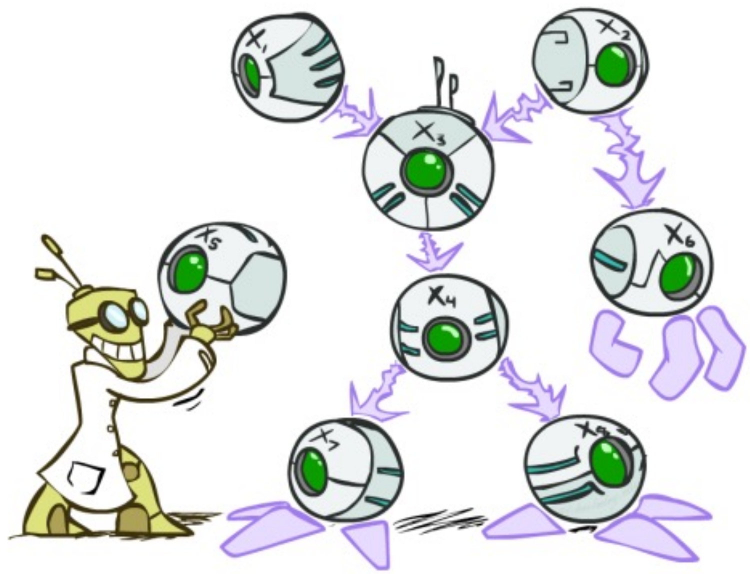
AI: Representation and Problem Solving

Bayes Nets II: Modeling



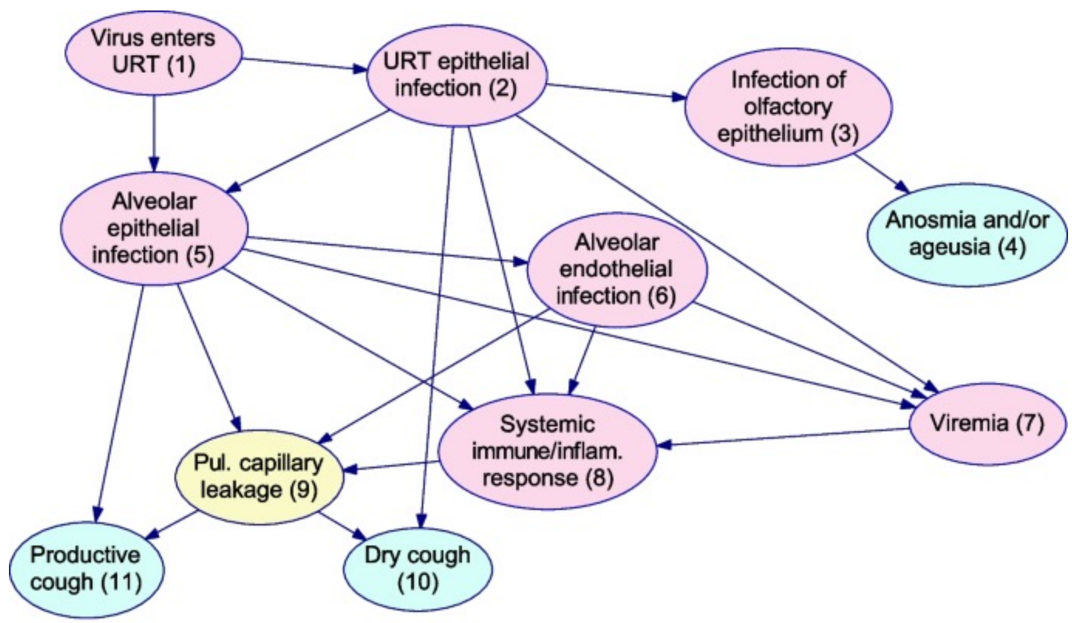
Instructors: Nihar Shah and Tuomas Sandholm

Slide credits: CMU AI and ai.berkeley.edu

Recap

Example: COVID modeling

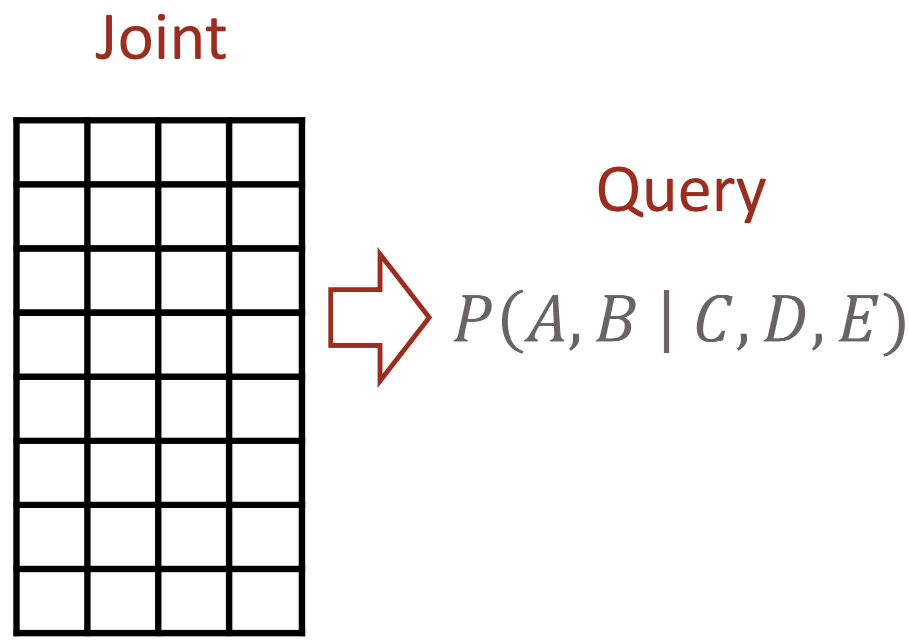
What is P(URT elipthelial infection = yes | dry cough=yes, productive cough=no, anosmia=yes)?



