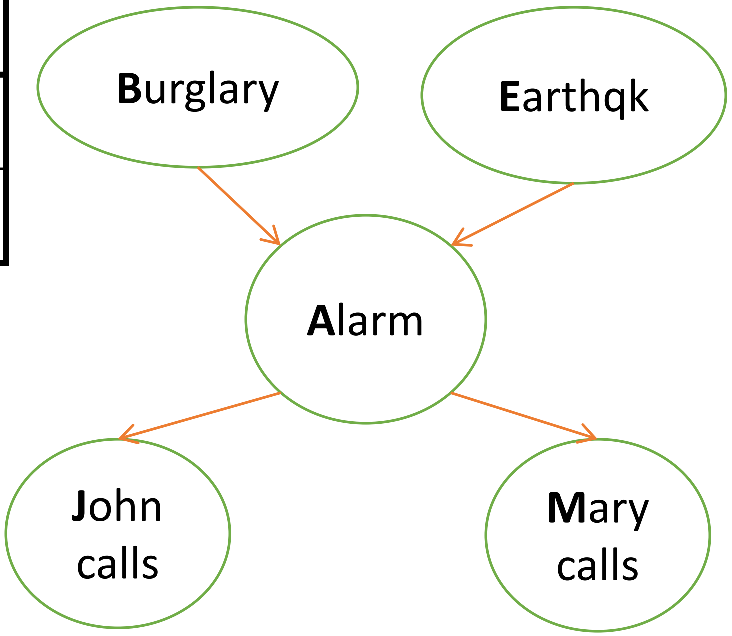
$P(T|R)$

+r	+t	3/4
	-t	1/4
-r	+t	1/2
	-t	1/2

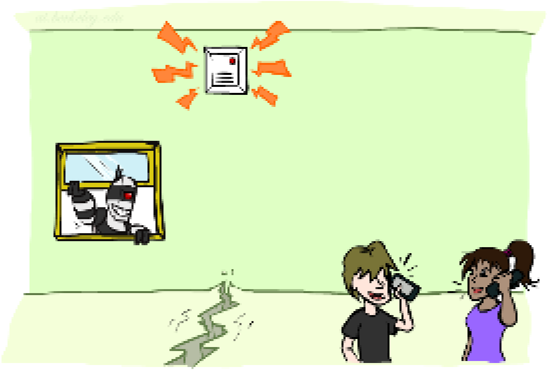
$$P(+r,-t) = \underbrace{P(r+)}_{1/4} \cdot \underbrace{P(-t|r+)}_{1/2} = \frac{1}{8}$$

Example: Alarm Network

B	P(B)
+b	0.001
-b	0.999



E	P(E)
+e	0.002
-e	0.998



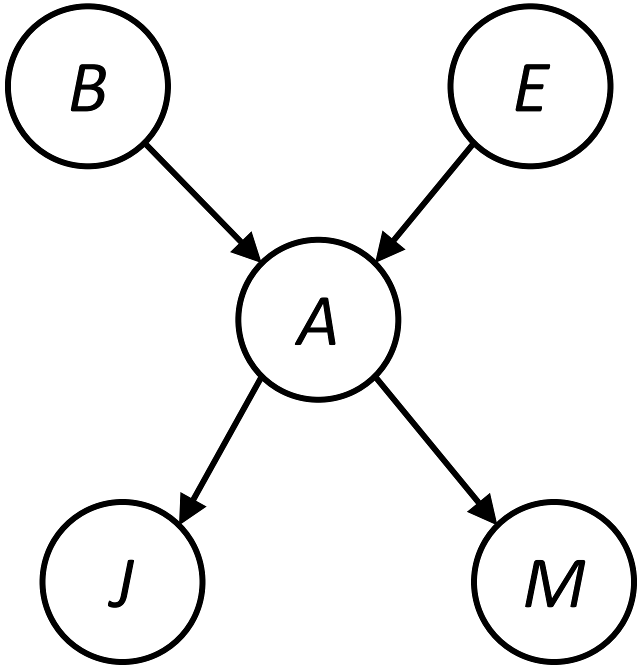
A	J	P(J|A)
+a	+j	0.9
+a	-j	0.1
-a	+j	0.05
-a	-j	0.95

