Lecture 10: Bayes Nets: Probabilistic Models

Shuai Li

John Hopcroft Center, Shanghai Jiao Tong University

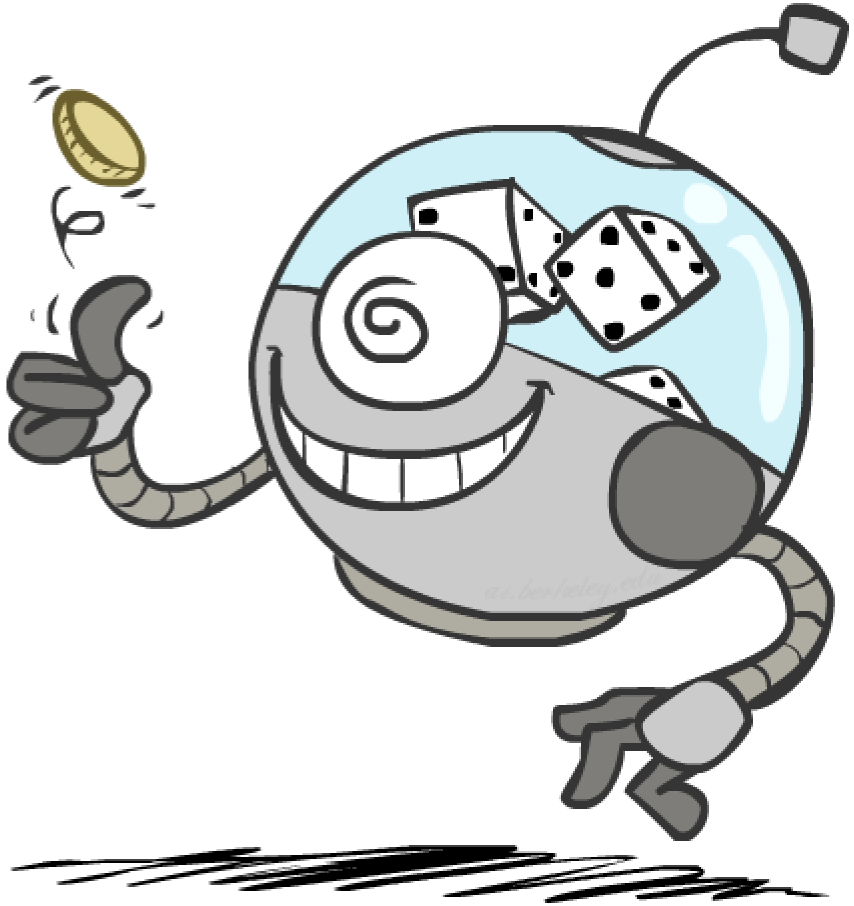
https://shuaili8.github.io

https://shuaili8.github.io/Teaching/CS410/index.html

Part of slide credits: CMU AI & http://ai.berkeley.edu

Background Part

Probability

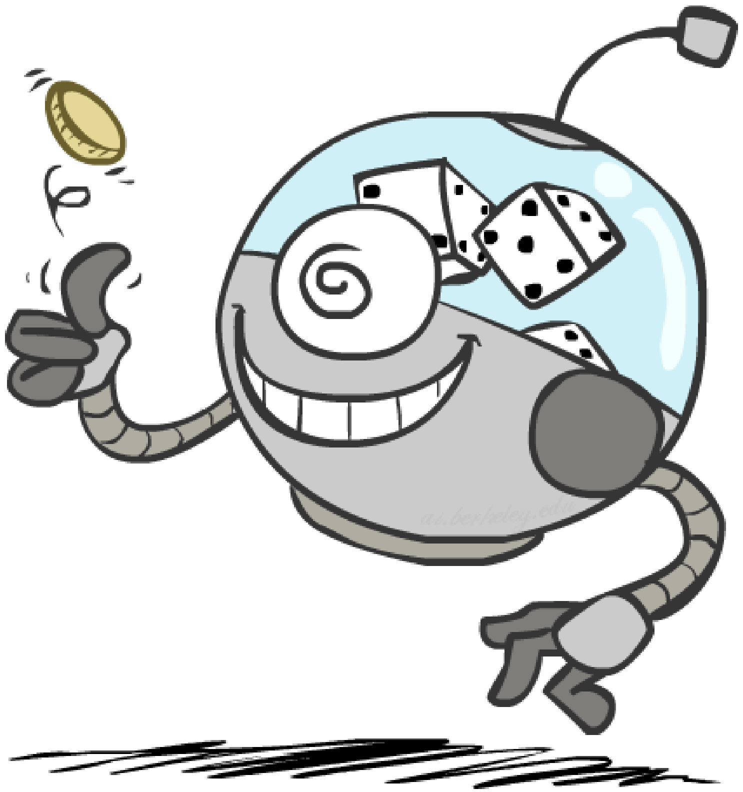


Random Variables

- A random variable is some aspect of the world about which we (may) have uncertainty
  - R = Is it raining?
  - T = Is it hot or cold?
  - D = How long will it take to drive to work?
  - L = Where is the ghost?

- We denote random variables with capital letters

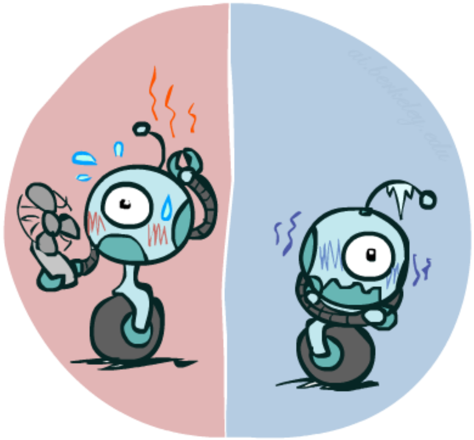
- Like variables in a CSP, random variables have domains
  - R in {true, false}   (often write as {+r, -r})
  - T in {hot, cold}
  - D in [0, ∞)
  - L in possible locations, maybe {(0,0), (0,1), …}



Probability Distributions

- Associate a probability with each value

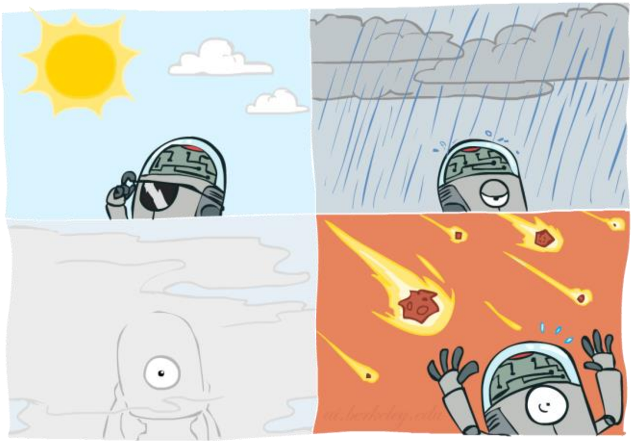
  - Temperature:



$$P(T)$$

T	P
hot	0.5
cold	0.5

  - Weather:



$$P(W)$$

W	P
sun	0.6
rain	0.1
fog	0.3
meteor	0.0

Probability Distributions 2

- Unobserved random variables have distributions

$P(T)$

T	P
hot	0.5
cold	0.5

$P(W)$

W	P
sun	0.6
rain	0.1
fog	0.3
meteor	0.0

- A distribution is a TABLE of probabilities of values
- A probability (lower case value) is a single number

$$P(W = rain) = 0.1$$

- Must have: $\forall x \;\; P(X = x) \geq 0$ and $\sum_x P(X = x) = 1$

Shorthand notation:

$$P(hot) = P(T = hot),$$
$$P(cold) = P(T = cold),$$
$$P(rain) = P(W = rain),$$
$$\ldots$$

OK *if* all domain entries are unique

Joint Distributions

- A *joint distribution* over a set of random variables: $X_1, X_2, \ldots X_n$ specifies a real number for each assignment (or *outcome*):

$$P(X_1 = x_1, X_2 = x_2, \ldots X_n = x_n)$$

$$P(x_1, x_2, \ldots x_n)$$

  - Must obey: $P(x_1, x_2, \ldots x_n) \geq 0$

$$\sum_{(x_1, x_2, \ldots x_n)} P(x_1, x_2, \ldots x_n) = 1$$

- Size of distribution if n variables with domain sizes d?
  - For all but the smallest distributions, impractical to write out!

$P(T, W)$

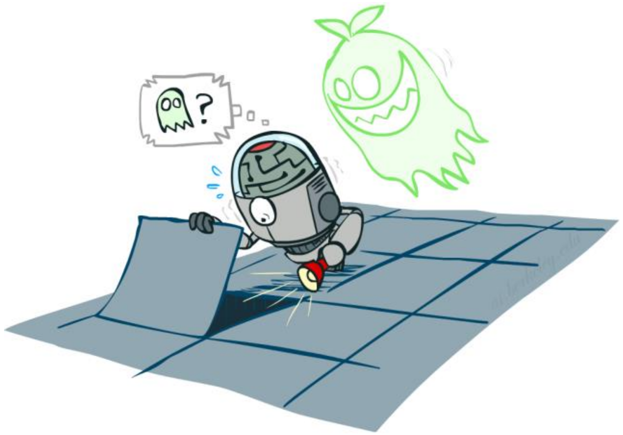
T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

Probabilistic Models

- A probabilistic model is a joint distribution over a set of random variables
- Probabilistic models:
  - (Random) variables with domains
  - Assignments are called *outcomes*
  - Joint distributions: say whether assignments (outcomes) are likely
  - *Normalized:* sum to 1.0
  - Ideally: only certain variables directly interact
- Constraint satisfaction problems:
  - Variables with domains
  - Constraints: state whether assignments are possible
  - Ideally: only certain variables directly interact

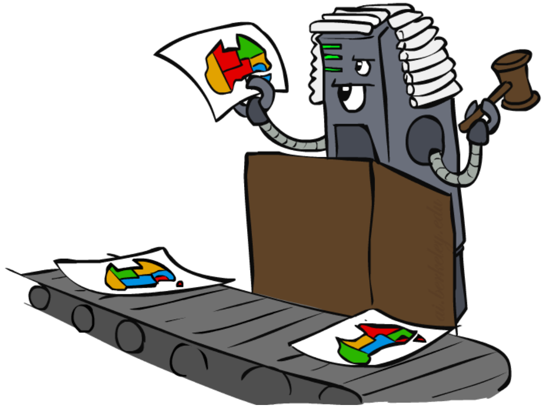
Distribution over T,W

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3



Constraint over T,W

T	W	P
hot	sun	T
hot	rain	F
cold	sun	F
cold	rain	T



Events

- An *event* is a set E of outcomes

$$P(E) = \sum_{(x_1 \ldots x_n) \in E} P(x_1 \ldots x_n)$$

- From a joint distribution, we can calculate the probability of any event
  - Probability that it's hot AND sunny?
  - Probability that it's hot?
  - Probability that it's hot OR sunny?
- Typically, the events we care about are *partial assignments*, like P(T=hot)

$$P(T, W)$$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

Quiz: Events

- P(+x, +y) ?

- P(+x) ?

- P(-y OR +x) ?

$$P(X, Y)$$

X	Y	P
+x	+y	0.2
+x	-y	0.3
-x	+y	0.4
-x	-y	0.1

Quiz: Events 2

- P(+x, +y) ?

  .2

- P(+x) ?

  .2+.3=.5

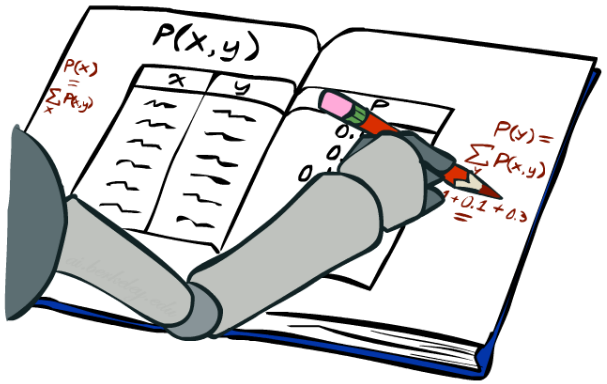
- P(-y OR +x) ?

  .1+.3+.2=.6

$$P(X, Y)$$

X	Y	P
+x	+y	0.2
+x	-y	0.3
-x	+y	0.4
-x	-y	0.1

Marginal Distributions



- Marginal distributions are sub-tables which eliminate variables
- Marginalization (summing out): Combine collapsed rows by adding

$$P(T,W)$$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

$$P(t) = \sum_s P(t,s)$$

$$P(T)$$

T	P
hot	0.5
cold	0.5

$$P(s) = \sum_t P(t,s)$$

$$P(W)$$

W	P
sun	0.6
rain	0.4

$$P(X_1 = x_1) = \sum_{x_2} P(X_1 = x_1, X_2 = x_2)$$

Quiz: Marginal Distributions

$P(X, Y)$

X	Y	P
+x	+y	0.2
+x	-y	0.3
-x	+y	0.4
-x	-y	0.1

$$P(x) = \sum_y P(x, y)$$

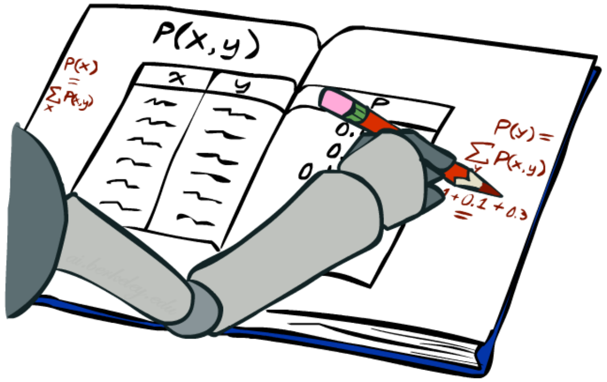
$$P(y) = \sum_x P(x, y)$$

$P(X)$

X	P
+x	
-x	

$P(Y)$

Y	P
+y	
-y	



Quiz: Marginal Distributions 2

$P(X,Y)$

X	Y	P
+x	+y	0.2
+x	-y	0.3
-x	+y	0.4
-x	-y	0.1

$$P(x) = \sum_y P(x,y)$$

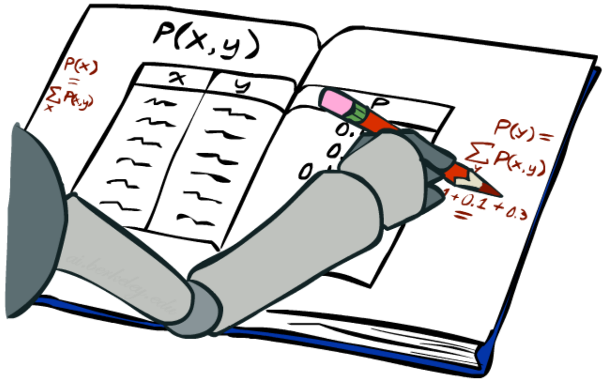
$$P(y) = \sum_x P(x,y)$$

$P(X)$

X	P
+x	.5
-x	.5

$P(Y)$

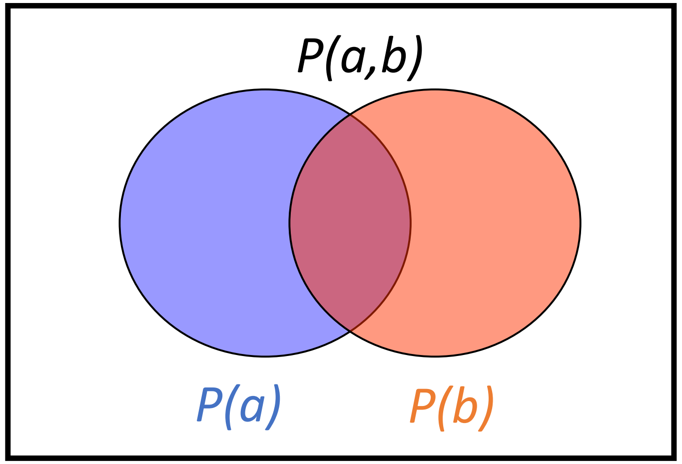
Y	P
+y	.6
-y	.4



Conditional Probabilities

- A simple relation between joint and conditional probabilities
  - In fact, this is taken as the *definition* of a conditional probability

$$P(a|b) = \frac{P(a,b)}{P(b)}$$



$$P(T,W)$$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

$$P(W=s|T=c) = \frac{P(W=s,T=c)}{P(T=c)} = \frac{0.2}{0.5} = 0.4$$

$$= P(W=s,T=c) + P(W=r,T=c)$$

$$= 0.2 + 0.3 \quad = 0.5$$

Quiz: Conditional Probabilities

- P(+x | +y) ?
- P(-x | +y) ?
- P(-y | +x) ?

$$P(X,Y)$$

X	Y	P
+x	+y	0.2
+x	-y	0.3
-x	+y	0.4
-x	-y	0.1

Quiz: Conditional Probabilities 2

- P(+x | +y) ?

  .2/.6=1/3

- P(-x | +y) ?

  .4/.6=2/3

- P(-y | +x) ?

  .3/.5=.6

$$P(X, Y)$$

X	Y	P
+x	+y	0.2
+x	-y	0.3
-x	+y	0.4
-x	-y	0.1

Conditional Distributions

- Conditional distributions are probability distributions over some variables given fixed values of others

Conditional Distributions

$P(W|T)$

$P(W|T = hot)$

W	P
sun	0.8
rain	0.2

$P(W|T = cold)$

W	P
sun	0.4
rain	0.6

Joint Distribution

$P(T, W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

Normalization Trick

$$P(T,W)$$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

$$\begin{aligned} P(W=s|T=c) &= \frac{P(W=s,T=c)}{P(T=c)} \\ &= \frac{P(W=s,T=c)}{P(W=s,T=c)+P(W=r,T=c)} \\ &= \frac{0.2}{0.2+0.3} = 0.4 \end{aligned}$$

$$\longrightarrow$$

$$\begin{aligned} P(W=r|T=c) &= \frac{P(W=r,T=c)}{P(T=c)} \\ &= \frac{P(W=r,T=c)}{P(W=s,T=c)+P(W=r,T=c)} \\ &= \frac{0.3}{0.2+0.3} = 0.6 \end{aligned}$$

$$P(W|T=c)$$

W	P
sun	0.4
rain	0.6

Normalization Trick 2

$$P(W=s|T=c) = \frac{P(W=s,T=c)}{P(T=c)}$$
$$= \frac{P(W=s,T=c)}{P(W=s,T=c)+P(W=r,T=c)}$$
$$= \frac{0.2}{0.2+0.3} = 0.4$$

$P(T,W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

SELECT the joint probabilities matching the evidence →

$P(c,W)$

T	W	P
cold	sun	0.2
cold	rain	0.3

NORMALIZE the selection (make it sum to one) →

$P(W|T=c)$

W	P
sun	0.4
rain	0.6

$$P(W=r|T=c) = \frac{P(W=r,T=c)}{P(T=c)}$$
$$= \frac{P(W=r,T=c)}{P(W=s,T=c)+P(W=r,T=c)}$$
$$= \frac{0.3}{0.2+0.3} = 0.6$$

Normalization Trick 3

$P(T,W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

SELECT the joint probabilities matching the evidence

$P(c,W)$

T	W	P
cold	sun	0.2
cold	rain	0.3

NORMALIZE the selection (make it sum to one)

$P(W|T=c)$

W	P
sun	0.4
rain	0.6

- Why does this work? Sum of selection is P(evidence)! (P(T=c), here)

$$P(x_1|x_2) = \frac{P(x_1,x_2)}{P(x_2)} = \frac{P(x_1,x_2)}{\sum_{x_1} P(x_1,x_2)}$$

Quiz: Normalization Trick

- P(X | Y=-y) ?

$P(X, Y)$

X	Y	P
+x	+y	0.2
+x	-y	0.3
-x	+y	0.4
-x	-y	0.1

SELECT the joint probabilities matching the evidence →

NORMALIZE the selection (make it sum to one) →

Quiz: Normalization Trick 2

- P(X | Y=-y) ?

$P(X,Y)$

X	Y	P
+x	+y	0.2
+x	-y	0.3
-x	+y	0.4
-x	-y	0.1

SELECT the joint probabilities matching the evidence →

X	Y	P
+x	-y	0.3
-x	-y	0.1

NORMALIZE the selection (make it sum to one) →

X	P
+x	0.75
-x	0.25

To Normalize

- (Dictionary) To bring or restore to a normal condition

All entries sum to ONE

- Procedure:
  - Step 1: Compute Z = sum over all entries
  - Step 2: Divide every entry by Z
- Example 1

W	P
sun	0.2
rain	0.3

Normalize → Z = 0.5

W	P
sun	0.4
rain	0.6

- Example 2

T	W	P
hot	sun	20
hot	rain	5
cold	sun	10
cold	rain	15

Normalize → Z = 50

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

Probabilistic Inference

Inference by Enumeration

- P(W)?

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

Inference by Enumeration 2

- P(W)?

  P(sun)=.3+.1+.1+.15=.65

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

Inference by Enumeration 3

- P(W)?

  P(sun)=.3+.1+.1+.15=.65
  P(rain)=1-.65=.35

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

Inference by Enumeration 4

- P(W | winter, hot)?

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

Inference by Enumeration 5

- $P(W \mid winter, hot)$?

  $P(sun|winter,hot) \propto .1$
  $P(rain|winter,hot) \propto .05$

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

Inference by Enumeration 6

- P(W | winter, hot)?

  P(sun|winter,hot) ∝ .1
  P(rain|winter,hot) ∝ .05
  P(sun|winter,hot)=2/3
  P(rain|winter,hot)=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

Inference by Enumeration 7

- P(W | winter)?

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

Inference by Enumeration 8

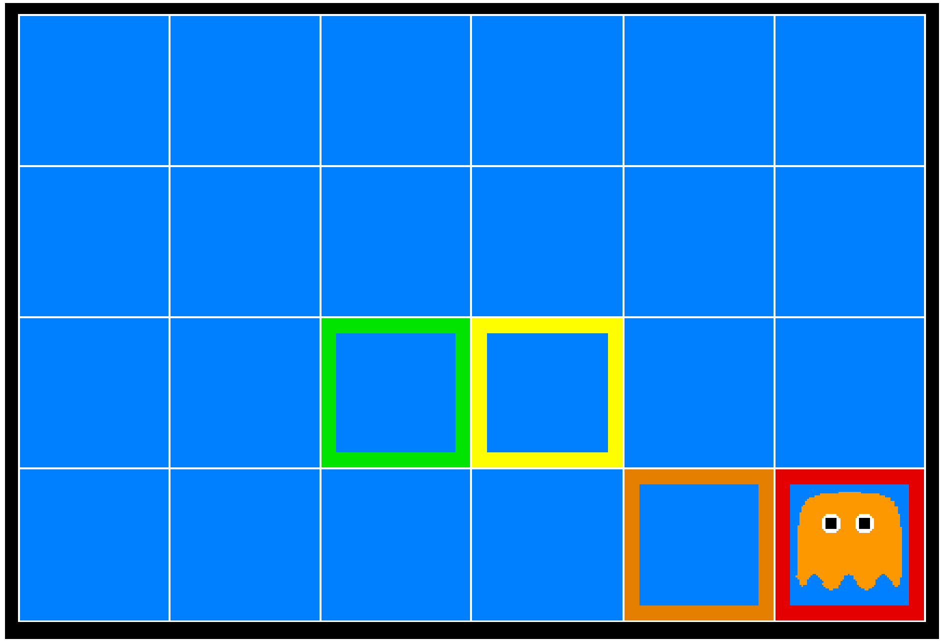
S	T	W	P
summer	hot	sun	0.30
summer	hot	rain	0.05
summer	cold	sun	0.10
summer	cold	rain	0.05
winter	hot	sun	0.10
winter	hot	rain	0.05
winter	cold	sun	0.15
winter	cold	rain	0.20

- P(W | winter)?