https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-023-01856-1

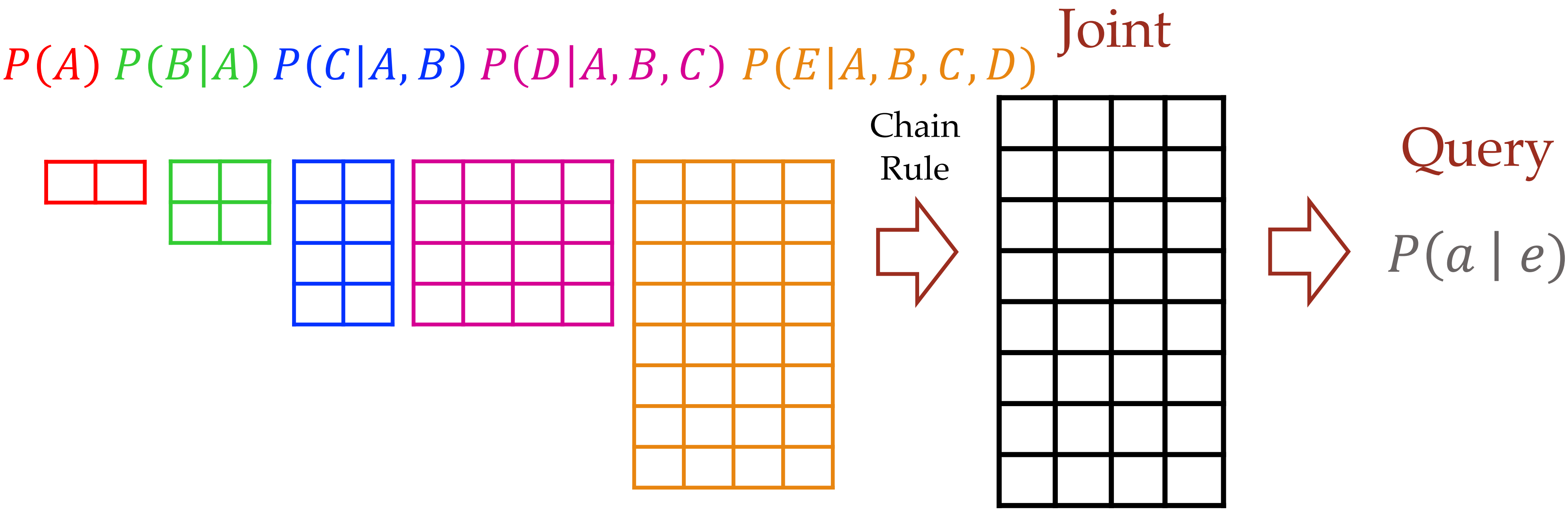
How to answer queries?

- **Joint distributions are the best!**
  - Allow us to answer all marginal or conditional queries
  - However…



- Often we don't have the joint table. Only know some set of conditional probability tables (CPTs)

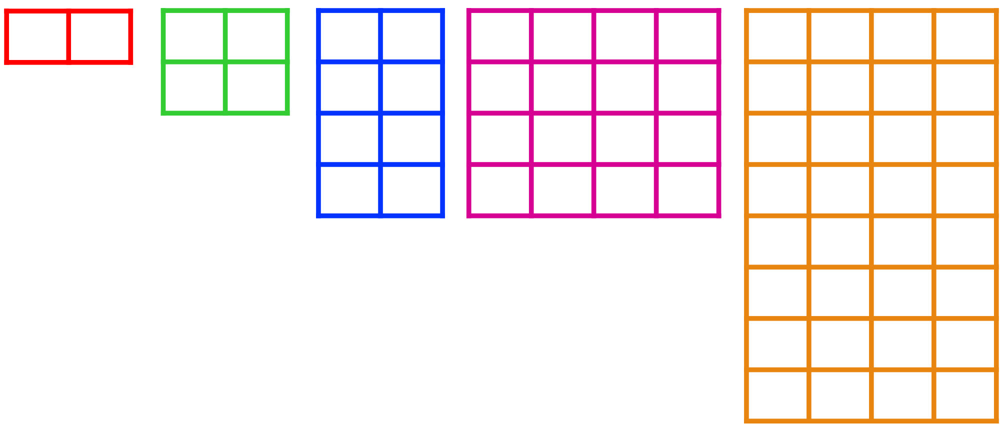
Construct joint from marginals/conditionals



$$P(X_1, \ldots, X_N) = \prod_{n=1}^{N} P(X_n \mid X_1, \ldots, X_{n-1})$$

Answering queries from CPTs: **Problem**

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



- If there are n variables taking d values each
- $\boldsymbol{d^n}$ **entries!!**
- Even the conditional $P(X_n|X_1,\ldots,X_{n-1})$ needs d^n entries

Today

- Addressing this issue by simplifying conditional distributions
  - Conditional independence assumptions
- Constructing the Bayes net
- Answering certain questions
  - "Bayes ball"

Sometimes, distributions have simpler structure

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

- Suppose $P(E|A,B,C,D)$=$P(E|A,B)$ and $P(D|A,B,C)$= $P(D|A,B)$
- **“Conditional independence”** (more on this soon)
- Then$P(A)\ P(B|A)\ P(C|A,B)\ P(D|A,B,C)\ P(E|A,B,C,D)$
$= P(A)\ P(B|A)\ P(C|A,B)\ P(D|A,B)\ P(E|A,B)$
- Needs less data to estimate conditionals (e.g., $P(E|A,B)$ is easier to estimate than $P(E|A,B,C,D)$)
- Needs less computation and storage to answer other queries

What is this “Independence”?

