Warm-up as you walk in

When does a probability table sum to 1?

$$P(\mathrm{c} \mid A)$$

$$P(\mathrm{A}, \mathrm{B}, \mathrm{c})$$

$$P(\mathrm{a} \mid B, C)$$

$$P(\mathrm{A} \mid b)$$

$$P(\mathrm{a}, \mathrm{b} \mid c)$$

$$P(\mathrm{A}, \mathrm{B} \mid c)$$

Announcements

Assignments:

- HW9 (written)
  - Due Tue 4/2, 10 pm

Optional Probability (online)

Midterm:

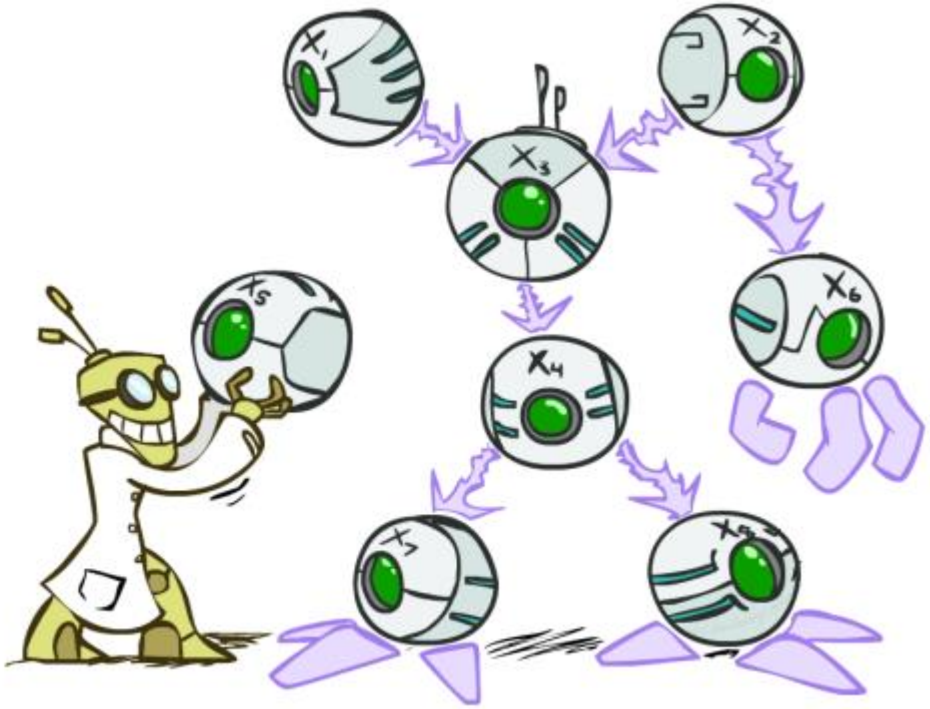
- Mon 4/8, in-class

Course Feedback:

- See Piazza post for mid-semester survey

AI: Representation and Problem Solving

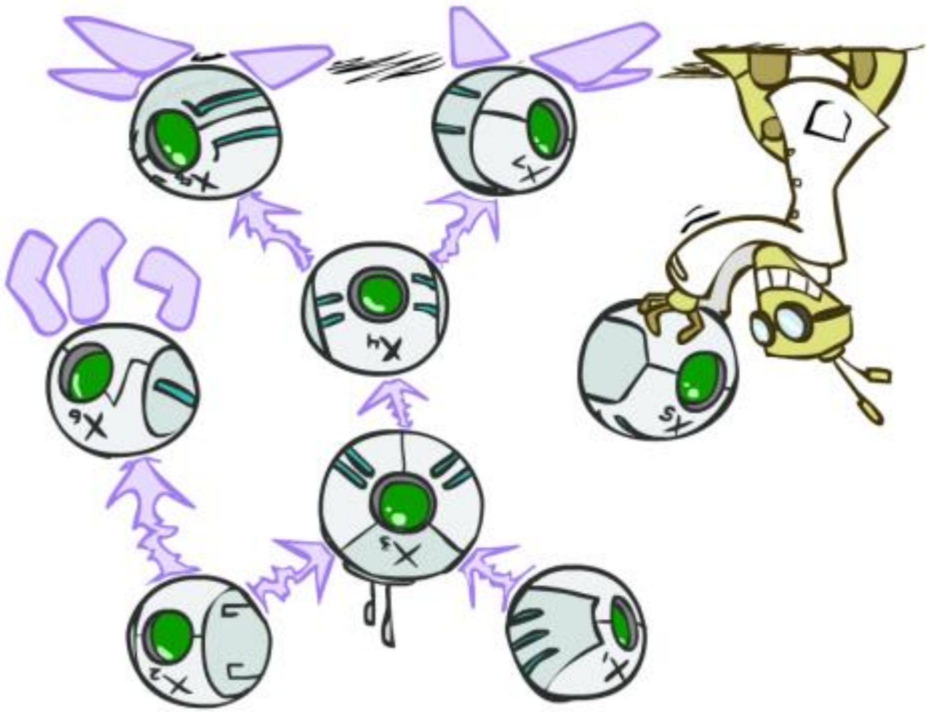
Bayes Nets



Instructors: Pat Virtue & Stephanie Rosenthal

Slide credits: CMU AI and http://ai.berkeley.edu

AI-pril Fool's!



Trouble maker credit: Arnav & Pranav

Course Survey

Please fill out on Piazza!

Warm-up as you walk in

When does a probability table sum to 1?

$P(\mathrm{c} \mid A)$

$P(\mathrm{A}, \mathrm{B}, \mathrm{c})$

$P(\mathrm{a} \mid B, C)$

$P(\mathrm{A} \mid b)$

$P(\mathrm{a}, \mathrm{b} \mid c)$

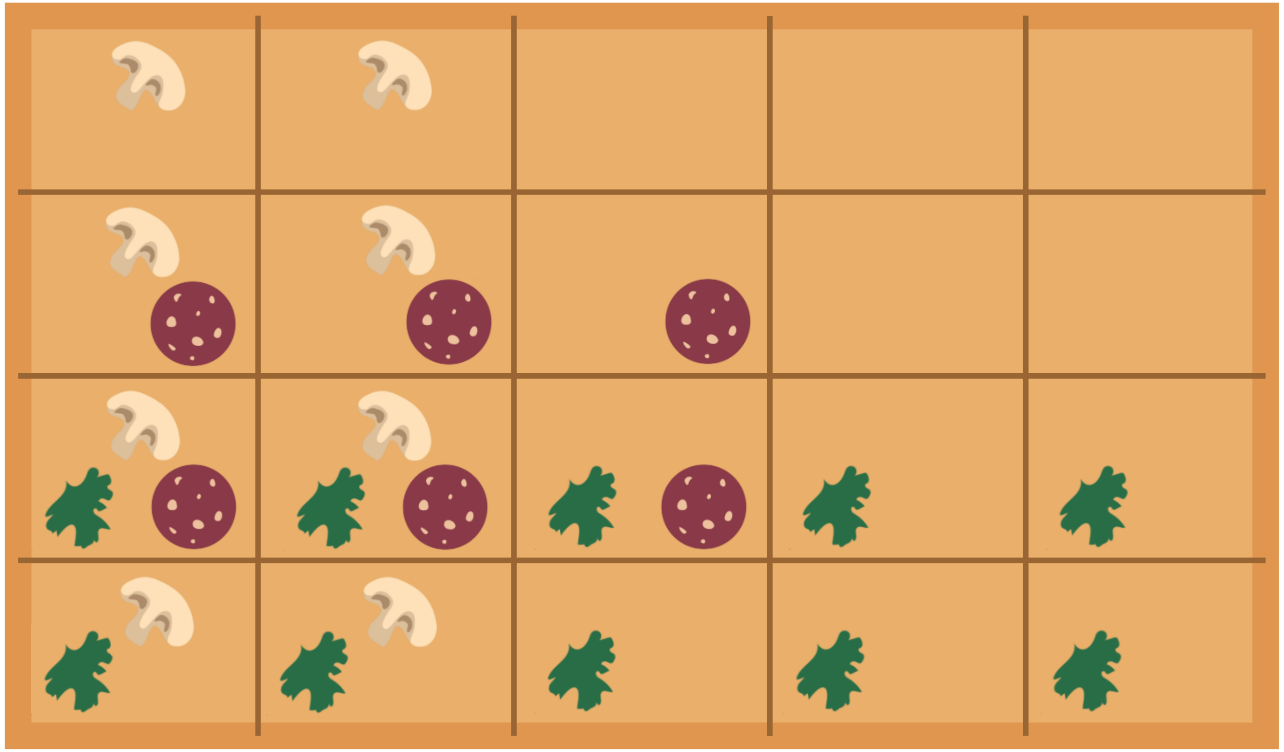
$P(\mathrm{A}, \mathrm{B} \mid c)$

Answer Any Query from Joint Distribution

What is the probability of getting a slice with:

1) No mushrooms
2) Spinach and no mushrooms
3) Spinach, when asking for slice with no mushrooms

- Mushrooms
- Spinach
- No spinach
- No spinach and mushrooms
- No spinach when asking for no mushrooms
- No spinach when asking for mushrooms
- Spinach when asking for mushrooms
- No mushrooms and no spinach



Answer Any Query from Joint Distribution

You can answer all of these questions:

$P(M)$

m_1	12/20
m_2	

$P(S)$

s_1	
s_2	

$P(M,S)$

m_1	s_1	
m_1	s_2	6/20
m_2	s_1	
m_2	s_2	

$P(M|s_1)$

m_1	
m_2	

$P(M|s_2)$

m_1	
m_2	

$P(S|m_1)$

s_1	
s_2	6/12

$P(S|m_2)$

s_1	
s_2	

Answer Any Query from Joint Distribution

P(Weather)?