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

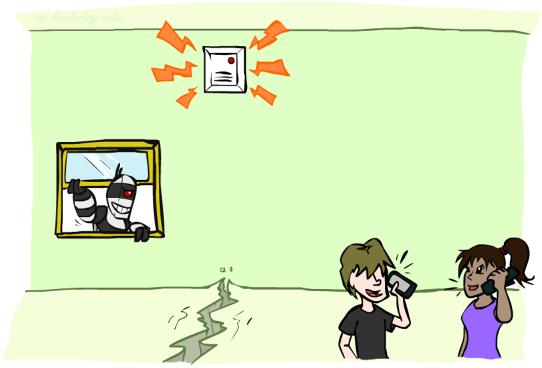
B	E	A	P(A|B,E)
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999

Example: Alarm Network

B	P(B)
+b	0.001
-b	0.999



E	P(E)
+e	0.002
-e	0.998



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

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

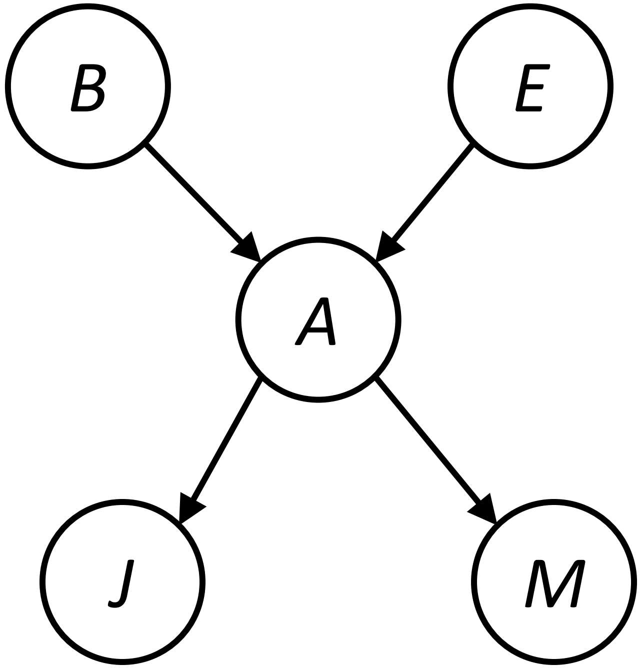
$$P(\underline{+b}, \underline{-e}, \underline{+a}, \underline{-j}, \underline{+m}) = P(+b)\,P(-e)\,P(+a \mid +b, -e) \times$$

$$\underline{P(-j \mid \cancel{+b}, \cancel{-e}, +a)}\; \underline{P(+m \mid \cancel{-j}, \cancel{+b}, \cancel{-e}, +a)}$$

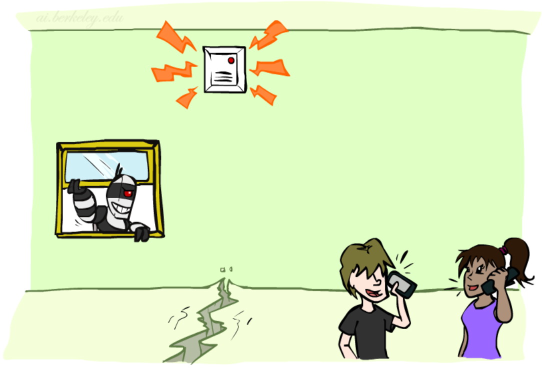
$$= P(+b) \cdot P(-e) \cdot P(+a \mid +b, -e) \cdot P(-j \mid +a) \cdot P(+m \mid +a)$$

Example: Alarm Network

B	P(B)
+b	0.001
-b	0.999



E	P(E)
+e	0.002
-e	0.998



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

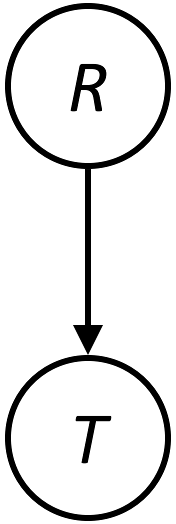
A	M	P(M|A)
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99

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

$$P(+b, -e, +a, -j, +m) =$$
$$P(+b)P(-e)P(+a|+b, -e)P(-j|+a)P(+m|+a) =$$
$$0.001 \times 0.998 \times 0.94 \times 0.1 \times 0.7$$