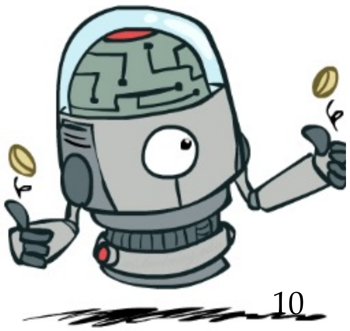
I roll two fair dice…

- What is the probability that the first roll is 5?
- What is the probability that the second roll is 5?
- What is the probability that both rolls are 5?
- If the first roll is 5, what is the probability that the second roll is 5?

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

  - $P(Roll_2=5 \mid Roll_1=5) \;=\; P(Roll_2=5) = 1/6$

- Independence and conditional independence!



Independence

Two random variables X and Y are *independent* if

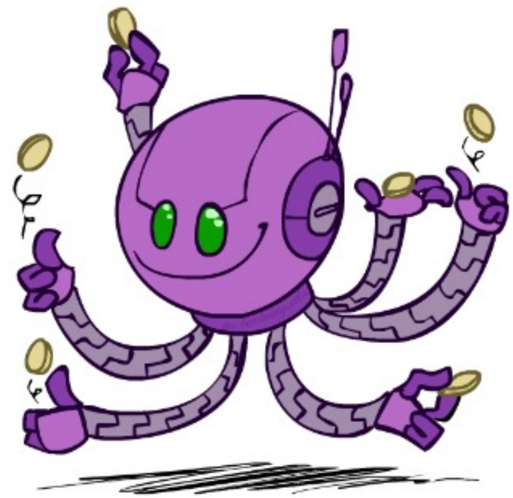
$$\forall x,y \qquad P(x, y) = P(x)\, P(y)$$

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

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

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

joint distribution is simply the product

Question

- Are Temperature and Wetness independent?

$P(T,W)$

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

$P(T)$

T	P
hot	0.5
cold	0.5

$P(W)$

W	P
sun	0.6
rain	0.4

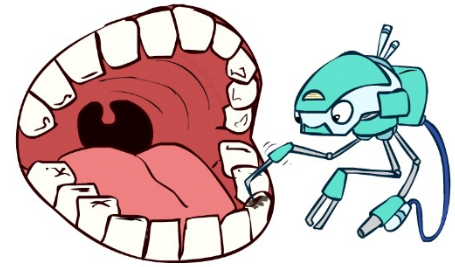
Conditional independence

- X and Y are independent if P(X | Y) = P(X)

- X and Y are **conditionally independent given Z** if
  - P(X, Y | Z) = P(X | Z) P(Y | Z)
  - P(X | Y, Z) = P(X | Z)

- Notation: $X \perp\!\!\!\perp Y \mid Z$

Conditional independence



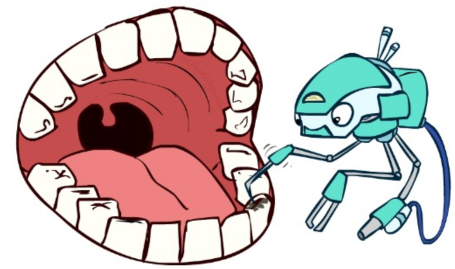
- P(Toothache, Cavity, (p)Robe)

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

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

- Probe is *conditionally independent* of Toothache given Cavity:
  - P(R | T, C) = P(R | C)

Conditional independence



Equivalent statements:

- P(Toothache | Probe , Cavity) = P(Toothache | Cavity)
- P(Toothache, Probe | Cavity) = P(Toothache | Cavity) P(Probe | Cavity)

Have we seen conditional independence in previous lectures?

MDPs

"Markov" generally means that given the present state, the future and the past are independent

For Markov decision processes, "Markov" means action outcomes depend only on the current state

$$P(S_{t+1} = s' | S_t = s_t, A_t = a_t, S_{t-1} = s_{t-1}, A_{t-1}, \ldots S_0 = s_0)$$

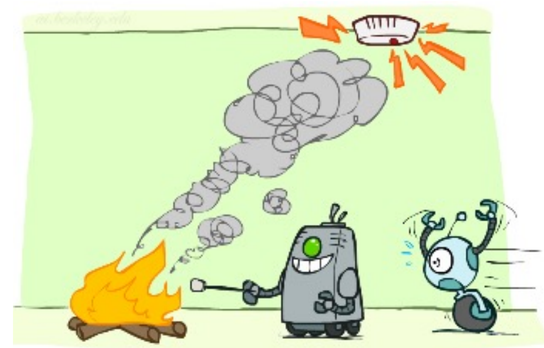
$$=$$

$$P(S_{t+1} = s' | S_t = s_t, A_t = a_t)$$

Andrey Markov
(1856-1922)