P(Weather | winter)?

P(Weather | winter, hot)?

Season	Temp	Weather	P(S, T, W)
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

Answer Any Query from Joint Distribution

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 Distribution

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)

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

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

Answer Any Query from Joint Distribution

P(Weather)?

P(Weather | winter)?

P(Weather | winter, hot)?

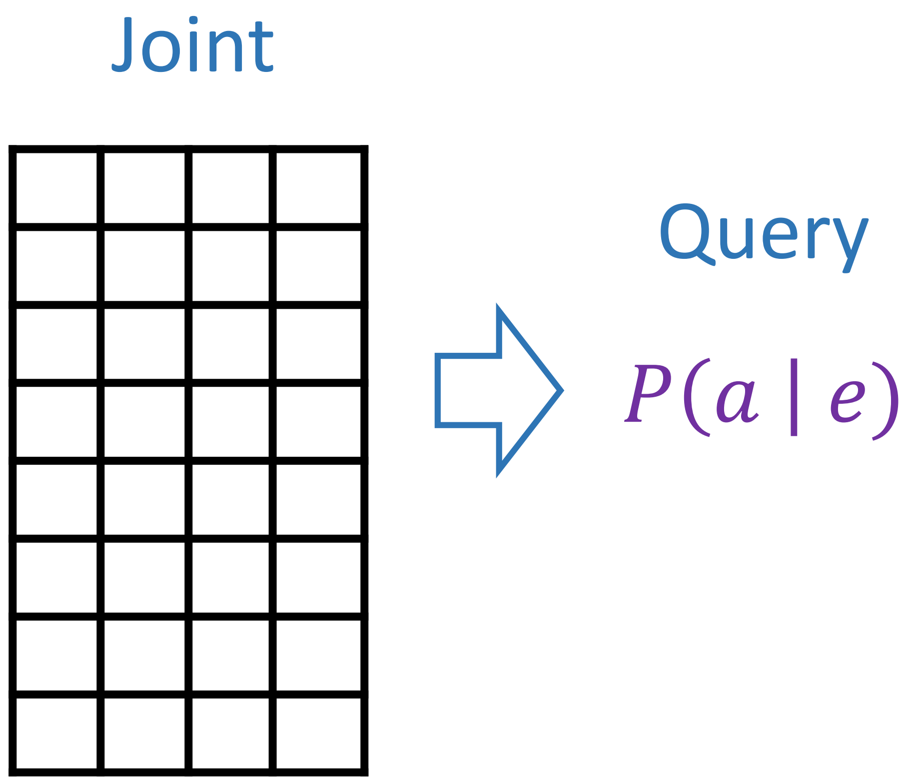
Season	Temp	Weather	P(S, T, W)
summer	hot	sun	0.30
summer	hot	rain	0.05
summer	cold	sun	0.10
summer	cold	rain	0.05
winter	hot	sun	0.10
winter	hot	rain	0.05
winter	cold	sun	0.15
winter	cold	rain	0.20

Answer Any Query from Joint Distribution

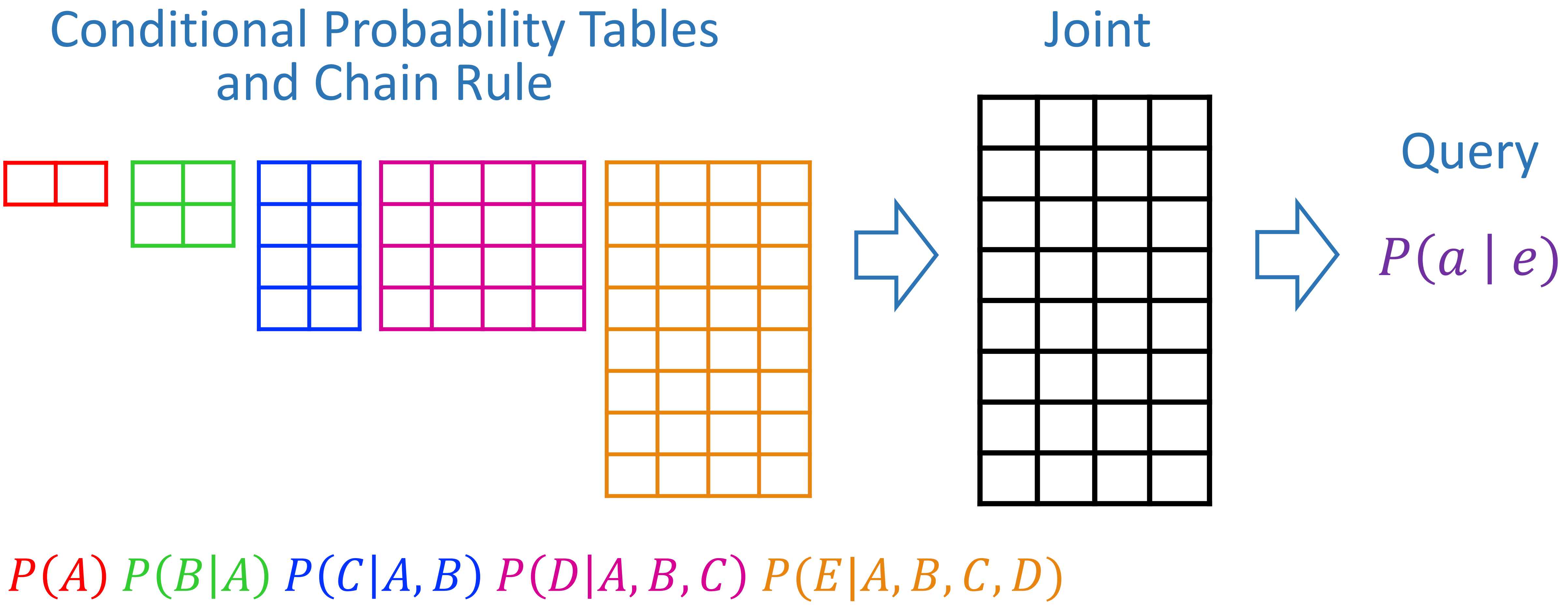
Joint distributions are the best!

Problems with joints

- Huge
  - n variables with d values
  - d^n entries
- We aren't given the joint table
  - Usually some set of
    conditional probability tables



Build Joint Distribution Using Chain Rule



Build Joint Distribution Using Chain Rule

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

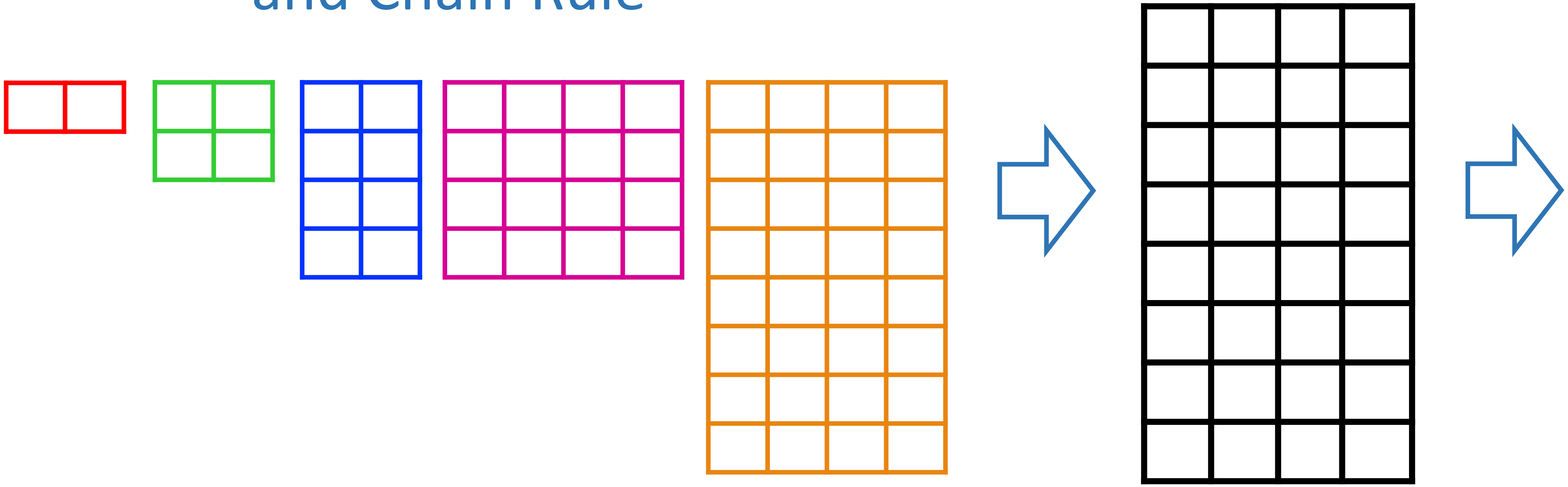
Answer Any Query from Condition Probability Tables