  $P(sun|winter) \propto .1+.15=.25$
  $P(rain|winter) \propto .05+.2=.25$
  $P(sun|winter)=.5$
  $P(rain|winter)=.5$

Main Part

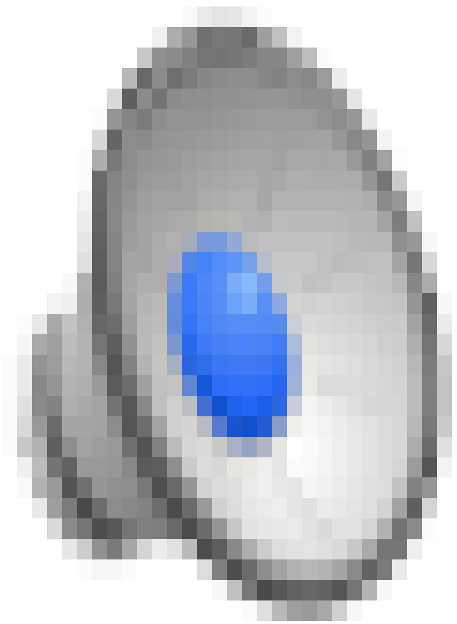
Inference in Ghostbusters



- A ghost is in the grid somewhere
- Sensor readings tell how close a square is to the ghost
  - On the ghost: red
  - 1 or 2 away: orange
  - 3 or 4 away: yellow
  - 5+ away: green
- Sensors are noisy, but we know P(Color | Distance)

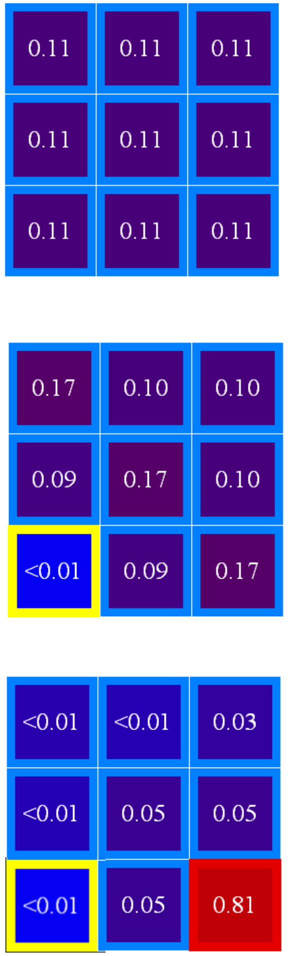
P(red | 3)	P(orange | 3)	P(yellow | 3)	P(green | 3)
0.05	0.15	0.5	0.3

Video of Demo Ghostbuster – No probability



Uncertainty

- General situation:
  - **Observed variables (evidence)**: Agent knows certain things about the state of the world (e.g., sensor readings or symptoms)
  - **Unobserved variables**: Agent needs to reason about other aspects (e.g. where an object is or what disease is present)
  - **Model**: Agent knows something about how the known variables relate to the unknown variables
- Probabilistic reasoning gives us a framework for managing our beliefs and knowledge



Probabilistic Inference

- Probabilistic inference: compute a desired probability from other known probabilities (e.g. conditional from joint)

- We generally compute conditional probabilities
  - P(on time | no reported accidents) = 0.90
  - These represent the agent's *beliefs* given the evidence

- Probabilities change with new evidence:
  - P(on time | no accidents, 5 a.m.) = 0.95
  - P(on time | no accidents, 5 a.m., raining) = 0.80
  - Observing new evidence causes *beliefs to be updated*

Inference by Enumeration

- General case:
  - Evidence variables: $E_1 \dots E_k = e_1 \dots e_k$
  - Query* variable: Q
  - Hidden variables: $H_1 \dots H_r$

$\left.\begin{array}{l} E_1 \dots E_k = e_1 \dots e_k \\ Q \\ H_1 \dots H_r \end{array}\right\} X_1, X_2, \dots X_n$

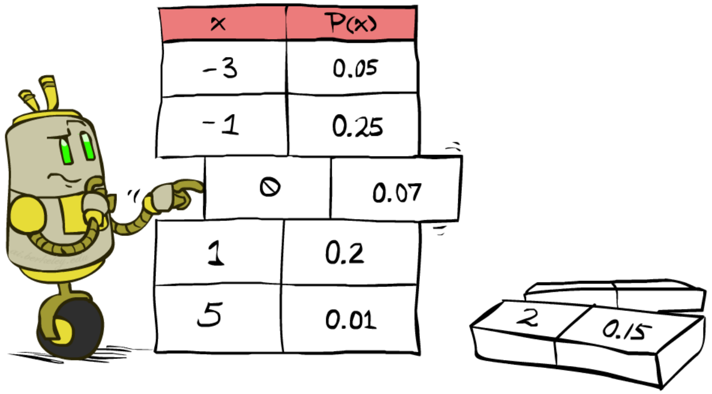
All variables

- We want:

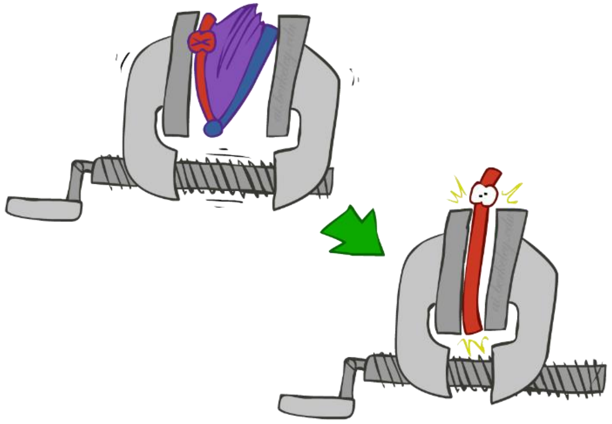
** Works fine with multiple query variables, too*

$$P(Q|e_1 \dots e_k)$$

- Step 1: Select the entries consistent with the evidence



- Step 2: Sum out H to get joint of Query and evidence



$$P(Q, e_1 \dots e_k) = \sum_{h_1 \dots h_r} P(\underbrace{Q, h_1 \dots h_r, e_1 \dots e_k}_{X_1, X_2, \dots X_n})$$

- Step 3: Normalize

$$\times \frac{1}{Z}$$

$$Z = \sum_q P(Q, e_1 \cdots e_k)$$

$$P(Q|e_1 \cdots e_k) = \frac{1}{Z} P(Q, e_1 \cdots e_k)$$

Answer Any Query from Joint Distributions

- Two tools to go from joint to query

1. Definition of conditional probability

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

2. Law of total probability (marginalization, summing out)

$$P(A) = \sum_b P(A,b)$$

$$P(Y \mid U,V) = \sum_x \sum_z P(x,Y,z \mid U,V)$$

Answer Any Query from Joint Distributions 2

- Two tools to go from joint to query
- Joint: $P(H_1, H_2, Q, E)$
- Query: $P(Q \mid e)$

1. Definition of conditional probability

$$P(Q|e) = \frac{P(Q,e)}{P(e)}$$

2. Law of total probability (marginalization, summing out)

$$P(Q,e) = \sum_{h_1} \sum_{h_2} P(h_1, h_2, Q, e)$$

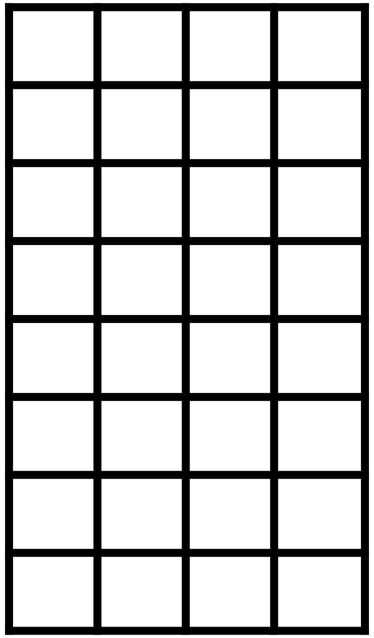
$$P(e) = \sum_{q} \sum_{h_1} \sum_{h_2} P(h_1, h_2, q, e)$$

Only need to compute $P(Q,e)$ then normalize

Answer Any Query from Joint Distributions 3

- Joint distributions are the best!

Joint

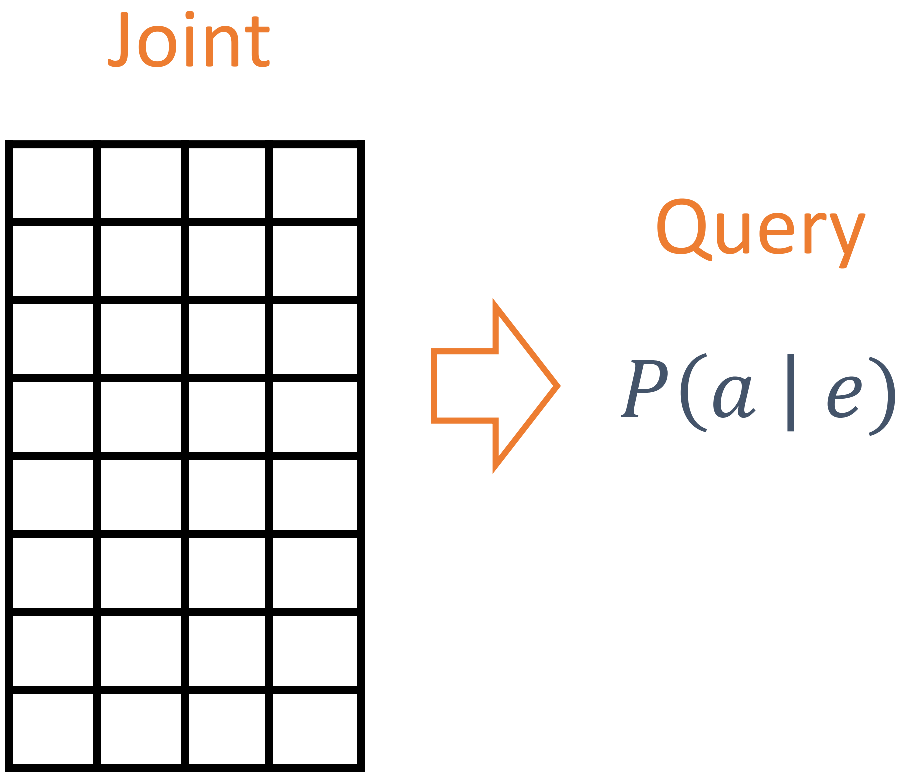


Query

$$P(q_1, q_2 \mid e_1, e_2, e_3)$$

Answer Any Query from Joint Distributions 4

- Joint distributions are the best!
- Problems with joints
  - We aren't given the joint table
  - Usually some set of conditional probability tables
- Problems with inference by enumeration
  - Worst-case time complexity $O(d^n)$
  - Space complexity $O(d^n)$ to store the joint distribution

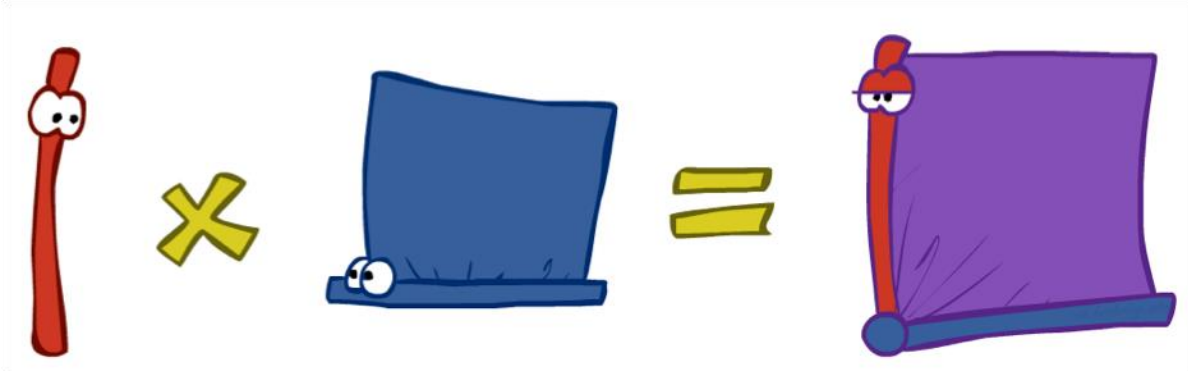


Build Joint Distribution Using Chain Rule

The Product Rule

- Sometimes have conditional distributions but want the joint

$$P(y)P(x|y) = P(x,y) \quad \Longleftrightarrow \quad P(x|y) = \frac{P(x,y)}{P(y)}$$



The Product Rule 2

$$P(y)P(x|y) = P(x,y)$$

- Example:

$P(W)$

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

$P(D,W)$

D	W	P
wet	sun	
dry	sun	
wet	rain	
dry	rain	

The Chain Rule

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

$$P(x_1, x_2, x_3) = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2)$$

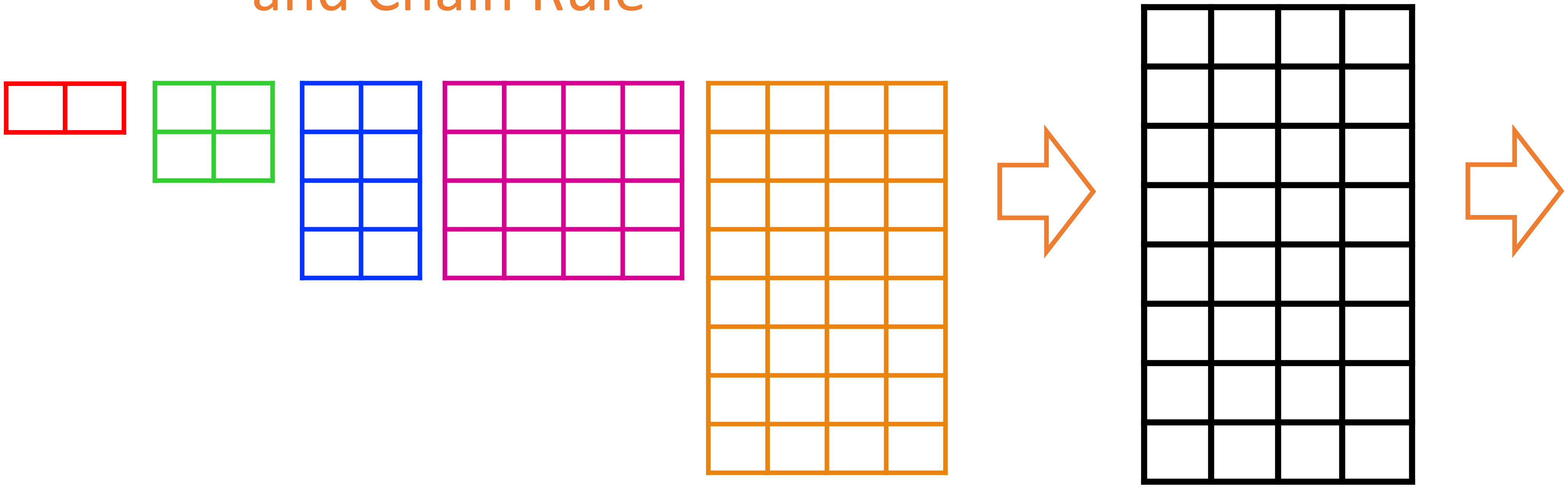
$$P(x_1, x_2, \ldots x_n) = \prod_i P(x_i|x_1 \ldots x_{i-1})$$

Build Joint Distribution Using Chain Rule

Conditional Probability Tables
and Chain Rule

Joint

Query



$P(a \mid e)$

$P(A)\ P(B|A)\ P(C|A,B)\ P(D|A,B,C)\ P(E|A,B,C,D)$

Build Joint Distribution Using Chain Rule 2

- Two tools to construct joint distribution

1. Product rule

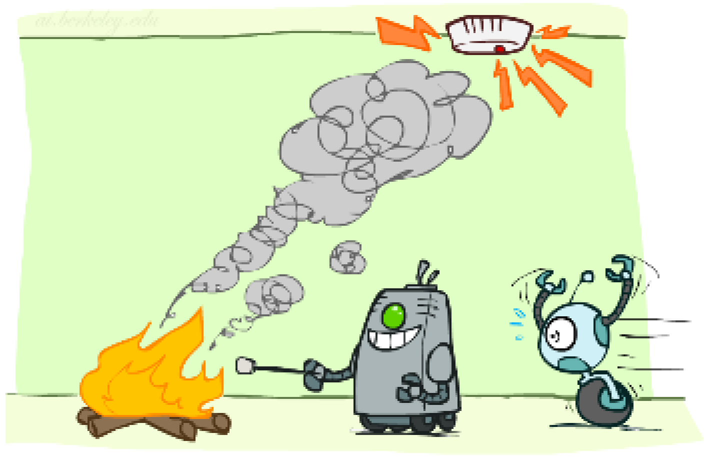
- $P(A,B) = P(A \mid B)P(B)$
- $P(A,B) = P(B \mid A)P(A)$

2. Chain rule

- $P(X_1, X_2, \ldots, X_n) = \prod_i P(X_i \mid X_1, \ldots, X_{i-1})$
- $P(A,B,C) = P(A)P(B \mid A)P(C \mid A,B)$ for ordering A, B, C
- $P(A,B,C) = P(A)P(C \mid A)P(B \mid A,C)$ for ordering A, C, B
- $P(A,B,C) = P(C)P(B \mid C)P(A \mid C,B)$ for ordering C, B, A
- …

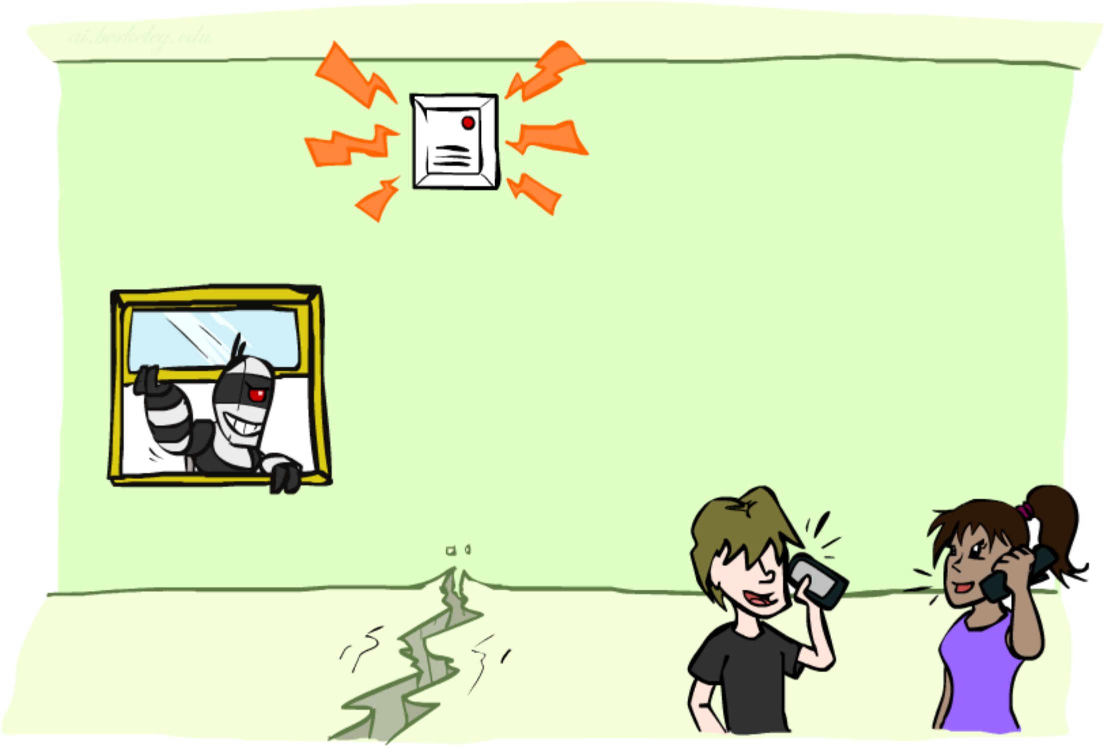
Example

- Binary random variables
  - Fire
  - Smoke
  - Alarm



Quiz

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



How many different ways can we write the chain rule?

A. 1

B. 5

C. 5 *choose* 5

D. 5!

E. 5^5

Answer Any Query from Condition Probability Tables

- Process to go from (specific) conditional probability tables to query

1. Construct the joint distribution
   1. Product Rule or Chain Rule
2. Answer query from joint
   1. Definition of conditional probability
   2. Law of total probability (marginalization, summing out)

Answer Any Query from Condition Probability Tables 2

- Bayes' rule as an example
- Given: $P(E|Q),\ P(Q)$ Query: $P(Q \mid e)$

1. Construct the joint distribution
   1. Product Rule or Chain Rule

$$P(E,Q) = P(E|Q)P(Q)$$

2. Answer query from joint
   1. Definition of conditional probability

$$P(Q \mid e) = \frac{P(e,Q)}{P(e)}$$

   2. Law of total probability (marginalization, summing out)

$$P(Q \mid e) = \frac{P(e,Q)}{\sum_q P(e,q)}$$

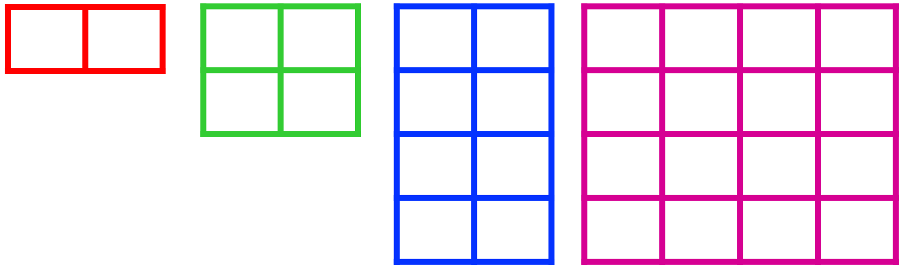
Only need to compute $P(e,Q)$ then normalize

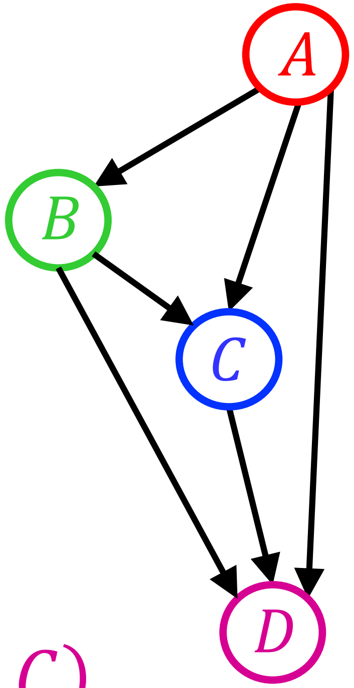
Bayesian Networks

Bayesian Networks

Bayes net

- One node per random variable, DAG
- One conditional probability table (CPT) per node: P(node | *Parents*(node))





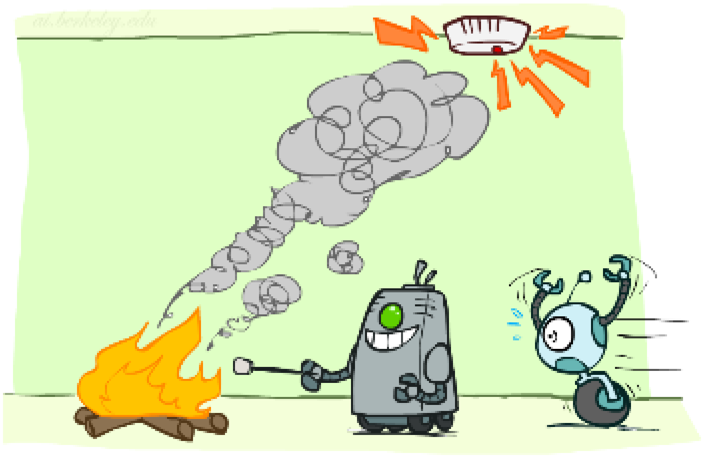
$$P(A,B,C,D) = P(A)\,P(B|A)\,P(C|A,B)\,P(D|A,B,C)$$

Encode joint distributions as product of conditional distributions on each variable

$$P(X_1, \ldots, X_N) = \prod_i P(X_i|\,Parents(X_i))$$

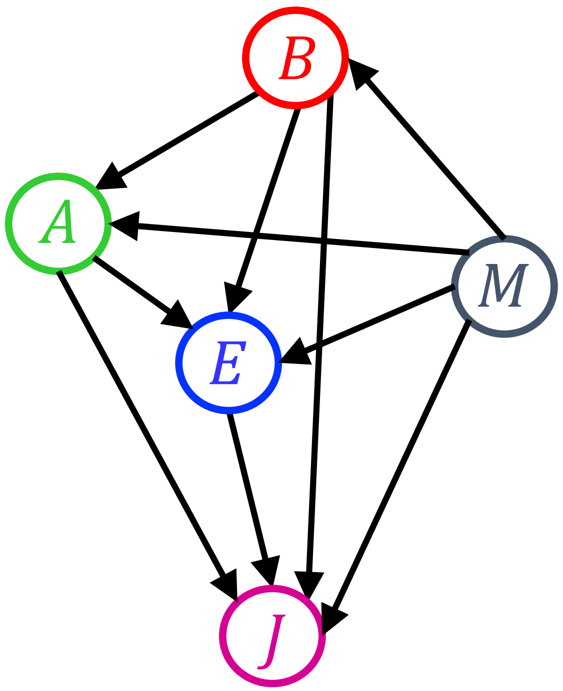
Build Bayes Net Using Chain Rule

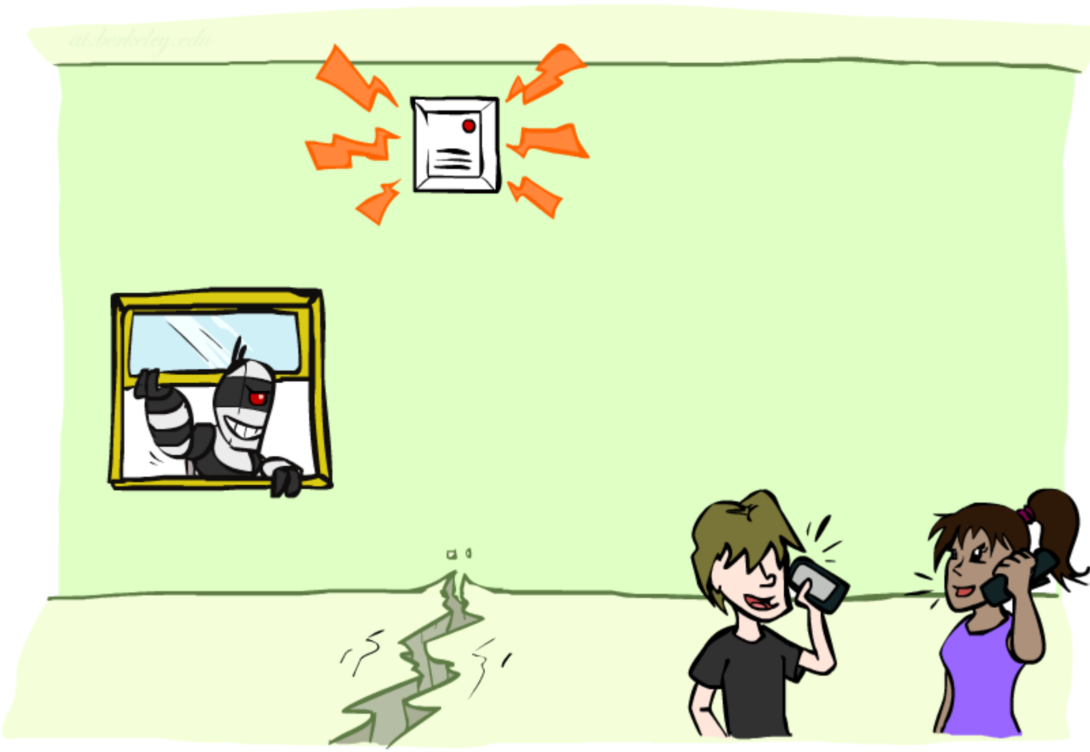
- Binary random variables
  - Fire
  - Smoke
  - Alarm