Moving back to Bayes nets via a cute example

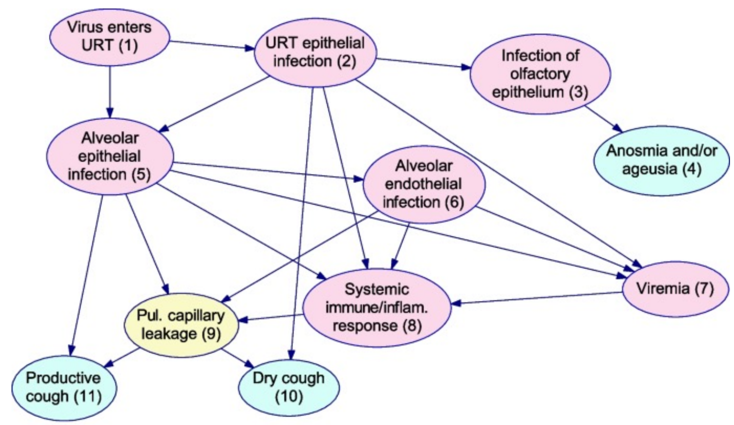
- Fire, Smoke, Alarm
  - **What is P(Fire | Alarm = yes)?**



- Joint distribution: P(S, F, A) = P(F) P(S | F) **P(A | S, F)**
- Estimate each term in the right hand side from some data
- P(A|S,F) involves estimating 4 distributions (corresponding to S,F=yes,yes; S,F=yes,no; S,F=no,yes; S,F=no,no)
- But we may assume that given there is (or isn't) smoke, the ringing of the alarm doesn't depend on whether there is fire
- P(A|S,F)=P(A,S) Conditional independence!
  - P(S, F, A) = P(F) P(S | F) **P(A | S)**
  - P(A|S) involves estimating and storing only 2 distributions

Bayes nets

- Graphical representation of conditional probability tables
- One node per random variable
- Directed acyclic graph
- Exploit conditional independence



Bayes nets

- Recall chain rule: $P(X_1, \dots, X_N) = \prod_{i=1}^{N} P(X_i | X_1, \dots, X_{i-1})$

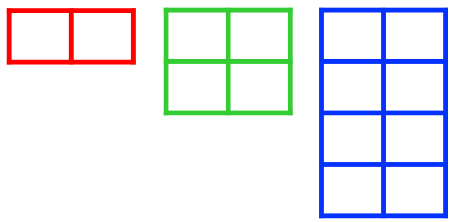
- Exploit conditional independences
  - E.g., suppose you know (or can assume) that $P(X_i | X_1, \dots, X_{i-1}) = P(X_i | \text{some subset of } X_1, \dots, X_{i-1})$
  - The subset of $X_1, \dots, X_{i-1}$ on the right hand side will be parents of X_i

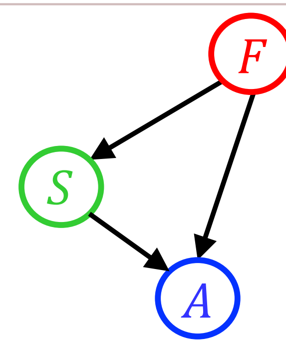
- Encode joint distributions as product of conditional distributions on each variable P(node | parents (node))