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

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

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)

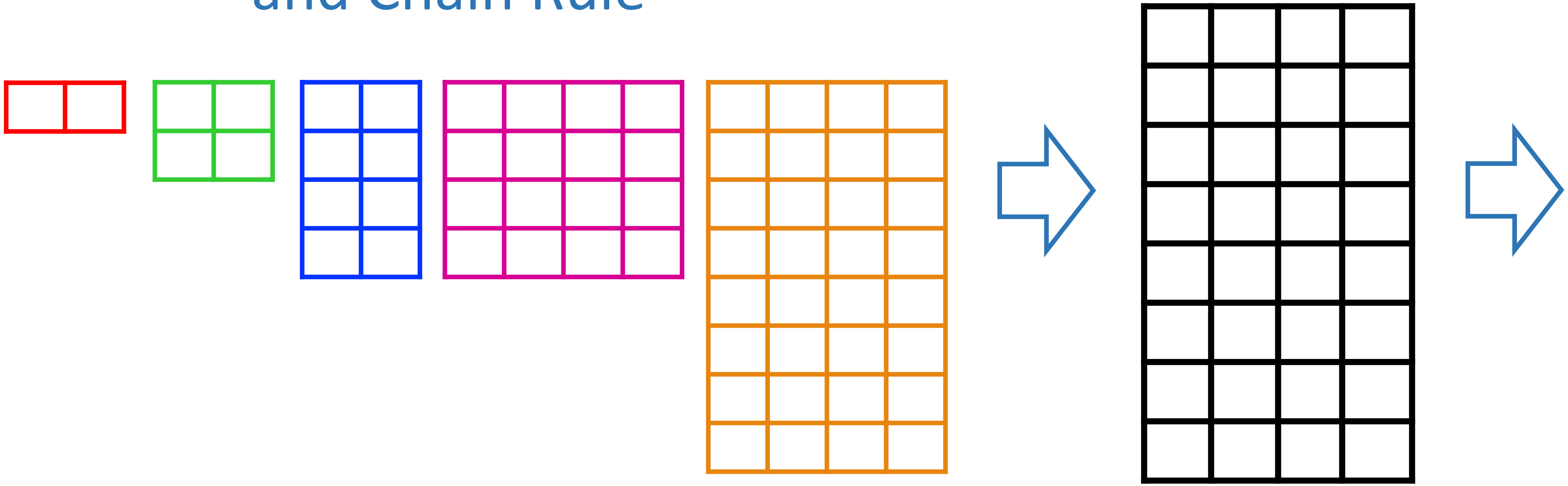
$$P(Q \mid e) = \frac{P(e,Q)}{\sum_q P(e,q)}$$

Answer Any Query from Condition Probability Tables

Conditional Probability Tables
and Chain Rule

Joint

Query

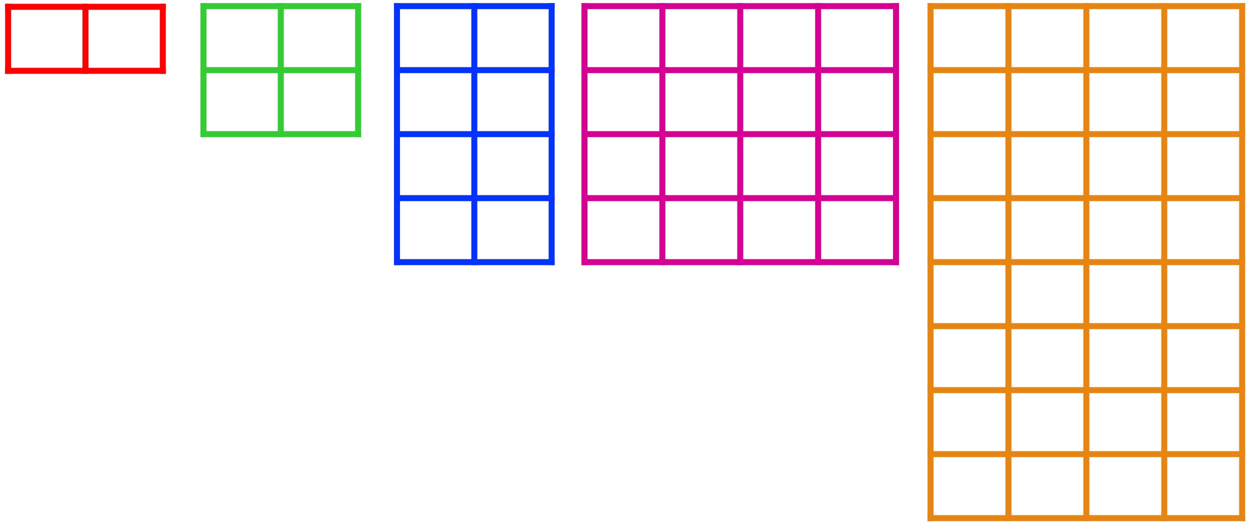


$P(a \mid e)$

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

Answer Any Query from Condition Probability Tables

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

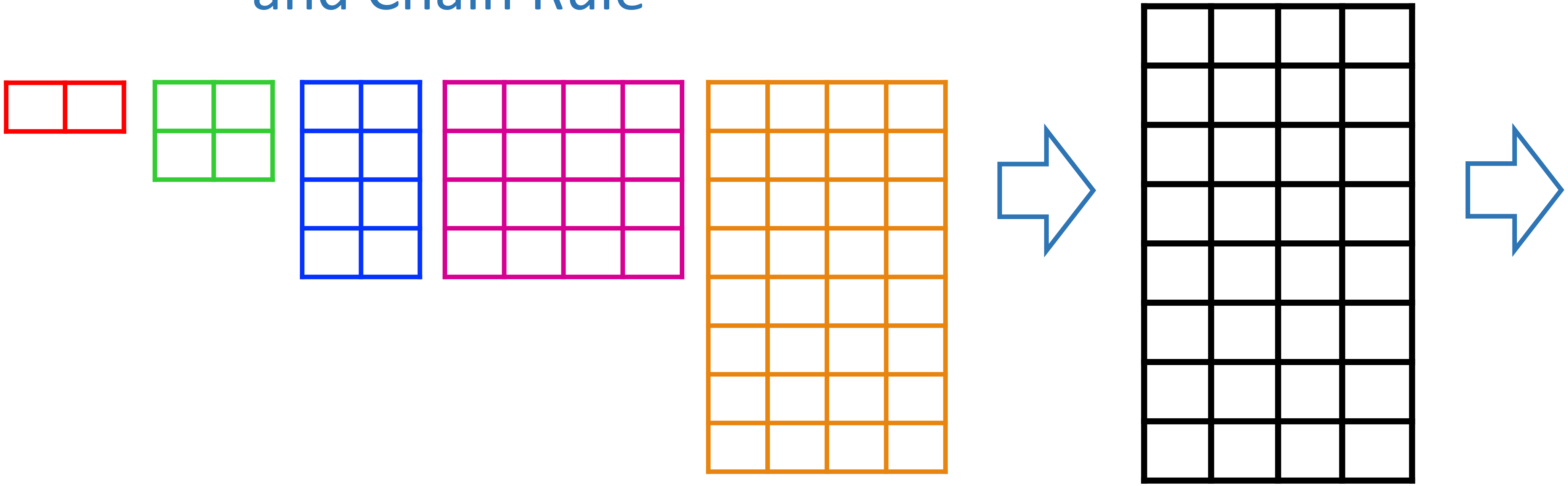
Answer Any Query from Condition Probability Tables

