CS188: Announcements

Self Grade Drop: You have 1 for the semester.

See Piazza

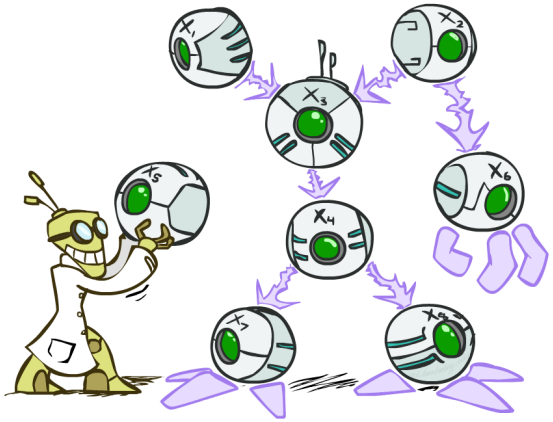
Homework due tonight.

Project 3 on Friday.

Discussing new attendance policy.

CS 188: Artificial Intelligence

http://bit.ly/cs188bnr



Bayes' Nets

Probability Recap

http://bit.ly/cs188bnr

Conditional probability: $P(x|y) = \frac{P(x,y)}{P(y)}$.

Product rule: $P(x,y) = P(x|y)P(y)$.

Bayes Rule: $P(y|x) = \frac{P(x|y)P(y)}{P(x)}$.

Chain rule:

$$\begin{aligned} P(x_1, x_2, \ldots, x_n) &= P(x_1)P(x_2|x_1)P(x_3|x_1, x_2)\ldots \\ &= \Pi_{i=1}^{n} P(x_i|x_1, \ldots, x_{i-1}). \end{aligned}$$

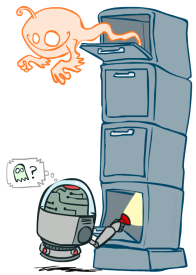
X, Y independent if and only if: $\forall x, y : P(x,y) = P(x)P(y)$.

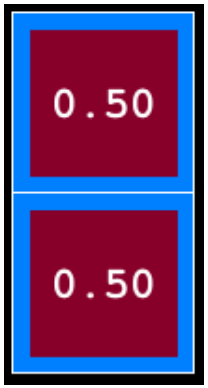
X and Y are conditionally independent given Z if and only if:

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

Ghostbusters Chain Rule

http://bit.ly/cs188bnr
[Demo: Ghostbuster – with probability (L12D2)]





T	B	G	P(T,B,G)
+t	+b	+g	0.16
+t	+b	-g	0.16
+t	-b	+g	0.24
+t	-b	-g	0.04
+t	+b	+g	0.04
-t	+b	-g	0.24
-t	-b	+g	0.06
-t	-b	-g	0.06

Each sensor depends only on where the ghost is.

That means, the two sensors are conditionally independent, given the ghost position.

T: Top square is red.
B: Bottom square is red.
G: Ghost is in the top.

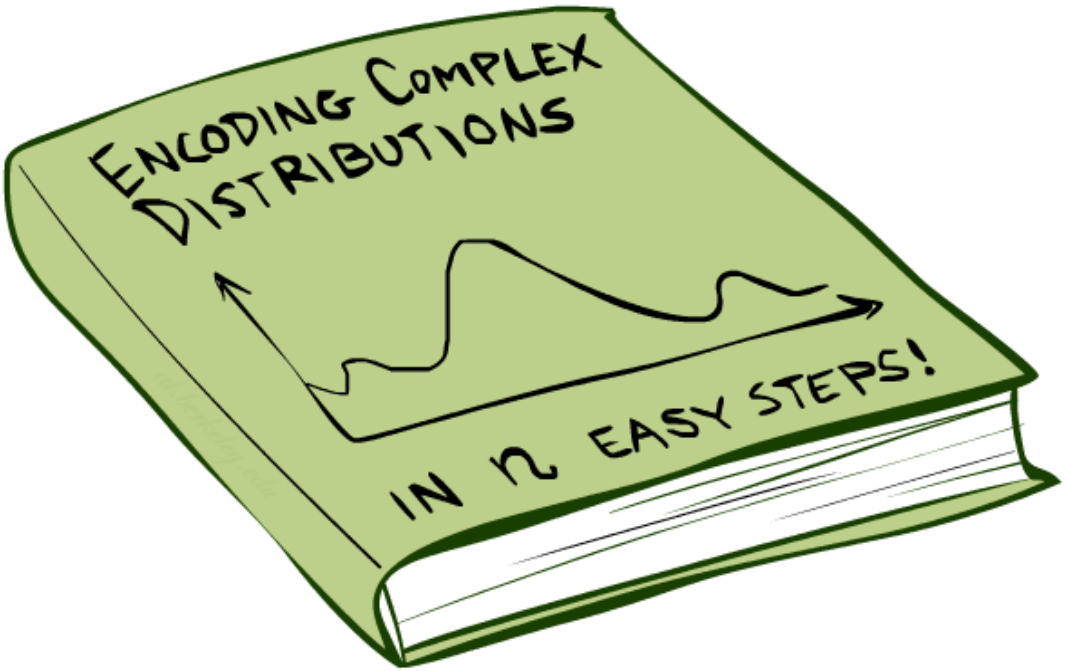
Givens:

P(+g) = 0.5
P(-g) = 0.5
P(+t | +g) = 0.8
P(+t | -g) = 0.4
P(+b | +g) = 0.4
P(+b | -g) = 0.8

$$P(T,B,G) = P(G)P(T|G)P(B|G)$$

Bayes'Nets: Big Picture

http://bit.ly/cs188bnr



Bayes' Nets: Big Picture

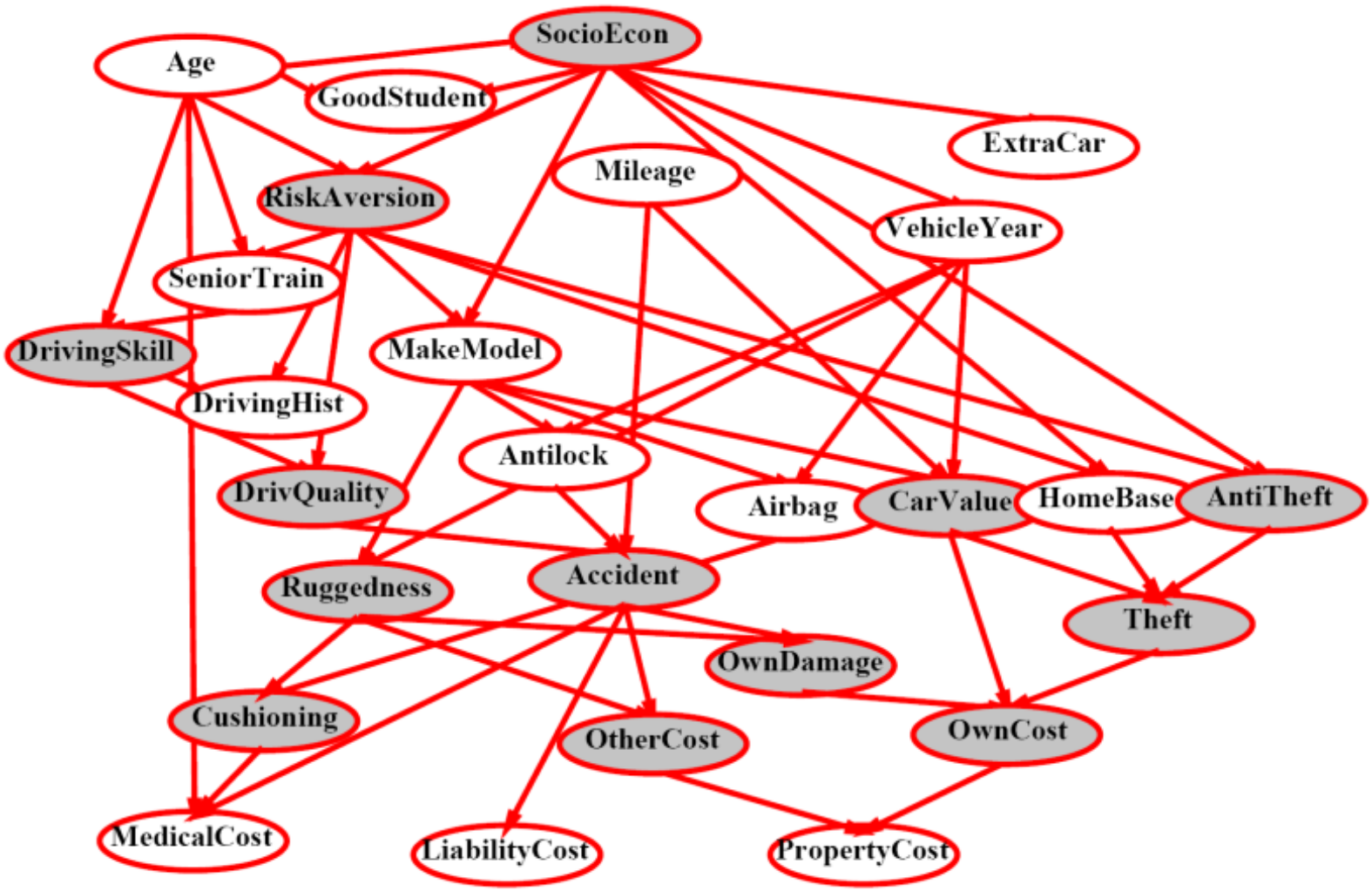
Two problems with using full joint distribution tables as our probabilistic models:

- 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

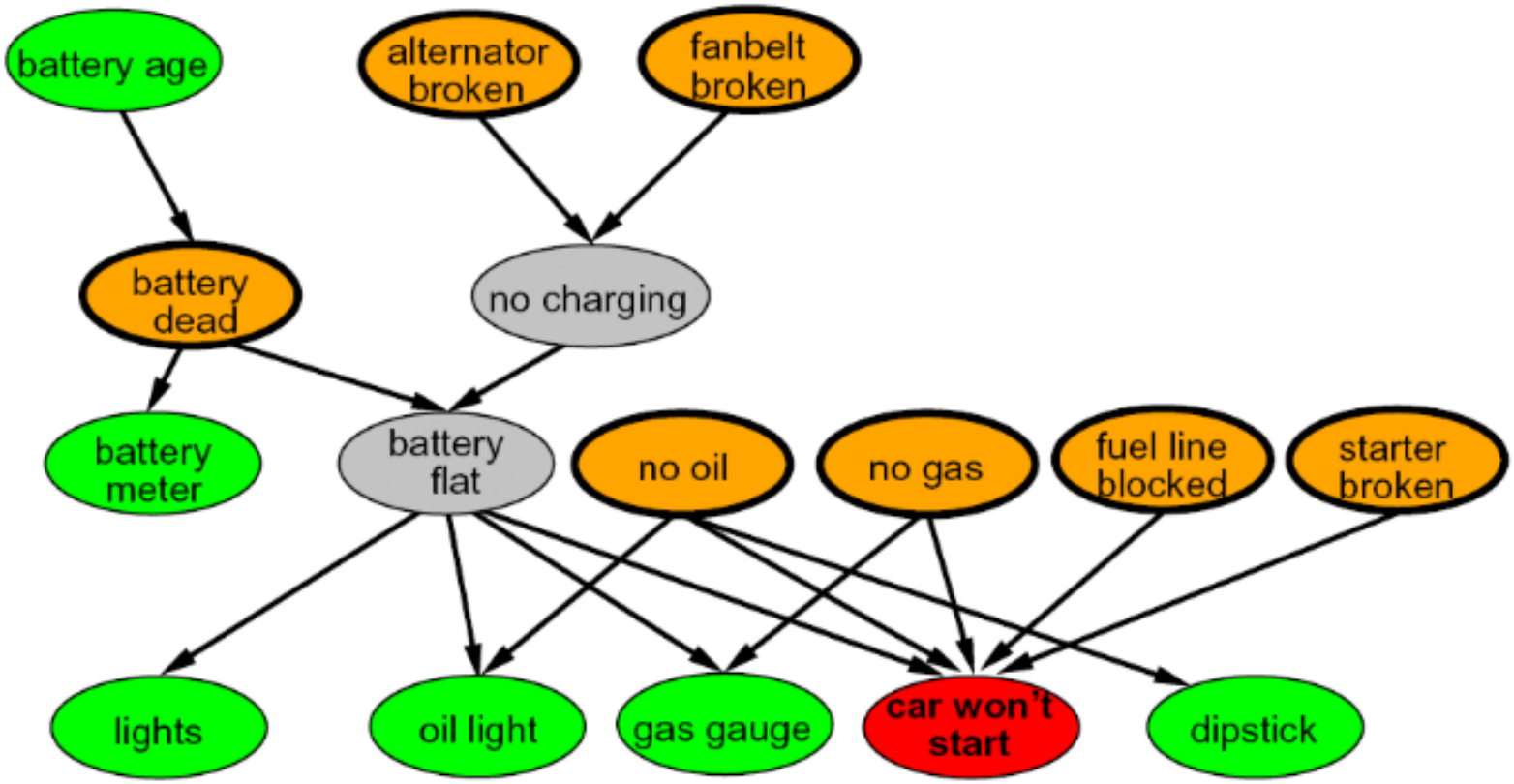
Bayes' nets: a technique for describing complex joint distributions (models) using simple, local distributions (conditional probabilities)

