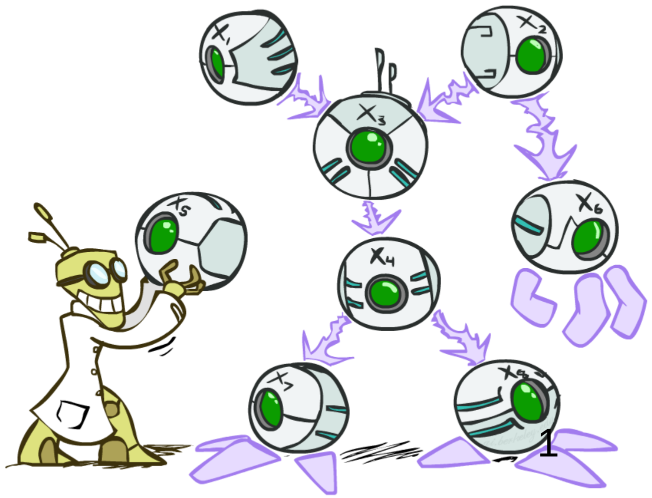
CSE 573 PMP: Artificial Intelligence

Hanna Hajishirzi

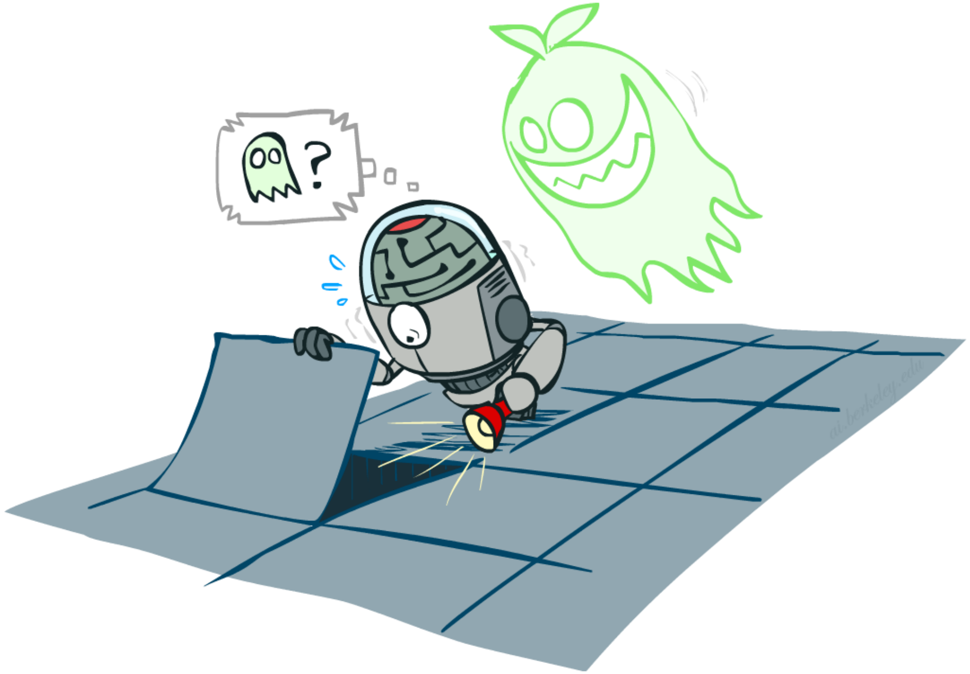
Uncertainty and Bayes Nets

slides adapted from
Dan Klein, Pieter Abbeel ai.berkeley.edu
And Dan Weld, Luke Zettlemoyer



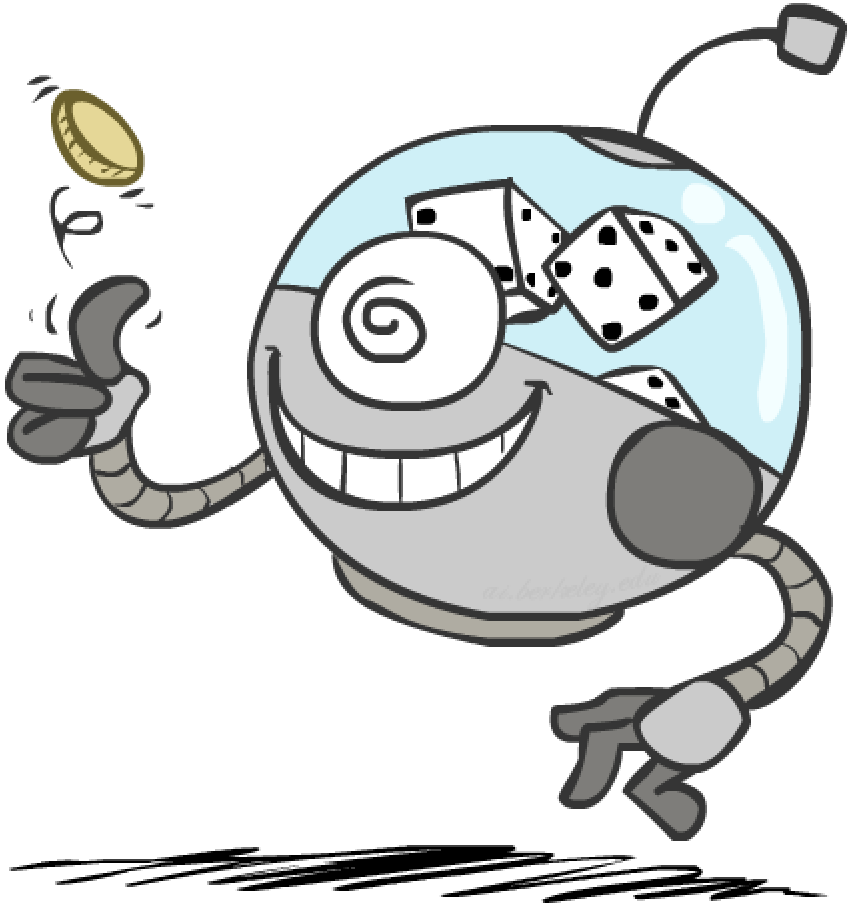
Our Status in CSE573

- We' re done with Search and planning
- We are done with learning to make decisions
- Probabilistic Reasoning and Machine Learning
  - Diagnosis
  - Speech recognition
  - Tracking objects
  - Robot mapping
  - Genetics
  - Error correcting codes
  - … lots more!



Today

- Probability
- Bayes Nets
- You'll need all this stuff for the next few weeks, so make sure you go over it now!

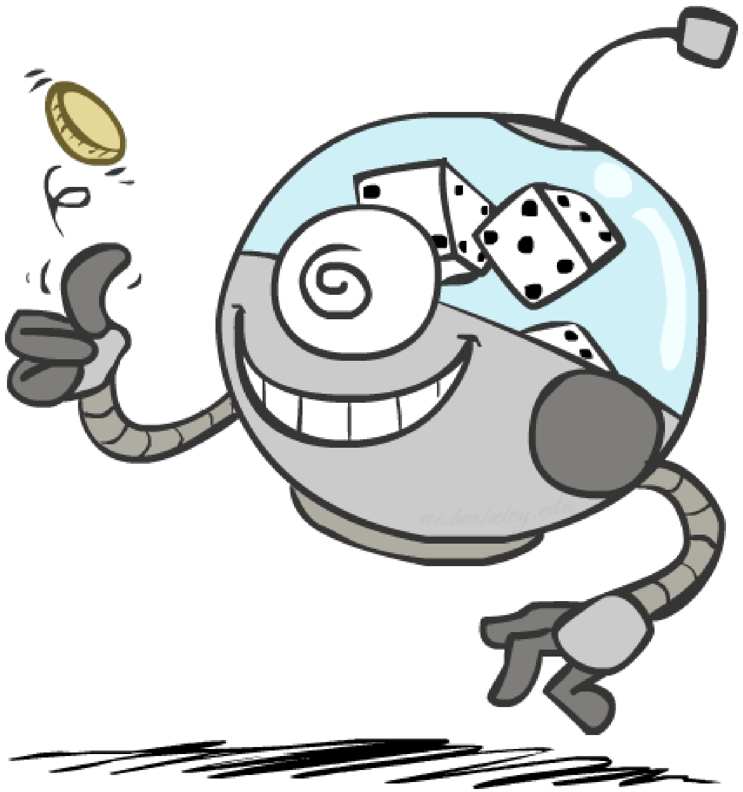


Random Variables

- 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

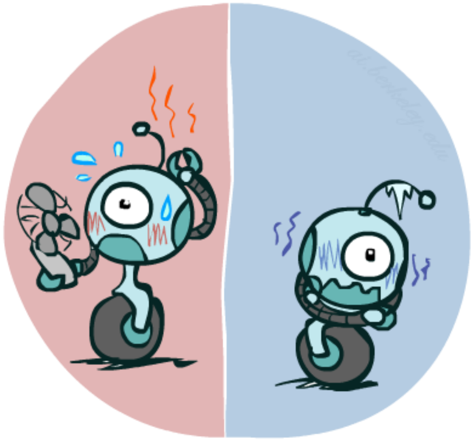
- Random variables have domains
  - R in {true, false}   (often write as {+r, -r})
  - T in {hot, cold}
  - D in $[0, \infty)$
  - L in possible locations, maybe {(0,0), (0,1), …}



Probability Distributions

- Associate a probability with each outcome

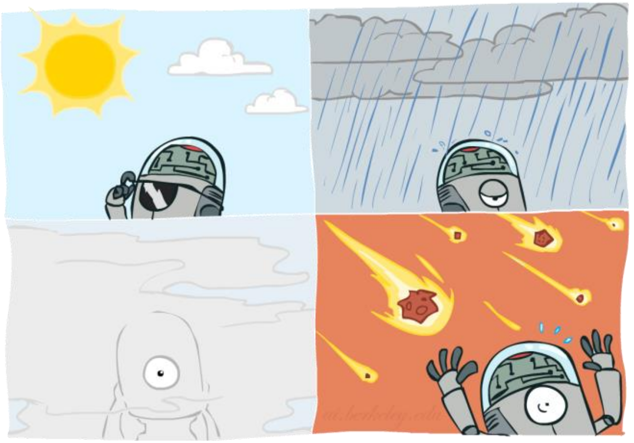
  - Temperature:



$P(T)$

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

- 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

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

- A probability (lower case value) is a single number

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

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

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

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

$$P(T, W)$$

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

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

$n \quad d$

$X_1 \ldots X_n$

$d \times d \times \ldots \times d = d^n$

Events

- An *event* is a set E of outcomes

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

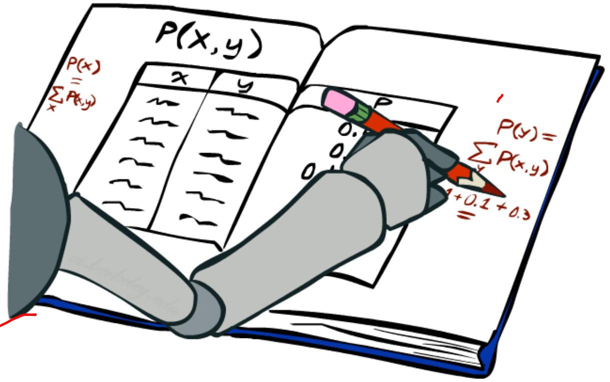
- From a joint distribution, we can calculate the probability of any event
  - Probability that it's hot AND sunny? 0.4
  - Probability that it's hot? 0.5
  - Probability that it's hot OR sunny? 0.7
- Typically, the events we care about are *partial assignments*, like P(T=hot)