- Thus we have $P(X_1, \dots, X_N) = \prod_{i=1}^{N} P(X_i | Parents(X_i))$

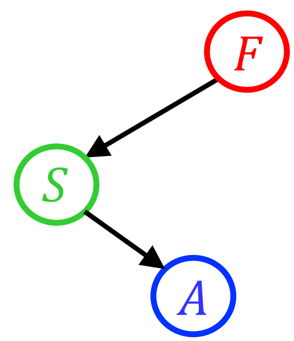
Bayes nets

$$P(X_1, \ldots, X_N) = \prod_{i=1}^{N} P(X_i \mid Parents(X_i))$$





$$P(S, F, A) = P(F)\, P(S|F)\, P(A|S, F)$$

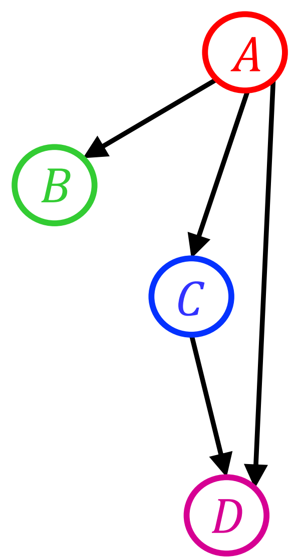


$$P(S, F, A) = P(F)\, P(S|F)\, P(A|S)$$

Question

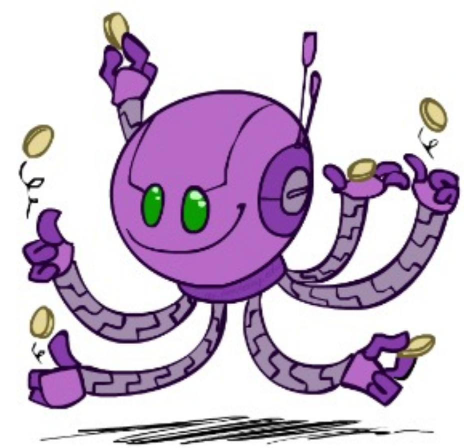
Write down the Bayes net for

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



Another example: Coin Flips

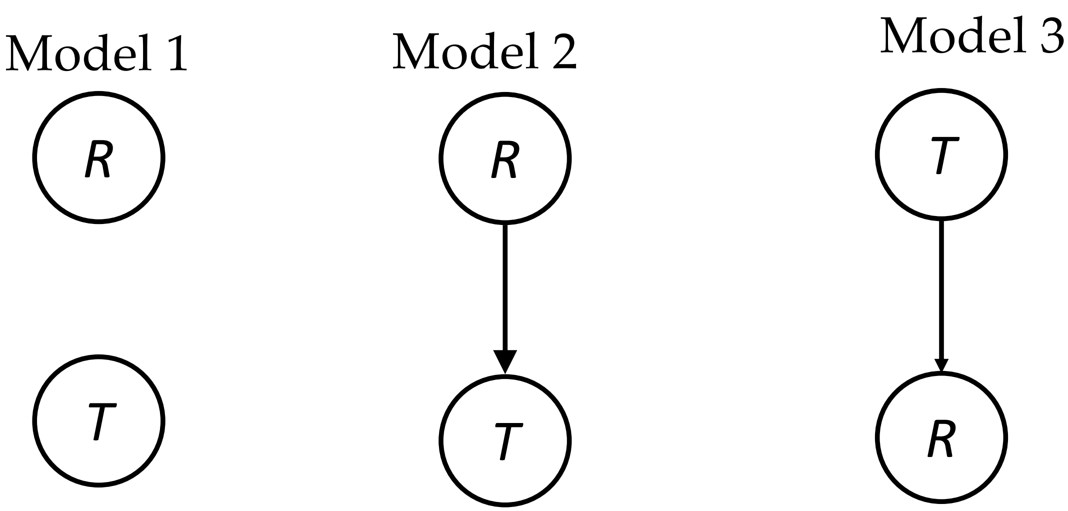
- N independent coin flips
- What is the Bayes net?



- $P(X_1,\ldots,X_n) = P(X_1)\,P(X_2)\ldots P(X_n)$
- All variables are independent

Example: Traffic

- Variables:
  - R: It rains
  - T: There is traffic
  - Which of the following is a better model?

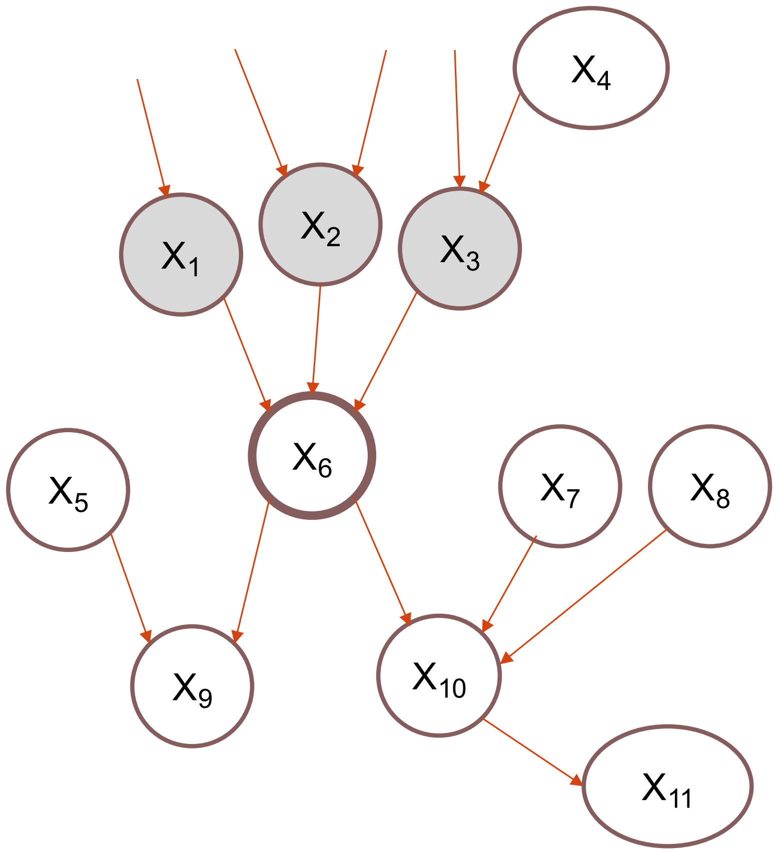


Conditional independence questions

- We wanted to answer questions of the form "What is P(infected | cough)?" or "What is P(infected)?"
- An important special case is to identify if variables are (conditionally) independent. Examples:
  - Is the probability of stock price going up tomorrow independent of global factors given domestic factors?
  - Is air pollution in a city independent of traffic patterns given amount of factory smoke?
  - A company may wish to know whether performance of an intern is independent of pre-req courses given their 281 grade

Conditional independence in Bayes nets

Every variable is conditionally independent of its non-descendants given its parents

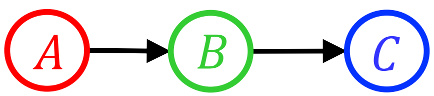


- In this example, is X_6 independent of X_4 given X_1, X_2, X_3?
- Yes…why?
- By definition in the Bayes net!
  - Recall: $P(X_1, \ldots, X_N) = \prod_{i=1}^{N} P(X_i | \, Parents(X_i))$

- **But what about other relations?**
  - E.g., is X_6 is independent of X_{11} given X_7, X_8, X_9?