Build Bayes Net Using Chain Rule 2

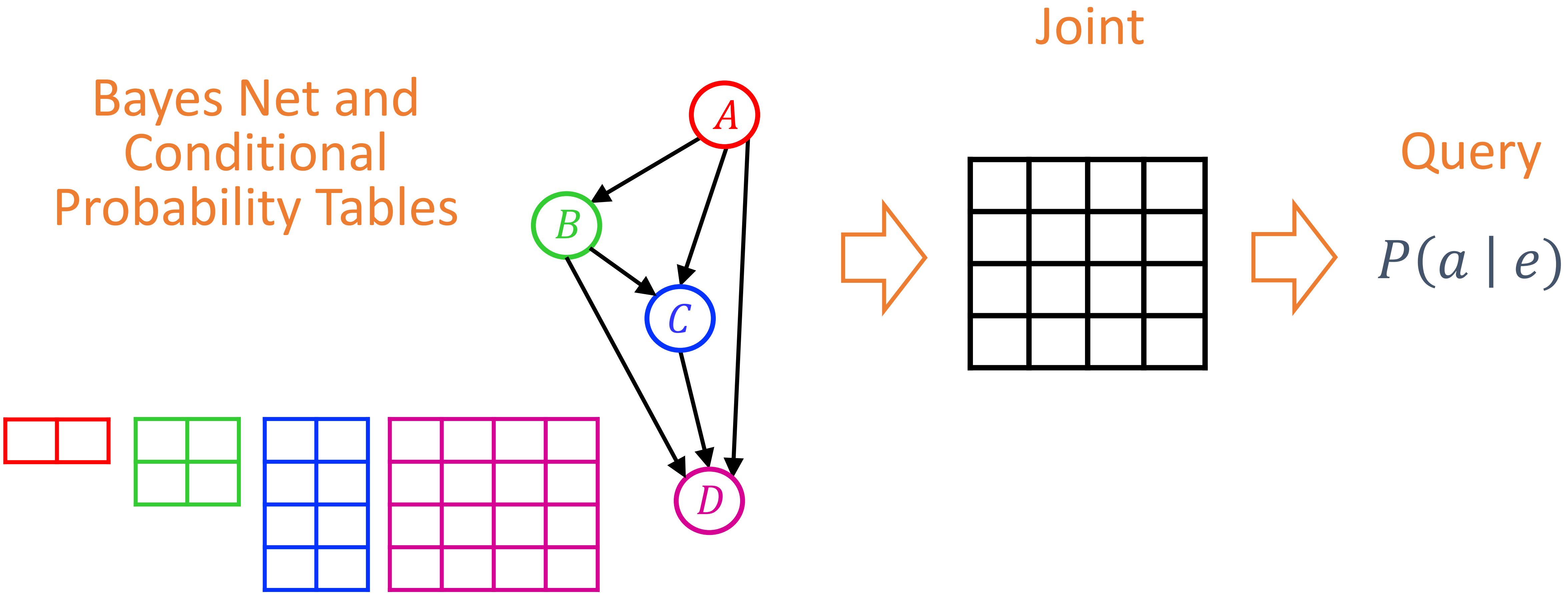
- Variables
  - B: Burglary
  - A: Alarm goes off
  - M: Mary calls
  - J: John calls
  - E: Earthquake!





Given the Bayes net, write the joint distribution?

Answer Any Query from Bayes Net

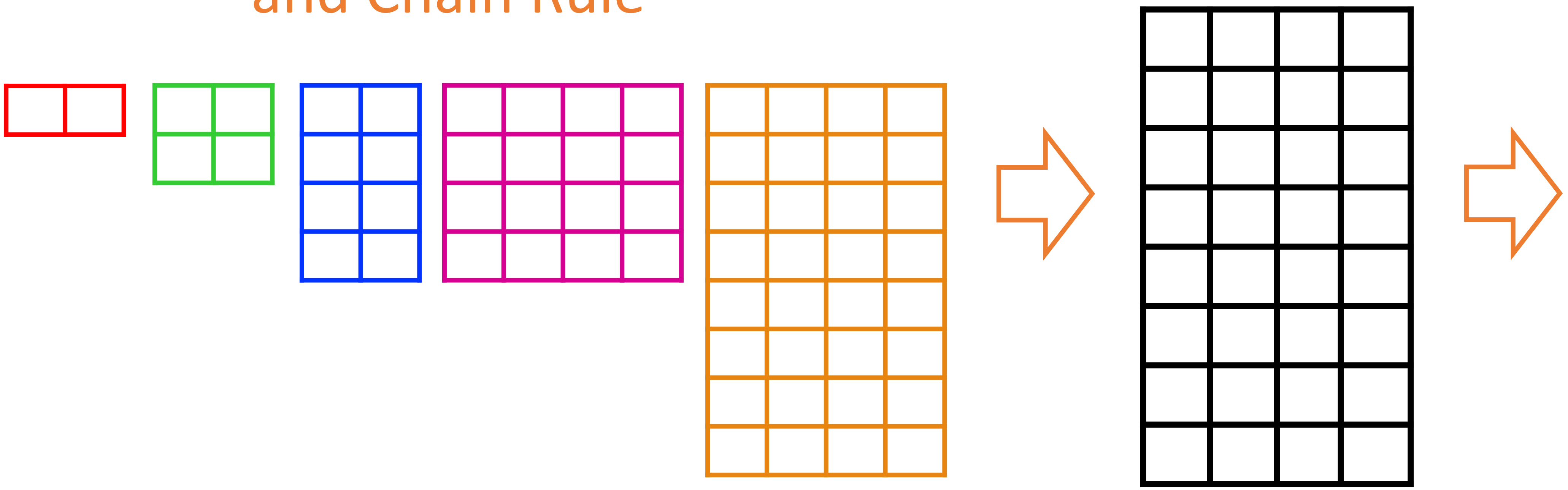


Answer Any Query from Condition Probability Tables

Conditional Probability Tables and Chain Rule

Joint

Query

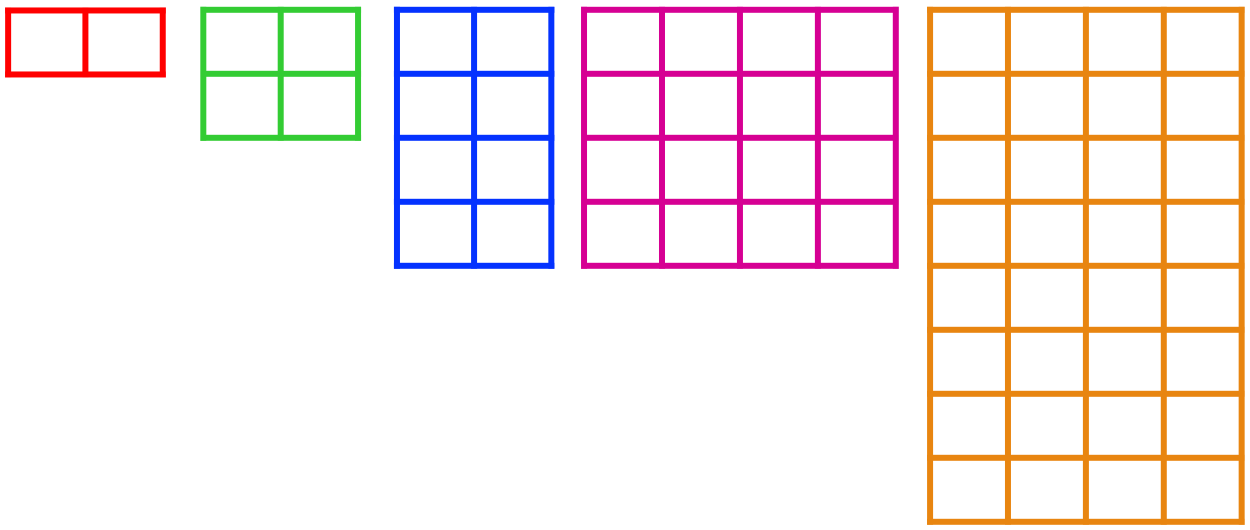


$P(a \mid e)$

$P(A)$ $P(B|A)$ $P(C|A,B)$ $P(D|A,B,C)$ $P(E|A,B,C,D)$

Answer Any Query from Condition Probability Tables 2

Conditional Probability Tables and Chain Rule

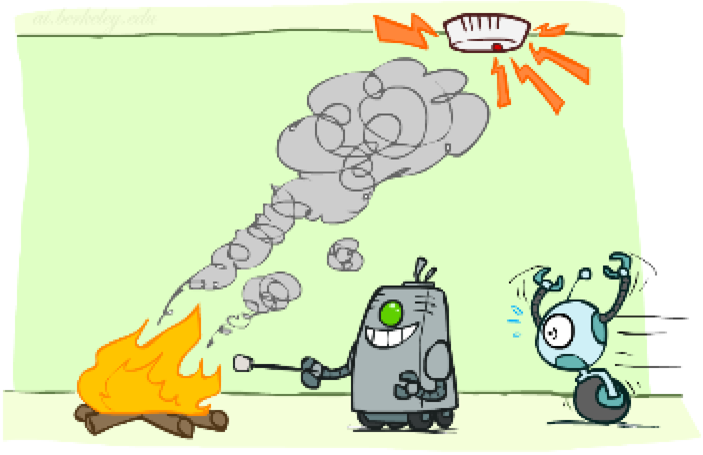


$P(A)\ P(B|A)\ P(C|A,B)\ P(D|A,B,C)\ P(E|A,B,C,D)$

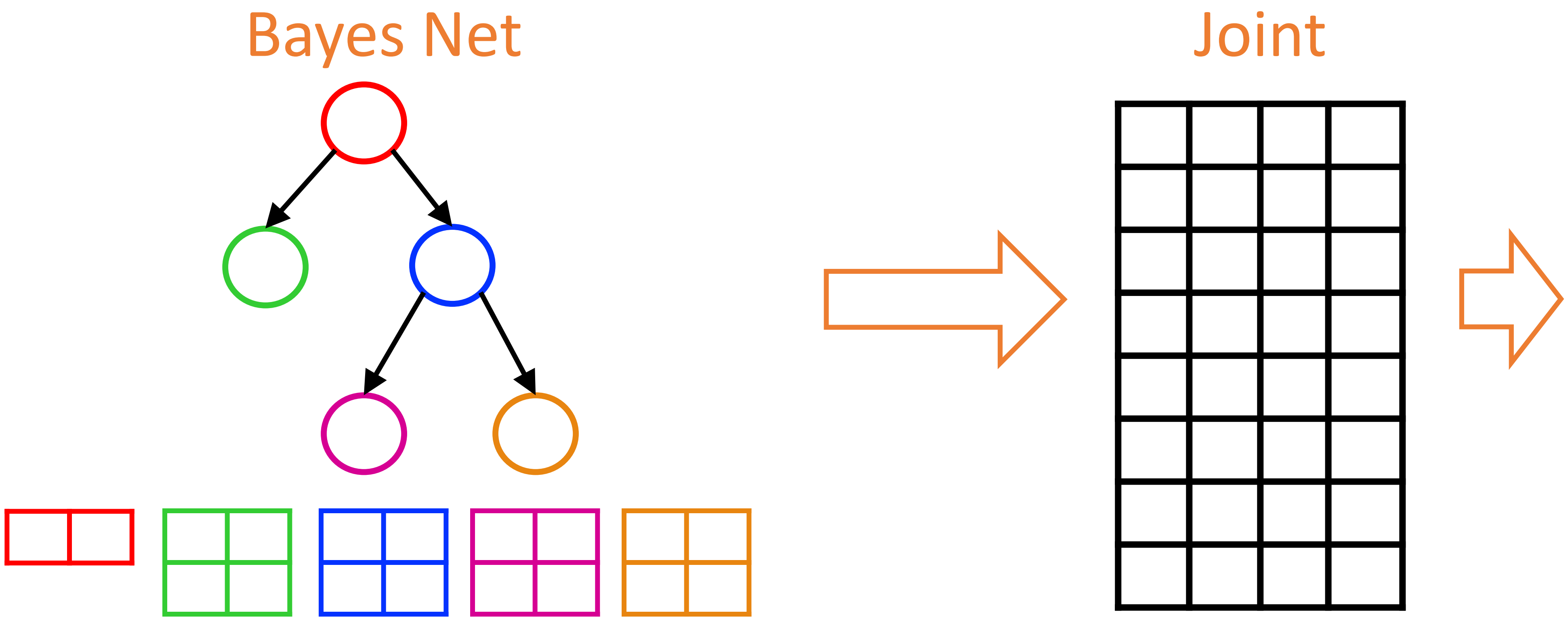
- Problems
  - Huge
    - n variables with d values
    - d^n entries
  - We aren't given the right tables

Do We Need the Full Chain Rule?

- Binary random variables
  - Fire
  - Smoke
  - Alarm



Answer Any Query from Condition Probability Tables



Query

$P(a \mid e)$

$P(A)$ $P(B|A)$ $P(C|A)$ $P(D|C)$ $P(E|C)$

$$P(X_1, \dots, X_N) = \prod_i P(X_i | \, Parents(X_i))$$

Probabilistic Models

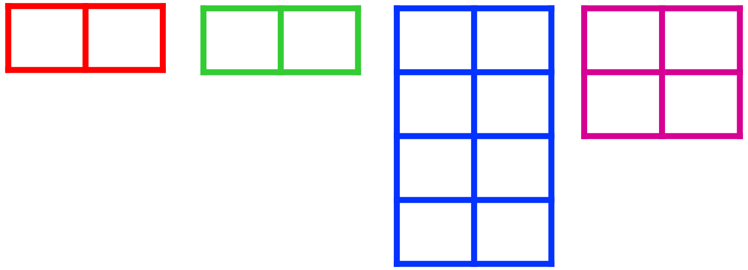
- Models describe how (a portion of) the world works

- Models are always simplifications
  - May not account for every variable
  - May not account for all interactions between variables
  - "All models are wrong; but some are useful."
    – George E. P. Box

- What do we do with probabilistic models?
  - We (or our agents) need to reason about unknown variables, given evidence
  - Example: explanation (diagnostic reasoning)
  - Example: prediction (causal reasoning)
  - Example: value of information

(General) Bayesian Networks

- One node per random variable, DAG
- One conditional probability table (CPT) per node: P(node | *Parents*(node))

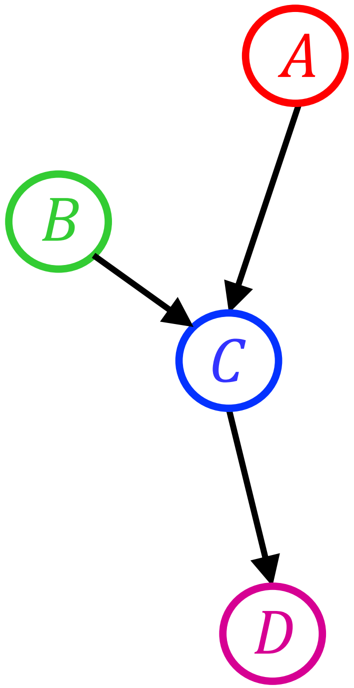


$$P(A,B,C,D) = P(A)\,P(B)\,P(C|A,B)\,P(D|C)$$

Encode joint distributions as product of conditional distributions on each variable

$$P(X_1,\dots,X_N) = \prod_i P(X_i|\,Parents(X_i))$$

Bayes net



Conditional Independence

Independence

- Two variables are *independent* if:

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

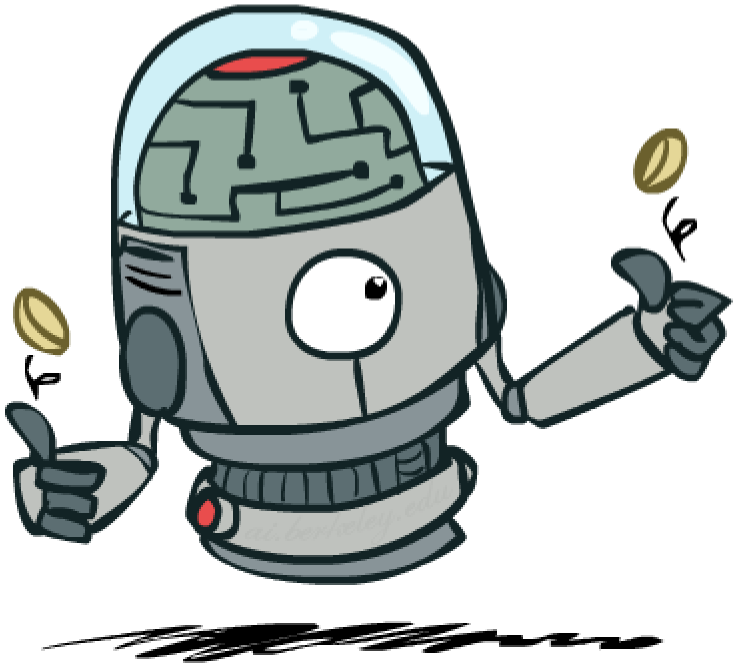
  - This says that their joint distribution *factors* into a product two simpler distributions
  - Another form:

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

  - We write:

$$X \perp\!\!\!\perp Y$$

- Independence is a simplifying *modeling assumption*
  - *Empirical* joint distributions: at best "close" to independent
  - What could we assume for {Weather, Traffic, Cavity, Toothache}?



Example: Independence?

$P(T)$

T	P
hot	0.5
cold	0.5

$P_1(T, W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

$P_2(T, W)$

T	W	P
hot	sun	0.3
hot	rain	0.2
cold	sun	0.3
cold	rain	0.2

$P(W)$

W	P
sun	0.6
rain	0.4

Example: Independence

- N fair, independent coin flips:

$P(X_1)$

H	0.5
T	0.5

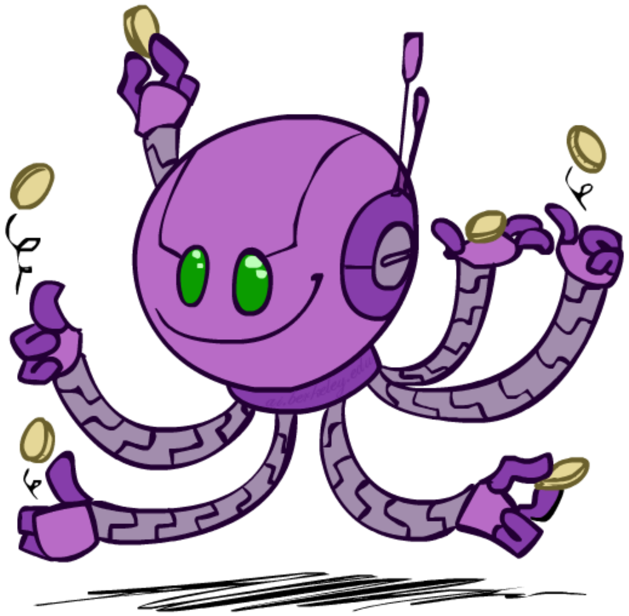
$P(X_2)$

H	0.5
T	0.5

$\cdots$

$P(X_n)$

H	0.5
T	0.5



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

2^n

Conditional Independence

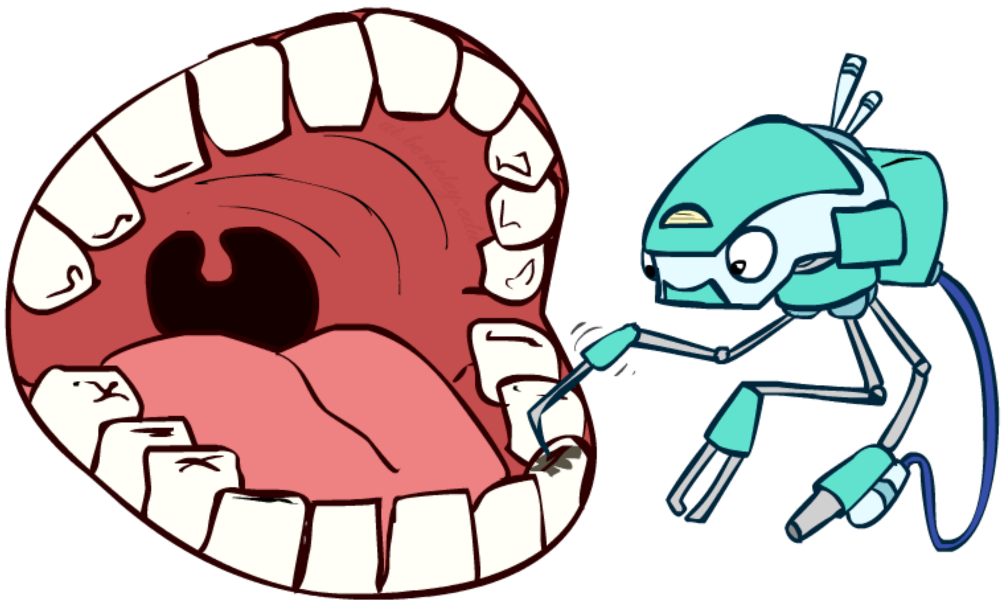
- P(Toothache, Cavity, Catch)

- If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:
  - P(+catch | +toothache, +cavity) = P(+catch | +cavity)

- The same independence holds if I don't have a cavity:
  - P(+catch | +toothache, -cavity) = P(+catch| -cavity)

- Catch is *conditionally independent* of Toothache given Cavity:
  - P(Catch | Toothache, Cavity) = P(Catch | Cavity)

- Equivalent statements:
  - P(Toothache | Catch , Cavity) = P(Toothache | Cavity)
  - P(Toothache, Catch | Cavity) = P(Toothache | Cavity) P(Catch | Cavity)
  - One can be derived from the other easily



Conditional Independence (cont.)

- P(Birds, Sunny, Sunglasses)
- If it is sunny, the probability that birds are out doesn't depend on whether you wear sunglasses:
  - P(+birds | +sunglasses, +sunny) = P(+birds | +sunny)
- The same independence holds if it isn’t sunny:
  - P(+birds | +sunglasses, -sunny) = P(+birds | -sunny)
- Birds is *conditionally independent* of Sunglasses given Sunny:
  - P(Birds | Sunglasses, Sunny) = P(Birds | Sunny)

Conditional Independence (cont.)

- Unconditional (absolute) independence very rare (why?)
- *Conditional independence* is our most basic and robust form of knowledge about uncertain environments.
- 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

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

Conditional Independence (cont.)