$$P(T, W)$$

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

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

$$P(t) = \sum_s P(t, \overset{w}{\cancel{s}})$$

$$P(\overset{w}{\cancel{s}}) = \sum_t P(t, \overset{w}{\cancel{s}})$$

$P(T)$

T	P
hot	0.5
cold	0.5

hot 0.5
cold 0.5

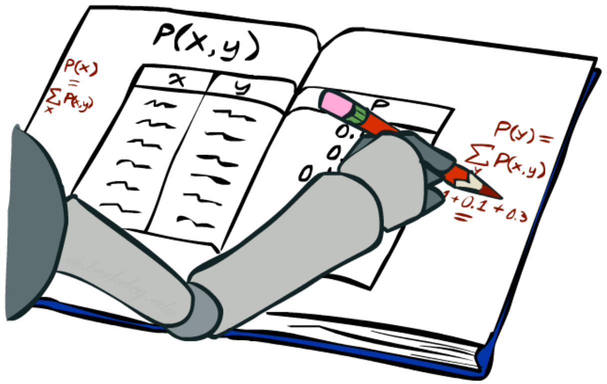
$P(W)$

W	P
sun	0.6
rain	0.4

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

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

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

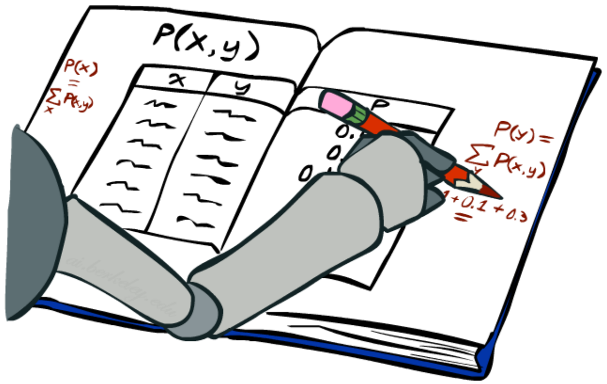
$P(X)$

X	P
+x	
-x	

$P(Y)$

Y	P
+y	
-y	

Quiz: Marginal Distributions



$P(X,Y)$

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

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

$P(X)$

X	P
+x	.5
-x	.5

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

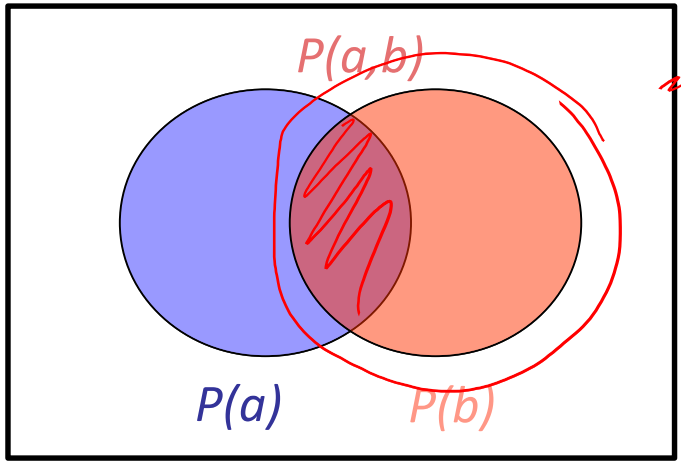
$P(Y)$

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

$P(X, Y)$

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

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

$$\frac{P(-x,+y)}{P(+y)} = \frac{0.4}{0.6}$$

- P(-y | +x) ?

Quiz: Conditional Probabilities

$$P(X, Y)$$

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

- P(+x | +y) ?

  .2/.6=1/3

- P(-x | +y) ?

  .4/.6=2/3

- P(-y | +x) ?

  .3/.5=.6

Conditional Distributions

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

{sun, rain}

Conditional Distributions

P(sun | hot)
p(rain | hot)

$P(W|T)$

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

W	P
sun	
rain	

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

W	P
sun	
rain	

Joint Distribution

$$P(T, W)$$

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

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

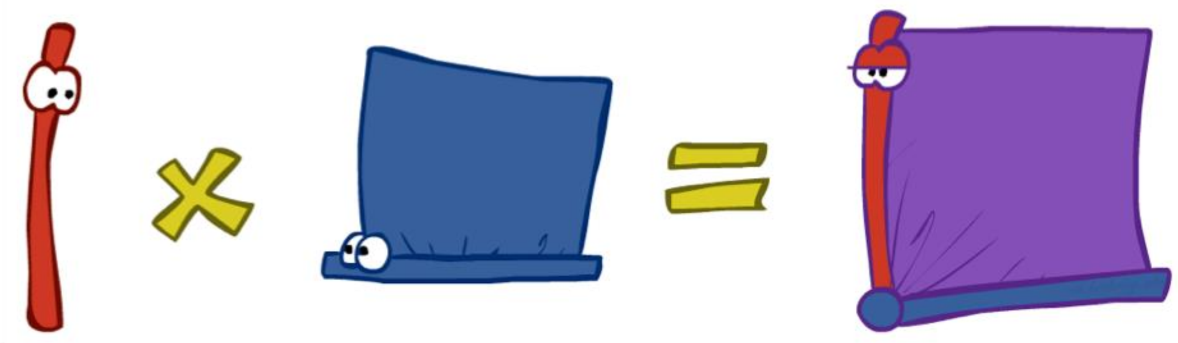
$$P(T, W)$$

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

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

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

- 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	

$P(w|s)\,P(s) = P(w,s)$

The Chain Rule

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

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

$= \underbrace{P(x_1)P(x_2|x_1)}_{P(x_1, x_2)}\, P(x_3|x_2, x_1)$

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

$P(x_1)P(x_2|x_1)P(x_3|x_1x_2) \ldots P(x_i|x_1 \ldots x_{i-1})$ $P(x_1, x_2, x_3)$

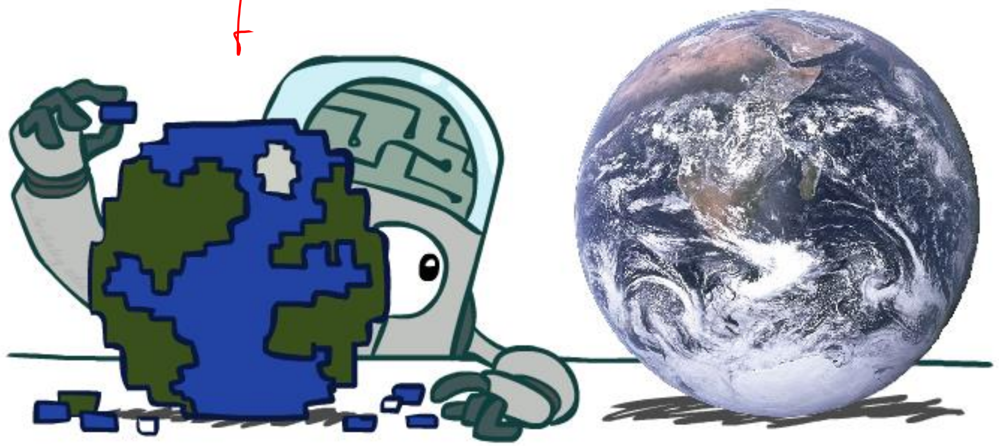
x_2, x_1, x_3, x_4

$P(x_2)P(x_1|x_2)P(x_3|x_1, x_2)P(x_4|x_1 x_2 x_3)$

Probabilistic Models

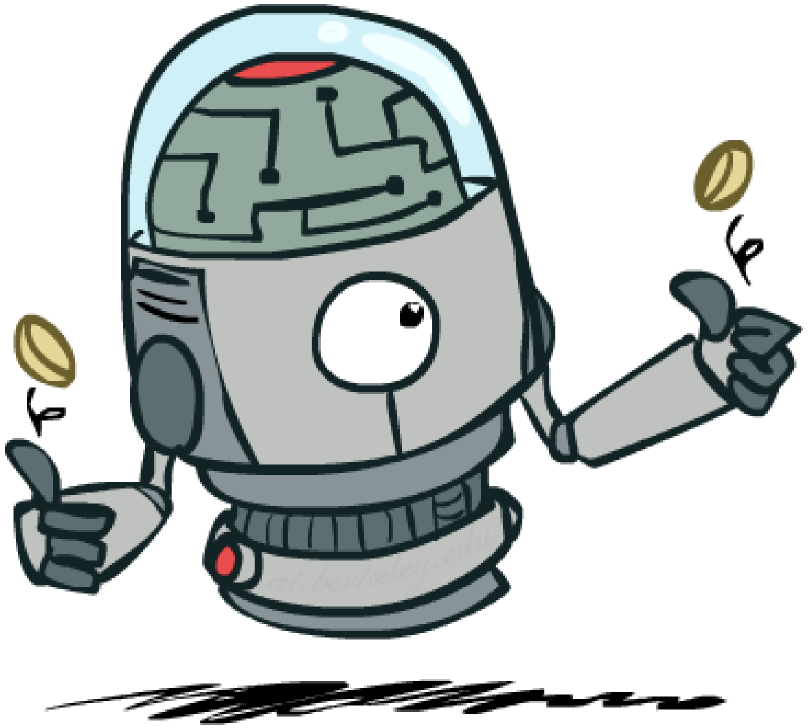
- Models describe how (a portion of) the world works

- 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)

Independence



Independence

- Two variables are *independent* if:

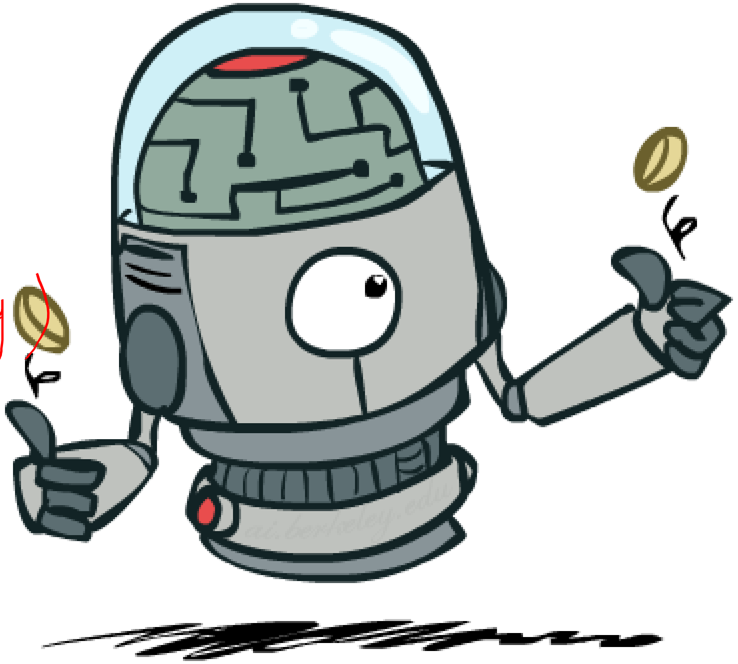
$$\forall x, y : P(x, y) = P(x)P(y)$$

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

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

  - We write: $X \perp\!\!\!\perp Y$

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



- 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

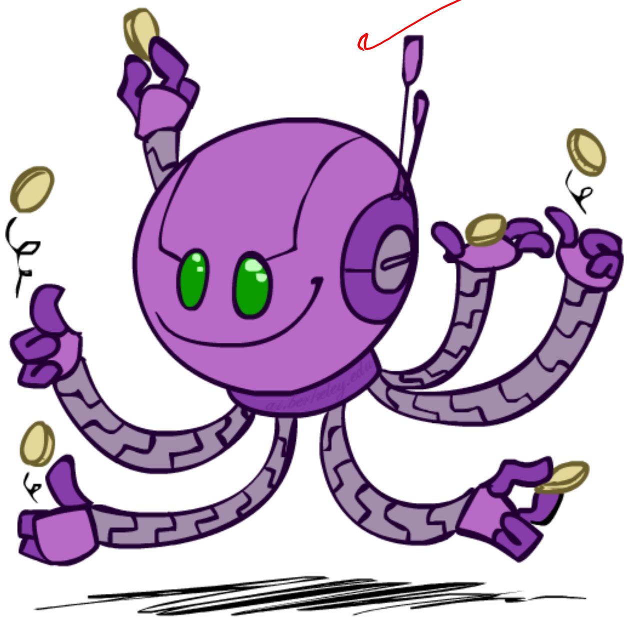
$P(W)$

W	P
sun	0.6
rain	0.4

Example: Independence

- N fair, independent coin flips:

$2 \times n$



$P(X_1)$

H	0.5
T	0.5

$P(X_2)$

H	0.5
T	0.5

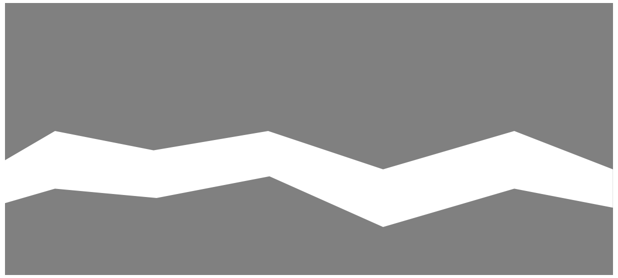
$\ldots$

$P(X_n)$

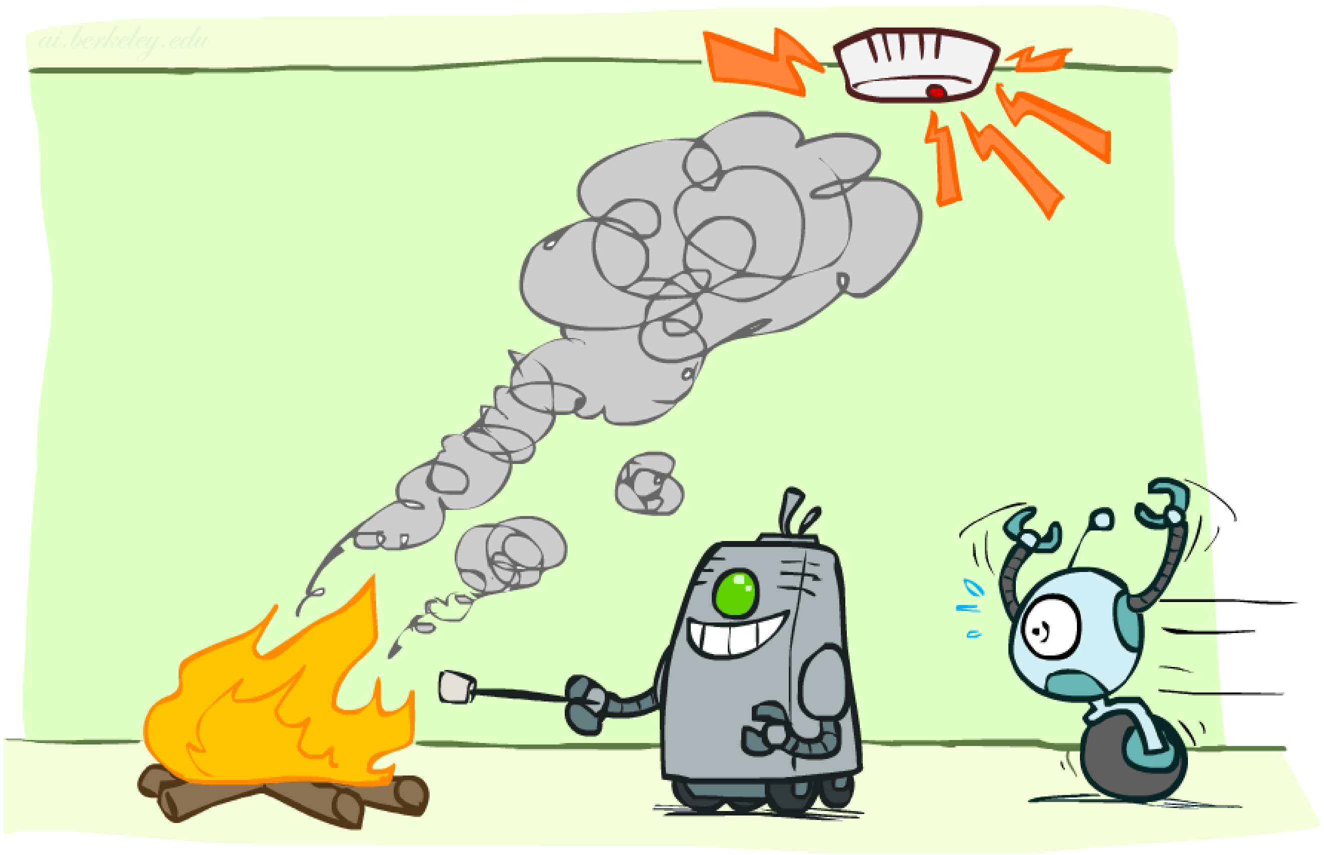
H	0.5
T	0.5

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

2^n

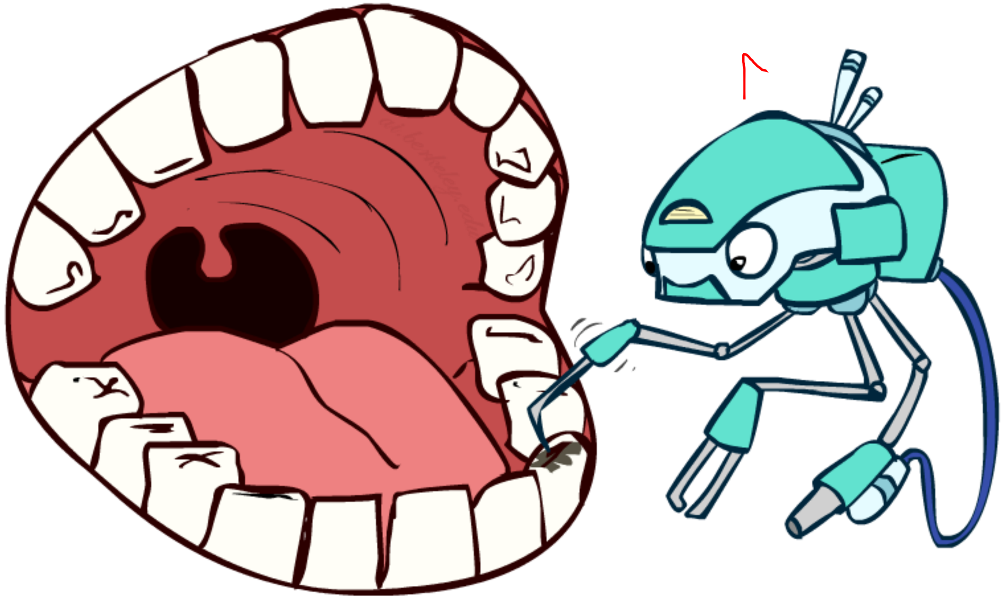


Conditional Independence



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 derive one 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 $X \perp\!\!\!\perp Y | Z$ $X \perp Y | Z$

if and only if:

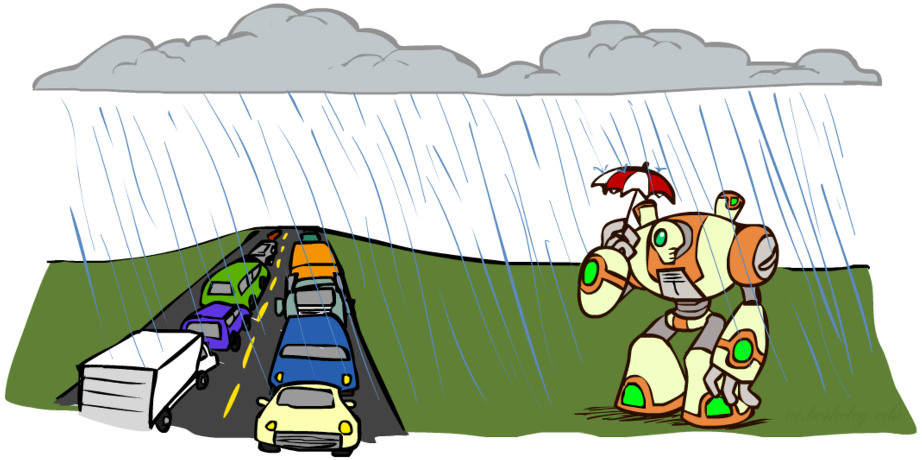
$$\forall x, y, z : P(x, y | z) = P(x|z) P(y|z)$$

or, equivalently, if and only if

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

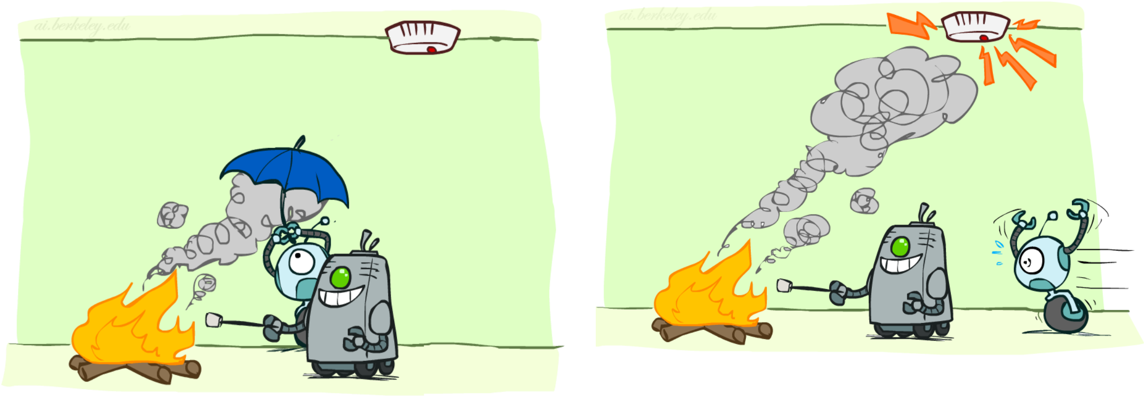
Conditional Independence

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



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

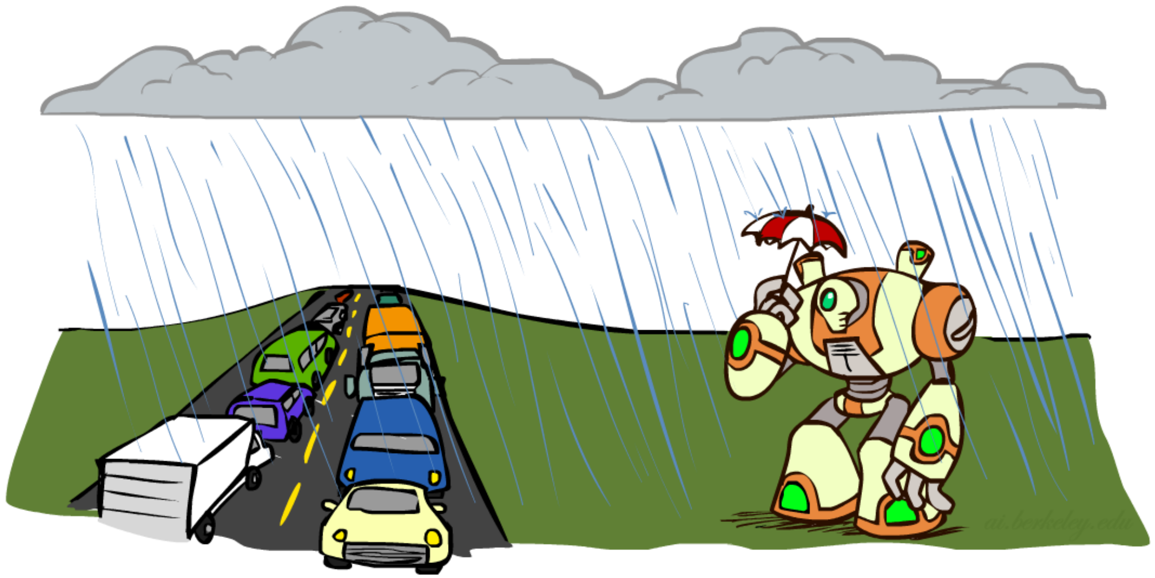
Smoke detects



Conditional Independence

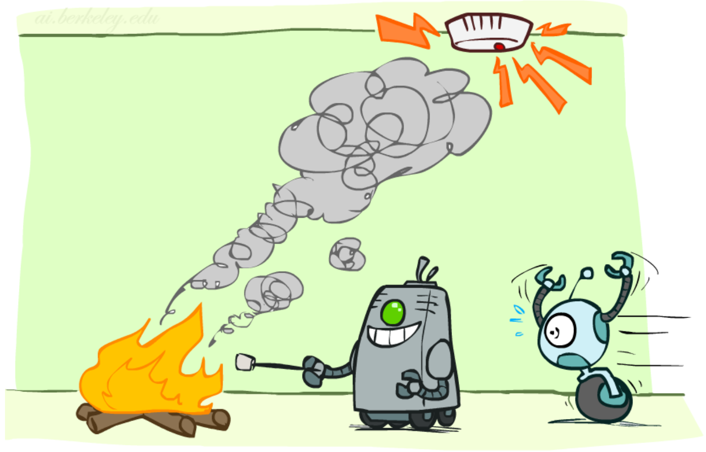
- What about this domain:
  - Traffic
  - Umbrella
  - Raining