Conditional Probability Tables
and Chain Rule

Joint

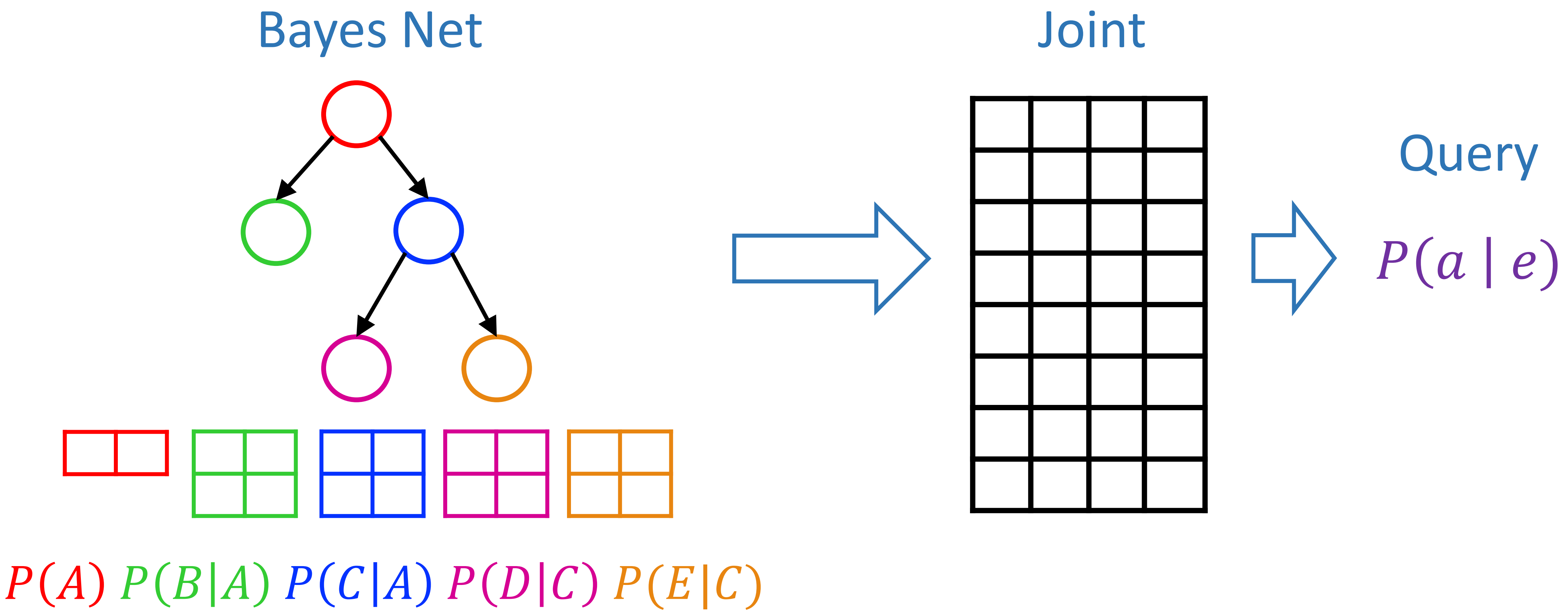
Query

$P(a \mid e)$



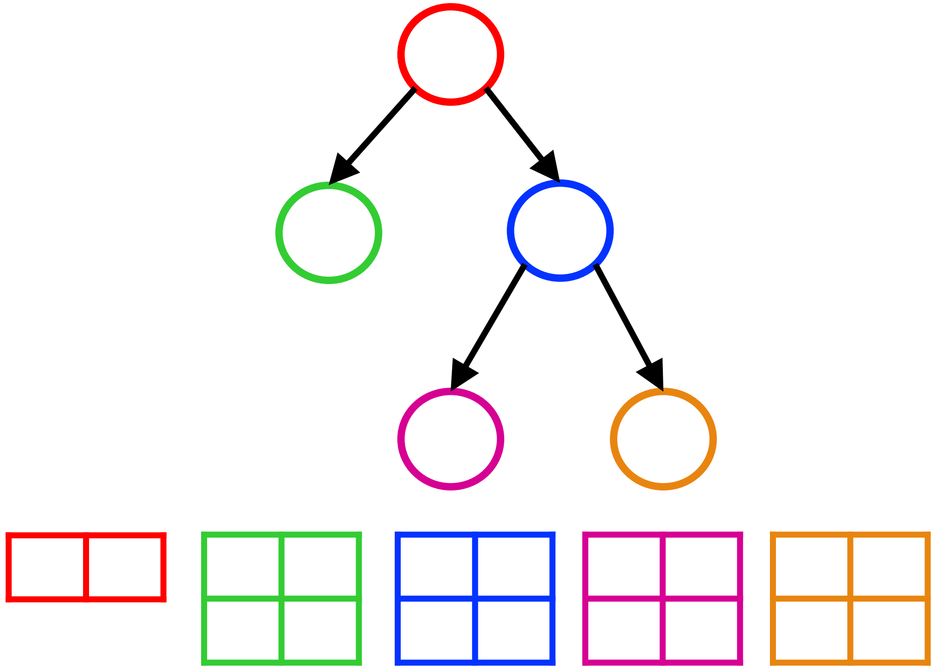
$P(A)$ $P(B|A)$ $P(C|A,B)$ $P(D|A,B,C)$ $P(E|A,B,C,D)$

Answer Any Query from Condition Probability Tables



Answer Any Query from Condition Probability Tables

Bayes Net



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

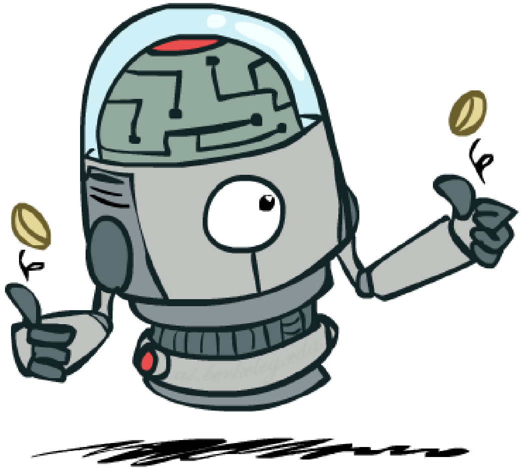
Query

$P(a \mid e)$

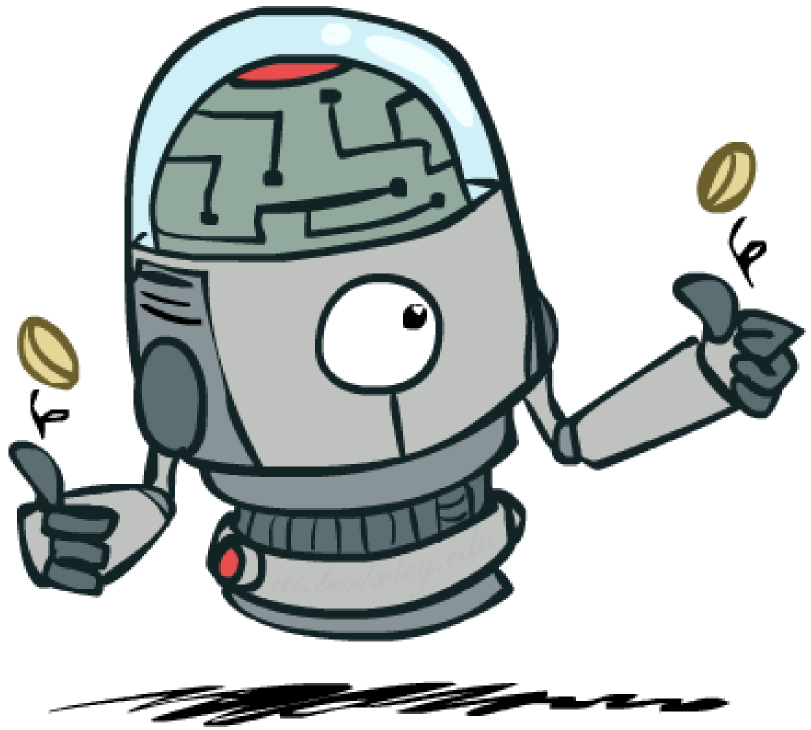
Build Joint Distribution Using Chain Rule

Chain rule

$$P(X_1, X_2, \ldots, X_n) = \prod_i P(X_i \mid X_1, \ldots, X_{i-1})$$



Independence



Independence

Two variables X and Y are (absolutely) ***independent*** if

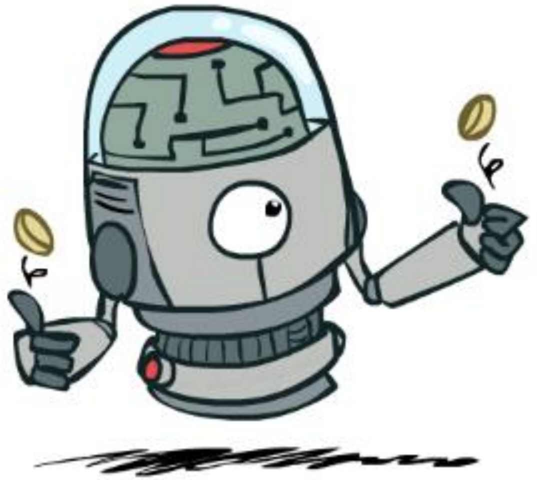
$$\forall x,y \qquad P(x, y) = P(x)\, P(y)$$

- This says that their joint distribution ***factors*** into a product of two simpler distributions
- Combine with product rule $P(x,y) = P(x|y)P(y)$ we obtain another form:

$$\forall x,y\ P(x \mid y) = P(x) \quad \text{or} \quad \forall x,y\ P(y \mid x) = P(y)$$

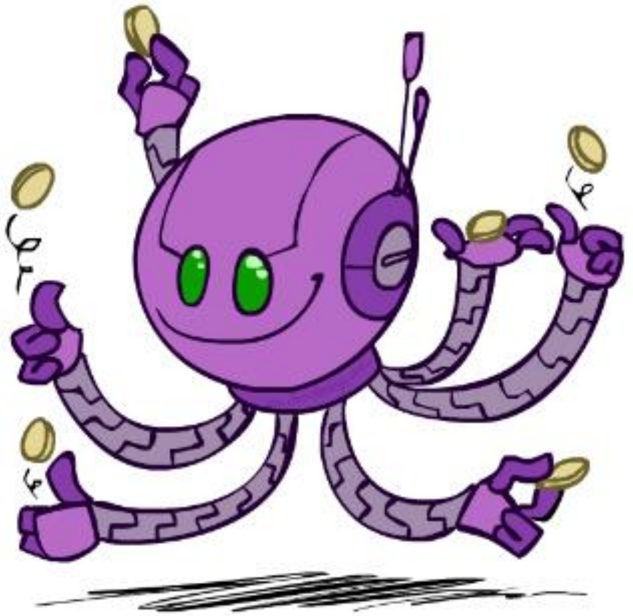
Example: two dice rolls $Roll_1$ and $Roll_2$

- $P(Roll_1{=}5, Roll_2{=}5) = P(Roll_1{=}5)\, P(Roll_2{=}5) = 1/6 \times 1/6 = 1/36$
- $P(Roll_2{=}5 \mid Roll_1{=}5) = P(Roll_2{=}5)$



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

. . .