Special cases that are useful building blocks

For the following Bayes nets, write down the conditional independence assumption being made



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

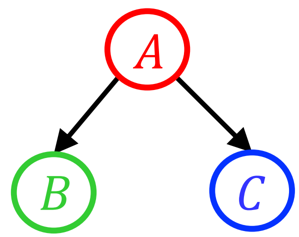
=

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

Assumption:

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

C is independent of A given B



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

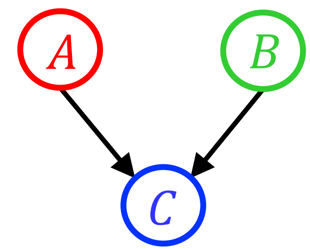
=

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

Assumption:

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

C is independent of 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 of B

Let’s dig deeper, and take a slightly different perspective

Proving conditional independence from the joint distribution

Causal Chain

Are X and Z independent given Y?

X: Low pressure Y: Rain Z: Traffic

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

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

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

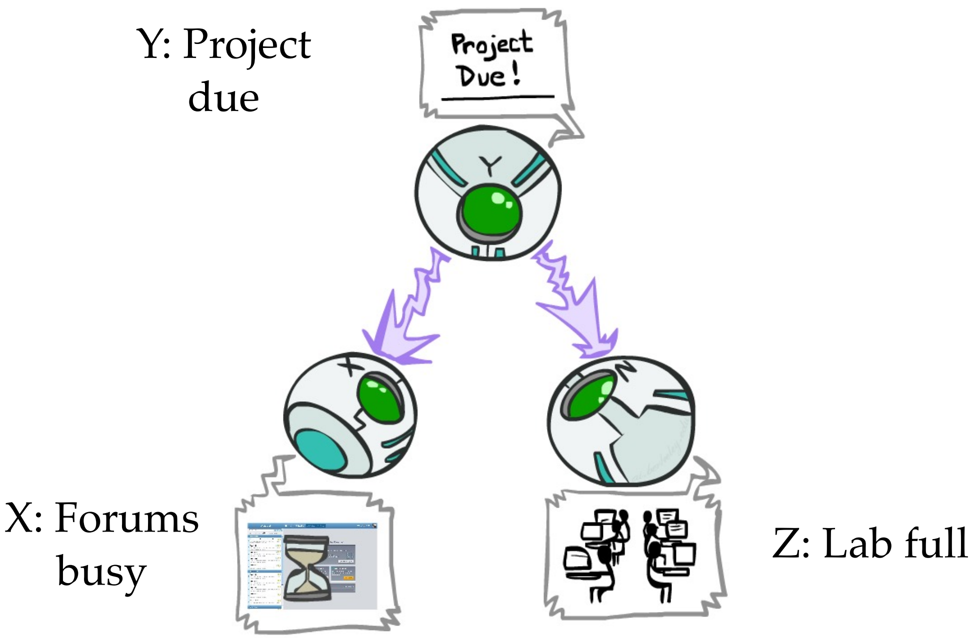
$$= P(z|y)$$

Yes!

We often say that *evidence* (of Y) along the chain *blocks* the influence (of X on Z)

Common Cause

Are X and Z independent given Y?



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

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

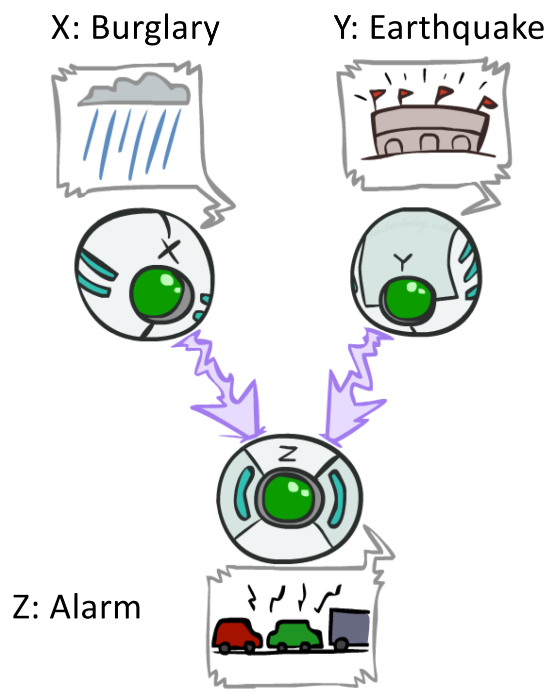
$$= \frac{P(y)P(x|y)P(z|y)}{P(y)P(x|y)}$$

$$= P(z|y)$$

Yes!

We often say that observing the *cause* (Y) *blocks* influence between effects X and Z.

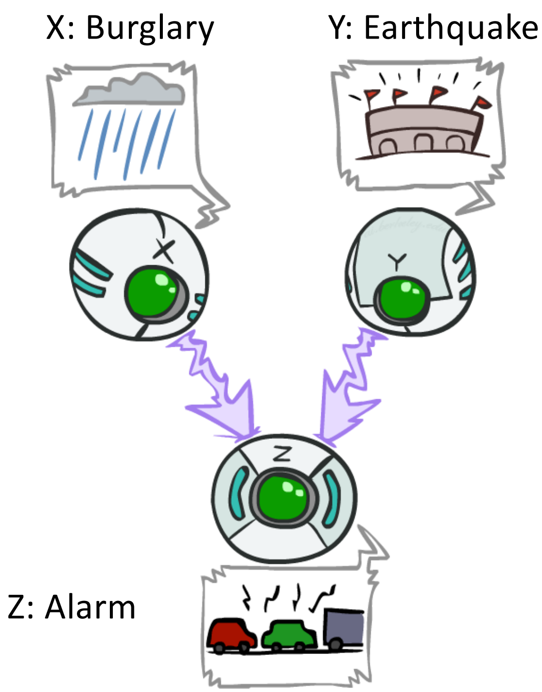
Common Effect



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

- Are X and Y independent?
  - *Yes*: the earthquake is independent of the burglar
  - Still need to prove they must be
  - Exercise: Given $P(x,y,z) = P(x)P(y)P(z|x,y)$ for all x, y, z, prove that X and Y are independent
- Are X and Y independent given Z?
  - *No*: if the alarm sounded and there was no earthquake, then there must have been a burglary.
- This is backwards from the other cases
  - We often say that observing an effect (Z) *activates* influence between possible *causes* (X and Y).