Traffic, Umbrella Raining



Conditional Independence

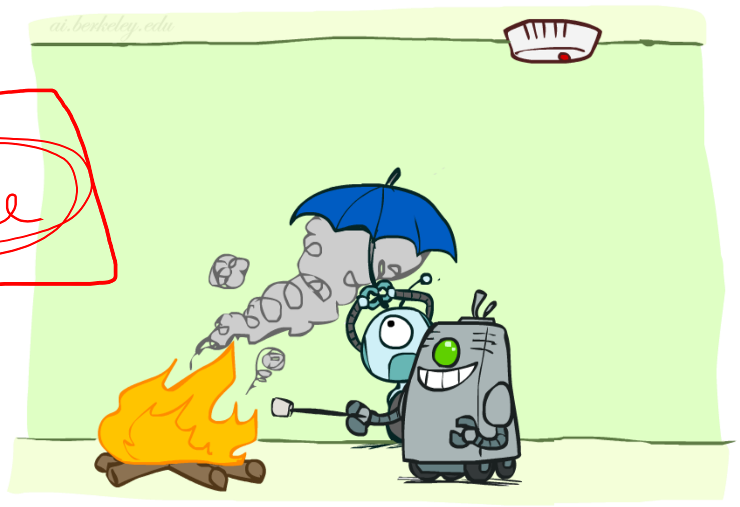
- What about this domain:
  - Fire
  - Smoke
  - Alarm



Smoke

Smoke, alarm | fire

fire, alarm | Smoke

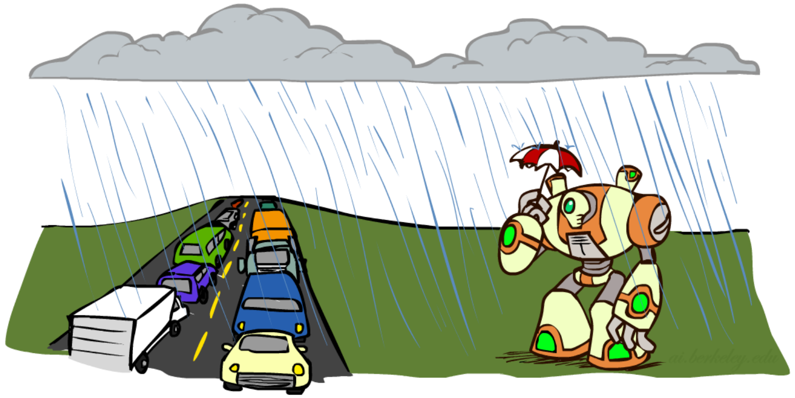


Conditional Independence and the Chain Rule

- Chain rule:

$$P(X_1, X_2, \dots X_n) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)\dots$$

- Trivial decomposition: Rain, traff, um



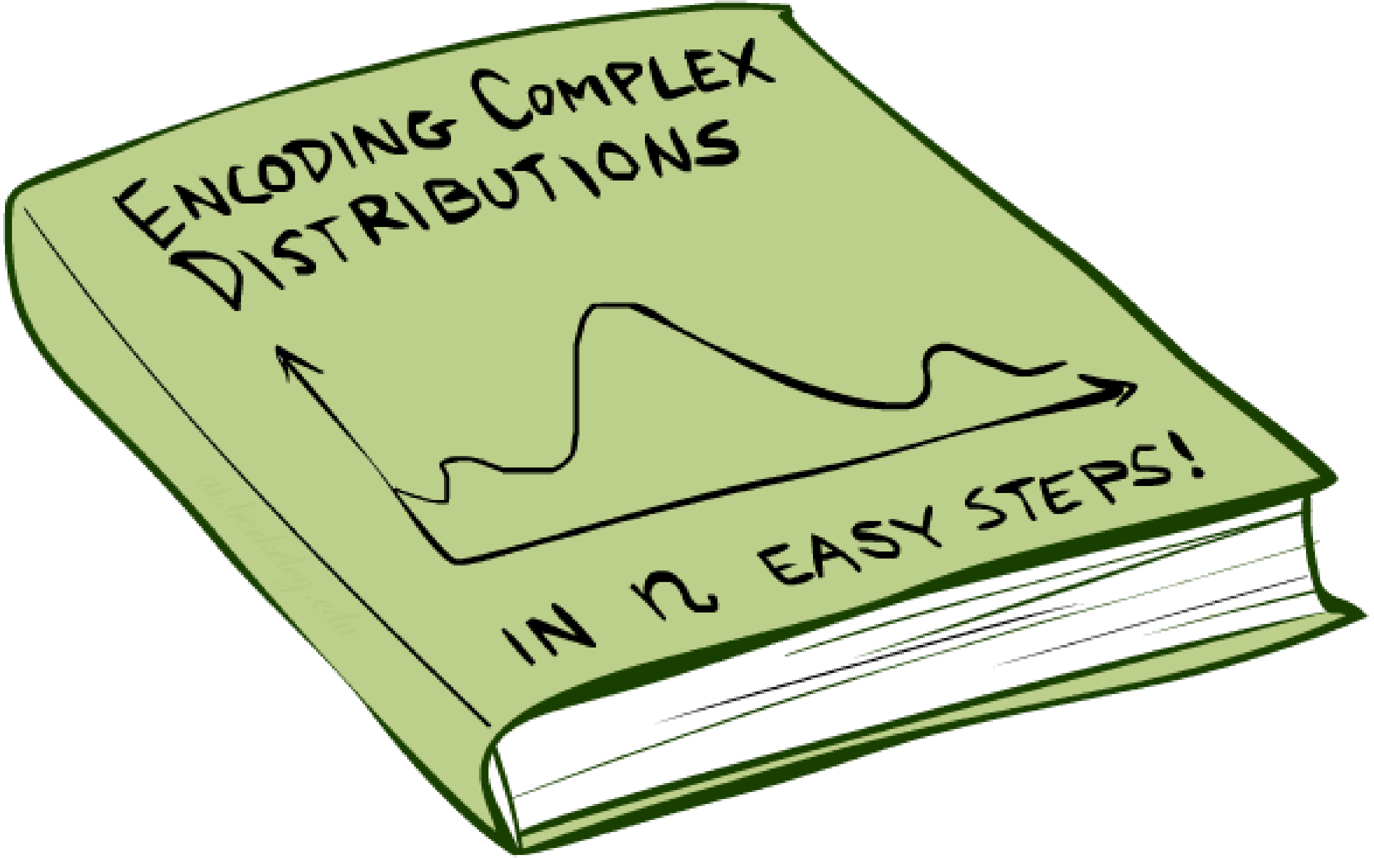
$$P(\text{Traffic}, \text{Rain}, \text{Umbrella}) =$$
$$P(\text{Rain})P(\text{Traffic}|\text{Rain})P(\text{Umbrella}|\text{Rain}, \text{Traffic})$$

- With assumption of conditional independence:

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

- We can represent joint distributions by multiplying these simpler local distributions.
- Bayes’nets / graphical models help us express conditional independence assumptions