Example: Traffic

- Causal direction



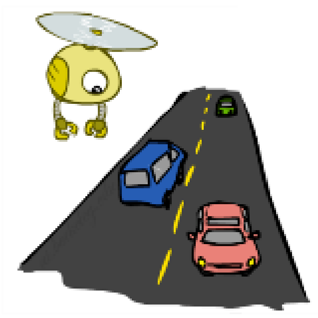
$P(R)$

+r	1/4
-r	3/4

$P(T|R)$

+r	+t	3/4
	-t	1/4

-r	+t	1/2
	-t	1/2

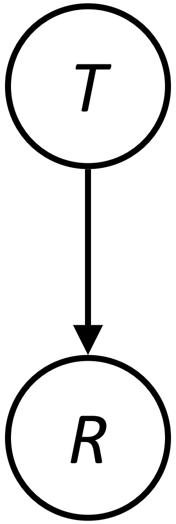


$\underline{P(T,R)} = P(T|R) \cdot P(R)$

+r	+t	3/16
+r	-t	1/16
-r	+t	6/16
-r	-t	6/16

Example: Reverse Traffic

- Reverse causality? $T \to R$



$P(T)$

+t	9/16
-t	7/16

$P(R|T)$

+t	+r	1/3
	-r	2/3

-t	+r	1/7
	-r	6/7

$P(T, R)$

+r	+t	3/16
+r	-t	1/16
-r	+t	6/16
-r	-t	6/16

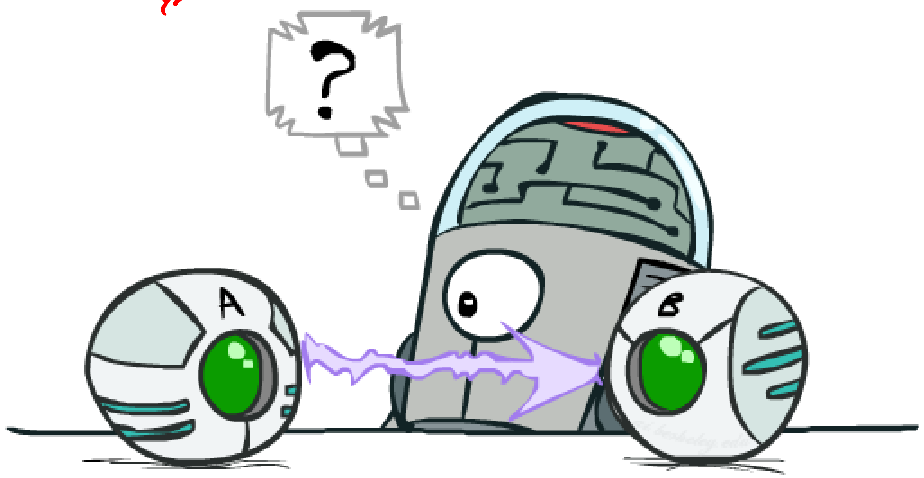
Causality?

- When Bayes' nets reflect the true causal patterns:
  - Often simpler (nodes have fewer parents)
  - Often easier to think about
  - Often easier to elicit from experts

- BNs need not actually be causal
  - Sometimes no causal net exists over the domain (especially if variables are missing)
  - E.g. consider the variables *Traffic* and *Drips*
  - End up with arrows that reflect correlation, not causation

- What do the arrows really mean?
  - Topology may happen to encode causal structure
  - Topology really encodes conditional independence

$$P(x_i|x_1, \ldots x_{i-1}) = P(x_i|parents(X_i))$$



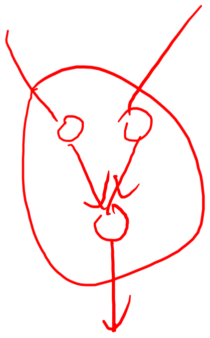
Size of a Bayes' Net

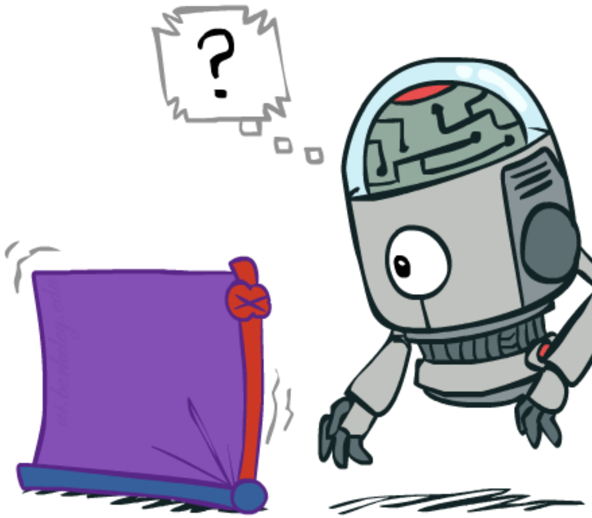
- How big is a joint distribution over N Boolean variables? (Table)

  2^N

- How big is an N-node net if nodes have up to k parents? $k \ll N$

  $O(N * 2^{k+1})$





- Both give you the power to calculate

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

- BNs: Huge space savings!
- Also easier to elicit local CPTs
- Also faster to answer queries (coming)