Common Effect



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

- Are X and Y independent given Z?
  - *No*: if the alarm sounded and there was no earthquake, then there must have been a burglary.
  - On the other hand, if the alarm sounded and there was an earthquake, there was probably no burglary.
- "Explaining away"
  - Suppose two causes positively influence an effect. Conditioned on the effect, further conditioning on one cause reduces the probability of the other cause

Recipe (“Bayes ball”)

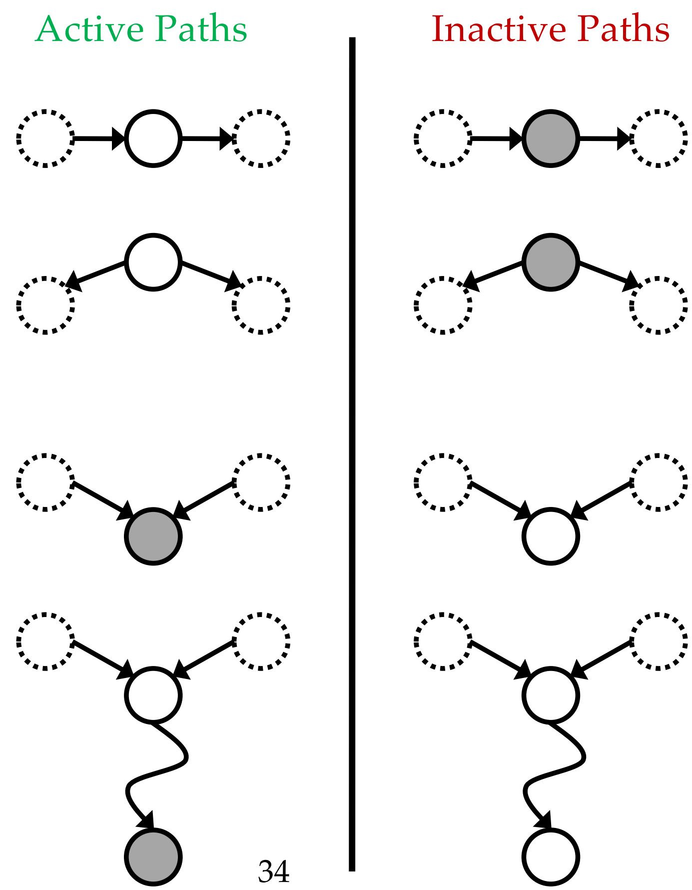
- **Question: Are X and Y conditionally independent given “evidence” variables {Z}?**
- Consider all (undirected) paths from X to Y
- A path is active if **each triple** is active
  - Causal chain A → B → C where B is unobserved (either direction)
  - Common cause A ← B → C where B is unobserved
  - Common effect A → B ← C where B *or one of its descendents* is observed

- No active paths => independence

Shachter, "Bayes-Ball: Rational Pastime (for Determining Irrelevance and Requisite Information in Belief Networks and Influence Diagrams)," *UAI* 1998.