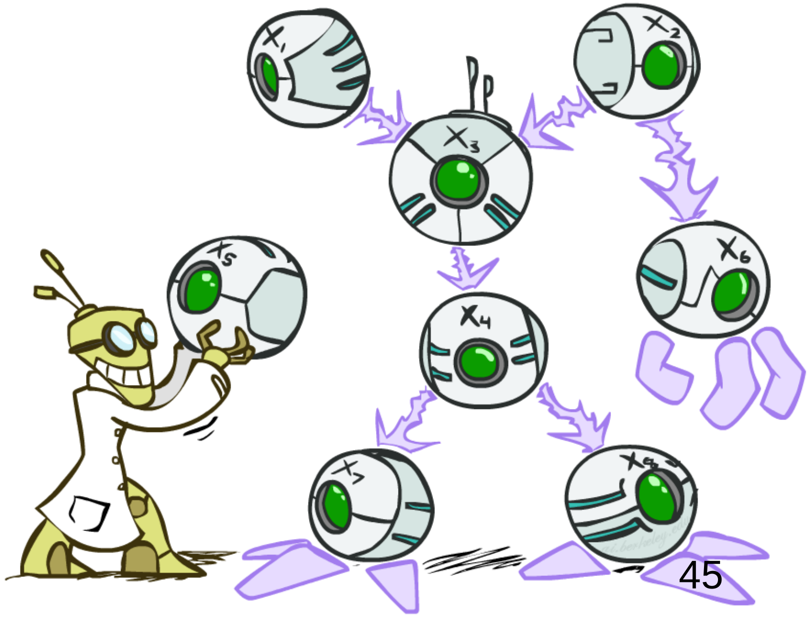
Bayes' Nets: Big Picture



Bayes' Nets: Big Picture

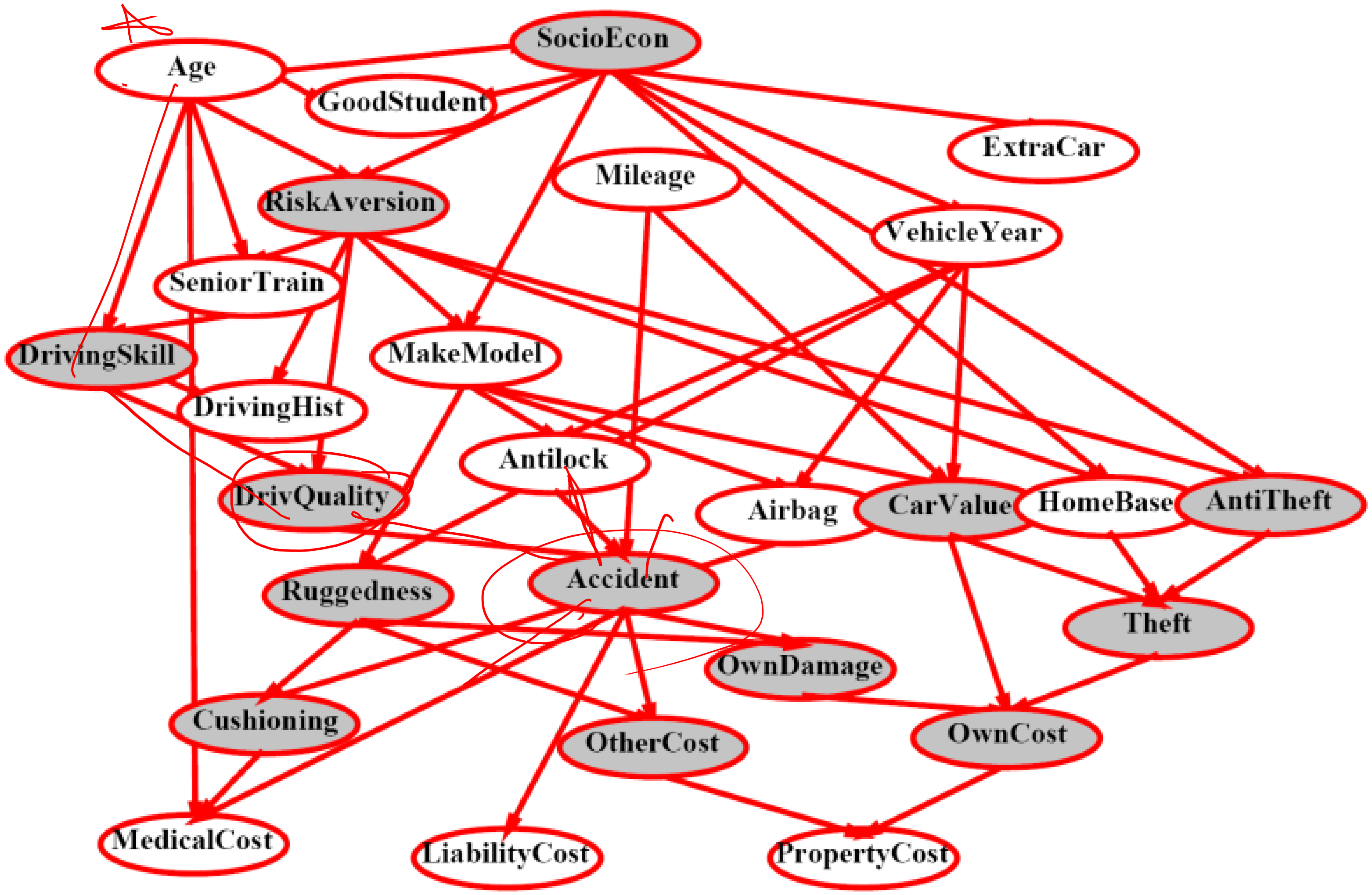
$O(d^n)$

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

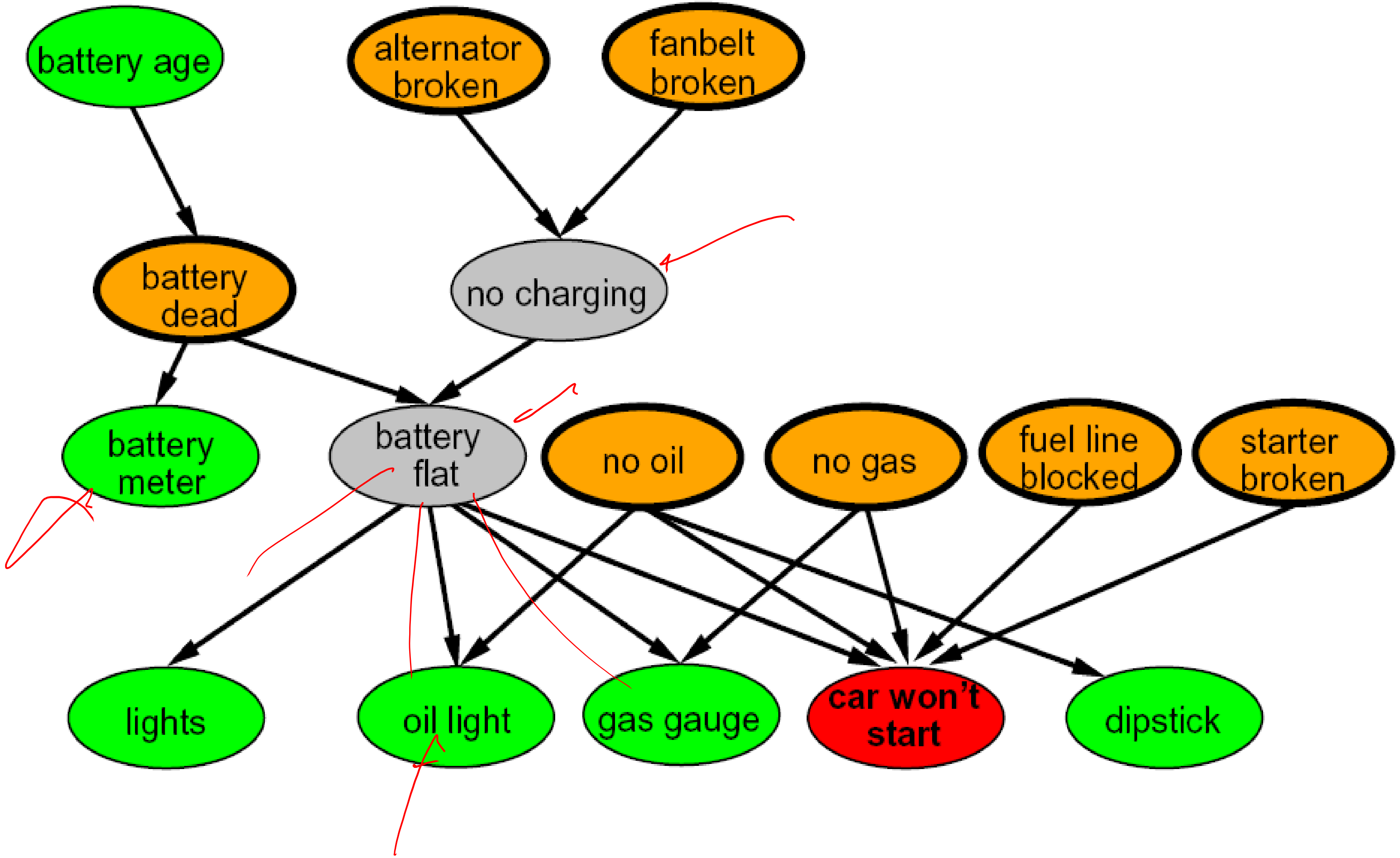


- Bayes' nets: a technique for describing complex joint distributions (models) using simple, local distributions (conditional probabilities)
  - More properly called graphical models
  - We describe how variables locally interact
  - ~~Local interactions~~ chain together to give global, indirect interactions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

Example Bayes' Net: Insurance



Example Bayes' Net: Car



Announcements

- Classes next week

Reviews

+

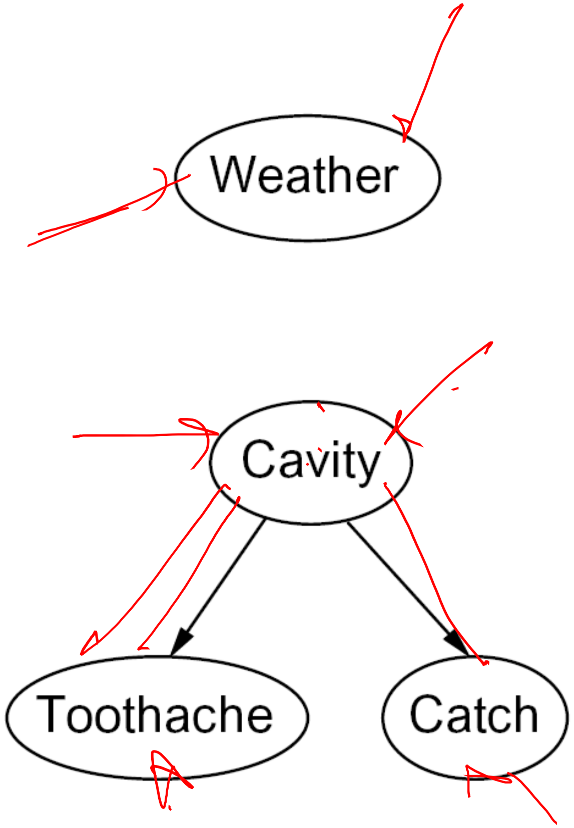
- lectures, slides, homework, assignments
- Instructor and TA knowledge
- remote classes

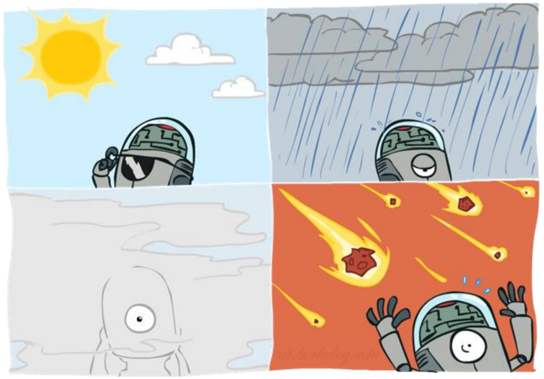
-

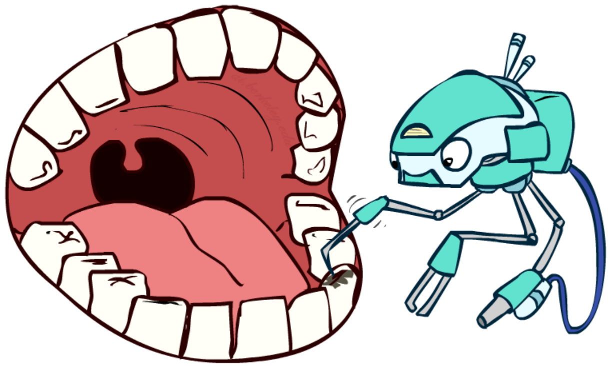
- workload
- Breaks in the middle of lectures
- lack of clarity about online vs. in-person
- lecture notes
- TA office hours or lecture breakout sessions be more organized
- remote classes

Graphical Model Notation

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

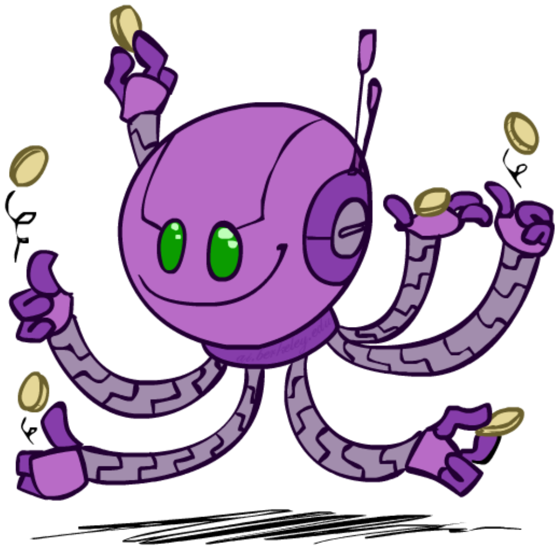






Example: Coin Flips

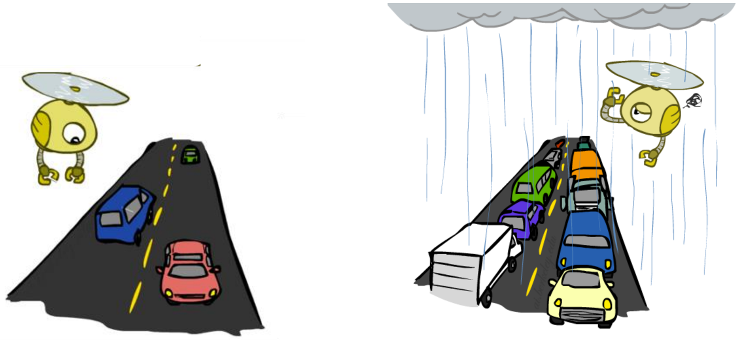
- N independent coin flips



X_1 X_2 $\cdots$ X_n

- No interactions between variables: absolute independence

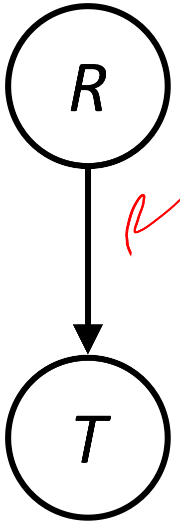
Example: Traffic



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

- Model 1: independence

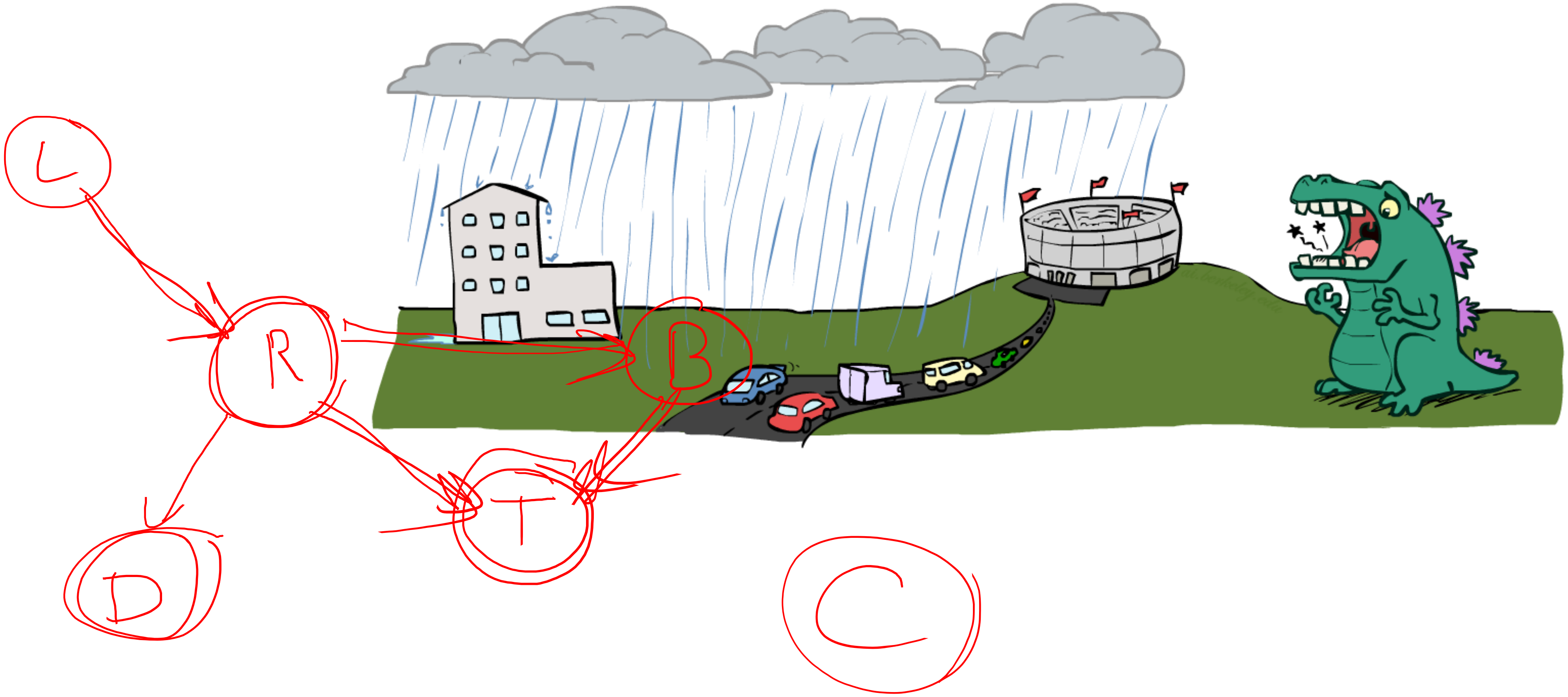
- Model 2: rain causes traffic



- Why is an agent using model 2 better?