- Unconditional (absolute) independence very rare (why?)
- *Conditional independence* is our most basic and robust form of knowledge about uncertain environments.
- X is conditionally independent of Y given Z $\quad X \perp\!\!\!\perp Y | Z$

  if and only if:

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

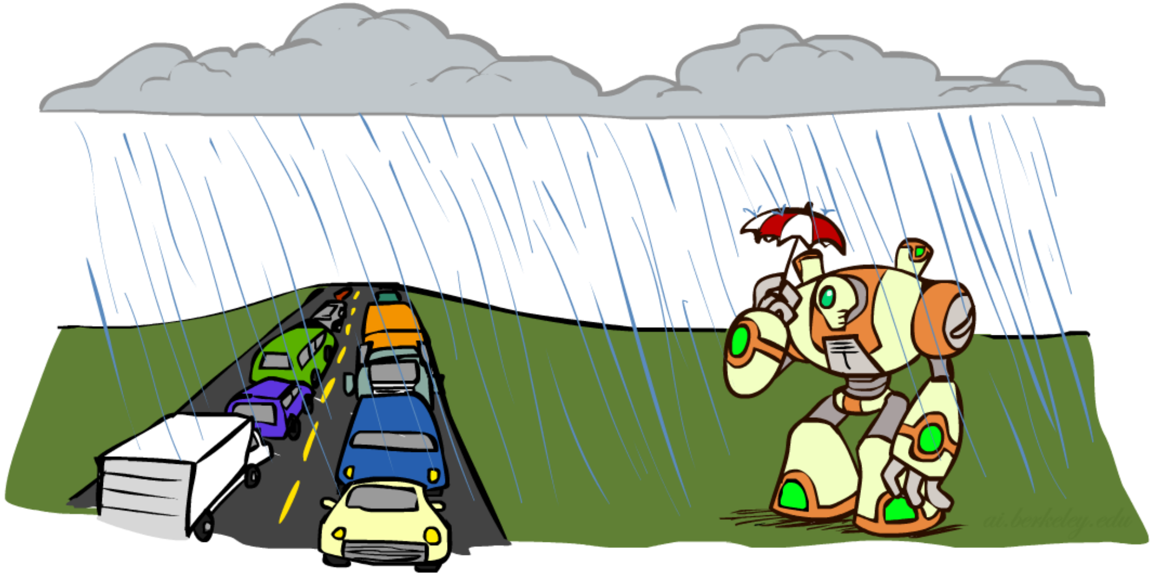
  or, equivalently, if and only if

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

$$\begin{aligned} P(x|z,y) &= \frac{P(x,z,y)}{P(z,y)} \\ &= \frac{P(x,y|z)P(z)}{P(y|z)P(z)} \\ &= \frac{P(x|z)P(y|z)P(z)}{P(y|z)P(z)} \end{aligned}$$

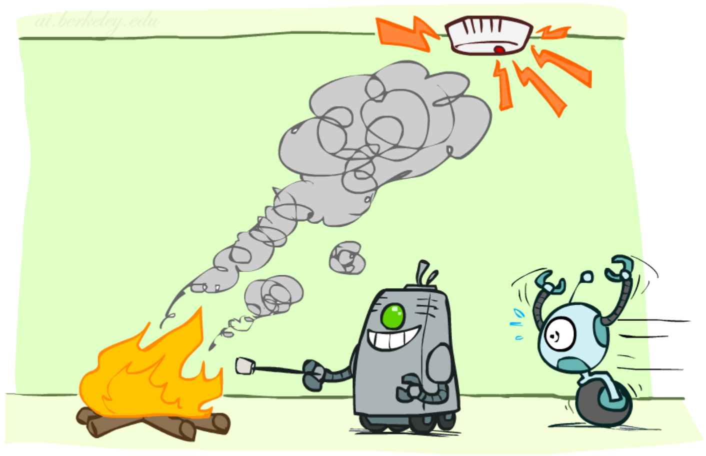
Conditional Independence (cont.)

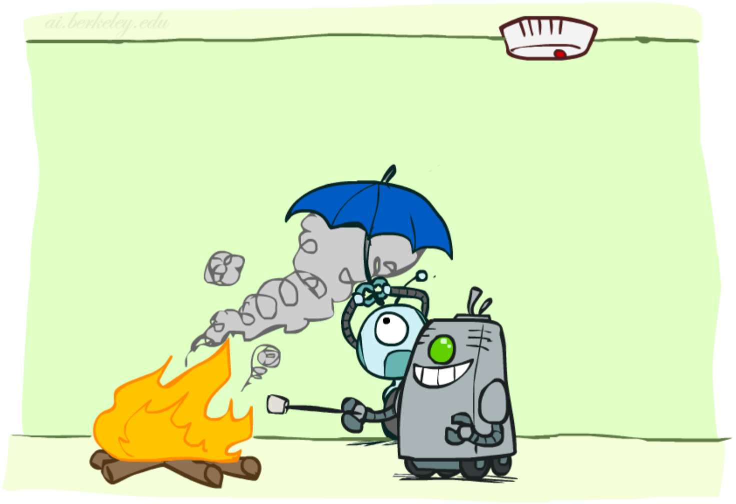
- What about this domain:
  - Traffic
  - Umbrella
  - Raining



Conditional Independence (cont.)

- What about this domain:
  - Fire
  - Smoke
  - Alarm



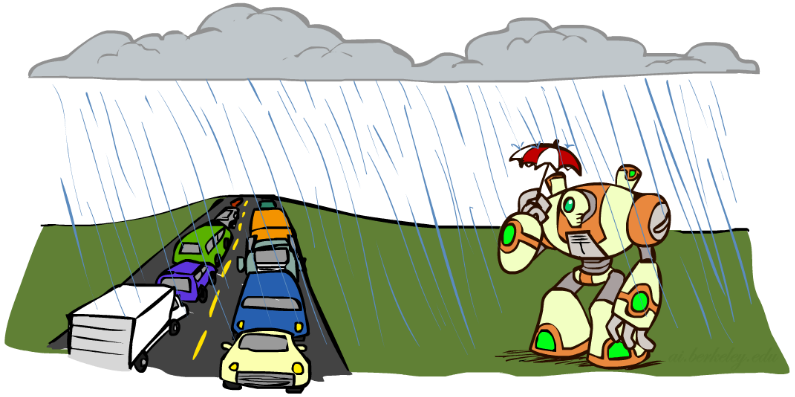


Conditional Independence and the Chain Rule

- 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:

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



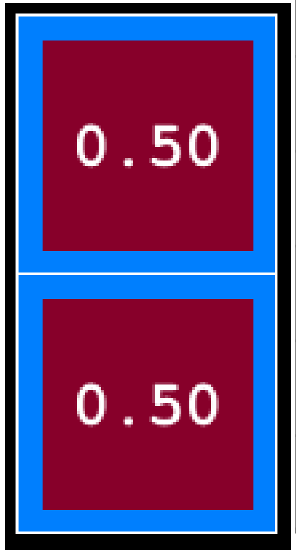
- With assumption of conditional independence:

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

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

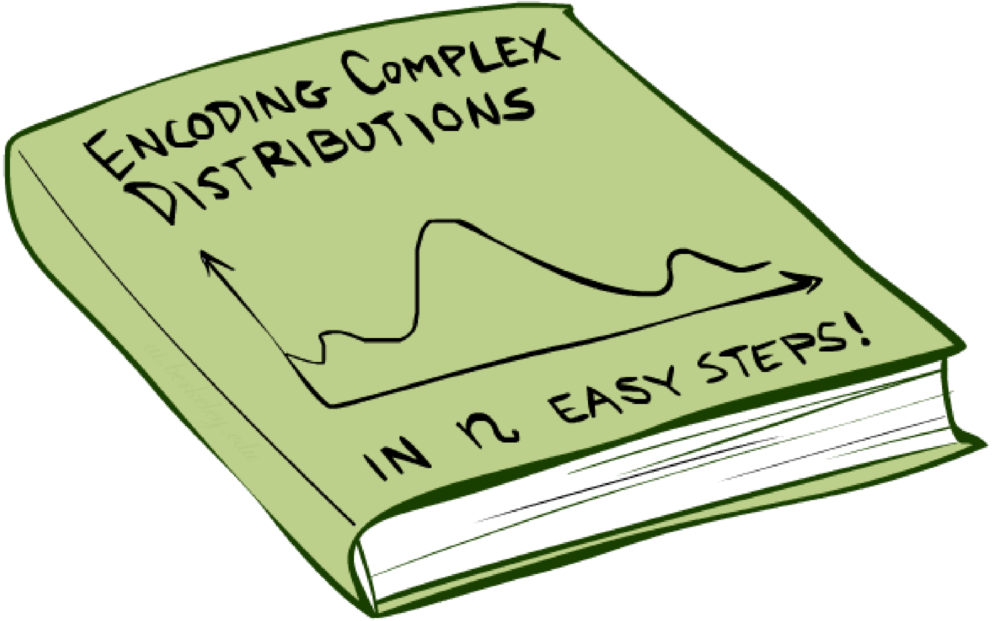
Ghostbusters Chain Rule

- 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:
  $P(+g) = 0.5$
  $P(-g) = 0.5$
  $P(+t \mid +g) = 0.8$
  $P(+t \mid -g) = 0.4$
  $P(+b \mid +g) = 0.4$
  $P(+b \mid -g) = 0.8$

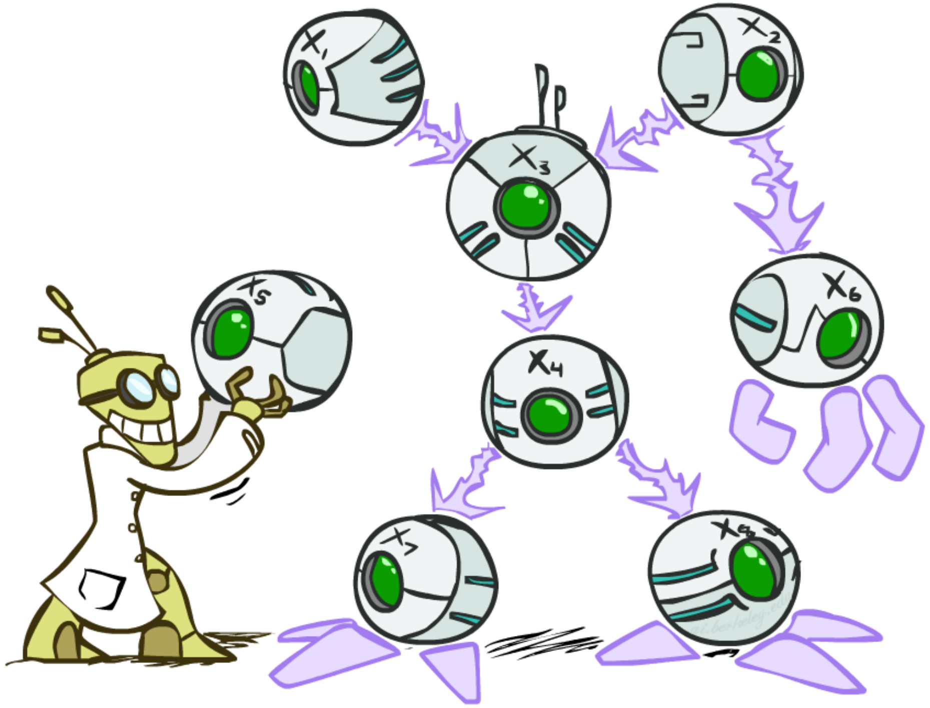


$$P(T,B,G) = P(G)\, P(T|G)\, P(B|G)$$

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

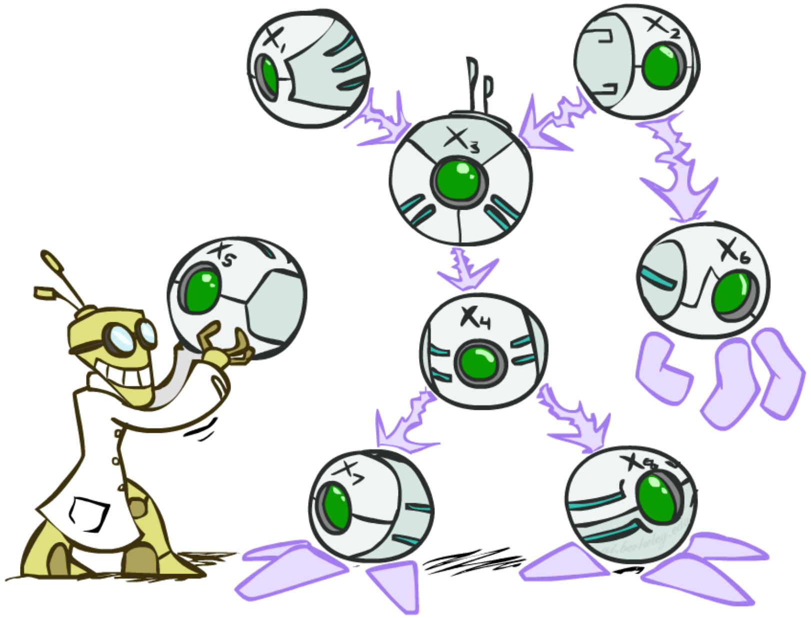


Bayes’ Nets



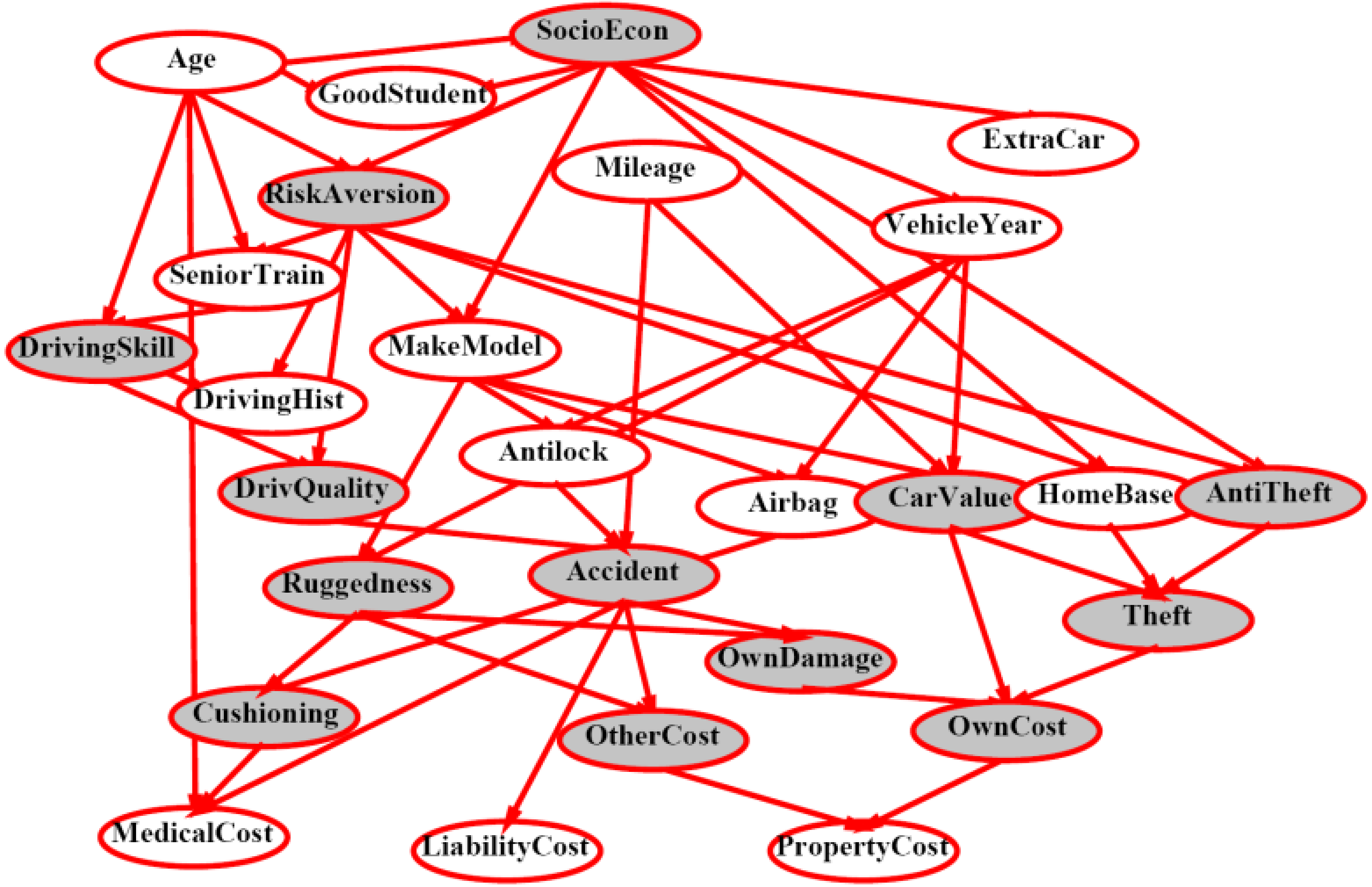
Bayes' Nets: Big Picture

- 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

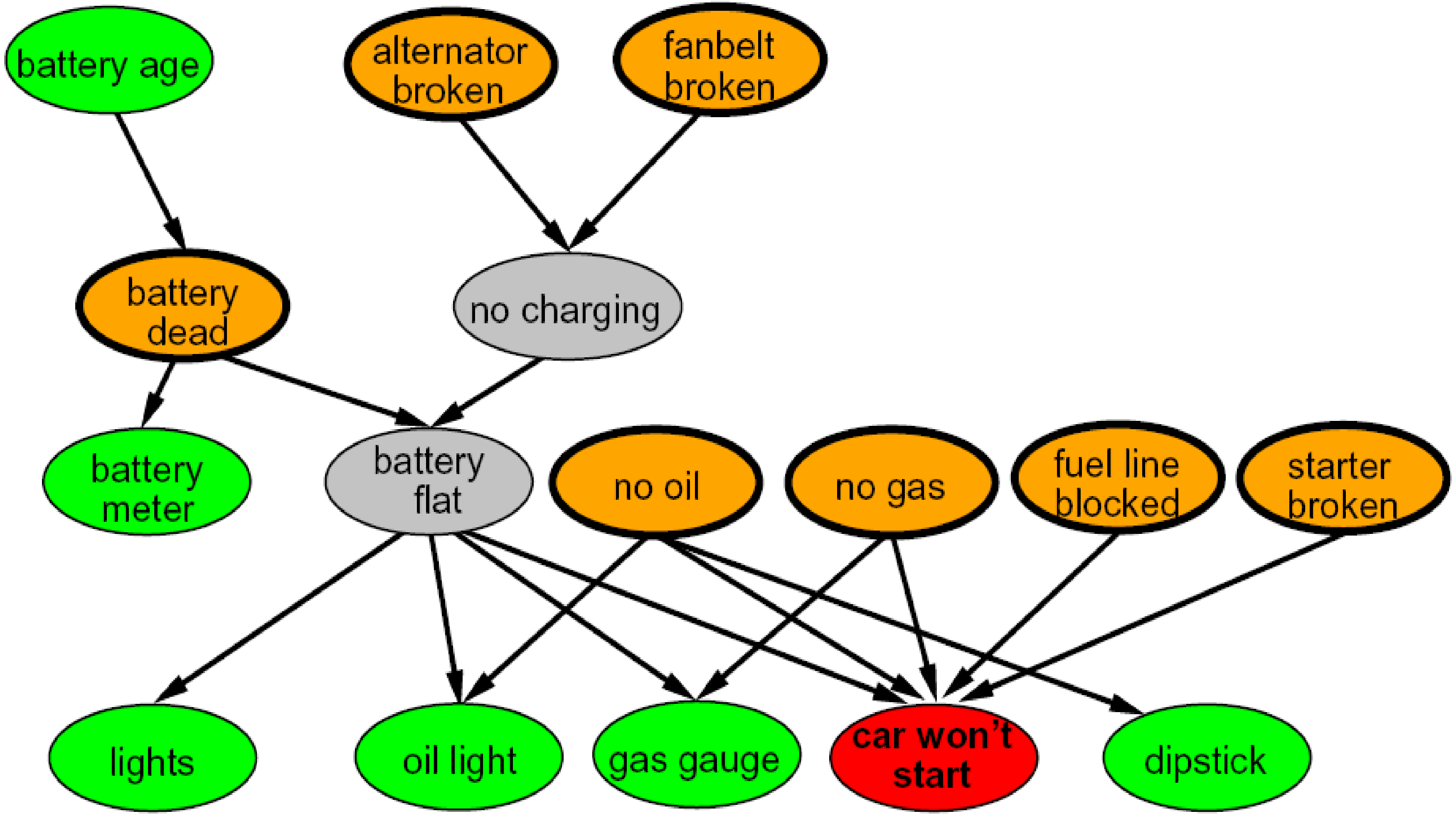


- Bayes' nets: a technique for describing complex joint distributions (models) using simple, local distributions (conditional probabilities)
  - More properly called graphical models
  - We describe how variables locally interact
  - Local interactions chain together to give global, indirect interactions
  - We first look at some examples

Example Bayes' Net: Insurance

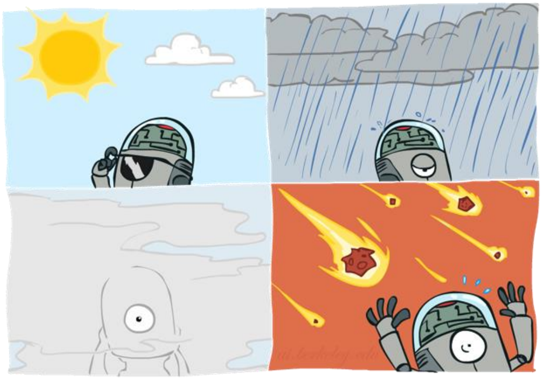


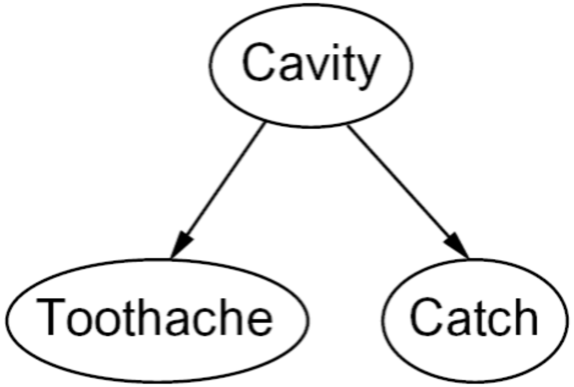
Example Bayes' Net: Car

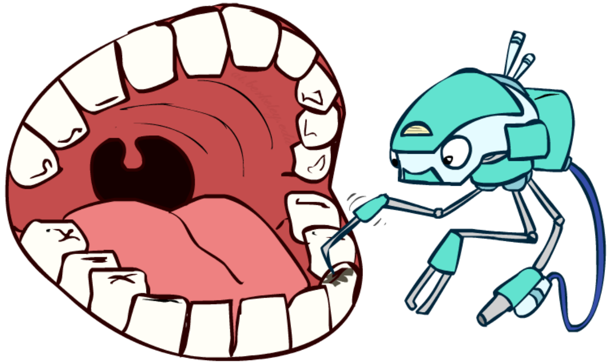


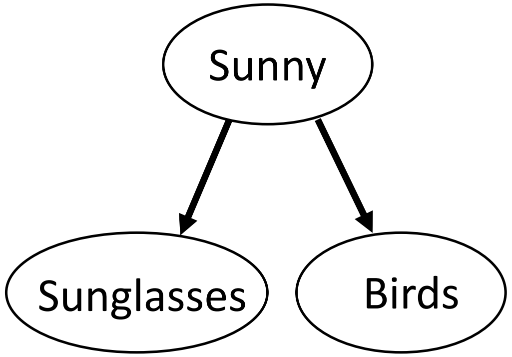
Graphical Model Notation

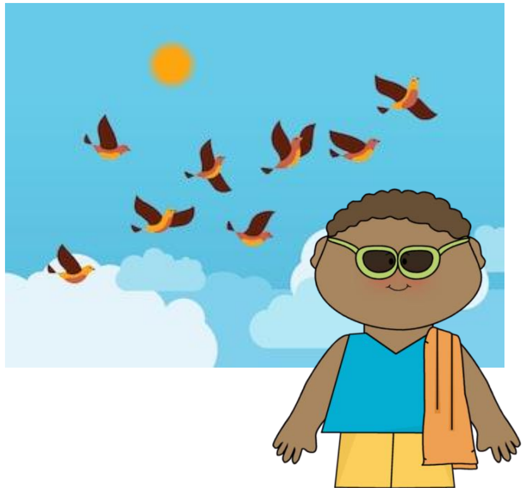
- Nodes: variables (with domains)
  - Can be assigned (observed) or unassigned (unobserved)
- Arcs: interactions
  - Similar to CSP constraints
  - Indicate “direct influence” between variables
  - Formally: encode conditional independence (more later)
- For now: imagine that arrows mean direct causation (in general, they don’t!)