- More properly called graphical models
- We describe how variables locally interact
- Local interactions chain together to give global, indirect interactions
- For now, we'll be vague about how these interactions are specified

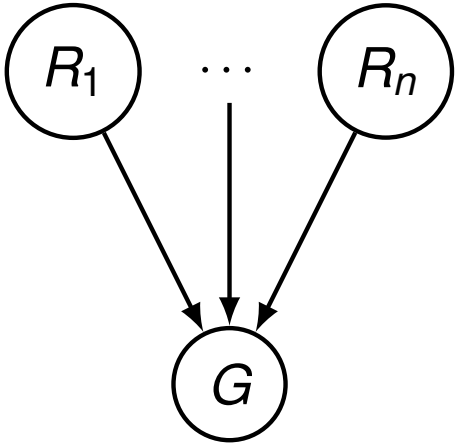
Example Bayes' Net: Insurance



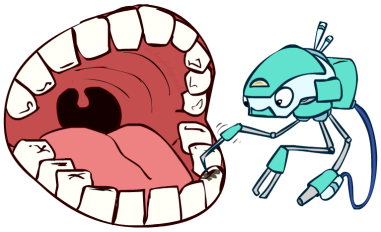
Example Bayes' Net: Car

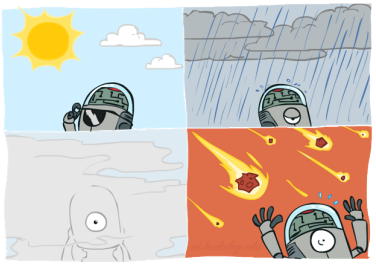


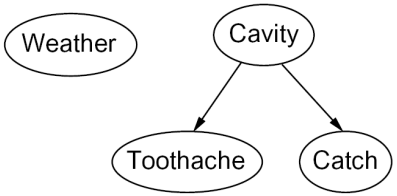
Ghostbusters Bayes Net



Graphical Model Notation







Nodes: variables (with domains)

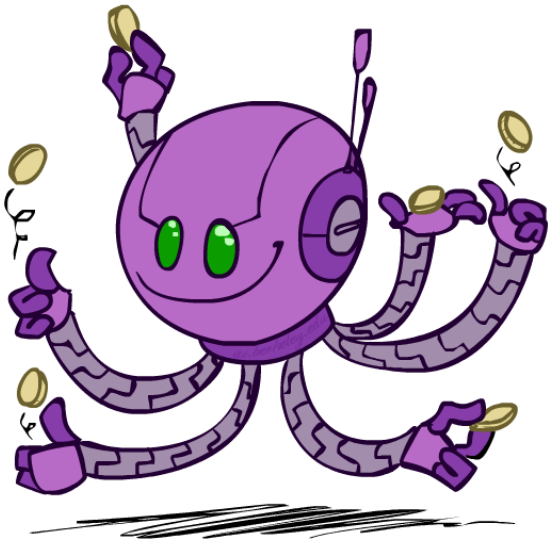
- Can be assigned (observed) or unassigned (unobserved)

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

No interactions between variables: absolute independence

X_1 X_2 $\cdots$ X_n

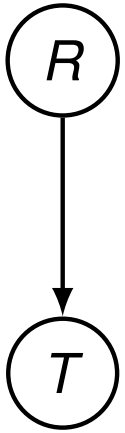
Example: Traffic

Variables:

- R: It rains
- T: There is traffic

Model 1: independence

Model 2: rain causes traffic



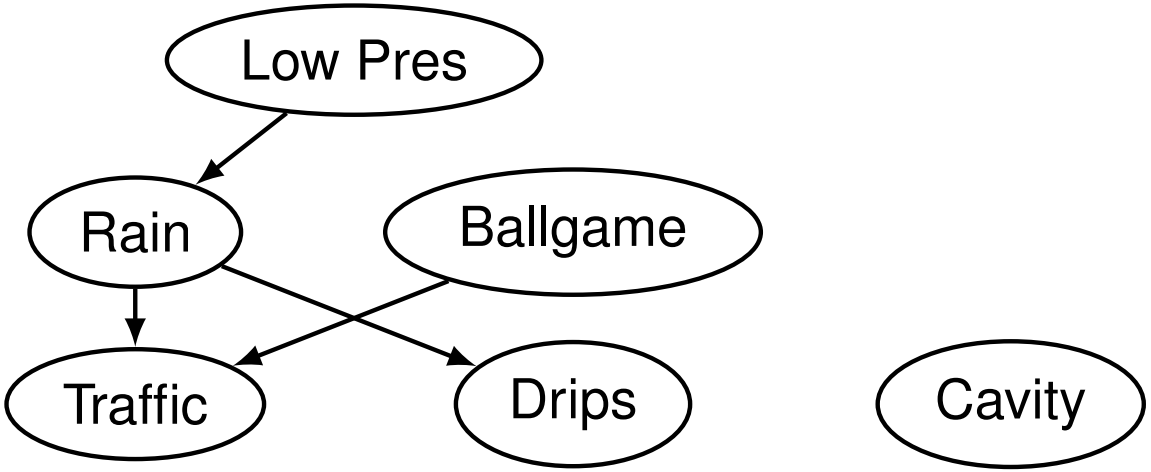
Why is an agent using model 2 better?

Example: Traffic II

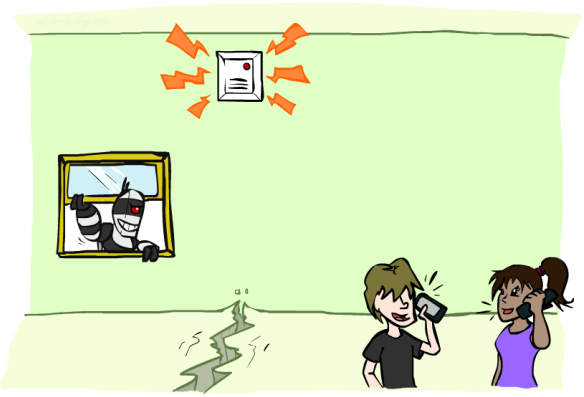
Let's build a causal graphical model!

Variables

- T: Traffic
- R: It rains
- L: Low pressure
- D: Roof drips
- B: Ballgame
- C: Cavity