$P(X_n)$

H	0.5
T	0.5

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

2^n

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

$$P_2(T, W) = P(T)P(W)$$

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

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:

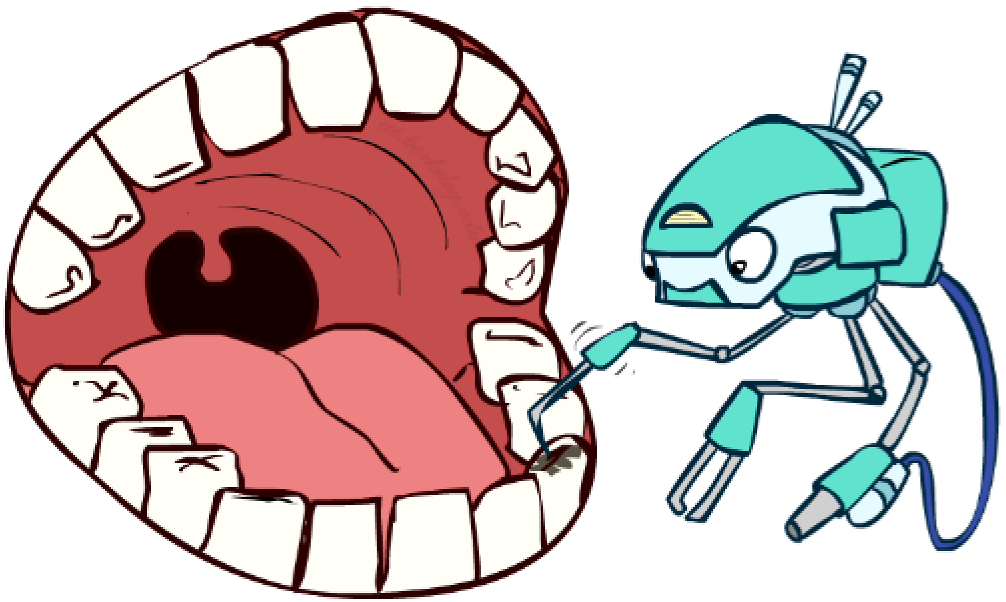
- $P(+catch \mid +toothache, +cavity) = P(+catch \mid +cavity)$

The same independence holds if I don't have a cavity:

- $P(+catch \mid +toothache, -cavity) = P(+catch \mid -cavity)$

Catch is *conditionally independent* of Toothache given Cavity:

- $P(Catch \mid Toothache, Cavity) = P(Catch \mid Cavity)$



Equivalent statements:

- $P(Toothache \mid Catch , Cavity) = P(Toothache \mid Cavity)$
- $P(Toothache, Catch \mid Cavity) = P(Toothache \mid Cavity)\, P(Catch \mid Cavity)$
- One can be derived from the other easily

Conditional Independence

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

if and only if:

$$\forall x,y,z \qquad P(x \mid y, z) = P(x \mid z)$$

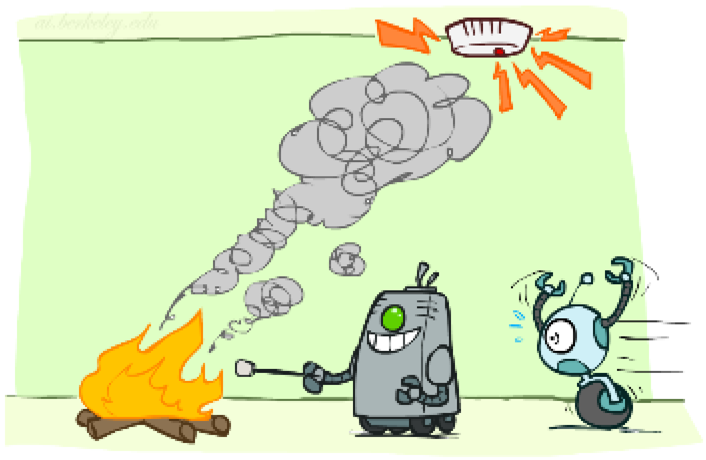
or, equivalently, if and only if

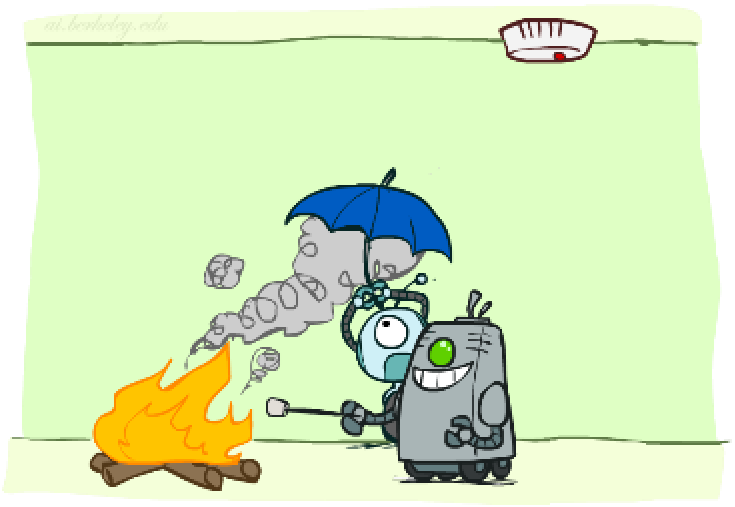
$$\forall x,y,z \qquad P(x, y \mid z) = P(x \mid z)\, P(y \mid z)$$

Conditional Independence

What about this domain:

- Fire
- Smoke
- Alarm





Conditional Independence

What about this domain:

- Traffic
- Umbrella
- Raining

Conditional Independence and the Chain Rule

Chain rule:

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

Trivial decomposition:

$P(Rain, Traffic, Umbrella) =$

With assumption of conditional independence:

$P(Rain, Traffic, Umbrella) =$

Conditional Independence and the Chain Rule

Chain rule:

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

Trivial decomposition:

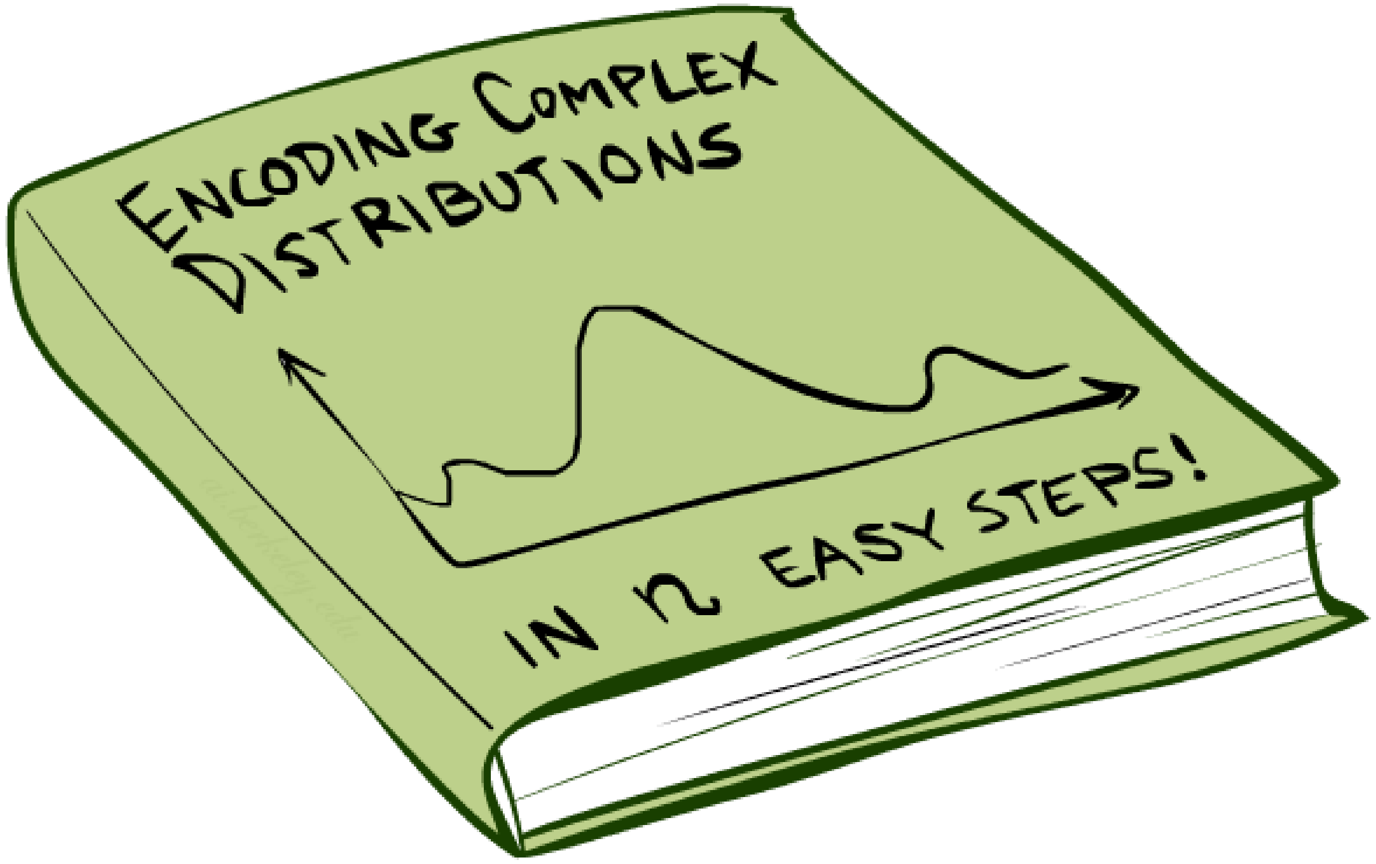
$P(Rain, Traffic, Umbrella) = P(Rain)\, P(Traffic \mid Rain)\, P(Umbrella \mid Rain, Traffic)$