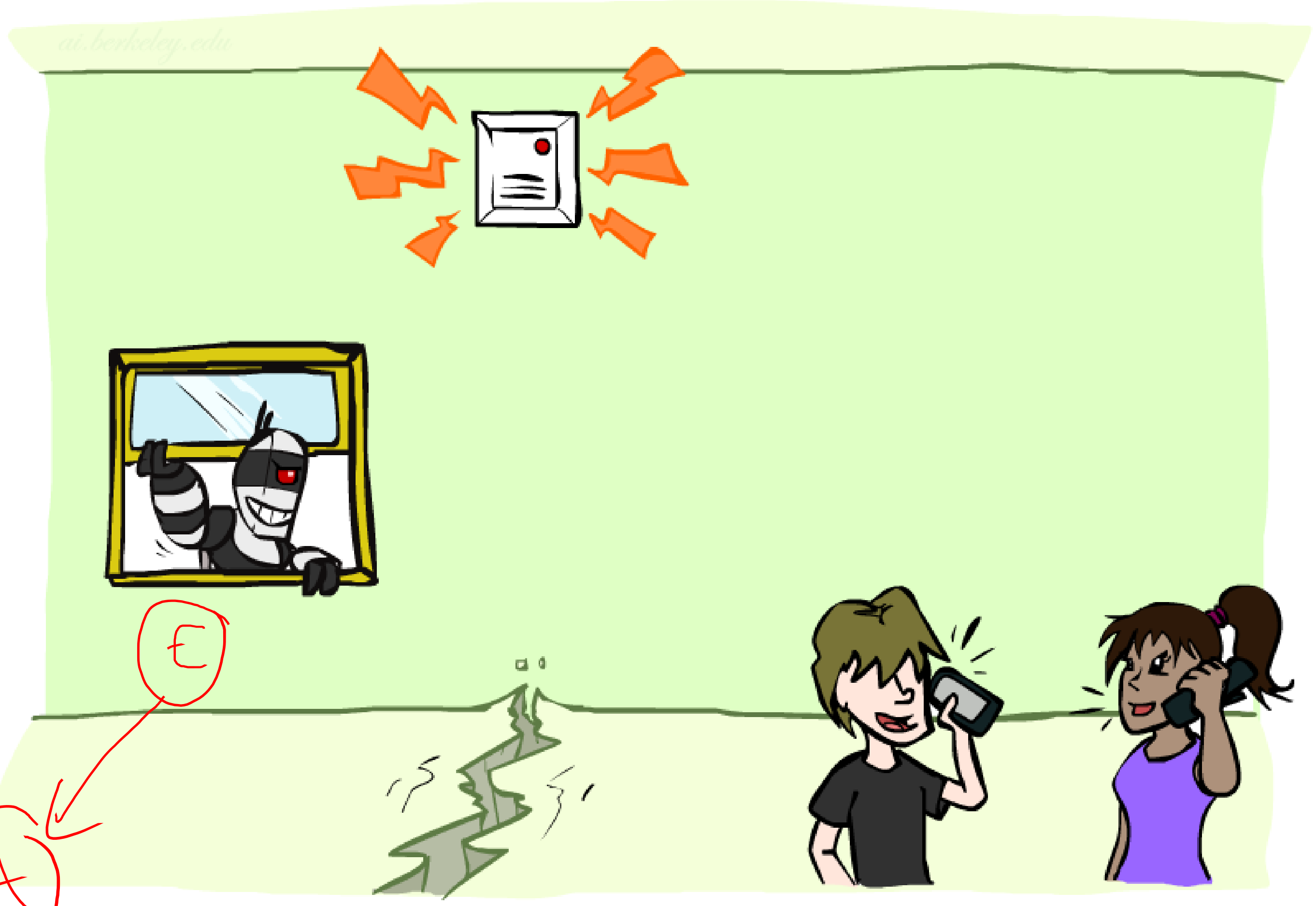
Example: Traffic II

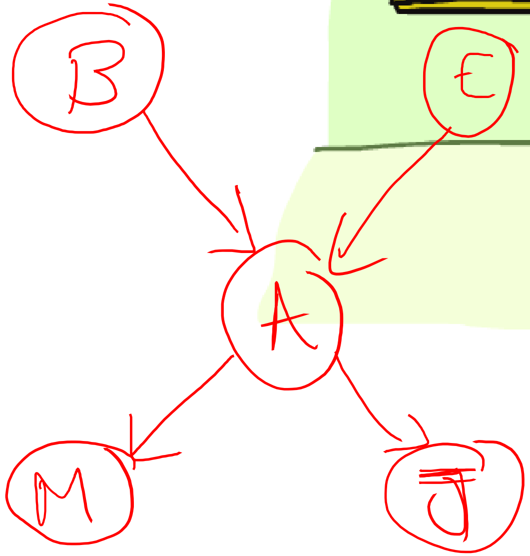
- Variables
  - T: Traffic
  - R: It rains
  - L: Low pressure
  - D: Roof drips
  - B: Ballgame
  - C: Cavity



Example: Alarm Network

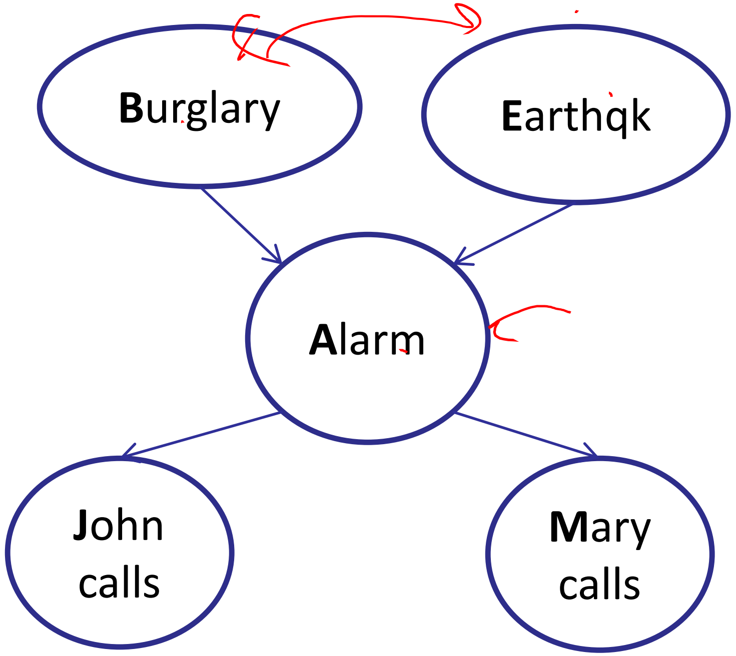
- Variables
  - B: Burglary
  - A: Alarm goes off
  - M: Mary calls
  - J: John calls
  - E: Earthquake!

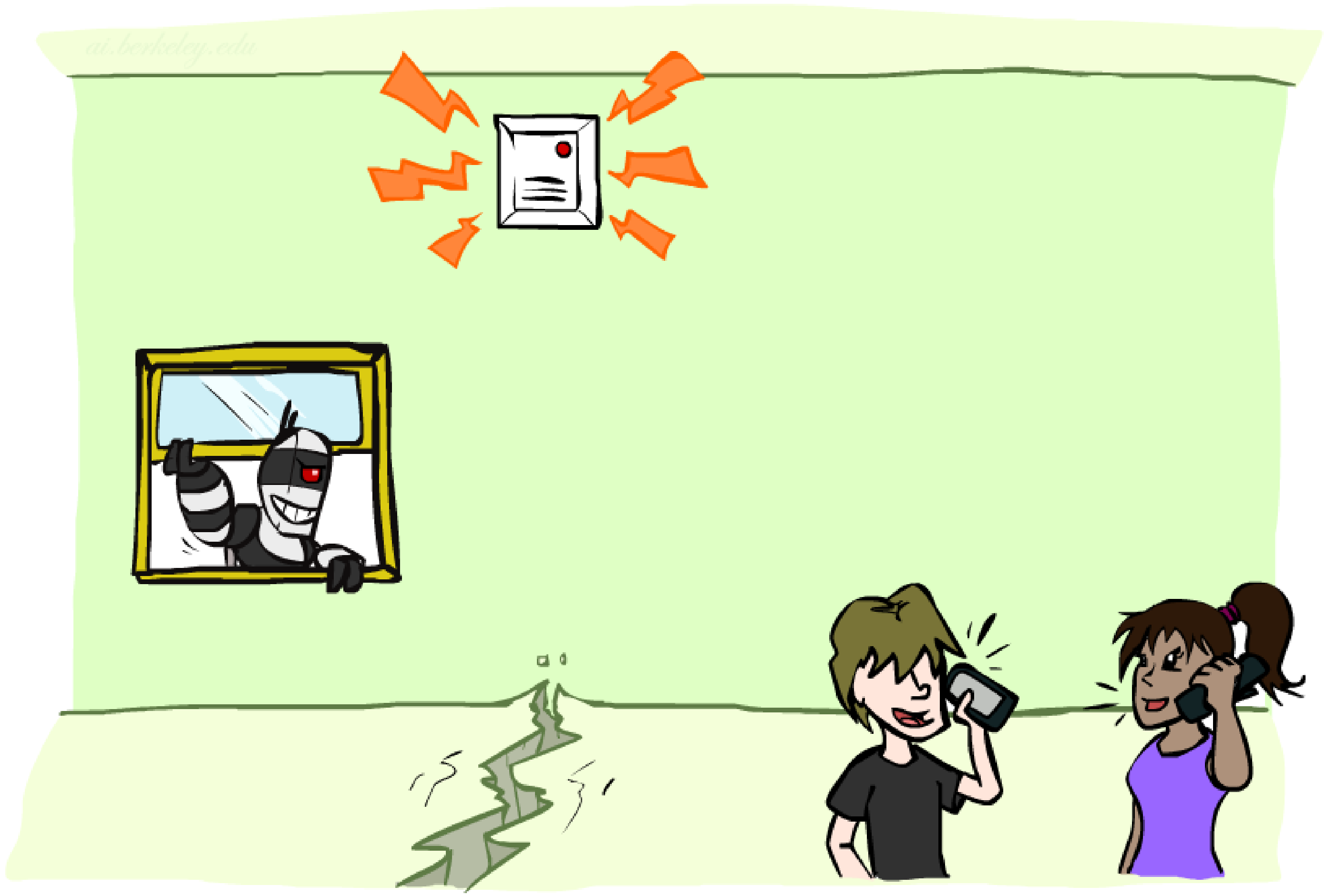




Example: Alarm Network

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





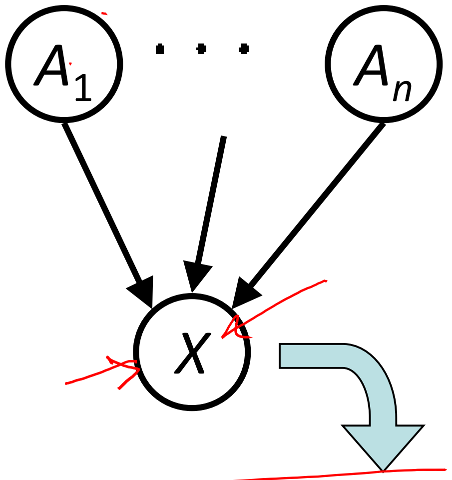
Bayes' Net Semantics

Bayes' Net Semantics

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

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

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



$$\boxed{P(X|A_1 \ldots A_n)}$$

$P(X|a_1 \ldots a_n)$

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

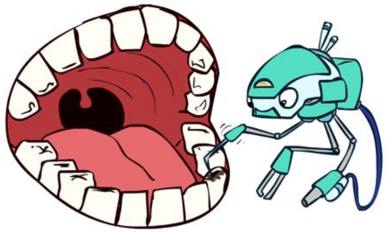
Probabilities in BNs

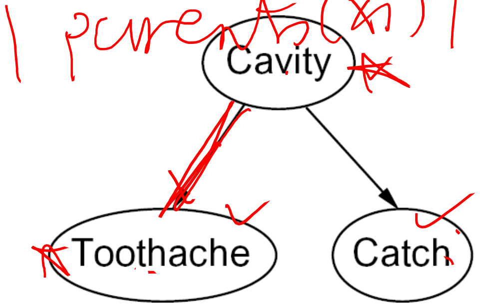
- Bayes' nets implicitly encode joint distributions
  - As a product of local conditional distributions
  - To see what probability a BN gives to a full assignment, multiply all the relevant conditionals together:

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

$P(x_1 | parents(x_1)) P(x_2 | Par \ldots)$

- Example:





$P(C$

$P(+cavity, +catch, -toothache) = P(+cavity) P(-t | +c) P(+catch | +cav)$

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

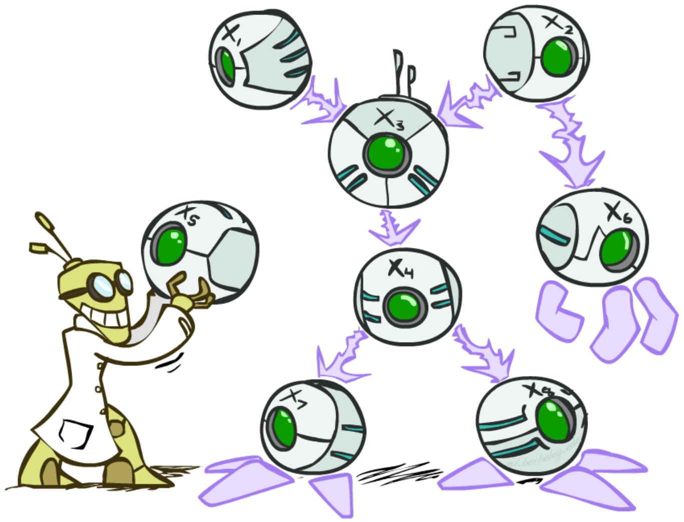
Bayes' Net Representation

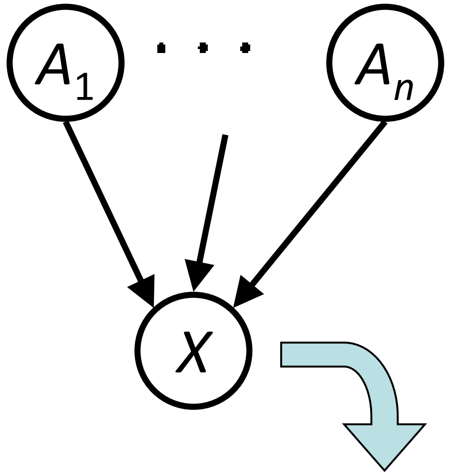
- A directed, acyclic graph, one node per random variable
- A conditional probability table (CPT) for each node
  - A collection of distributions over X, one for each combination of parents' values

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

- 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:

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





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

$P(X_1, X_2, X_3) = P(X_1) P(X_2|X_1) P(X_3|X_2 X_1)$

Probabilities in BNs

DAG

Conditional indepen

- Why are we guaranteed that setting

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

results in a proper joint distribution?

$\prod_{i=1}^{n} P(X_i|X_1 \ldots X_{i-1})$

parent(X_i)

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

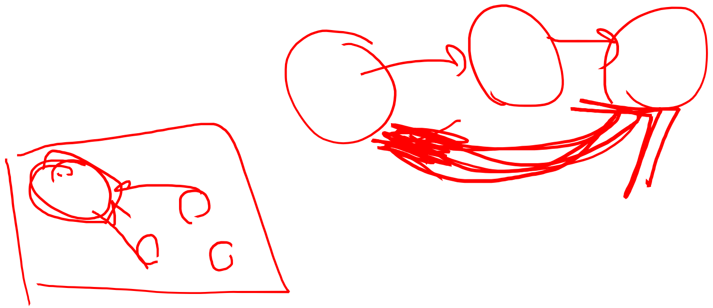
X_i

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

$X_1 \ldots X_n$

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

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



Example: Coin Flips

X_1 X_2 . . . X_n

$P(X_1)$

h	0.5
t	0.5

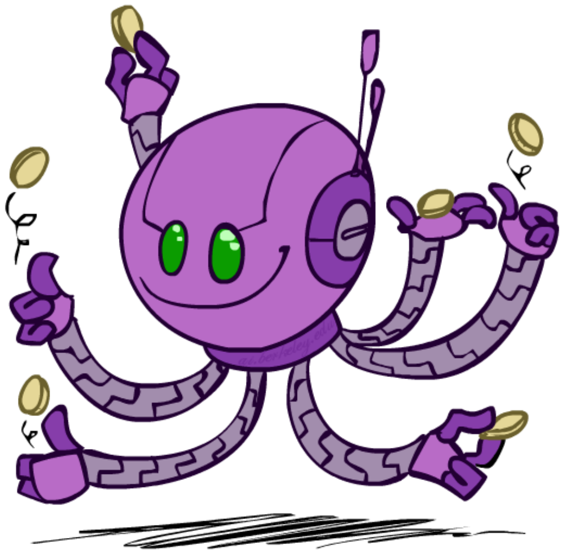
$P(X_2)$

h	0.5
t	0.5

. . .

$P(X_n)$

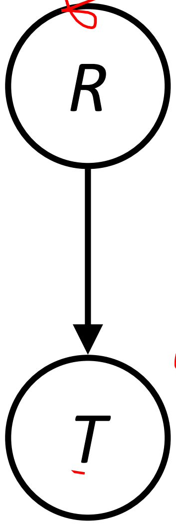
h	0.5
t	0.5



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

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

Example: Traffic



$P(R)$

+r	1/4
-r	3/4

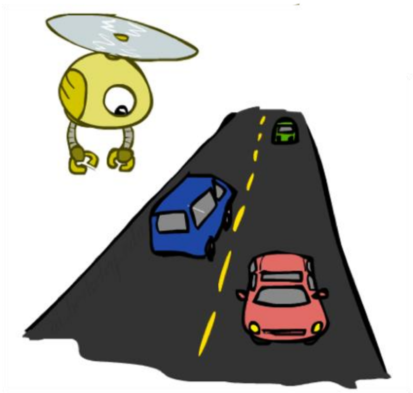
$P(T|R)$

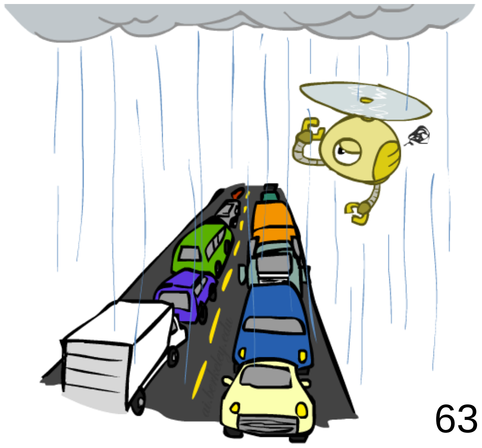
+r	+t	3/4
	-t	1/4
-r	+t	1/2
	-t	1/2

P(t|r)
P(-t|r)
P(t|-r)
P(-t|-r)

$P(+r, -t) =$ P(+r)P(-t|+r) = ¼*1/4

P(+r) P(-t|+r)

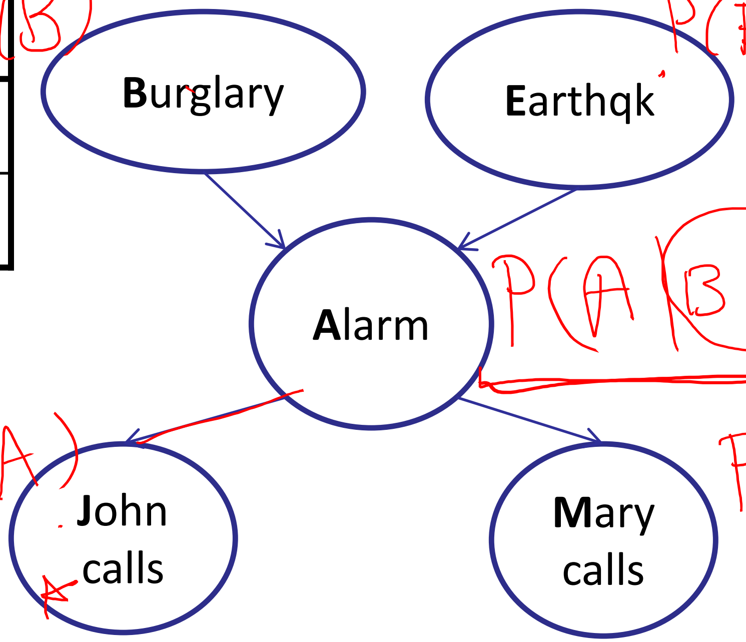




Example: Alarm Network

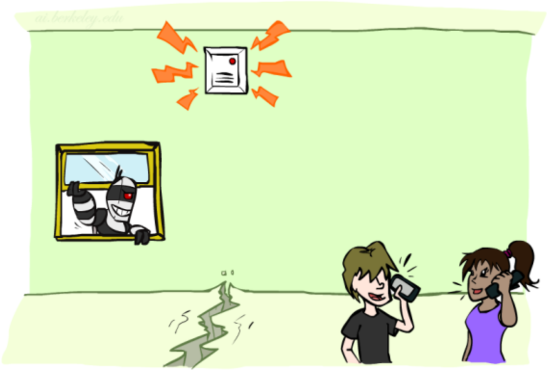
B	P(B)
+b	0.001
-b	0.999

P(B)



P(E)

E	P(E)
+e	0.002
-e	0.998



P(A|B,E)

P(J|A)

P(M|A)

2 2 2

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)P(E)

P(A|B,E)

P(J|A) P(M|A)

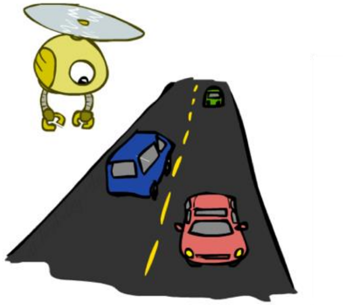
A	J	P(J|A)
+a	+j	0.9
+a	-j	0.1
-a	+j	0.05
-a	-j	0.95

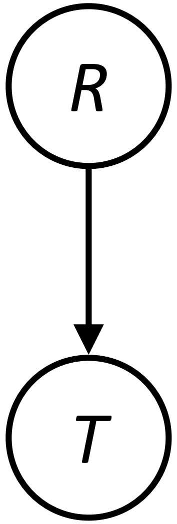
A	M	P(M|A)
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99