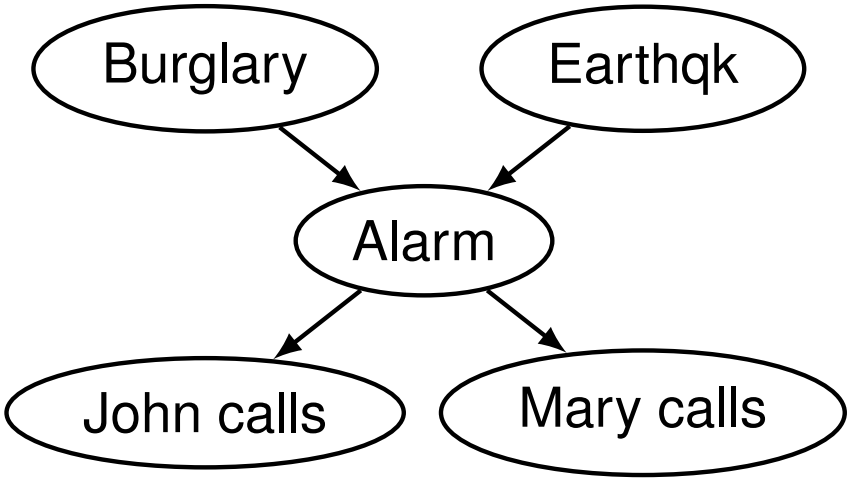


Example: Alarm Network



Variables

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



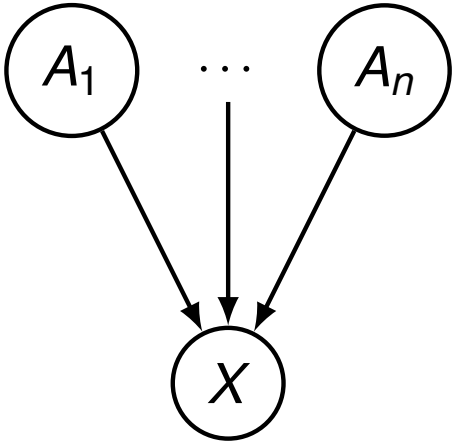
Bayes' Net Semantics

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

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

Probabilities in Bayesnets

Bayes' nets implicitly encode joint distributions

- As a product of local conditional distributions

$$P(x_1, \ldots, x_n) = \prod_i Pr(x_i | Parents(X_i))$$

- To see what probability a BN gives to a full assignment, multiply all the relevant conditionals together:

$$P(x_1, \ldots, x_n) = \prod_i Pr(x_i | Parents(X_i))$$

Example:

$$Pr(G = b | T = r, B = r) = Pr(G = r | T = r) P(S = s | B = r)$$

$$Pr(G = b | T = r, B = r) = Pr(G = r | T = r, B = r) P(S = s | B = r)$$

Probabilities in BNs

Why are we guaranteed that setting results is joint distribution?

Chain rule (valid for all distributions):

$$P(x_1, \ldots, x_n) = \prod_i P(x_i | x_1, \ldots, x_i)$$

Assume conditional independences:

$$P(x_1, \ldots, x_n) = \prod_i P(x_i | parents(X_i))$$

Consequence: Not every BN can represent every joint distribution.

- The topology enforces certain conditional independencies!

Example: Coin Flips

X_1

h	0.5
t	0.5

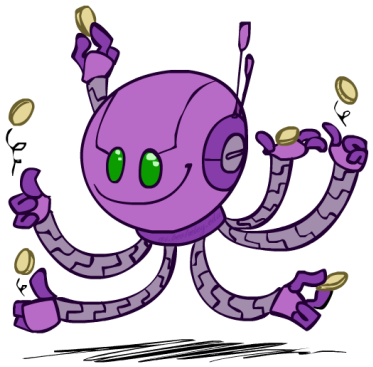
X_2

h	0.5
t	0.5

...