With assumption of conditional independence:

$P(Rain, Traffic, Umbrella) = P(Rain)\, P(Traffic \mid Rain)\, P(Umbrella \mid Rain)$

Bayes nets / graphical models help us express conditional independence assumptions

Bayes' Nets: Big Picture



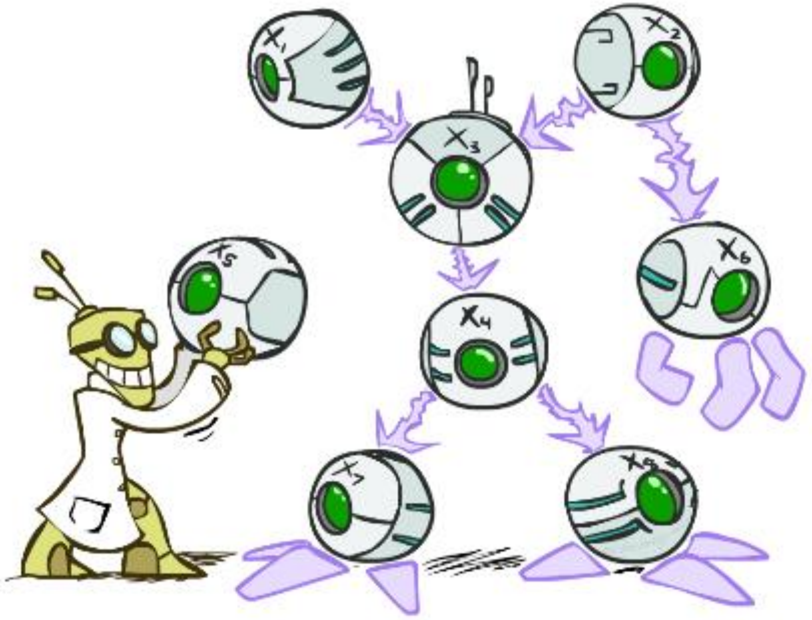
Bayes' Nets: Big Picture

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

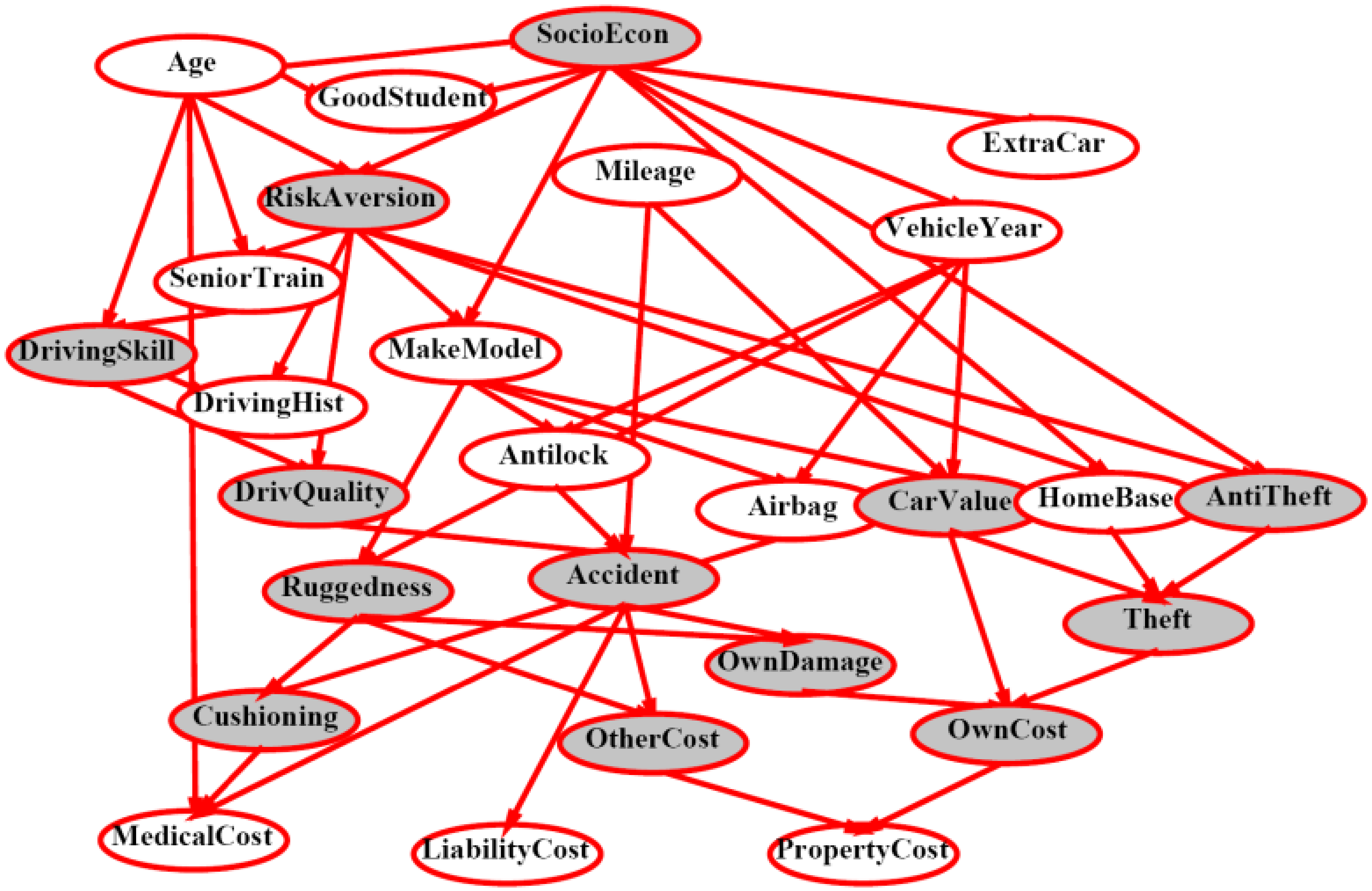
- Unless there are only a few variables, the joint is WAY too big to represent explicitly
- Hard to learn (estimate) anything empirically about more than a few variables at a time

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

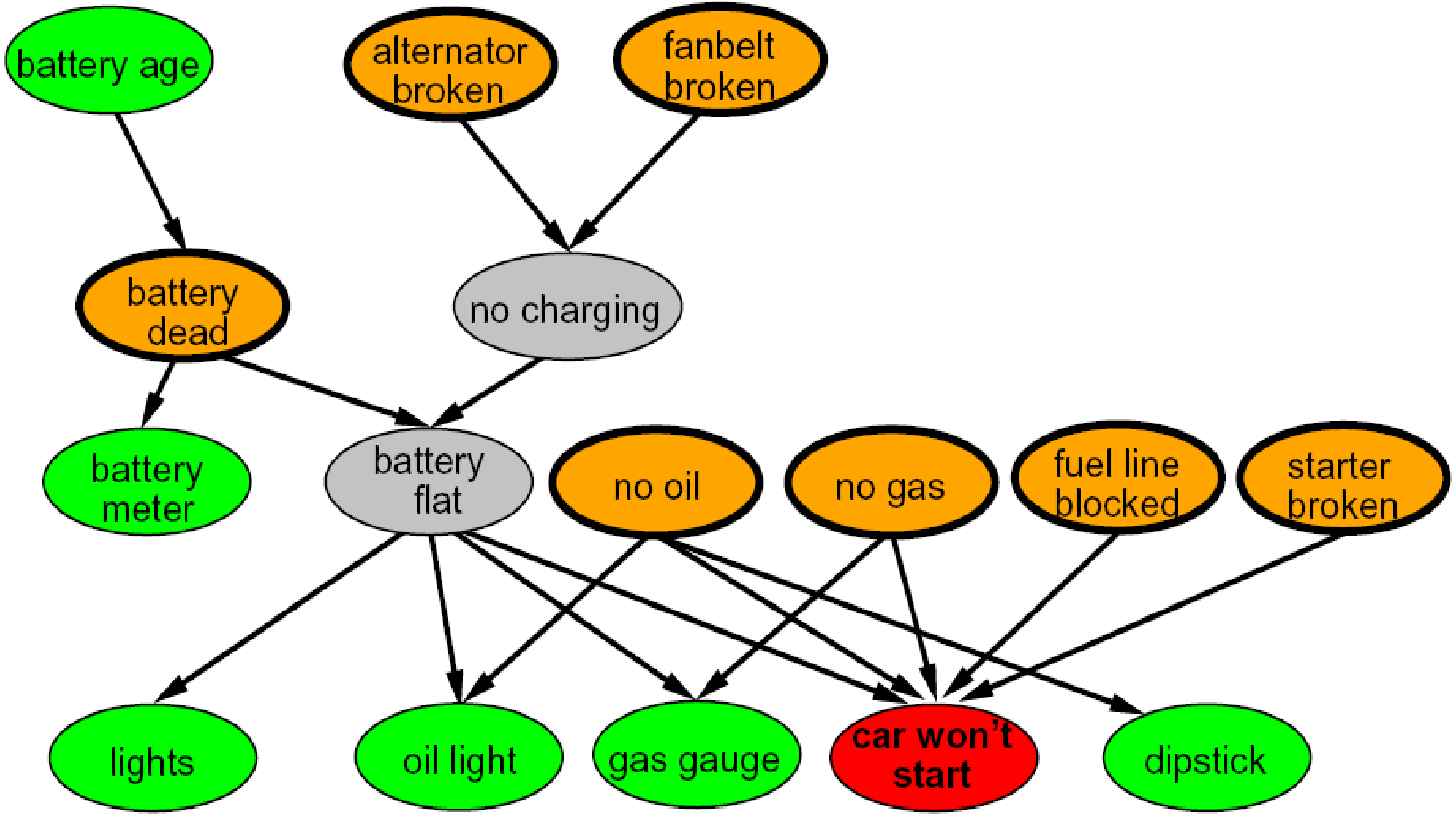
- More properly called graphical models
- We describe how variables locally interact
- Local interactions chain together to give global, indirect interactions