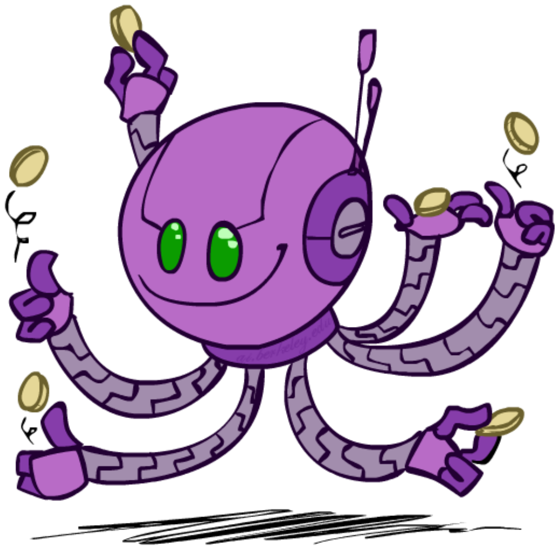






Example: Coin Flips

- N independent coin flips



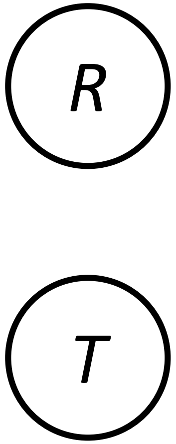
X_1 $\quad$ X_2 $\quad$ $\cdots$ $\quad$ X_n

- No interactions between variables: absolute independence

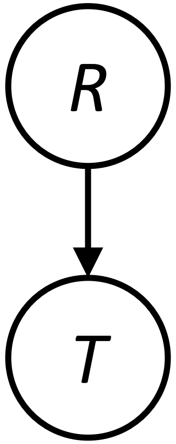
Example: Traffic

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

- Model 1: independence



- Model 2: rain causes traffic

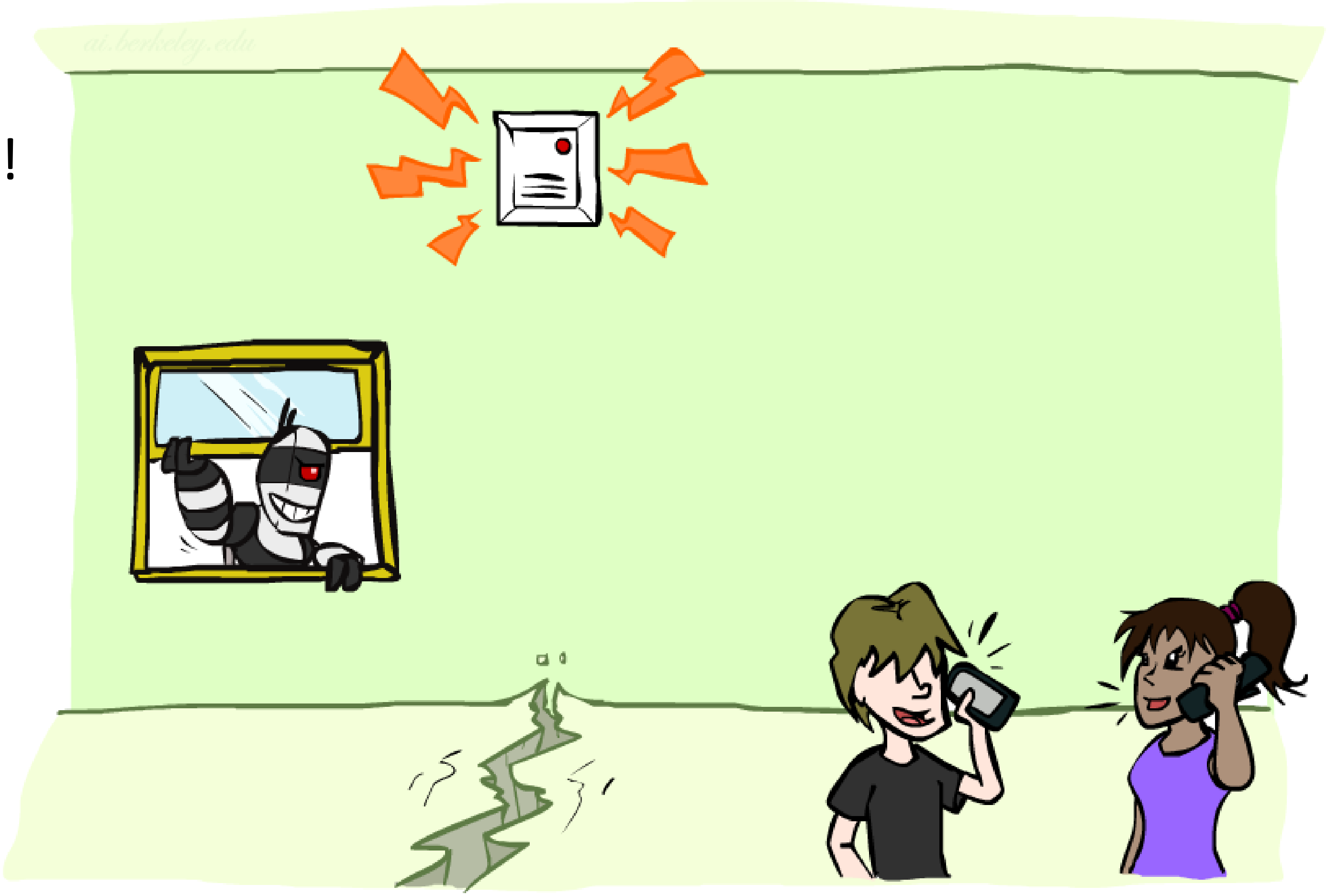


- Why is an agent using model 2 better?

Example: Alarm Network

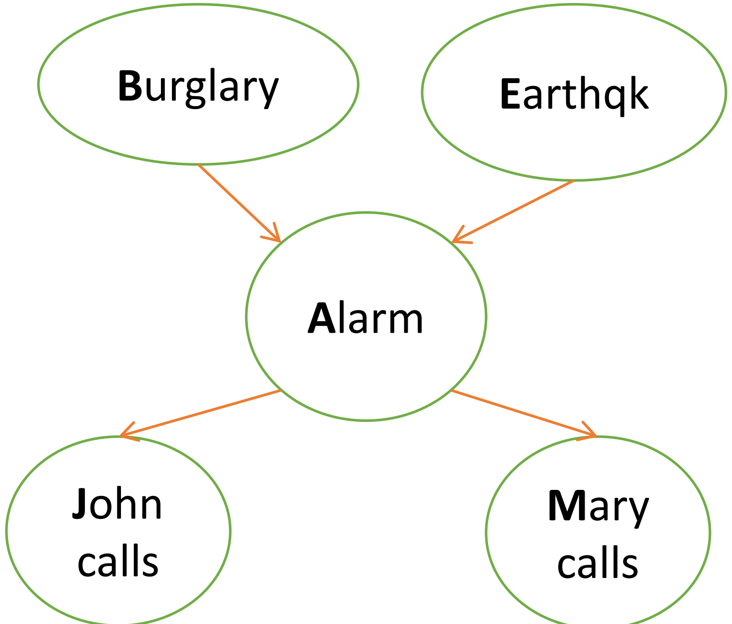
- Let’s build a causal graphical model!

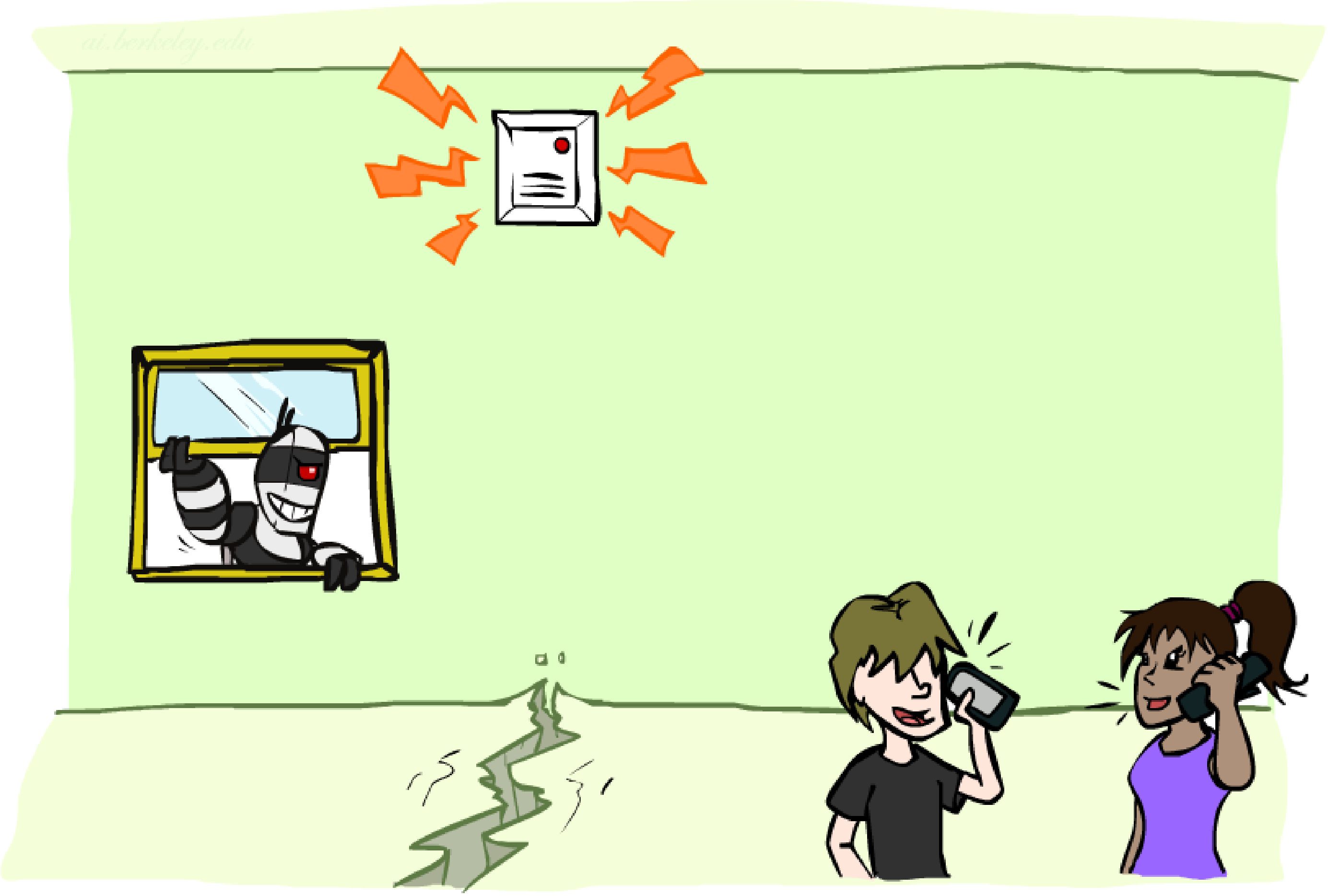
- Variables
  - B: Burglary
  - A: Alarm goes off
  - M: Mary calls
  - J: John calls
  - E: Earthquake!



Example: Alarm Network 2

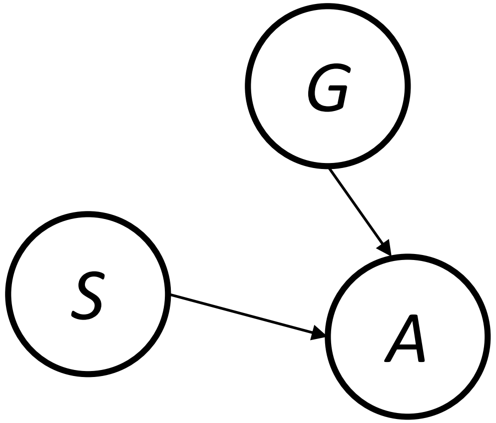
- Variables
  - B: Burglary
  - A: Alarm goes off
  - M: Mary calls
  - J: John calls
  - E: Earthquake!





Example: Humans

- G: human's goal / human's reward parameters
- S: state of the physical world
- A: human's action



Example: Traffic II

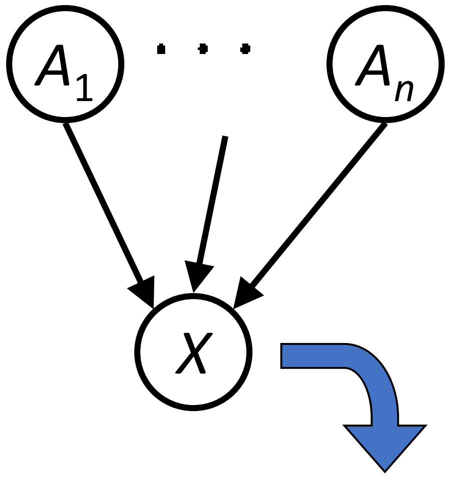
- Variables
  - T: Traffic
  - R: It rains
  - L: Low pressure
  - D: Roof drips
  - B: Ballgame
  - C: Cavity

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



$$P(X|A_1 \ldots A_n)$$

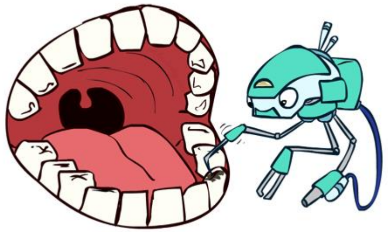
A Bayes net = Topology (graph) + Local Conditional Probabilities

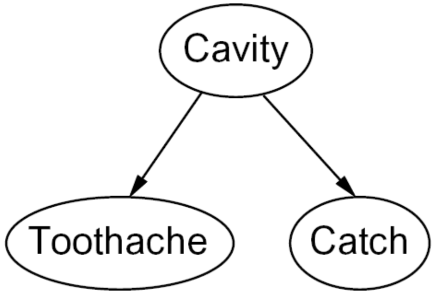
Probabilities in BNs

- 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:

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

  - Example:





$$P(\mathit{+cavity},\ \mathit{+catch},\ \mathit{-toothache})$$

=P(-toothache|+cavity)P(+catch|+cavity)P(+cavity)

Probabilities in BNs 2

- Why are we guaranteed that setting

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

  results in a proper joint distribution?

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

  → Consequence:

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

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

Example: Coin Flips

X_1 X_2 $\cdots$ X_n

$P(X_1)$

h	0.5
t	0.5

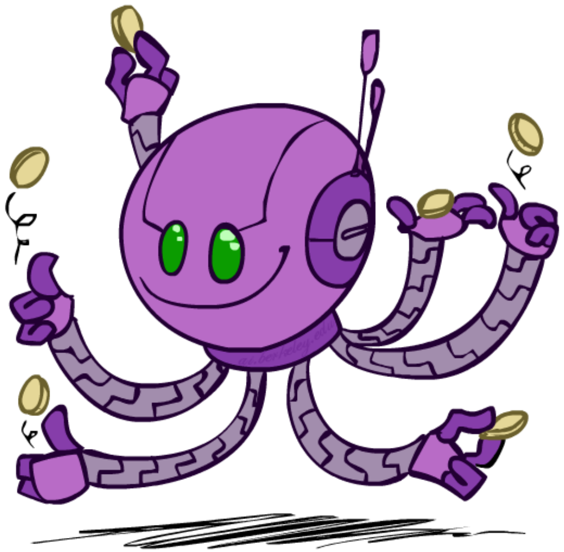
$P(X_2)$

h	0.5
t	0.5

$\cdots$

$P(X_n)$

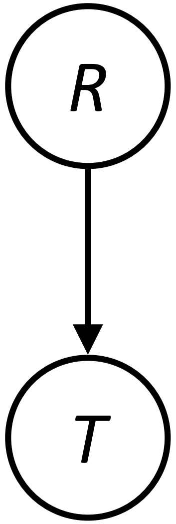
h	0.5
t	0.5



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

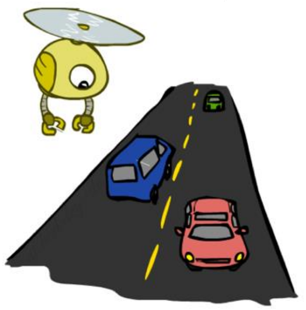
$P(T|R)$

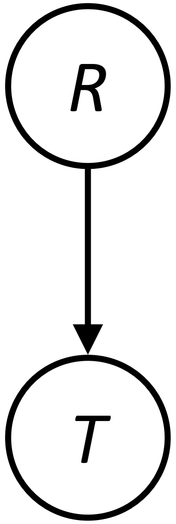
+r	+t	3/4
	-t	1/4
-r	+t	1/2
	-t	1/2

$P(+r, -t) =$ P(+r)P(-t|+r) = (1/4) *(1/4)

Example: Traffic 2

- Causal direction



$P(R)$

+r	1/4
-r	3/4

$P(T|R)$

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

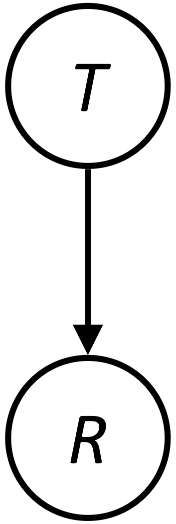
-r	+t	1/2
	-t	1/2

$P(T, R)$

+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
	-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

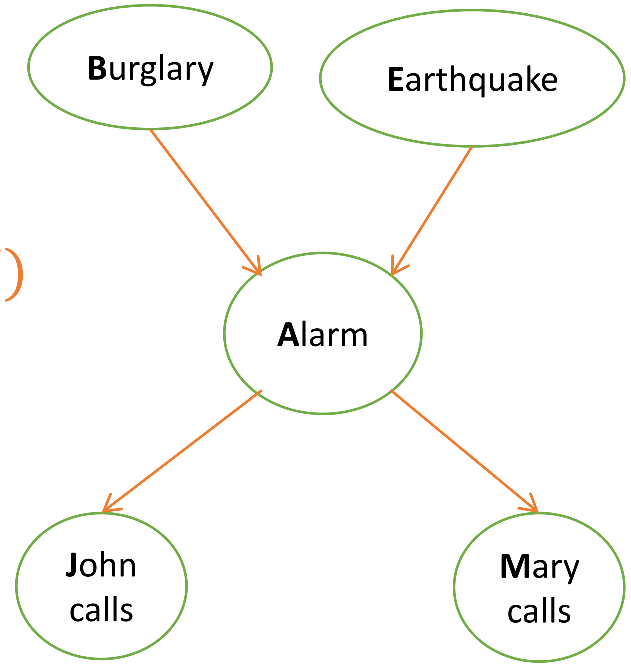
Example: Alarm Network

- Joint distribution factorization example

- Generic chain rule
  - $P(X_1 \ldots X_2) = \prod_i P(X_i | X_1 \ldots X_{i-1})$

$$P(B,E,A,J,M) = P(B)\,P(E|B)\,P(A|B,E)\,P(J|B,E,A)\,P(M|B,E,A,J)$$

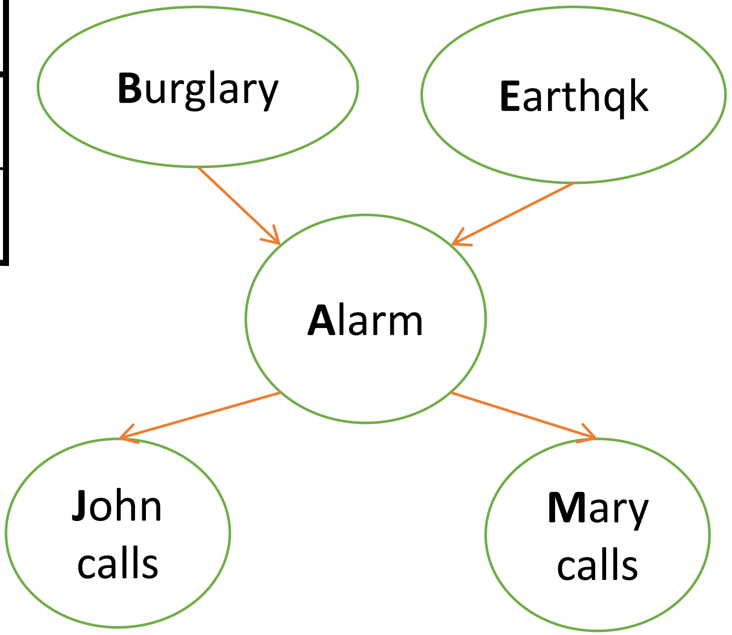
$$P(B,E,A,J,M) = P(B)\,P(E) \quad P(A|B,E)\,P(J|A) \qquad P(M|A)$$



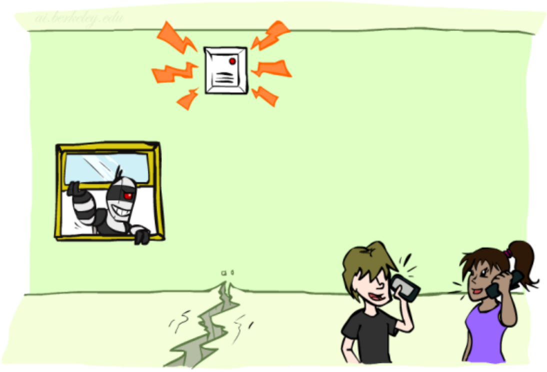
- Bayes nets
  - $P(X_1 \ldots X_2) = \prod_i P(X_i | Parents(X_i))$

Example: Alarm Network

B	P(B)
+b	0.001
-b	0.999



E	P(E)
+e	0.002
-e	0.998



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

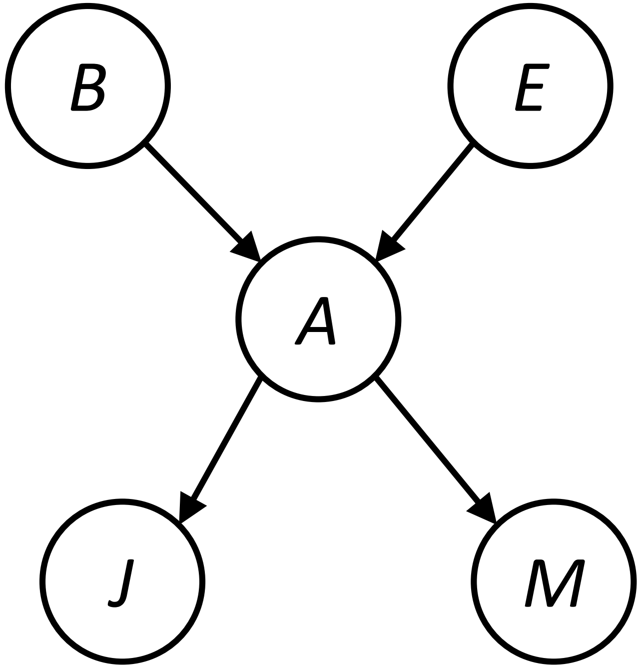
A	M	P(M|A)
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99

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

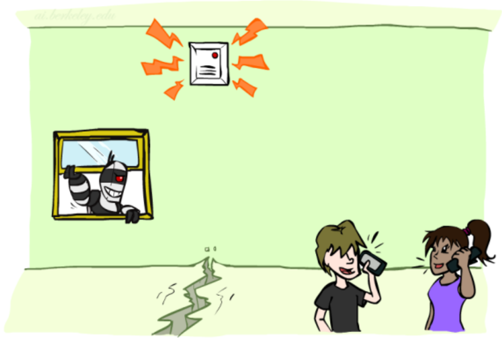
P(M|A)P(J|A)P(A|B,E)

Example: Alarm Network 2

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

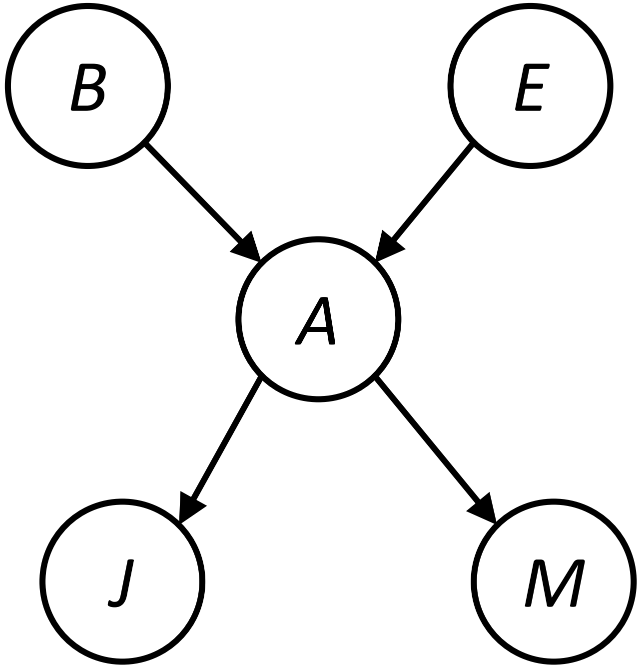
A	M	P(M|A)
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99

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

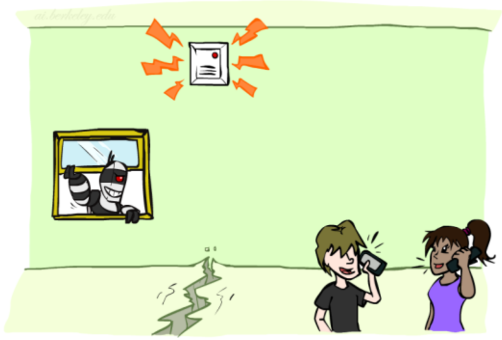
$$P(+b, -e, +a, -j, +m) =$$

Example: Alarm Network 3

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

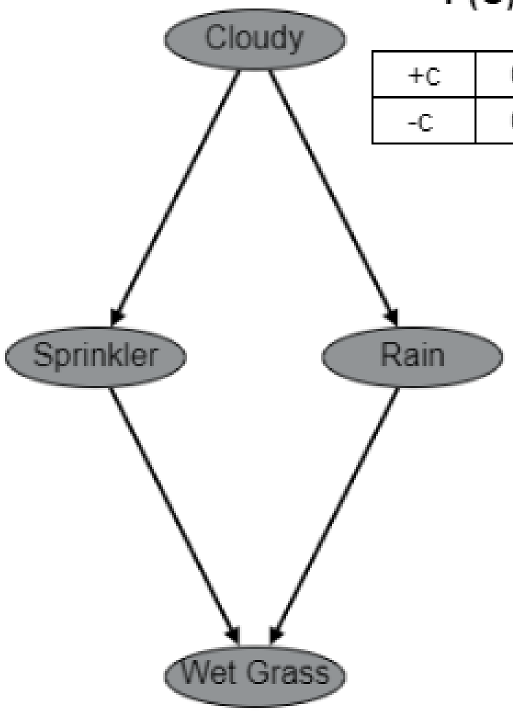
B	E	A	P(A|B,E)
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999

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

Quiz

- Compute $P(-c, +s, -r, +w)$

A. 0.0
B. 0.0004
C. 0.001
D. 0.036
E. 0.18
F. 0.198
G. 0.324



P(C)

+c	0.5
-c	0.5

P(S | C)

+c	+s	0.1
	-s	0.9
-c	+s	0.5
	-s	0.5

P(R | C)

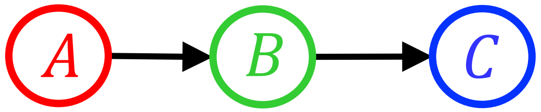
+c	+r	0.8
	-r	0.2
-c	+r	0.2
	-r	0.8

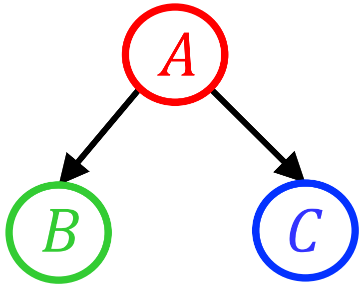
P(W | S, R)

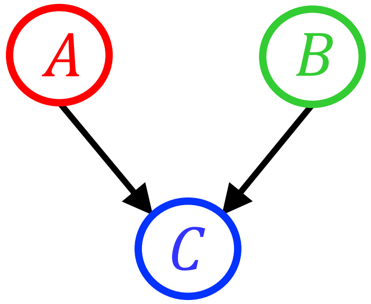
+s	+r	+w	0.99
		-w	0.01
	-r	+w	0.9
		-w	0.1
-s	+r	+w	0.9
		-w	0.1
	-r	+w	0.99
		-w	0.01

Quiz 2

- Match the product of CPTs to the Bayes net.