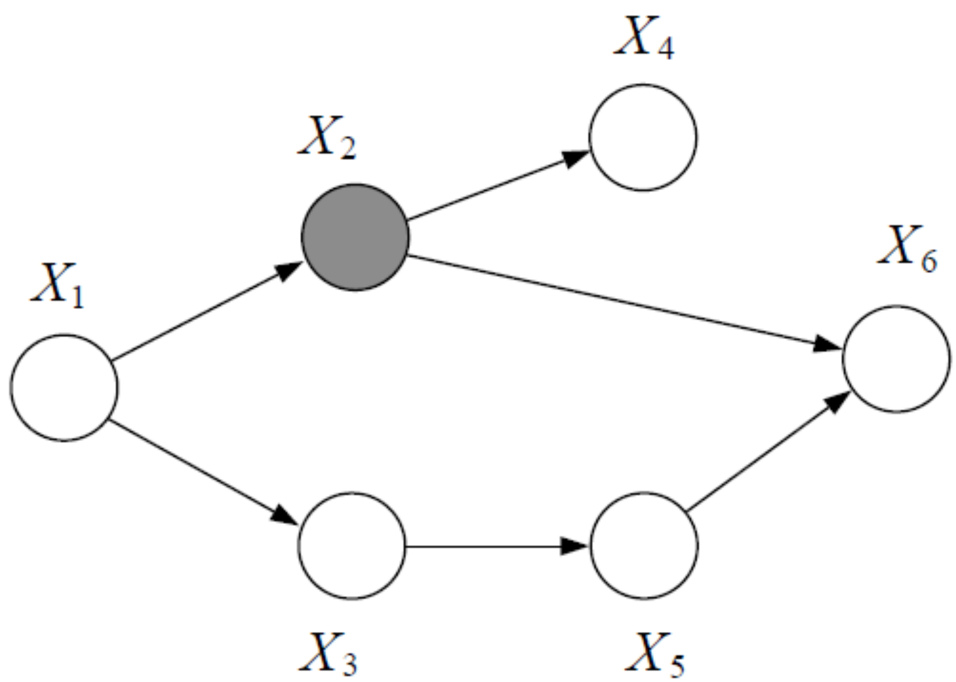
Bayes Ball in pictures

Given nodes ("evidence" nodes) are shaded

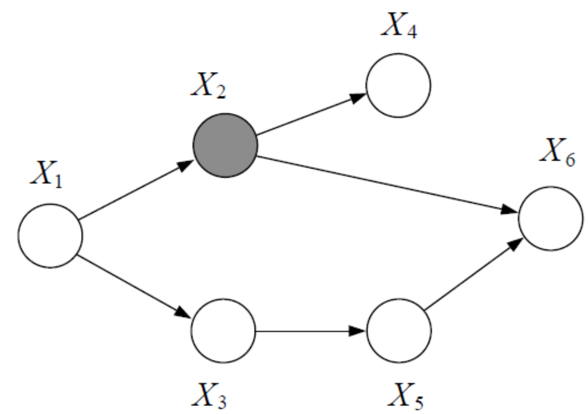


Question

- Is X_1 independent of X_6 given X_2?



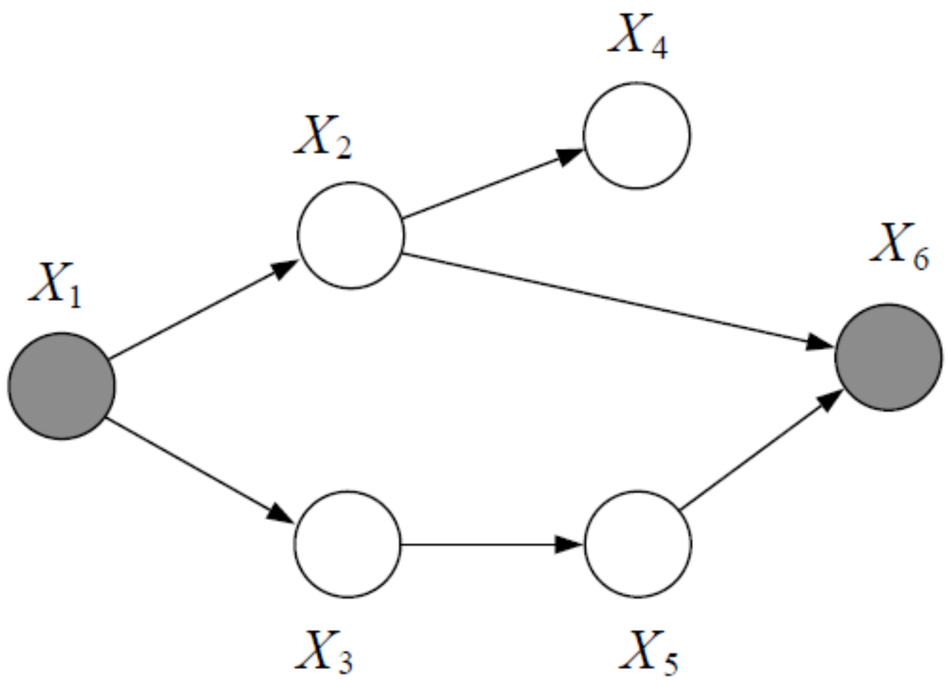
Question



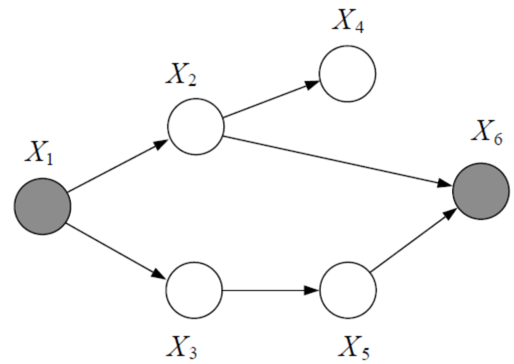
- Is X_1 independent of X_6 given X_2?
- Consider the path $X_1 - X_2 - X_6$
  - Causal chain where middle node is observed
  - Not active
- Consider the path $X_1 - X_3 - X_5 - X_6$
  - Each triplet is a causal chain where middle node is unobserved
  - This path is active
- Thus the answer is “No”

Another question

- Is X_2 independent of X_3 given X_1 and X_6?



Question



- Is X_2 independent of X_3 given X_1 and X_6?
- Consider the path $X_2 - X_1 - X_3$
  - Common cause where X_1 is observed. Thus not active
- Consider the path $X_2 - X_6 - X_5 - X_3$
  - Triplet $X_6 - X_5 - X_3$ is causal chain where middle node is unobserved. Thus this triple is active
  - Triplet $X_2 - X_6 - X_5$ is common effect where X_6 is observed. Thus this triple is not active
  - This path is also not active
- All paths are not active, and hence the answer is “Yes”

Important note

- We look at all paths using undirected edges
- But when going down a path and looking at triplets, we need to look at the direction of the edges
- Common cause and common effect induce opposite effects: observing parent causes independence, observing child causes dependence

Poll

Choose the true statement(s):

(A) If X and Y are conditionally independent given Z, then X and Y are independent

(B) If X and Y are independent, then X and Y are also conditionally independent given Z

(C) Neither is true