Example Bayes' Net: Insurance



Example Bayes' Net: Car



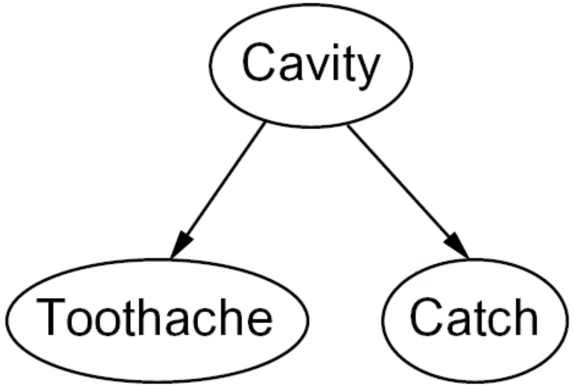
Graphical Model Notation

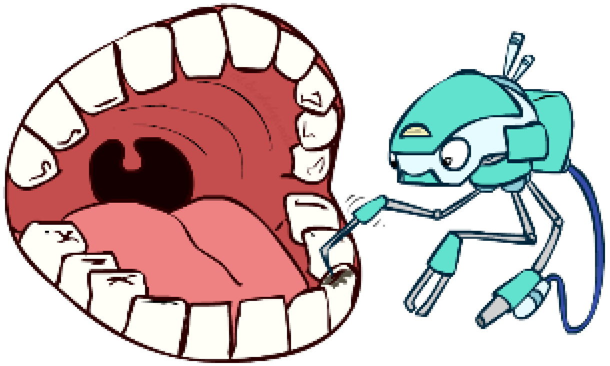
Nodes: variables (with domains)

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

Arcs: interactions

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





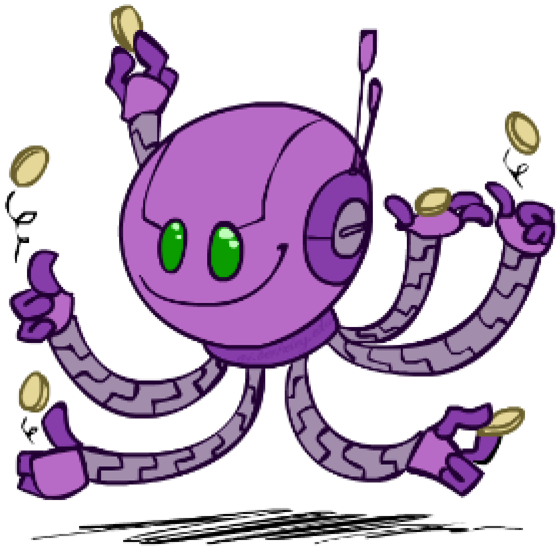
P(Cav, Catch, ToothAche)

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

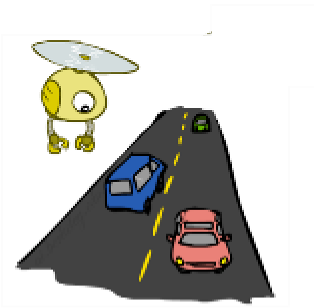


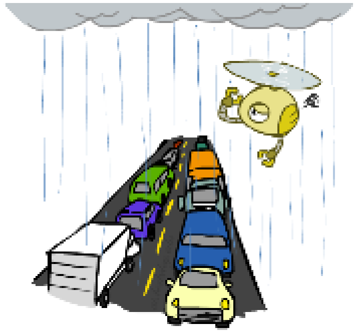
No interactions between variables: absolute independence

Example: Traffic

Variables:

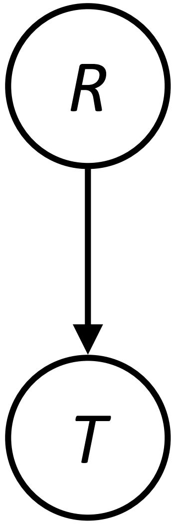
- R: It rains
- T: There is traffic





Model 1: independence

- Model 2: rain causes traffic



Why is an agent using model 2 better?

Example: Traffic II

Let's build a causal graphical model!