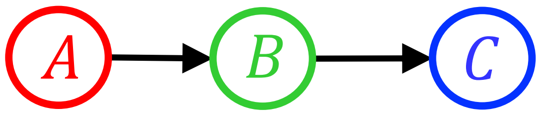


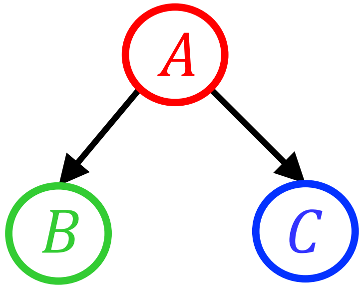


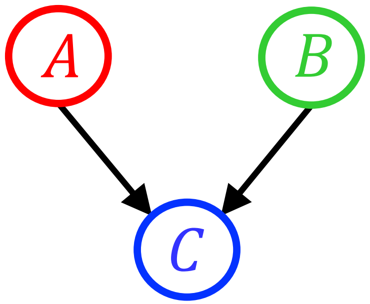
- I. $P(A)\,P(B|A)\,P(C|B)$ $P(A)\,P(B|A)\,P(C|A)$ $P(A)\,P(B)\,P(C|A,B)$
- II. $P(A)\,P(B|A)\,P(C|B)$ $P(A)\,P(B|A)\,P(C|B)$ $P(A)\,P(B|A)\,P(C|A)$
- III. $P(A)\,P(B|A)\,P(C|A)$ $P(A)\,P(B)\,P(C|A,B)$ $P(A)\,P(B|A)\,P(C|B)$

Conditional Independence Semantics

- For the following Bayes nets, write the joint $P(A, B, C)$
    1. Using the chain rule (with top-down order A,B,C)
    2. Using Bayes net semantics (product of CPTs)

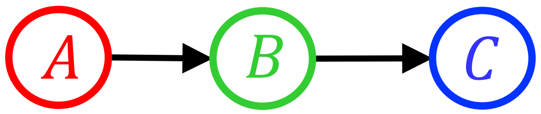






Conditional Independence Semantics 2

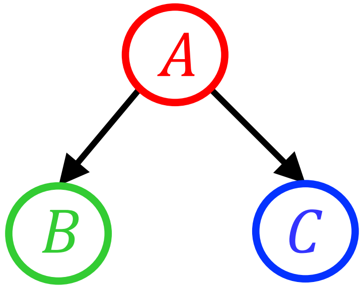
- For the following Bayes nets, write the joint $P(A, B, C)$
  1. Using the chain rule (with top-down order A,B,C)
  2. Using Bayes net semantics (product of CPTs)



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

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

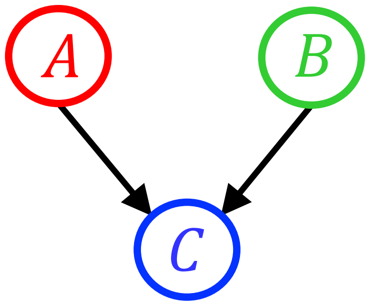
Assumption:
$P(C|A,B) = P(C|B)$
C is independent from A given B



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

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

Assumption:
$P(C|A,B) = P(C|A)$
C is independent from B given A



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

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

Assumption:
$P(B|A) = P(B)$
A is independent from B given { }

Causal Chains

- This configuration is a "causal chain"

X: Low pressure Y: Rain Z: Traffic

$$P(x, y, z) = P(x)P(y|x)P(z|y)$$

- Guaranteed X independent of Z ?
- *No!*
  - One example set of CPTs for which X is not independent of Z is sufficient to show this independence is not guaranteed.
  - Example:
    - Low pressure causes rain causes traffic, high pressure causes no rain causes no traffic
    - In numbers:

      P(+y | +x) = 1, P(-y | - x) = 1,
      P(+z | +y) = 1, P(-z | -y) = 1

Causal Chains 2

- This configuration is a "causal chain"

X: Low pressure

Y: Rain

Z: Traffic

$$P(x, y, z) = P(x)P(y|x)P(z|y)$$

- Guaranteed X independent of Z given Y?

$$P(z|x, y) = \frac{P(x, y, z)}{P(x, y)}$$

$$= \frac{P(x)P(y|x)P(z|y)}{P(x)P(y|x)}$$

$$= P(z|y)$$

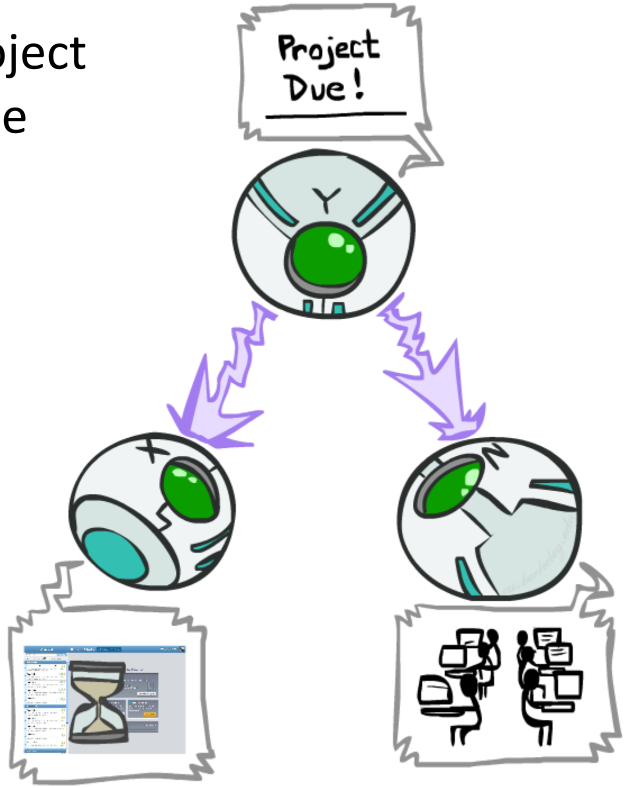
Yes!

- Evidence along the chain "blocks" the influence

Common Causes

- This configuration is a “common cause”

Y: Project due



X: Forums busy

Z: Lab full

$$P(x,y,z) = P(y)P(x|y)P(z|y)$$

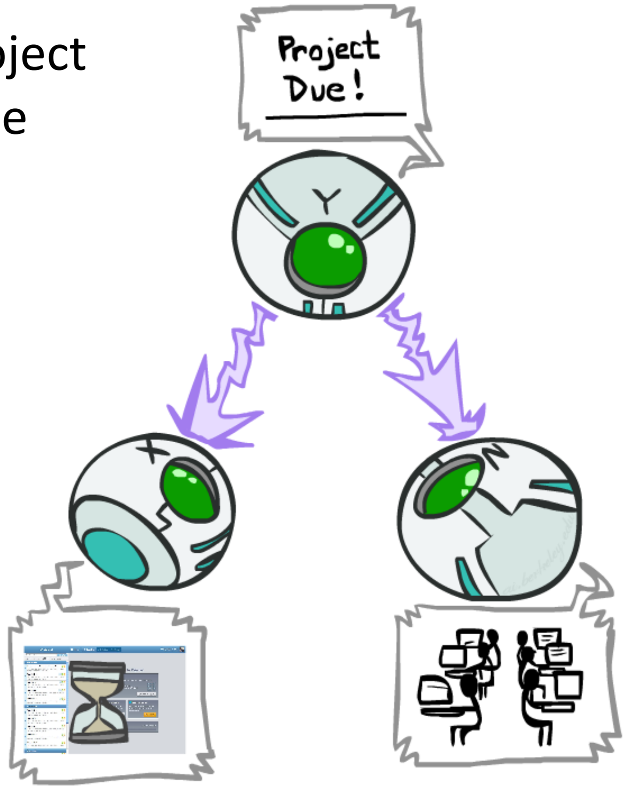
- Guaranteed X independent of Z ?
- *No!*
  - One example set of CPTs for which X is not independent of Z is sufficient to show this independence is not guaranteed.
  - Example:
    - Project due causes both forums busy and lab full
    - In numbers:

P(+x | +y) = 1, P(-x | -y) = 1,
P(+z | +y) = 1, P(-z | -y) = 1

Common Cause 2

- This configuration is a "common cause"

Y: Project due



X: Forums busy

Z: Lab full

$$P(x, y, z) = P(y)P(x|y)P(z|y)$$

- Guaranteed X and Z independent given Y?

$$P(z|x, y) = \frac{P(x, y, z)}{P(x, y)}$$

$$= \frac{P(y)P(x|y)P(z|y)}{P(y)P(x|y)}$$

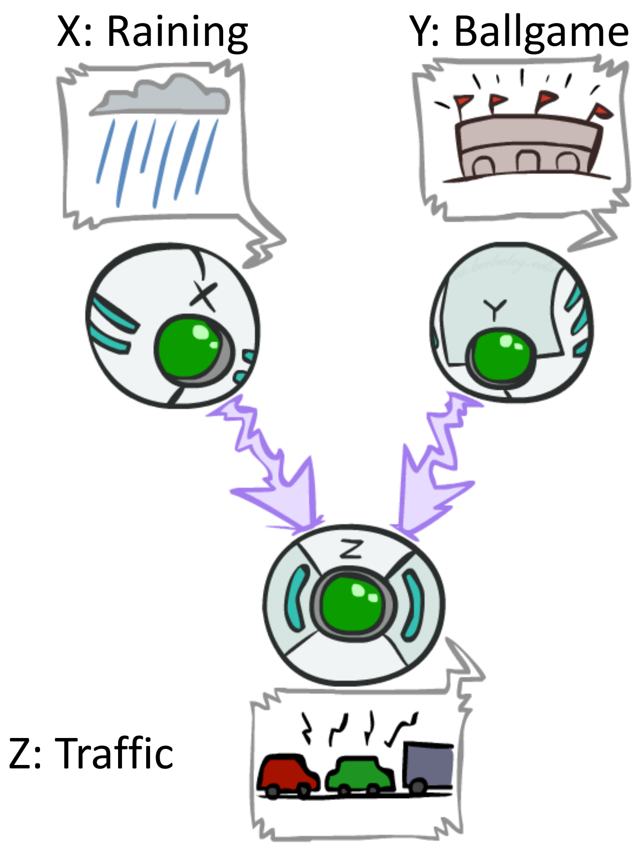
$$= P(z|y)$$

Yes!

- Observing the cause blocks influence between effects

Common Effect

- Last configuration: two causes of one effect (v-structures)



- Are X and Y independent?
  - *Yes*: the ballgame and the rain cause traffic, but they are not correlated
- Proof:

$$P(x,y) = \sum_z P(x,y,z)$$

$$= \sum_z P(x)P(y)P(z|x,y)$$

$$= P(x)P(y)\sum_z P(z|x,y)$$

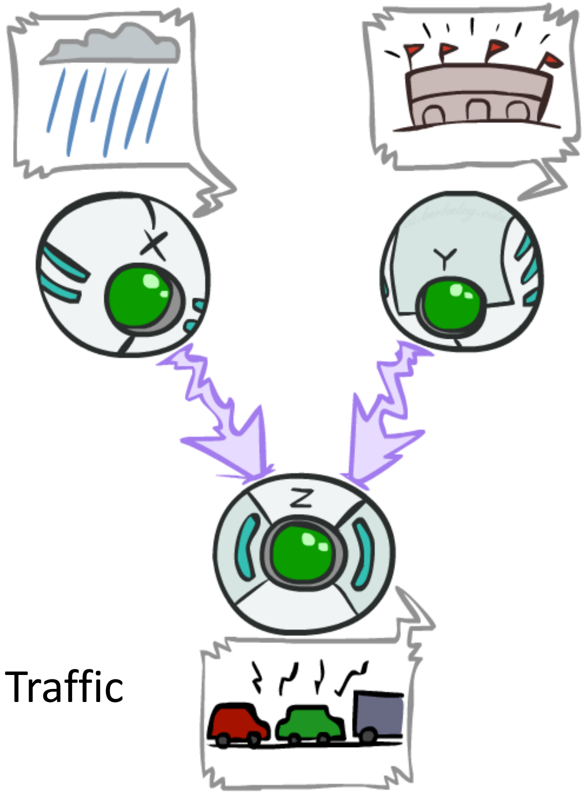
$$= P(x)P(y)$$

Common Effect 2

- Last configuration: two causes of one effect (v-structures)

X: Raining

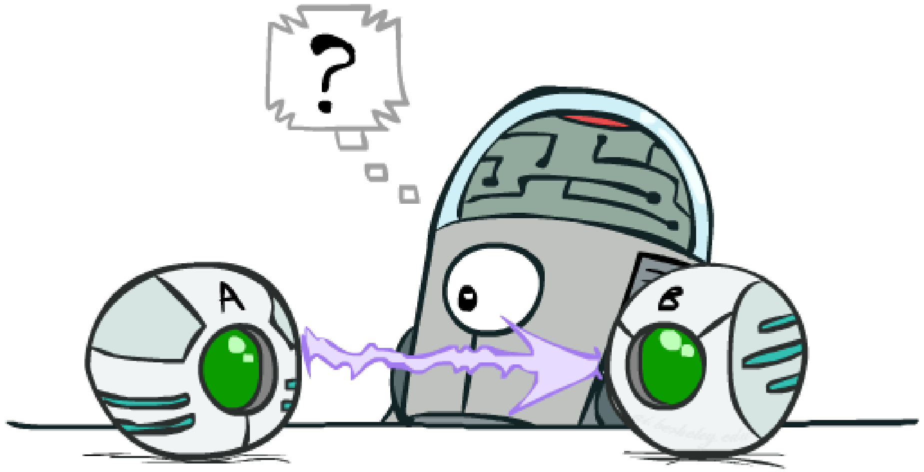
Y: Ballgame



Z: Traffic

- Are X and Y independent?
  - *Yes*: the ballgame and the rain cause traffic, but they are not correlated
  - (Proved previously)
- Are X and Y independent given Z?
  - *No*: seeing traffic puts the rain and the ballgame in competition as explanation.
- This is backwards from the other cases
  - Observing an effect activates influence between possible causes

Causality?



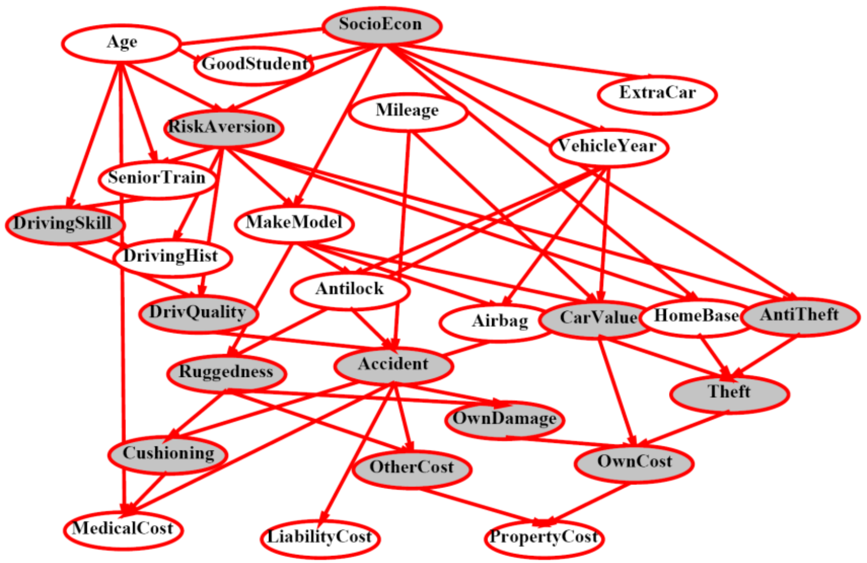
- When Bayes' nets reflect the true causal patterns:
  - Often simpler (nodes have fewer parents)
  - Often easier to think about
  - Often easier to elicit from experts
- BNs need not actually be causal
  - Sometimes no causal net exists over the domain (especially if variables are missing)
  - E.g. 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

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

Bayes Nets: Independence

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

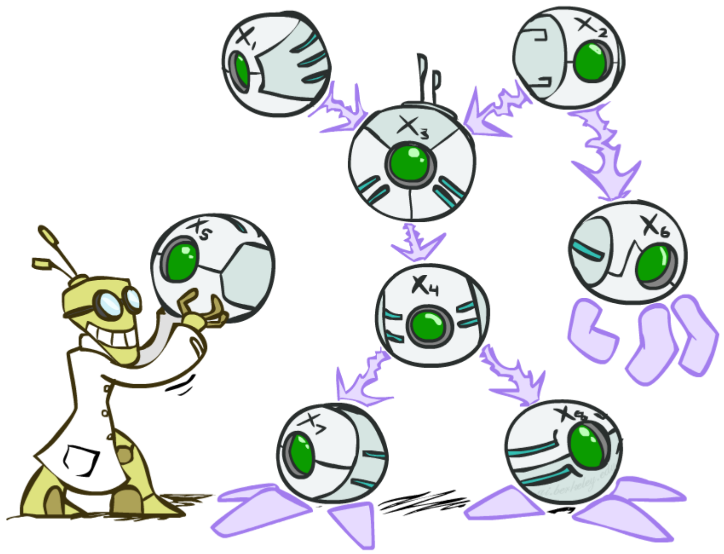
$$P(X|a_1 \ldots a_n)$$

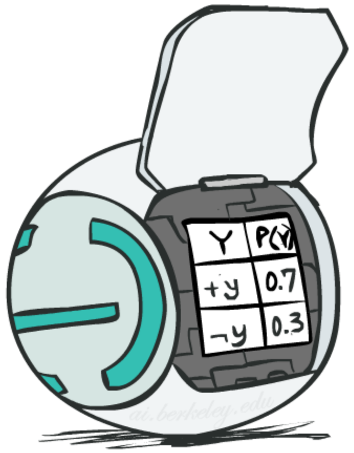
- Bayes' nets implicitly encode joint distributions

  - As a product of local conditional distributions

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

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





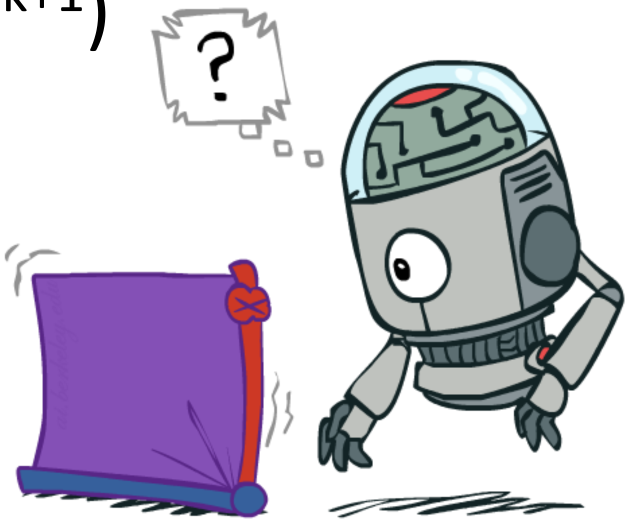
Size of a Bayes Net

- 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?

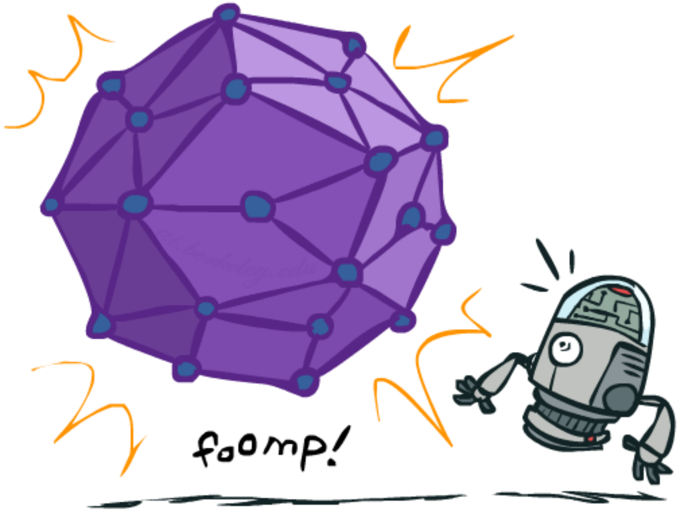
  $O(N * 2^{k+1})$



- Both give you the power to calculate

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

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

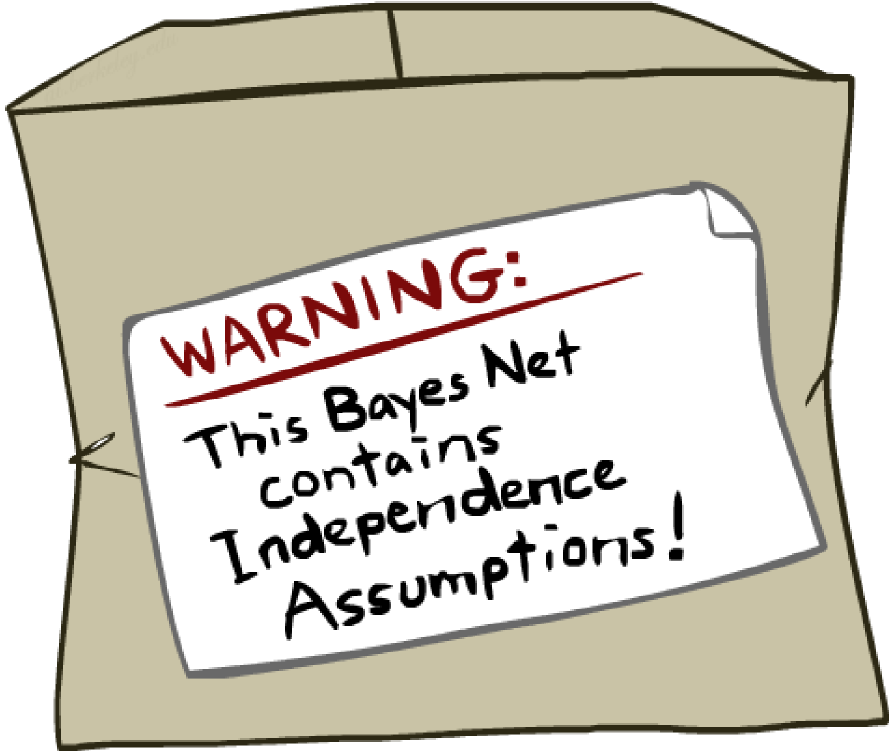


Bayes Nets: Assumptions

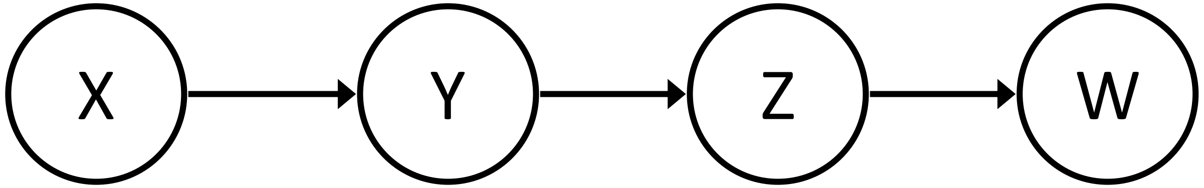
- Assumptions we are required to make to define the Bayes net when given the graph:

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

- Beyond those "chain rule → Bayes net" conditional independence assumptions
  - Often additional conditional independences
  - They can be read off the graph
- Important for modeling: understand assumptions made when choosing a Bayes net graph



Example



- Conditional independence assumptions directly from simplifications in chain rule:

$$P(x, y, z, w) = P(x)P(y|x)P(z|x, y)P(w|x, y, z)$$
$$= P(x)P(y|x)P(z|y)P(w|z)$$

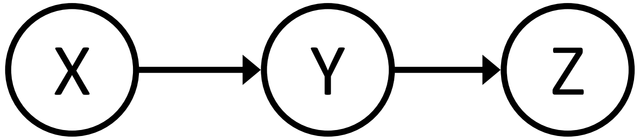
$$X \perp\!\!\!\perp Z|Y \qquad W \perp\!\!\!\perp \{X, Y\}|Z$$

- Additional implied conditional independence assumptions?

$$W \perp\!\!\!\perp X|Y$$ How?