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

Example: Traffic

- Causal direction





$P(R)$

+r	1/4
-r	3/4

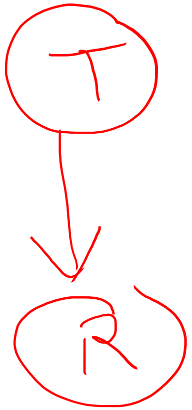
$P(T|R)$

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

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

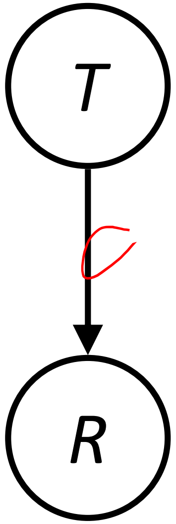
$P(T, R)$

+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

$P(R|T)$

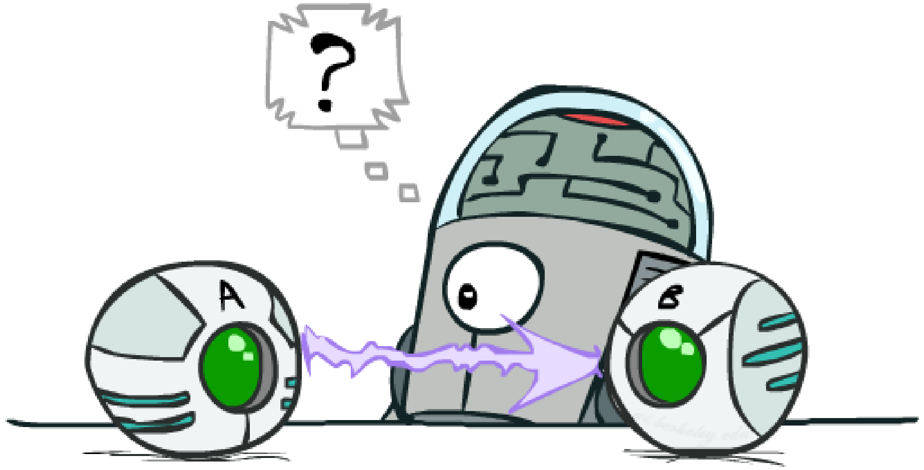
+t	+r	1/3
	-r	2/3
-t	+r	1/7
	-r	6/7

$P(T, R)$

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

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



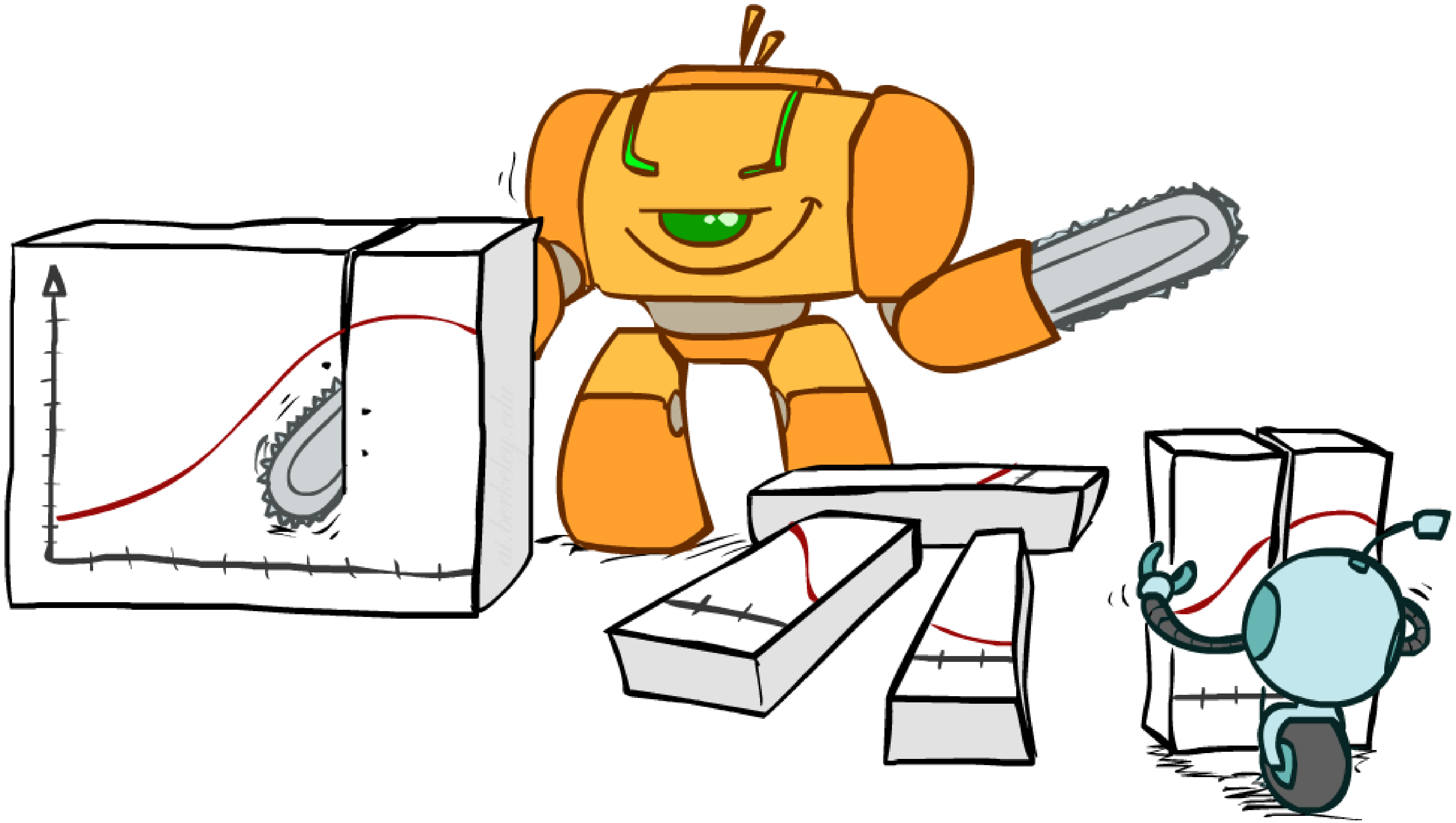
- BNs need not actually be causal
  - Sometimes no causal net exists over the domain (especially if variables are missing)
  - E.g. consider the variables *Traffic* and *Drips*
  - End up with arrows that reflect correlation, not causation

$$T \rightleftarrows D$$

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

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

Bayes Rule



Bayes' Rule

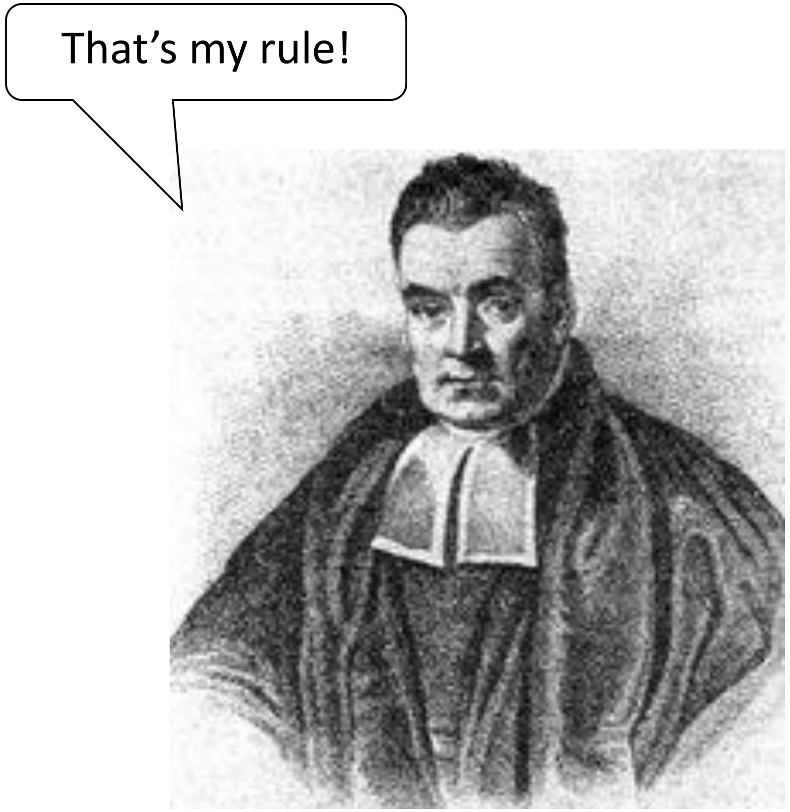
- Two ways to factor a joint distribution over two variables:

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

- Dividing, we get:

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

- Why is this at all helpful?
  - Lets us build one conditional from its reverse
  - Often one conditional is tricky but the other one is simple
  - Foundation of many systems we'll see later (e.g. ASR, MT)



- In the running for most important AI equation!

Inference with Bayes' Rule

s	m
-s	+m
+s	-m

$P(+s) = \sum_{M} P(s, M) = P(s,+m) + P(s,-m)$

$P(s)P(s|m)$ $\quad$ $p(s)p(s|-m)$

$P(m)p(m|s)$ $\quad$ $P(-m)P(-m|s)$

- Example: Diagnostic probability from causal probability:

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

- Example:
  - M: meningitis, S: stiff neck

$$\left.\begin{aligned} P(+m) &= 0.0001 \\ P(+s|+m) &= 0.8 \\ P(+s|-m) &= 0.01 \end{aligned}\right\} \text{Example givens}$$

$$P(+m|+s) = \frac{P(+s|+m)P(+m)}{P(+s)} = \frac{P(+s|+m)P(+m)}{P(+s|+m)P(+m) + P(+s|-m)P(-m)} = \frac{0.8 \times 0.0001}{0.8 \times 0.0001 + 0.01 \times 0.999}$$

  - Note: posterior probability of meningitis still very small
  - Note: you should still get stiff necks checked out! Why?

$P(+s) = P(+s,+m) + P(+s,-m)$

$= P(+s|+m)P(+m) + P(+s|-m)P(-m)$

Quiz: Bayes' Rule

$$P(W|D) = \frac{P(D|W)P(W)}{P(D)}$$

- Given:

$P(W)$

R	P
sun	0.8
rain	0.2

$P(D|W)$

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

- What is P(W | dry) ?

Quiz: Bayes' Rule

- Given:

$P(W)$

R	P
sun	0.8
rain	0.2

$P(D|W)$

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

- What is P(W | dry) ?

$= \frac{P(dry|sun)P(sun)}{P(dry)}$

P(sun|dry) ~ P(dry|sun)P(sun) = .9*.8 = .72 α
P(rain|dry) ~ P(dry|rain)P(rain) = .3*.2 = .06
P(sun|dry)=12/13
P(rain|dry)=1/13

$\frac{0.72}{0.72 + 0.06}$

$\frac{0.06}{0.06+0.72}$

Uncertainty Summary

- Conditional probability $\boxed{P(x|y) = \frac{P(x,y)}{P(y)}}$
- Product rule $P(x,y) = P(x|y)P(y)$
- Chain rule
$$\begin{aligned} P(X_1, X_2, \ldots X_n) &= P(X_1)P(X_2|X_1)P(X_3|X_1,X_2)\ldots \\ &= \prod_{i=1}^{n} P(X_i|X_1,\ldots,X_{i-1}) \end{aligned}$$
- 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: $X \perp\!\!\!\perp Y | Z$
$$\forall x, y, z : P(x,y|z) = P(x|z)P(y|z)$$

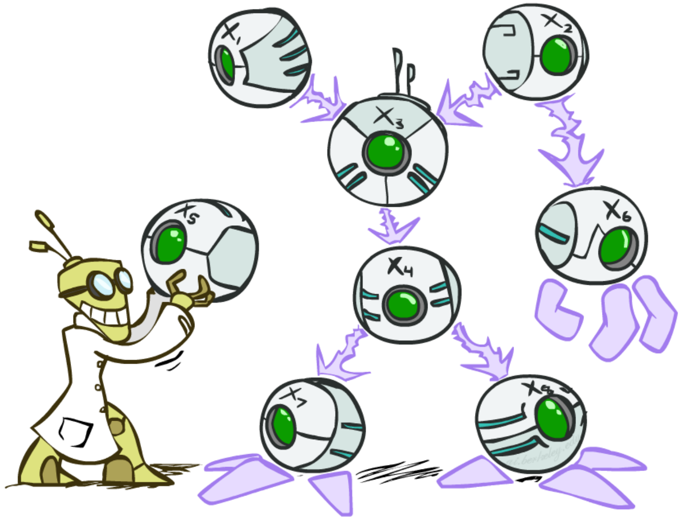
BN lecture

Bayes Rule

Bayes' Net Representation

- 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:

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