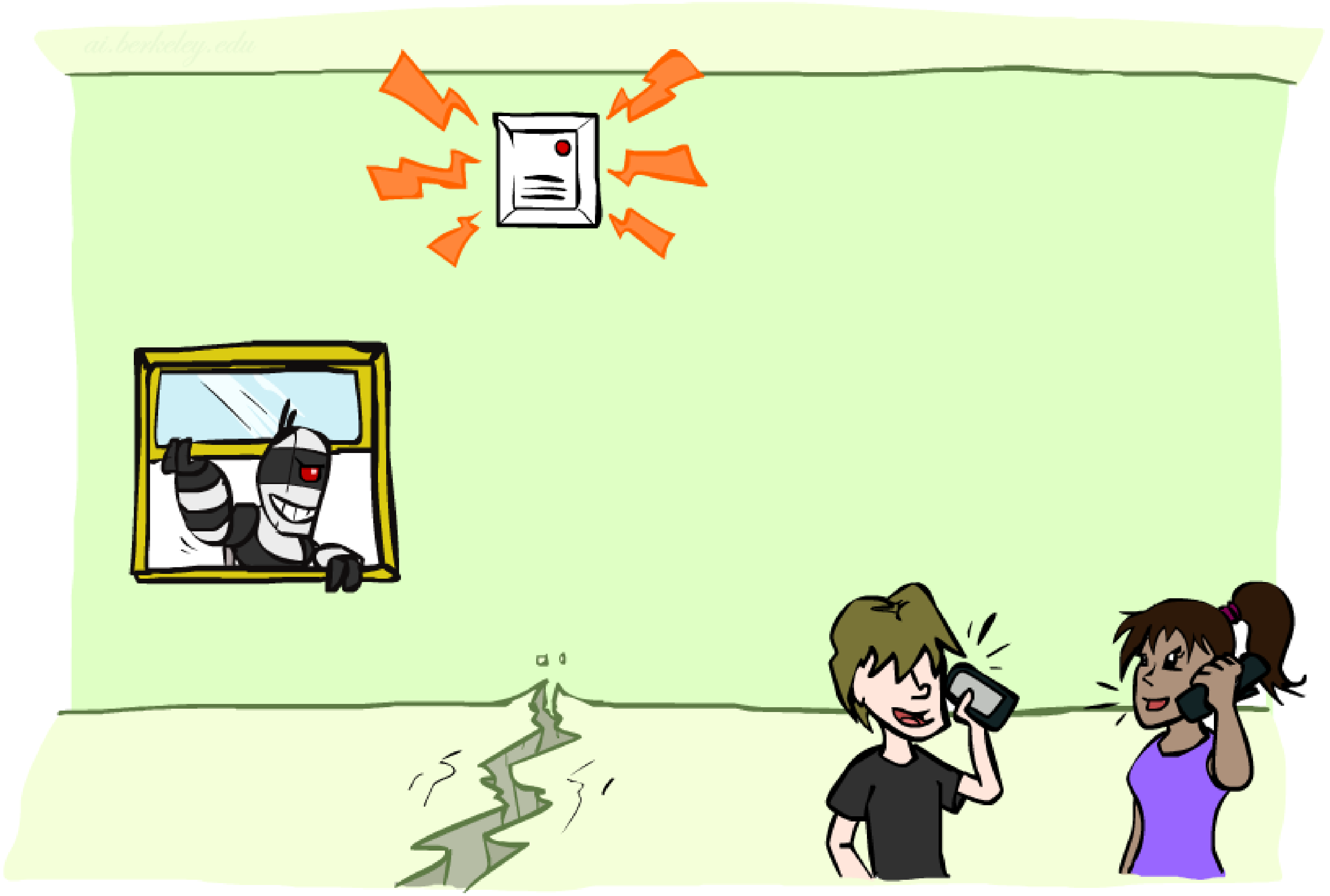
Variables

- 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

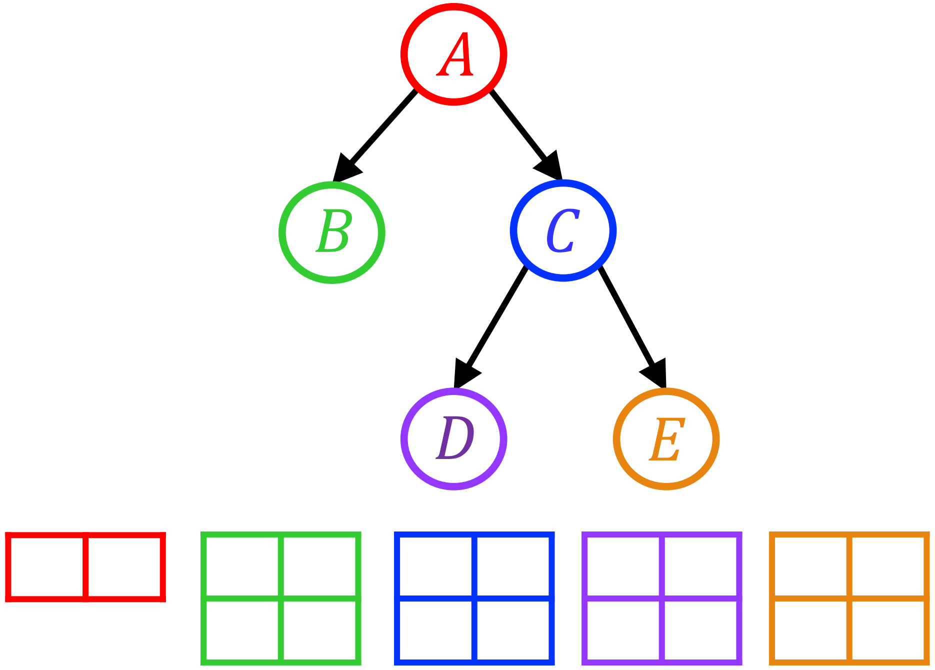
Bayes Nets Syntax Review

One node per random variable

DAG

One CPT per node: P(node | *Parents*(node))

Bayes net



Bayes Net Global Semantics

Bayes nets:

- Encode joint distributions as product of conditional distributions on each variable

$$P(X_1 \ldots X_2) = \prod_i P(X_i \mid Parents(X_i))$$

Semantics Example

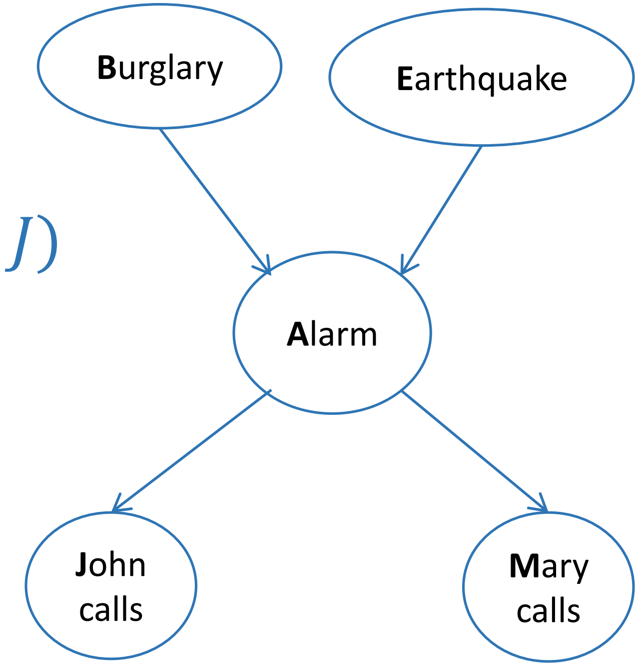
Joint distribution factorization example

Generic chain rule

- $P(X_1 \dots X_2) = \prod_i P(X_i | X_1 \dots X_{i-1})$

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

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



Bayes nets

- $P(X_1 \dots X_2) = \prod_i P(X_i | Parents(X_i))$

Example: Coin Flips

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

$P(X_1)$

h	0.5
t	0.5

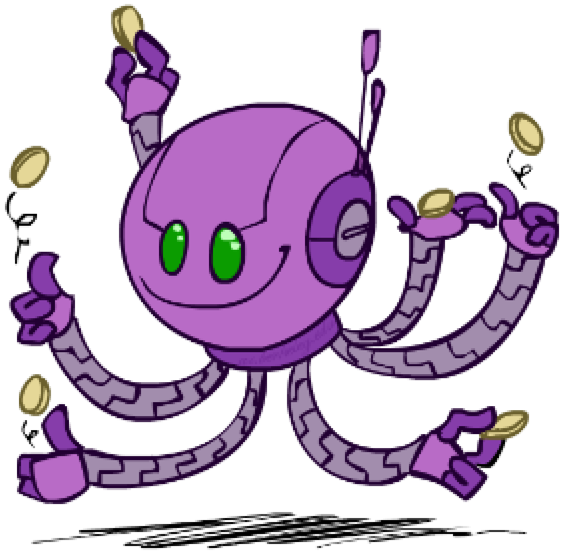
$P(X_2)$

h	0.5
t	0.5

$\cdots$

$P(X_n)$

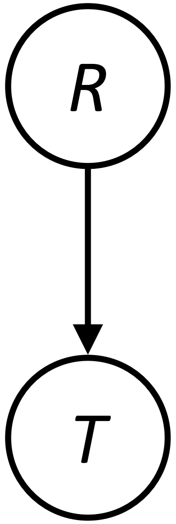
h	0.5
t	0.5



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

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

Example: Traffic



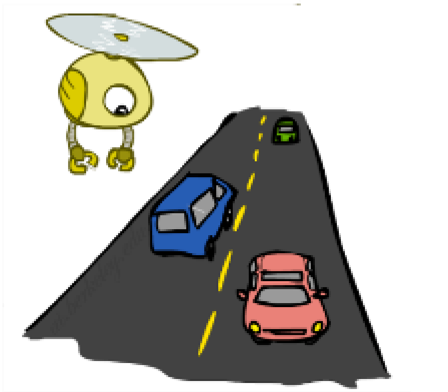
$$P(R)$$

+r	1/4
-r	3/4

$$P(T|R)$$

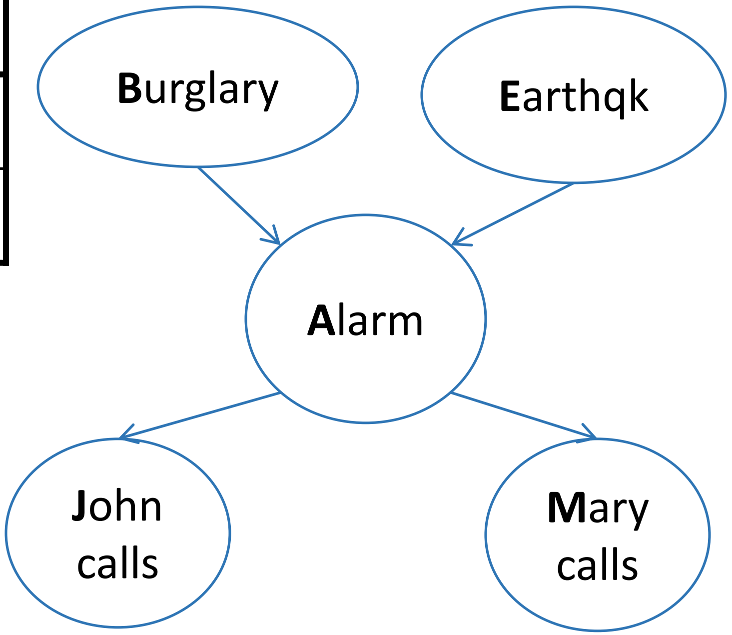
+r	+t	3/4
	-t	1/4
-r	+t	1/2
	-t	1/2

$$P(+r, -t) =$$

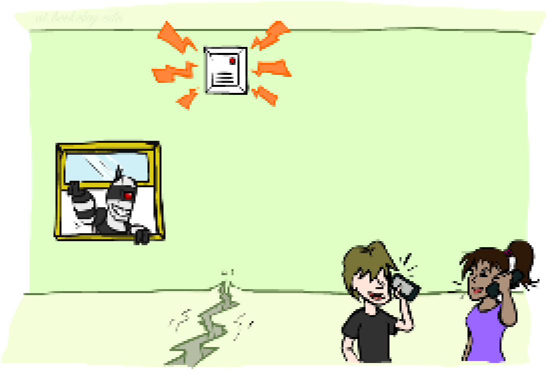


Example: Alarm Network

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