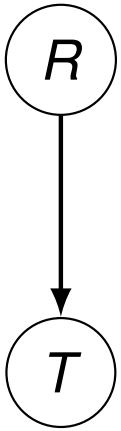
X_n

h	0.5
t	0.5



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

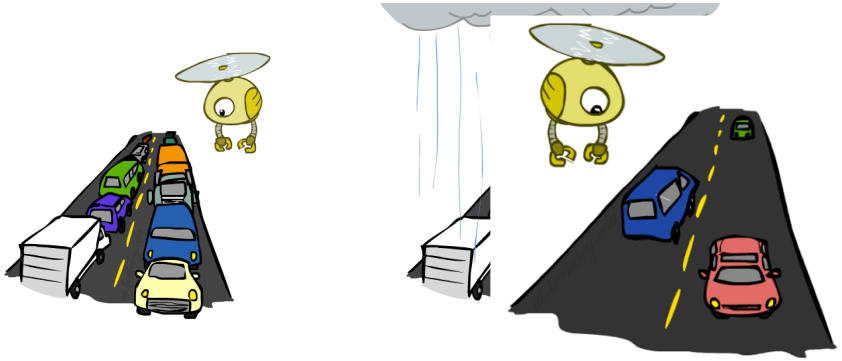
Example: Traffic



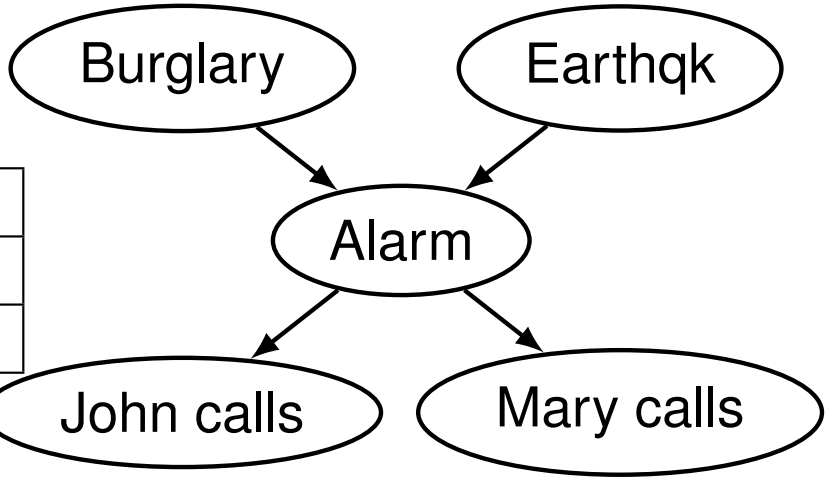
R	T
+r	1/4
-r	3/4

R	T	P(T|R)
+r	+t	3/4
+r	-t	1/4
-r	+t	1/2
-r	-t	1/2

$$P(+r,-t) = P(+r) \times P(-t \mid +r)$$
$$= (1/4) \times (1/4) = 1/16$$



Example: Alarm Network



B	P(B)
+b	0.001
-b	0.999

E	P(E)
+e	0.002
-e	0.998

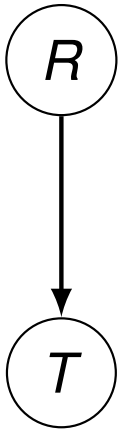
B	E	A	P(A|B,E)
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999

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

Example: Traffic

Causal direction



R	T
+r	1/4
-r	3/4

P(T|R)

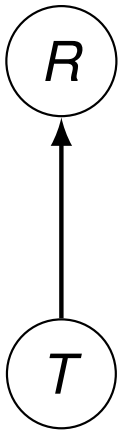
+r	+t	3/4
+r	-t	1/4
-r	+t	1/2
-r	-t	1/2

P(R,T)

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

Example: Reverse Traffic

Reverse causality?



P(T)

+t	9/16
-t	7/16

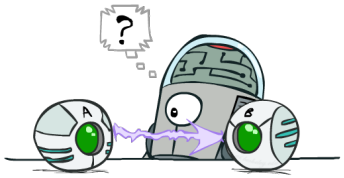
P(R|T)

+t	+r	1/3
+t	-r	2/3
-t	+r	1/7
-t	-r	6/7

P(R,T)

+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

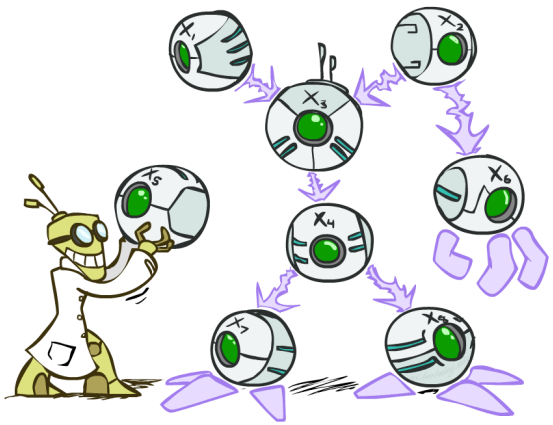
- Sometimes no causal net exists over the domain (especially if variables are missing)
- E.g. consider the variables Traffic and Drips
- End up with arrows that reflect correlation, not causation

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

What do the arrows really mean?

- Topology may happen to encode causal structure
- Topology really encodes conditional independence

Bayes' Nets



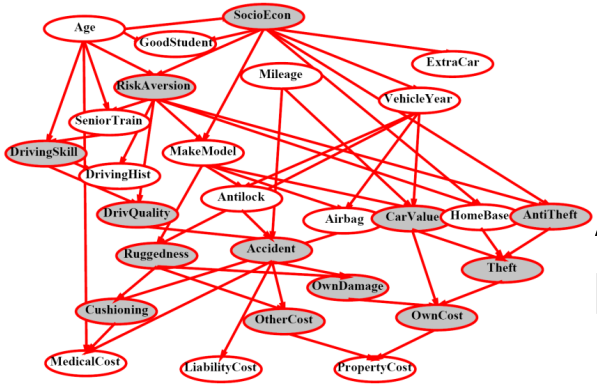
So far: how a Bayes' net encodes a joint distribution

Next: how to answer queries about that distribution

- So far:
- First assembled BNs using an intuitive notion of conditional independence as causality.
- Then saw that key property is conditional independence
- Main goal: answer queries about conditional independence and influence