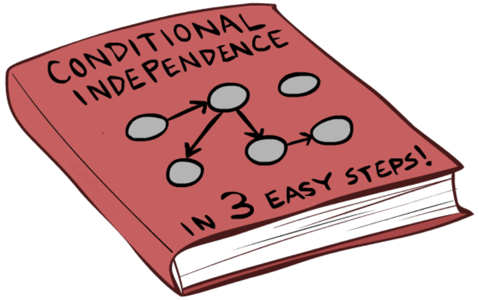
Independence in a BN

- Important question about a BN:
  - Are two nodes independent given certain evidence?
  - If yes, can prove using algebra (tedious in general)
  - If no, can prove with a counter example
  - Example:



  - Question: are X and Z necessarily independent?
    - Answer: no. Example: low pressure causes rain, which causes traffic.
    - X can influence Z, Z can influence X (via Y)
    - Addendum: they *could* be independent: how?

The General Case

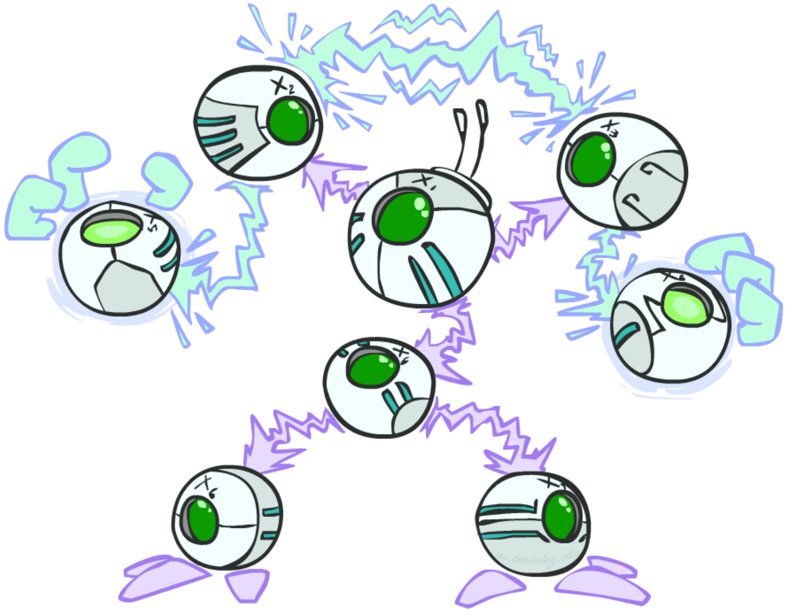


- General question: in a given BN, are two variables independent (given evidence)?

- Solution: analyze the graph

- Any complex example can be broken

  into repetitions of the three canonical cases

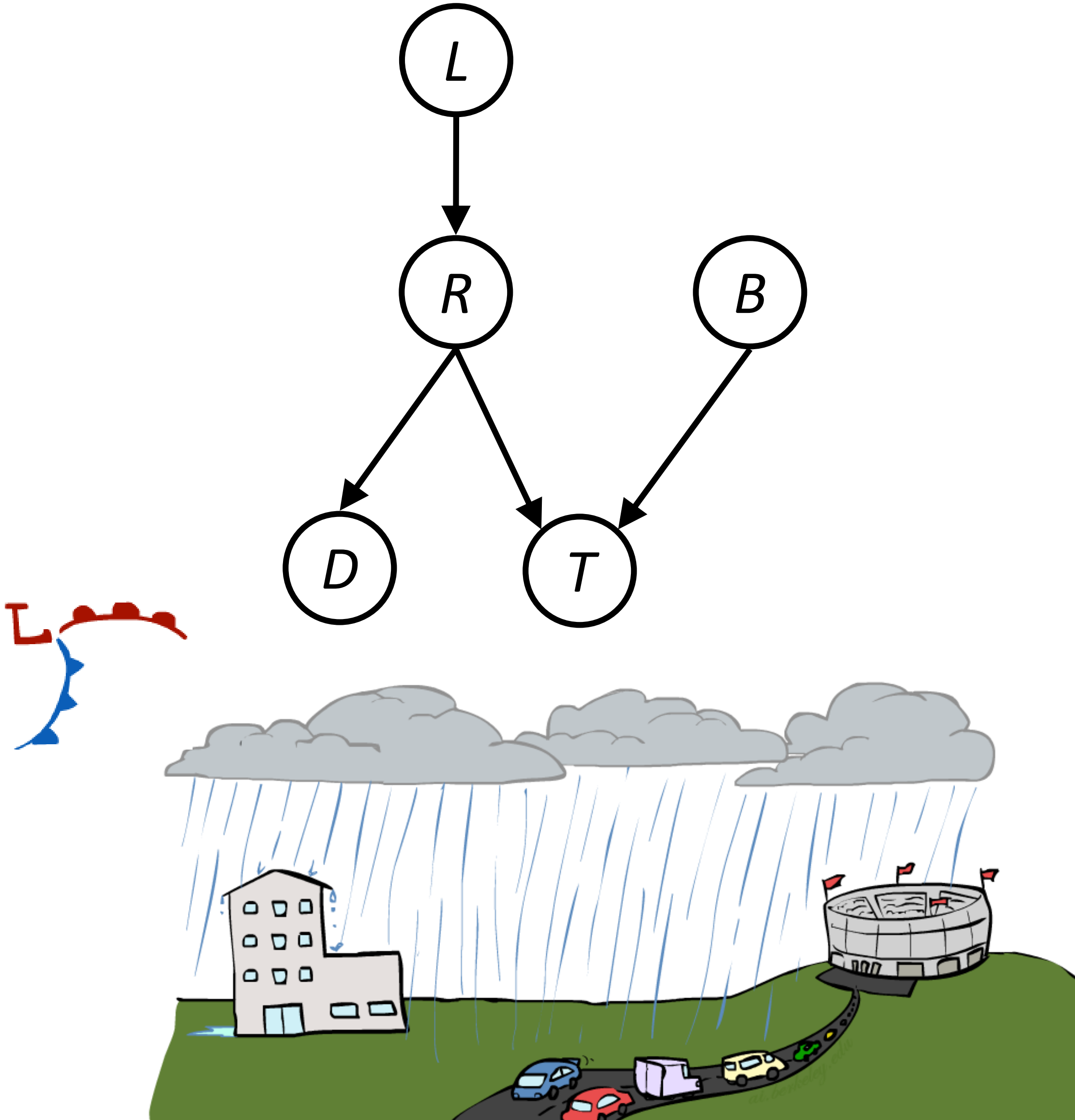


Reachability

- Recipe: shade evidence nodes, look for paths in the resulting graph

- Attempt 1: if two nodes are connected by any undirected path not blocked by a shaded node, they are **not** conditionally independent

- Almost works, but not quite
  - Where does it break?
  - Answer: the v-structure at T doesn't count as a link in a path unless "active"

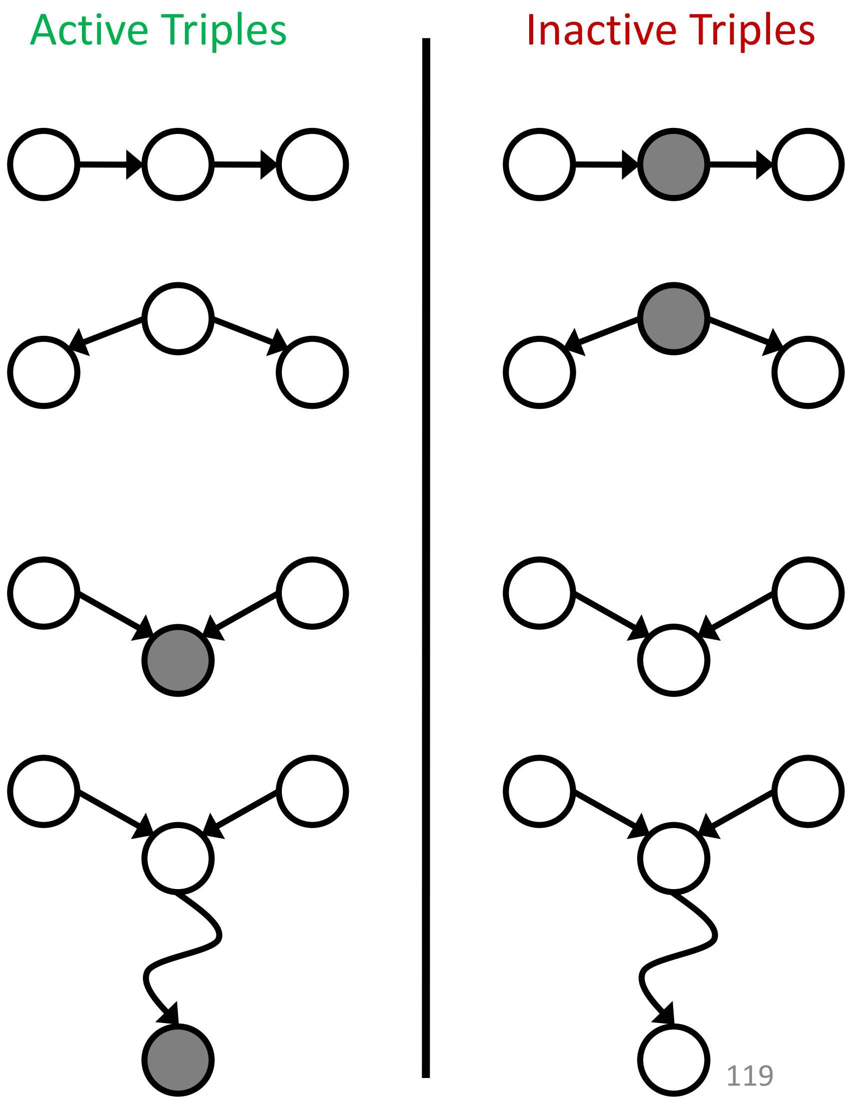


Active / Inactive Paths

- Question: Are X and Y conditionally independent given evidence variables {Z}?
  - Yes, if X and Y "d-separated" by Z
  - Consider all (undirected) paths from X to Y
  - No active paths = independence!

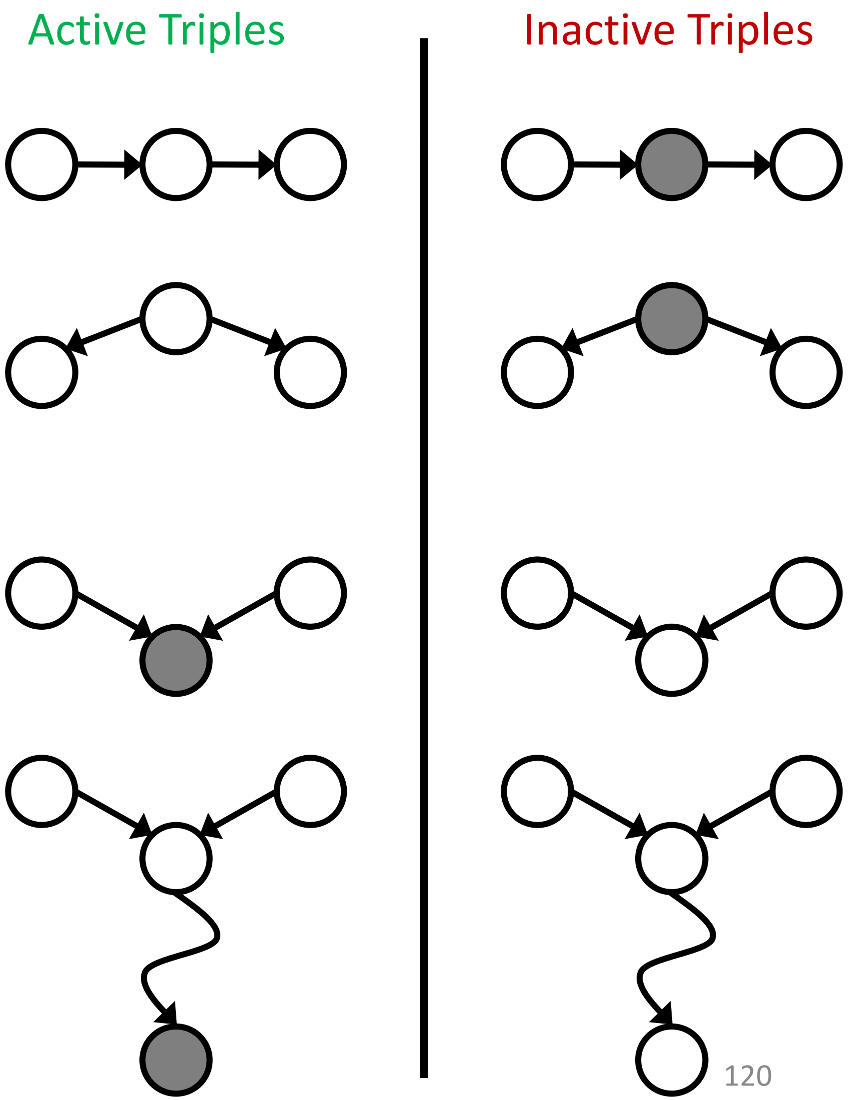
- A path is active if each triple is active:
  - Causal chain A -> B -> C where B is unobserved (either direction)
  - Common cause A <- B -> C where B is unobserved
  - Common effect (aka v-structure)
    A -> B <- C where B *or one of its descendants* is observed

- All it takes to block a path is a single inactive segment



Bayes Ball / D-separation

- Question: Are X and Y conditionally independent given evidence variables {Z}?

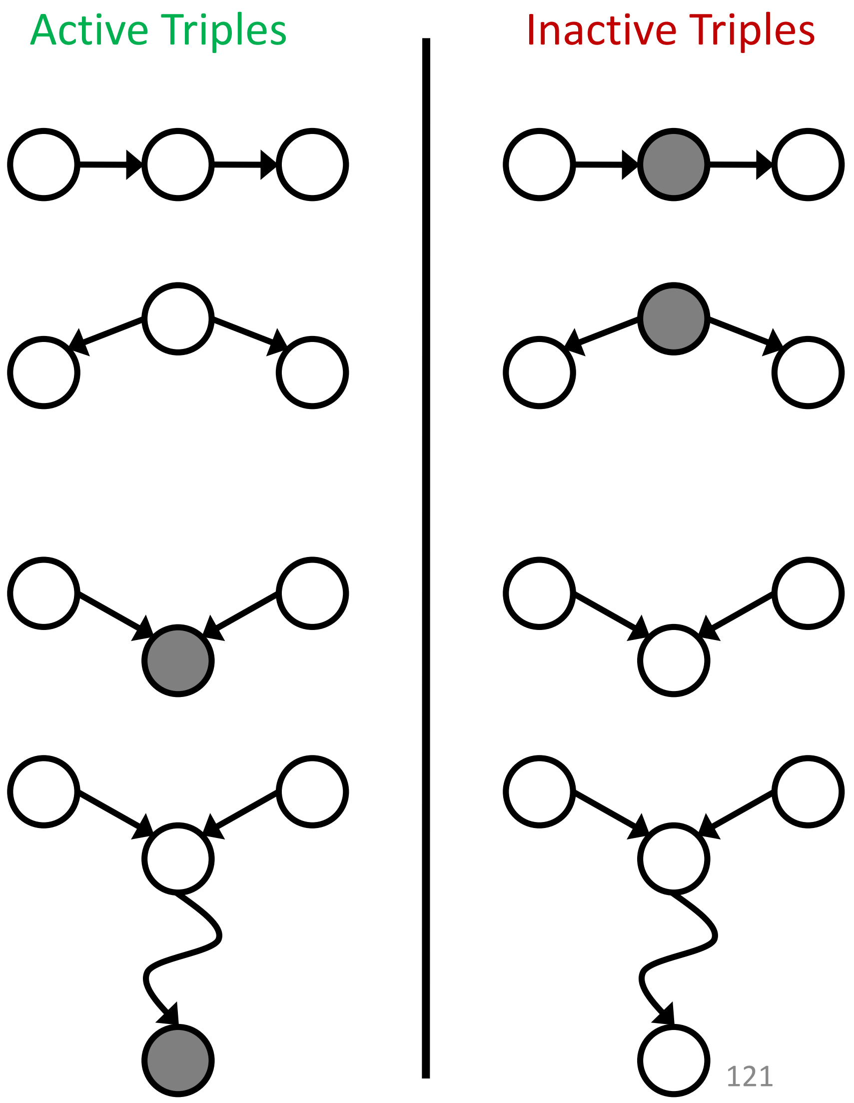


- Shachter, Ross D. "Bayes-Ball: Rational Pastime (for Determining Irrelevance and Requisite Information in Belief Networks and Influence Diagrams)." *Proceedings of the Fourteenth conference on Uncertainty in Artificial Intelligence.* 1998.

Bayes Ball 2

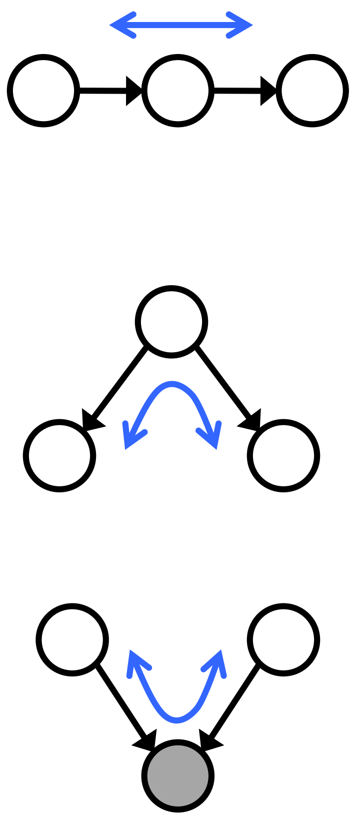
- Question: Are X and Y conditionally independent given evidence variables {Z}?

1. Shade in Z
2. Drop a ball at X
3. The ball can pass through any *active* path and is blocked by any *inactive* path (ball can move either direction on an edge)
4. If the ball reaches Y, then X and Y are NOT conditionally independent given Z

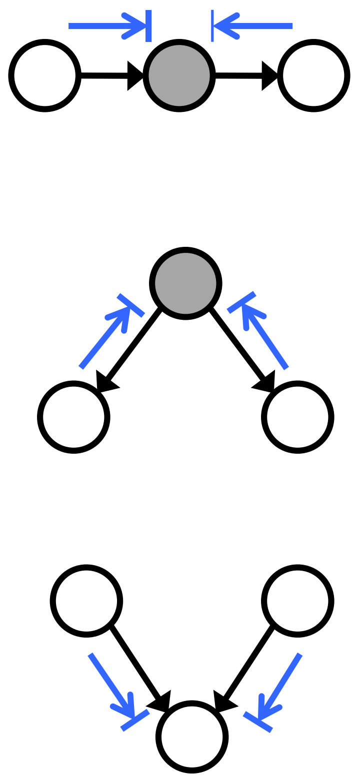


Bayes Ball 3

Active Paths



Inactive Paths



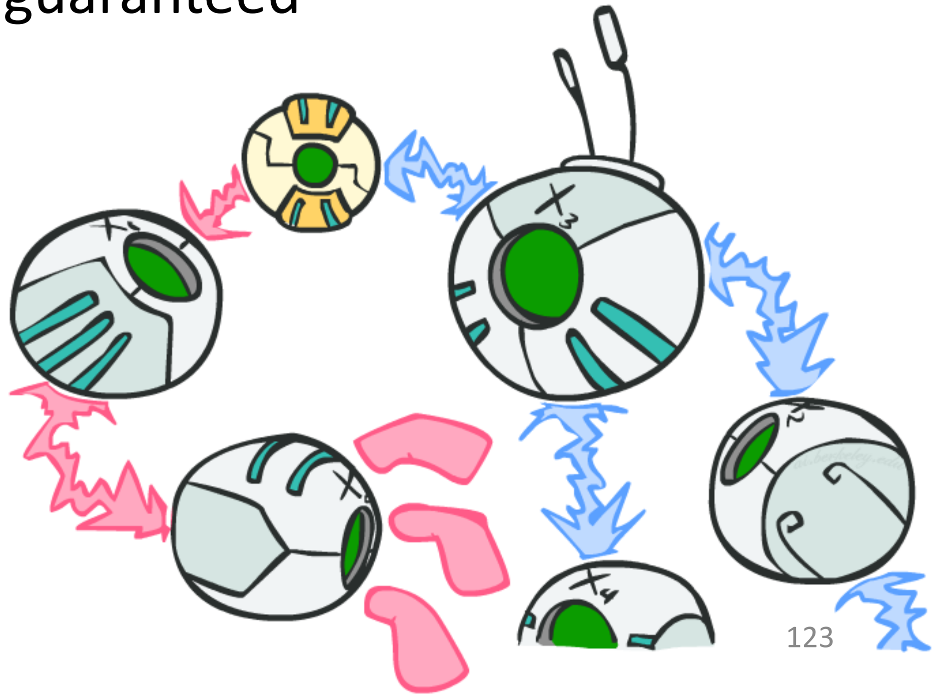
More Variables

- Query: $X_i \perp\!\!\!\perp X_j | \{X_{k_1}, ..., X_{k_n}\}$?
- Check all (undirected!) paths between X_i and X_j
  - If one or more active, then independence not guaranteed

$$X_i \not\perp\!\!\!\perp X_j | \{X_{k_1}, ..., X_{k_n}\}$$

  - Otherwise (i.e. if all paths are inactive), then independence is guaranteed

$$X_i \perp\!\!\!\perp X_j | \{X_{k_1}, ..., X_{k_n}\}$$

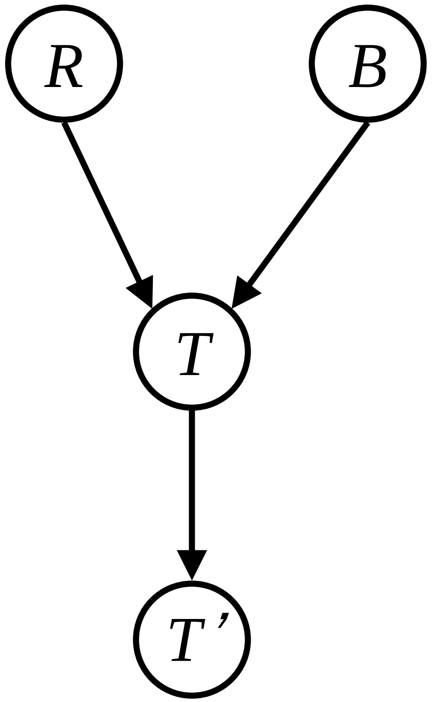


Example

$R \perp\!\!\!\perp B$ *Yes*

$R \perp\!\!\!\perp B | T$

$R \perp\!\!\!\perp B | T'$



Example 2

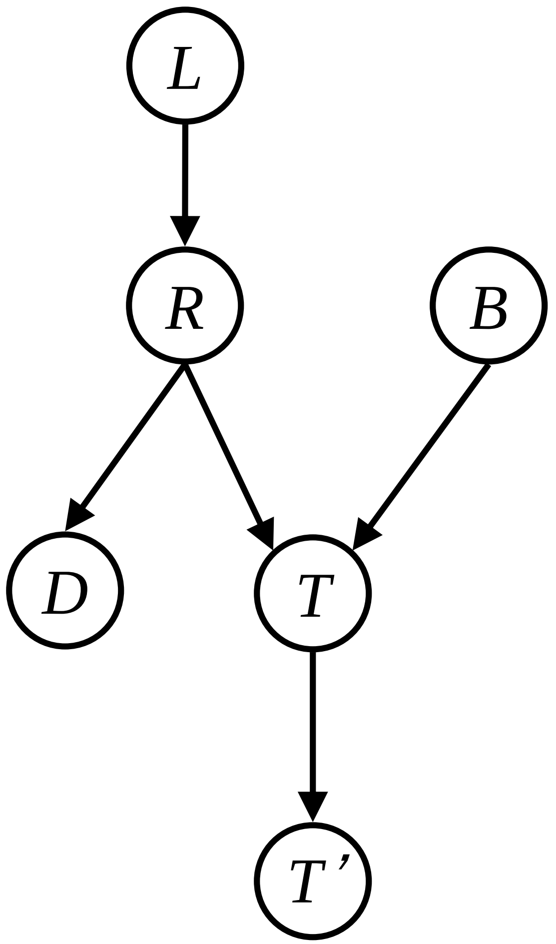
$L \perp\!\!\!\perp T' | T$ *Yes*

$L \perp\!\!\!\perp B$ *Yes*

$L \perp\!\!\!\perp B | T$

$L \perp\!\!\!\perp B | T'$

$L \perp\!\!\!\perp B | T, R$ *Yes*



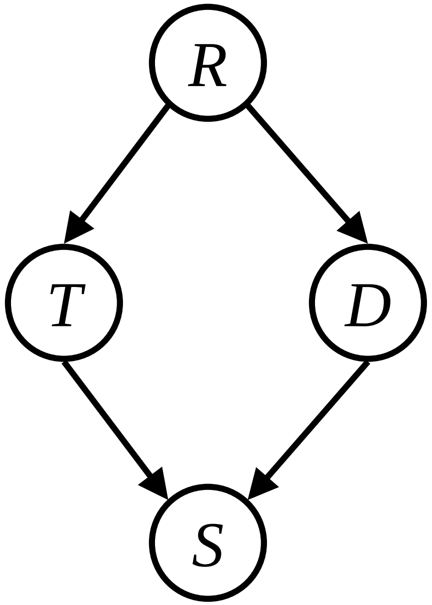
Example 3

- Variables:
  - R: Raining
  - T: Traffic
  - D: Roof drips
  - S: I'm sad
- Questions:

$$T \perp\!\!\!\perp D$$

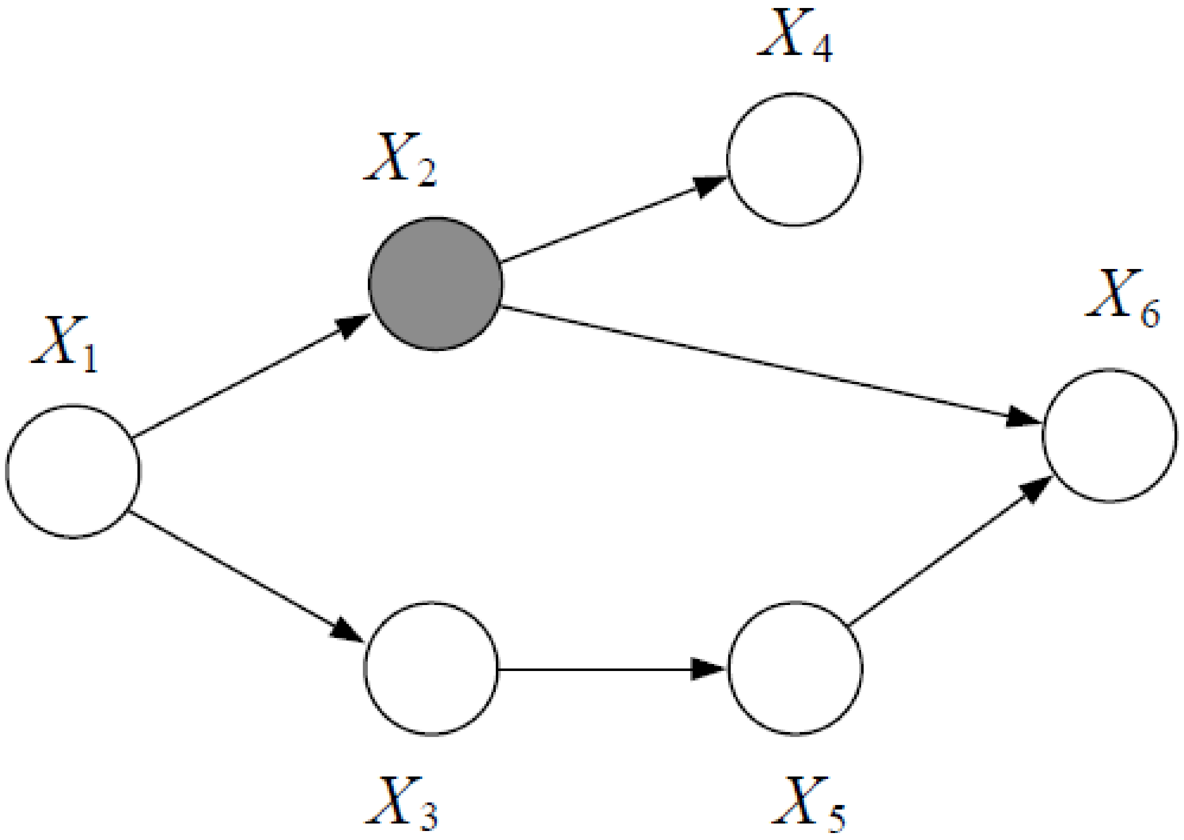
$$T \perp\!\!\!\perp D | R \qquad \textit{Yes}$$

$$T \perp\!\!\!\perp D | R, S$$



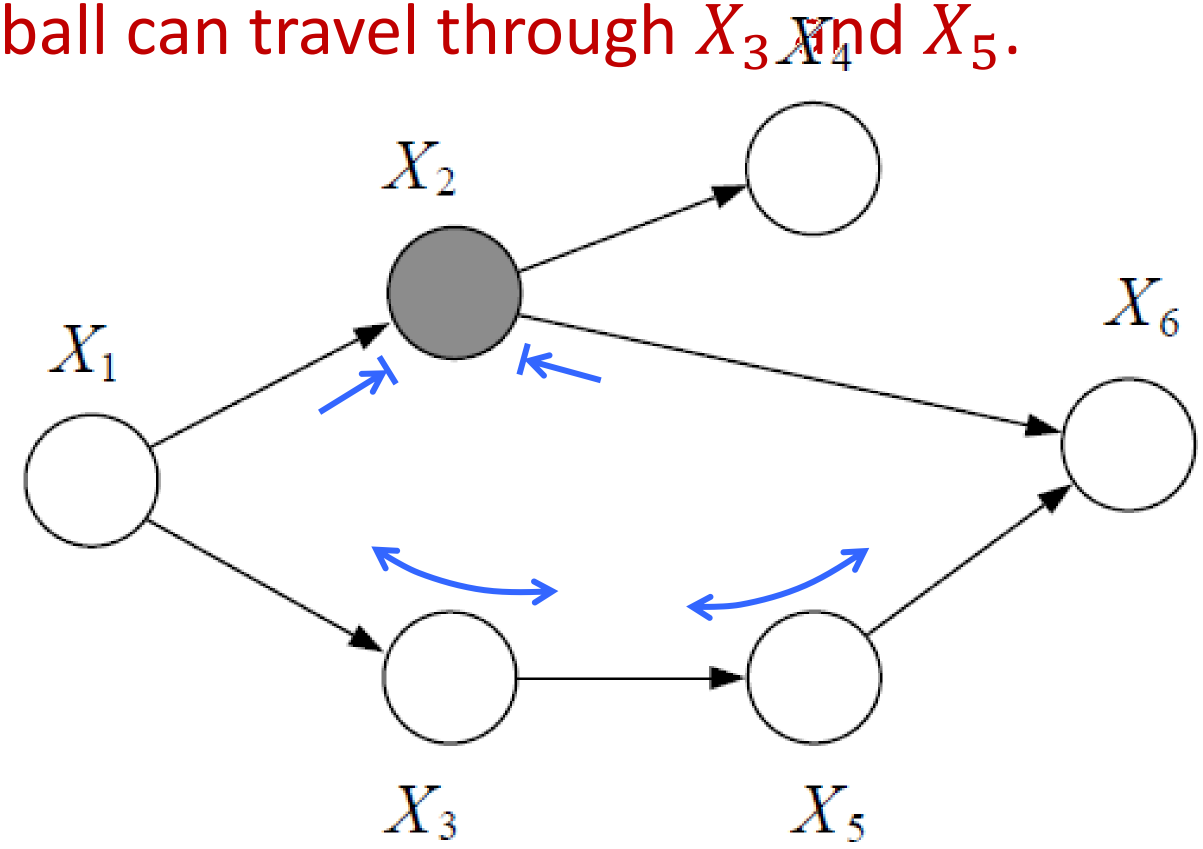
Quiz

- Is X_1 independent from X_6 given X_2?



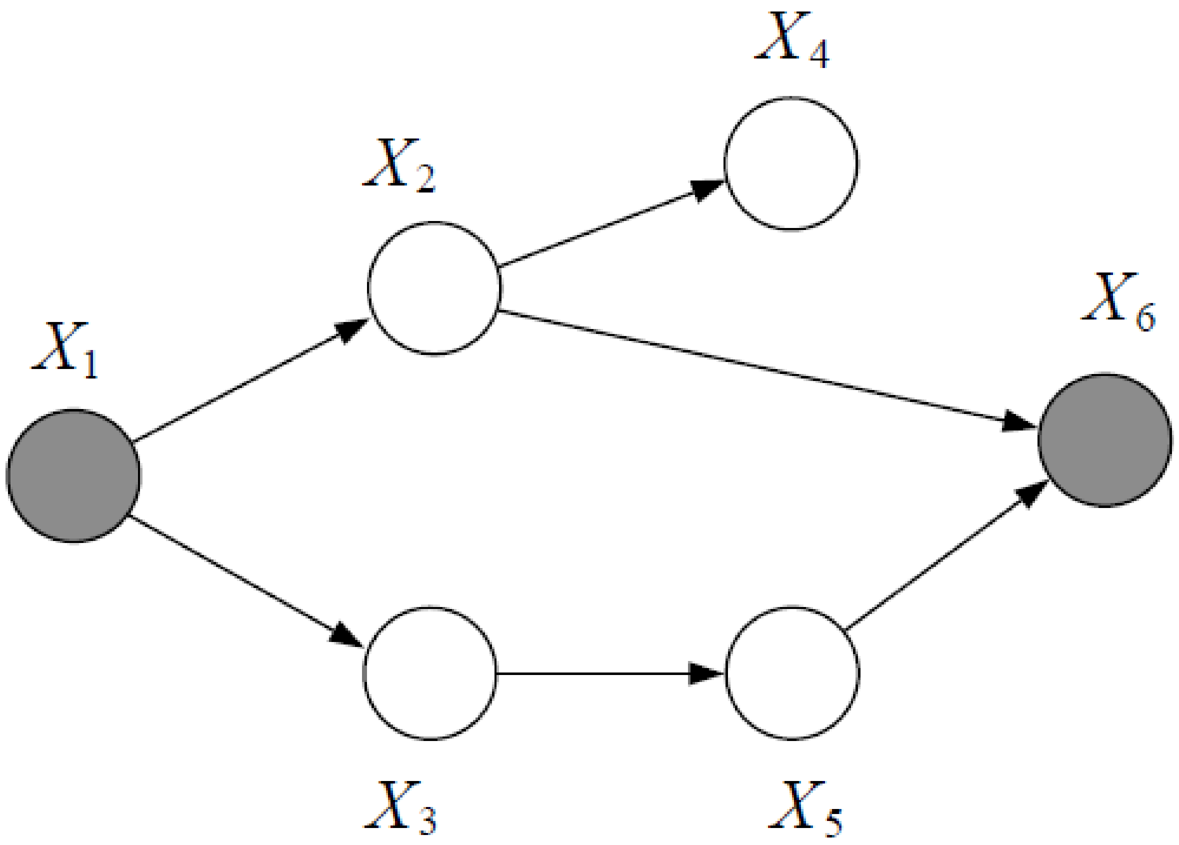
Quiz (cont.)

- Is X_1 independent from X_6 given X_2?
- No, the Bayes ball can travel through X_3 and X_5.



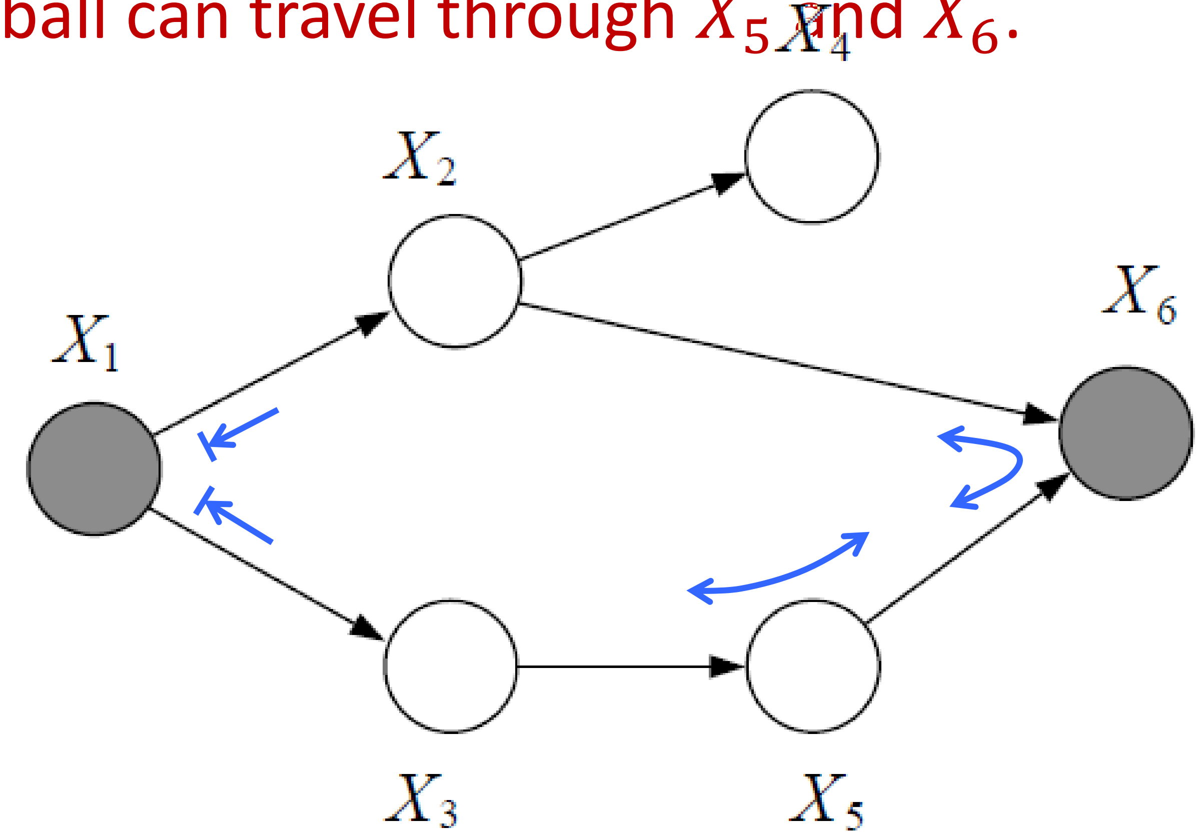
Quiz 2

- Is X_2 independent from X_3 given X_1 and X_6?



Quiz 2 (cont.)

- Is X_2 independent from X_3 given X_1 and X_6?
- No, the Bayes ball can travel through X_5 and X_6.

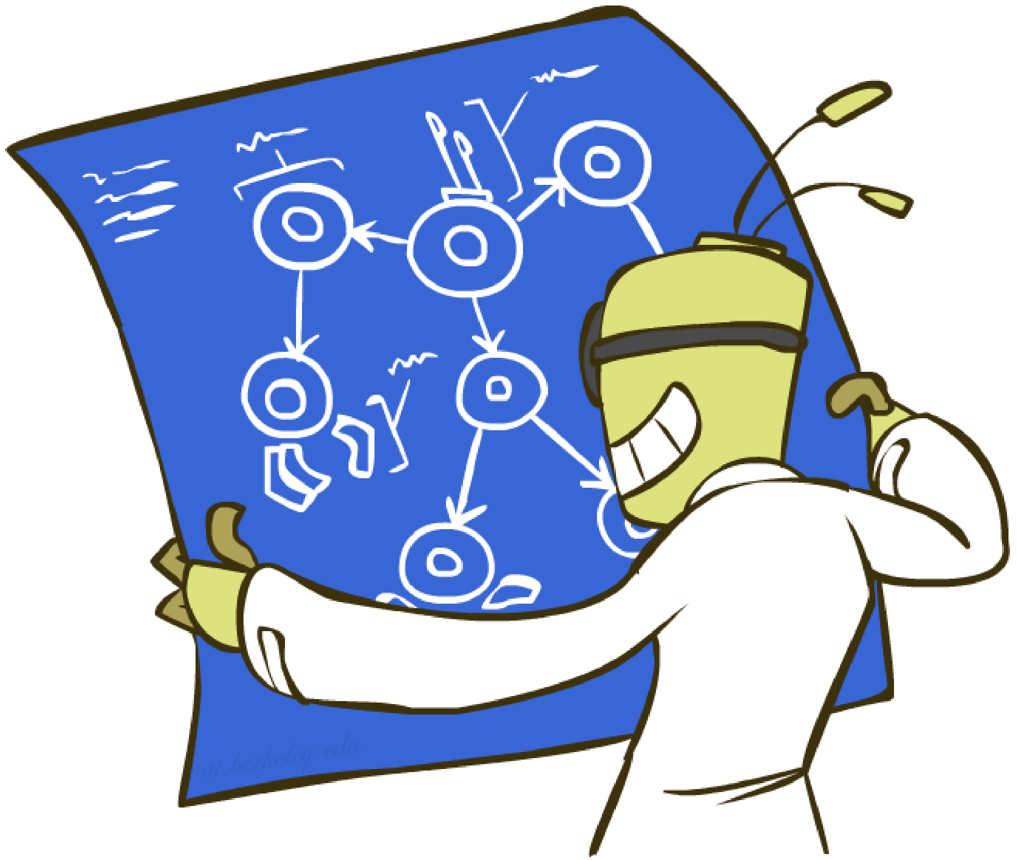


Structure Implications

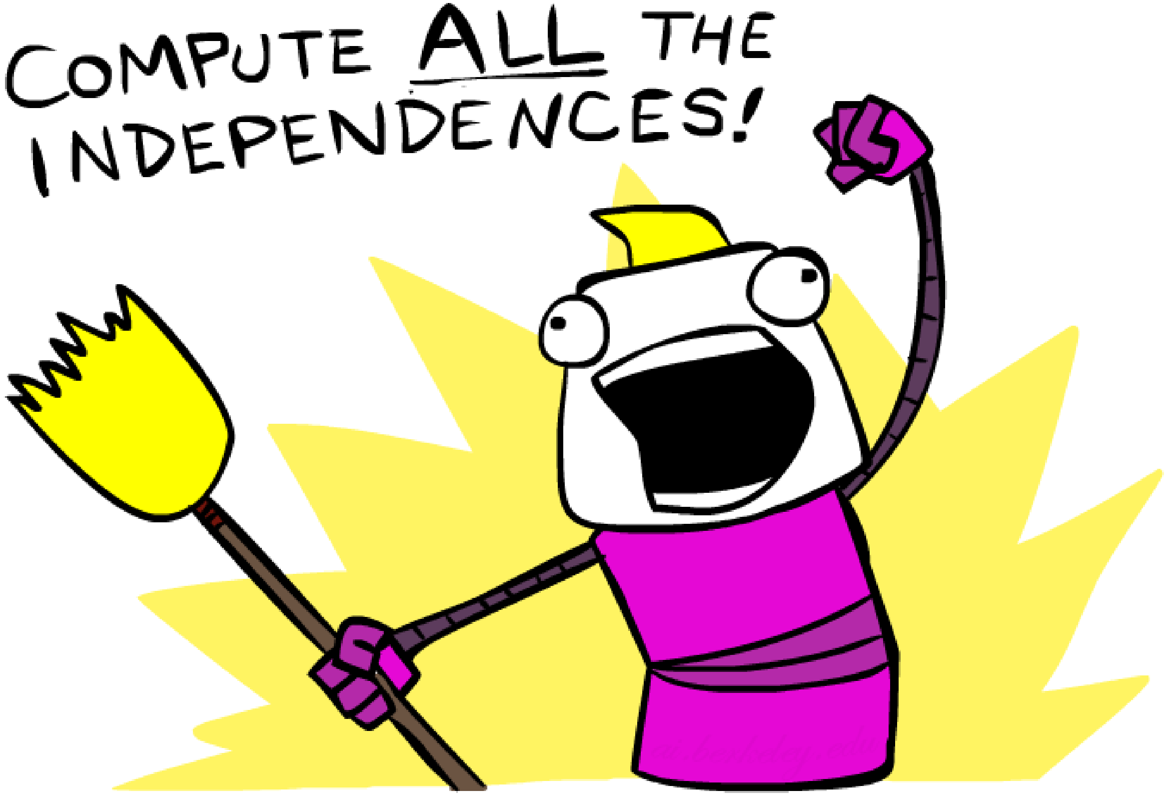
- Given a Bayes net structure, can run Bayes ball/d-separation algorithm to build a complete list of conditional independences that are necessarily true of the form

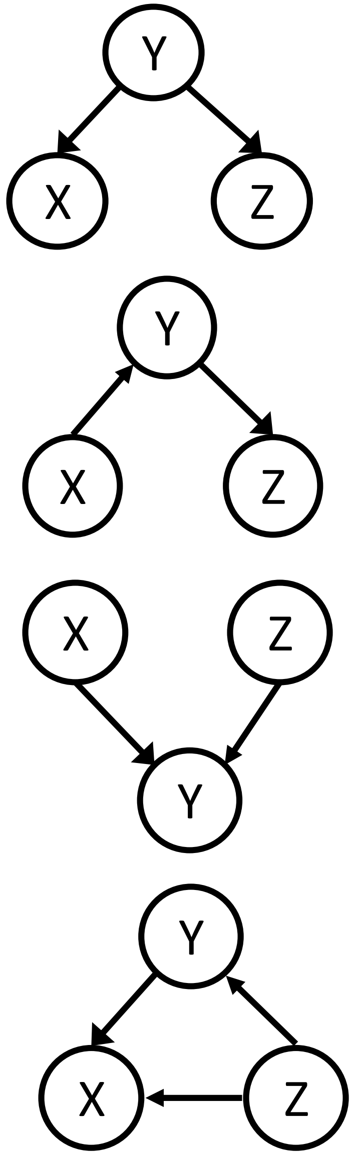
$$X_i \perp\!\!\!\perp X_j | \{X_{k_1}, ..., X_{k_n}\}$$

- This list determines the set of probability distributions that can be represented



Computing All Independences





Topology Limits Distributions

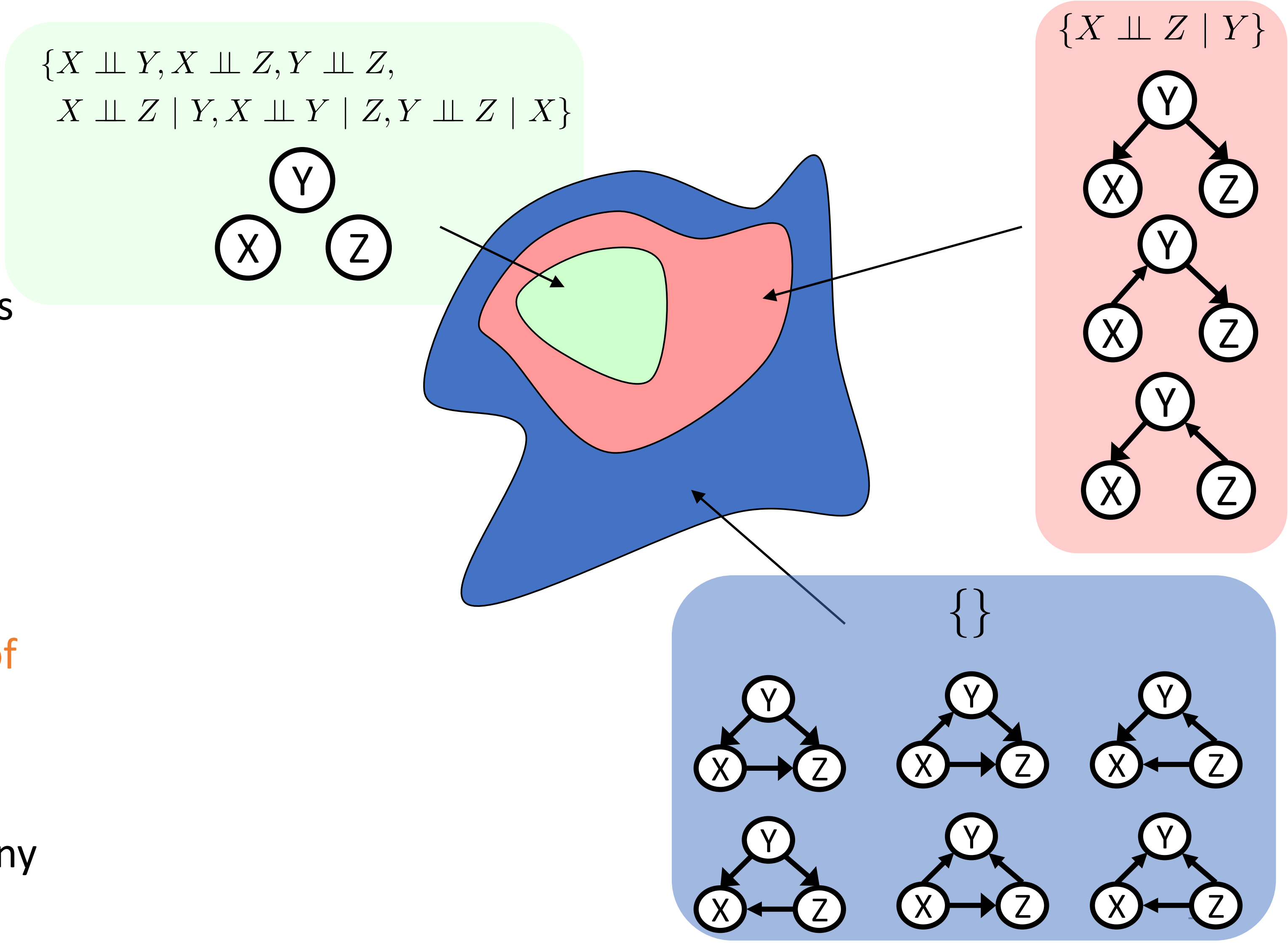
- Given some graph topology G, only certain joint distributions can be encoded

- The graph structure guarantees certain (conditional) independences

- (There might be more independence)

- Adding arcs increases the set of distributions, but has several costs

- Full conditioning can encode any distribution



Bayes Nets Representation Summary

- Bayes nets compactly encode joint distributions (by making use of conditional independences!)

- Guaranteed independencies of distributions can be deduced from BN graph structure

- Bayes ball/D-separation gives precise conditional independence guarantees from graph alone

- A Bayes net's joint distribution may have further (conditional) independence that is not detectable until you inspect its specific distribution

Summary

Shuai Li

https://shuaili8.github.io

- Probability
  - Joint/marginal/conditional probabilities
- Answer any query from joint distributions
- Build Joint Distribution Using Chain Rule
- Bayes Nets
- Conditional independence, Semantics
- Causality
- Bayes nets independence, Bayes Nets Representation

Questions?