After that: how to answer numerical queries (inference)

Bayes' Nets

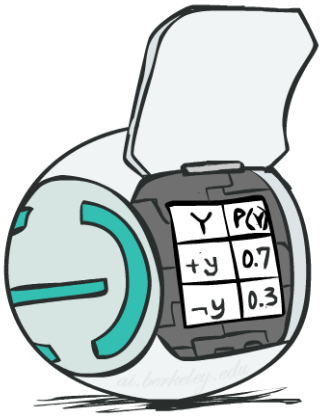


A Bayes' net is an efficient encoding of a probabilistic model of a domain

Questions we can ask:

- Inference: given a fixed BN, what is P(X | e)?
- Representation: given a BN graph, what kinds of distributions can it encode?
- Modeling: what BN is most appropriate for a given domain?

Bayes' Net Semantics

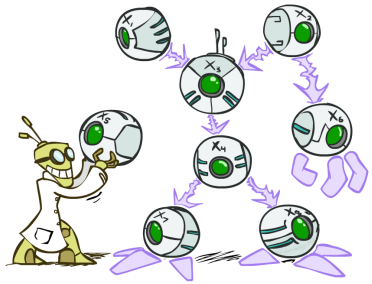


A directed, acyclic graph, one node per random variable

A conditional probability table (CPT) for each node

- A collection of distributions over X, one for each combination of parents' values

$$P(x_1|a_1,\ldots,a_n).$$

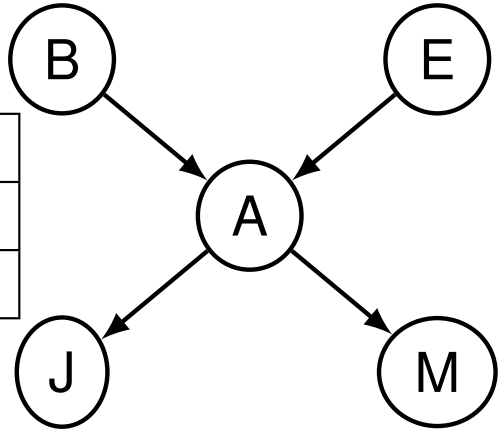


Bayes' nets implicitly encode joint distributions

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

$$P(x_1,\ldots,x_n) = \prod_i P(x_i|parents(X_i)),$$

Example: Alarm Network



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

P(+b,-e,+a,-j,+m) = P(+b) P(-e) P(+a| +b,-e) P(-j | +a) P(+m | +a)

$= -0.001 \times 0.998 \times 0.94 \times 0.1 \times 0.7$

Size of a Bayes' Net

How big is a joint distribution over N Boolean variables?

- 2^N

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

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

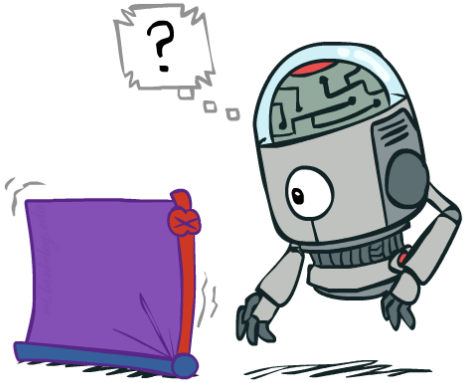
Both give you the power to calculate

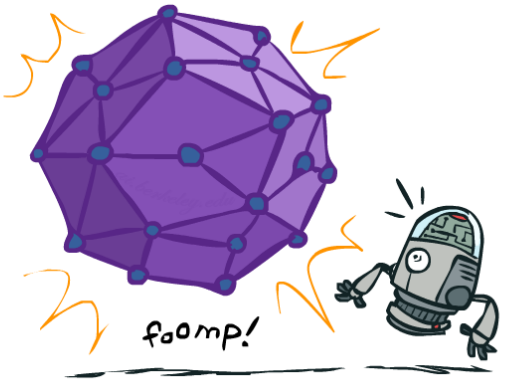
$$P(X_1, X_2, \cdots, X_N)$$

BNs: Huge space savings!

Also easier to elicit local CPTs

Also faster to answer queries (coming)





Bayes' Nets

Representation. ✔

Conditional Independences (Next.)

Probabilistic Inference

Learning Bayes' Nets from Data